Running head: The interoceptive accuracy scale

Page count: 33 (including references; 7499 words; 5 figures/tables)

Submission date: 07/10/2018

Validation of the interoceptive accuracy scale (IAS) supports distinction between self-reported interoceptive accuracy and awareness

Jennifer Murphy[1], Rebecca Brewer[2], David Plans[3], Sahib S Khalsa[4, 5], Caroline Catmur[6], Geoffrey Bird[1,3]

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London

[2]Department of Psychology, Royal Holloway, University of London

[3]Department of Experimental Psychology, University of Oxford

[4]Laureate Institute for Brain Research, Tulsa, USA

[5]Oxley College of Health Sciences, University of Tulsa, Tulsa, USA

[6]Department of Psychology, Institute of Psychiatry, Psychology and Neusroscience, King's College London

*Corresponding author: Jennifer.Murphy@kcl.ac.uk

+ 44 (0) 207 848 5410

Social, Genetic and Developmental Psychiatry Centre (MRC)

Institute of Psychiatry, Psychology and Neuroscience - PO80

De Crespigny Park,

Denmark Hill, London,

United Kingdom, SE5 8AF

Abstract

Measures of the perception of the state of one's own body ('interoception') can be categorized as one of several types. Most commonly, these measures are of: 1) the ability to form an accurate percept of the body's state, 2) confidence in the accuracy of a specific interoceptive percept at a particular point in time, and 3) trait differences in the degree to which one is aware of interoceptive information. At present, however, there is a paucity of measures designed to assess trait differences in self-perceived interoceptive accuracy. This paper reports on the development of such a measure, the Interoceptive Accuracy Scale (IAS). Across six studies we report on the IAS factor structure, test-retest reliability, and its relationship with measures of trait interoceptive awareness, accuracy of interoceptive percepts, confidence in the accuracy of specific interoceptive percepts, and metacognition with respect to interoceptive accuracy ('interoceptive insight'). Results support the distinction between individual differences in awareness of interoceptive information and in the accuracy of interoceptive perception, and suggest that both accuracy and awareness can be measured using both objective measures and self-report instruments such as the IAS.

Keywords: Interoception, Interoceptive Accuracy, Interoceptive Sensibility, Interoceptive Awareness, Self-report

1. Introduction

In recent years the contribution of interoception, the perception of the internal state of one's body (Craig, 2002; c.f. Khalsa et al., 2017), to aspects of higher-order cognition such as emotional abilities and learning and decision making has begun to be appreciated (e.g., Barrett & Simmons, 2015; Critchley & Harrison, 2013; Khalsa & Lapidus, 2016; Murphy, Brewer, Catmur, & Bird, 2017a; Quattrocki & Friston, 2014; Seth, 2013; Damasio, 1994; Heyes & Bird, 2008). In addition to these elegant theoretical models of how interoception may contribute to a raft of cognitive processes, a relatively small but growing body of literature has examined the degree to which individual differences in interoception predict performance in tests of abilities as diverse as decision making (Dunn et al., 2010; Sokol-Hessner, Hartley, Hamilton, & Phelps, 2015; Werner, Jung, Duschek, & Schandry, 2009), theory of mind (Shah, Catmur, & Bird, 2017), emotion recognition (Terasawa, Moriguchi, Tochizawa, & Umeda, 2014), and memory (Garfinkel et al., 2013). Empirical progress has been hampered, however, by a paucity of tests of interoceptive ability, and by problems with some of the existing measures (e.g., Brener & Ring, 2016; Murphy, Brewer, Hobson, Catmur, & Bird, 2018; Ring, Brener, Knapp, & Mailloux, 2015; Ring & Brener, 1996; Khalsa, Rudrauf, Sandesara, Olshansky, & Tranel, 2009).

Measures of individual differences in interoception (at least those aspects of interoception accessible for conscious report; Khalsa et al, 2017) can be organised according to a 2 x 2 dimensional structure. The first dimension reflects whether the accuracy of one's interoceptive percept (i.e. one's *ability* to perceive interoceptive signals accurately), or the degree to which one is aware of interoceptive information (i.e. one's propensity to attend to interoceptive information), is the target of measurement, and the second dimension reflects whether the measure provides an objective measure of interoceptive accuracy or awareness, or whether it is a self-report measure designed to obtain an individual's belief as to the accuracy of their interoceptive perception, or the degree to which they are aware of interoceptive information (Murphy et al., 2018a; based on Garfinkel & Critchley, 2013; Garfinkel, Seth, Barrett, Suzuki, & Critchley, 2015; Ceunen, Van Diest, & Vlaeyen, 2013). The distinction between accuracy and awareness reflects the fact that a person may report being very aware of their internal signals (e.g., they may report always feeling hunger or the need to urinate) but recognise that their perception of these signals is inaccurate (e.g., they may feel hungry even after eating a large meal, or realise that their urge to urinate is not a good indicator of the need to do so). Similarly, a person may be able to form a very accurate percept of their internal state when explicitly asked to do so (e.g., when taking part in an experiment), but that this may not reflect their propensity to be aware of internal signals in their everyday lives.

Existing measures of interoceptive ability are not evenly distributed across this 2 x 2 matrix, with the majority of tests providing objective measures of interoceptive accuracy (e.g., Dale & Anderson, 1978; Garfinkel et al., 2016; Murphy, Catmur, & Bird, 2018a; Schandry, 1981; Whitehead, Drescher, Heiman, & Blackwell, 1977; see Khalsa et al., 2017), or self-report measures of trait interoceptive awareness[1] (one's engagement by, and attention to, interoceptive signals; e.g., Porges, 1993; Mehling et al., 2012; sometimes referred to as sensibility; see Khalsa et al., 2017). The focus of this paper is on a less well-populated quadrant of the matrix; measurement of subjective beliefs concerning the accuracy of one's interoceptive perception (i.e. one's perception of one's own interoceptive accuracy). At present, such beliefs are obtained by asking participants to estimate their confidence in their responses during an objective test of interoceptive perception. For example, if asked to count how many times their heart beat in a period of time (a count which can be compared to an objective measure of heart beats to determine its accuracy), the participant might be asked to provide their degree of confidence in their count (e.g., Garfinkel et al., 2015). Although a very effective way of measuring beliefs concerning interoceptive accuracy, two features of this measure may mean that it is not always appropriate for addressing particular research questions. First, such a measure provides an estimate of beliefs in one interoceptive domain only. In the heartbeat counting example, one obtains a measure of the participant's beliefs concerning their ability to perceive accurately their heartbeats, but not their beliefs concerning their ability to perceive accurately signals in other interoceptive domains, such as the gut or respiratory system. Second, confidence judgements during objective measurement of interoceptive accuracy provide a moment-by-moment—presumably state-dependent—measurement of beliefs concerning interoceptive accuracy. For some research questions, however, it may be preferable to obtain a trait-based measure of a participant's general belief concerning their interoceptive accuracy, independent of specific states which may momentarily increase (e.g. exercise) or decrease (e.g. depressed mood) interoceptive accuracy (e.g., Pollatos, Traut-Mattausch, & Schandry, 2009; Schandry & Specht, 1981; Ring et al., 2015).

Whilst there are a number of measures that assess beliefs concerning the perception of interoceptive signals, these tend to conflate the accuracy of interoceptive perception and the degree of awareness of internal states, and some also include other aspects of bodily processing (e.g., Fiene, Ireland, & Brownlow, 2018; Mehling et al., 2012; Shields, Mallory, & Simon, 1989). As far as we are aware, only one previous measure has been developed which solely assesses self-perceived trait interoceptive accuracy, the Interoceptive Confusion Questionnaire (ICQ), a measure with good predictive validity but poor psychometric properties (Brewer, Cook, & Bird, 2016). Accordingly, this paper reports on the development and validation of the Interoceptive Accuracy Scale (IAS), a trait-based measure assessing global (i.e. interoceptive domain-general rather than domain-specific) beliefs concerning one's ability to accurately perceive interoceptive signals. In line with the 2 x 2 dimensional matrix specified above, it was predicted that little or no relationship would be observed between self-report measures of interoceptive accuracy and awareness, and that if self-report measures predict objectively-measured interoceptive accuracy, then the relationship would be specific to self-reported interoceptive accuracy rather than self-reported interoceptive awareness.

2. The interoceptive accuracy scale

The IAS is reproduced in Figure 1. The scale was constructed to include a number of items relating to physical sensations that have either been described as interoceptive (Khalsa & Lapidus, 2016; Khalsa et al., 2017) or are associated with activation in the insula (e.g., Critchley & Harrison, 2013; Langer, Beeli, & Jäncke, 2010; Mazzone, McLennan, McGovern, Egan, & Farrell, 2007), an area commonly associated with the processing of interoceptive signals (e.g., Craig, 2002; Khalsa et al., 2017). As far as possible we attempted to include signals where objective accuracy could be pertained (e.g., flatulence and eructation were individually specified rather than using broad statements like "gastric sensations" for alimentary interoception). Instructions also provided participants with specific examples of what would, and would not, constitute accurate internal perception (see Figure 1). The scale is comprised of 21 items rated on a scale from Strongly Agree (5) to Strongly Disagree (1), with scores ranging from 21-105. Higher scores indicate greater self-reported interoceptive accuracy.

2.1 Ethical approval and data availability

For all studies ethical approval was granted by the local ethics subcommittee. Participants provided informed consent and were fully debriefed following study completion. For all studies data is available at https://osf.io/3m5nh/?view_only=a68051df4abe4ecb992f22dc8c17f769.

3. Study 1: Exploratory Principal Components Analysis (PCA)

3.1 Method

An opportunity sample of 451 individuals (M_{age} = 25.77, SD_{age} = 8.37, Age range = 18-69, 313 Female, 133 Male, 5 Other) completed the IAS online via Qualtrics (Provo, UT) with entry into a prize draw (£25 Amazon voucher) offered as an incentive. Such a sample size is considered adequate for PCA (e.g., Field, 2005). Participants were recruited via pre-existing databases of individuals willing to take part in research and via social media platforms. Of these 451 individuals, 158 individuals indicated that English was their second language (ESL), and 75.4% of the sample did not have a current or previous diagnosis of a psychiatric condition.

3.2 Results

Cronbach's Alpha indicated good internal consistency of the IAS (α = .88). Initially the suitability of the data for analysis was examined. First, it was observed that all the 21 IAS items correlated with another item. Secondly, the Kaiser-Meyer-Olkin measure of sampling adequacy was .887, above the commonly recommended value of ~.5, and Bartlett's test of sphericity was significant (χ^2 (210) = 2695.69, $p < .05$) (Williams, Onsman, & Brown, 2010). The diagonals of the anti-image correlation matrix were also all over .848. Finally, the communalities were all above .210. As such, the data were deemed to be suitable for analysis including all 21 items.

PCA with varimax rotation[2] (Field, 2005) was used to examine the factor structure with coefficients <.40 suppressed (Pituch, & Stevens, 2015). Examination of the scree plot suggested a 2-factor solution; therefore, extraction was fixed to two factors explaining 19% and 18.6% of the variance, respectively, in the rotated solution. The rotated factor solution is shown in Table 1, with items arranged by their magnitude of factor loading.

As shown in Table 1, a clear 2-factor solution was observed. Whilst there is no obvious explanation for these groupings it is possible that the first factor reflects the perception of interoceptive signals and the second factor comprises signals that may be difficult to perceive using interoceptive information alone (e.g., bruising or low blood sugar may be indicated by the presence of a skin discolouration or shaking, respectively), or perturbations of bodily functions that are socially unacceptable.

4. Study 2: Test re-test reliability

4.1 Method

Having examined the structure of the IAS, the test re-test reliability of the IAS was examined over a period of 30 days in a second sample. In addition, the test re-test reliability of two other interoception questionnaires, the ICQ and the long awareness scale from the Porges body perception questionnaire (BPQ; Porges, 1993) was also assessed to determine whether the IAS had test re-test reliability comparable to existing questionnaires. Inclusion of these measures also enables their relationship with the IAS to be established. Whilst there are several other interoception questionnaires, these were selected for comparison with the IAS as they are the only measures which focus specifically on interoceptive awareness (BPQ) or accuracy (ICQ) in their entirety. In line with Murphy et al., (2018) it was expected that a relationship between the measures of accuracy (IAS and ICQ) would be observed, but no relationship would be observed between the measures of accuracy and the measure of awareness (BPQ).

Questionnaires were completed online via Qualtrics in a randomised order. Participants were recruited via pre-existing databases and social media platforms. A prize draw (£25 Amazon voucher) was offered as an incentive. Of the 194 participants that completed stage 1, 117 participants completed both testing sessions (M_{age} = 40.48, SD_{age} =

17.08, Range 18-90, 79 Female, 35 Male, 3 Other, N=16 ESL, 84 had no previous or current psychiatric conditions, 16 reported an existing condition and 17 declined to answer) approximately 30 days apart (M_{days} = 30.75, SD_{days} = 1.17, Range 29-35).

4.2 Results

Good test re-test reliability was found for the IAS ($r(115) = .754$, $p<.001$), BPQ ($r(115) = .684$, $p<.001$) and the ICQ ($r(115) = .814$, $p<.001$). Removal of ESL or individuals reporting diagnoses did not change the overall pattern of results (all rs>.66, all $p<.001$). The magnitude of the correlations reported here (and throughout the paper) were formally compared using Steiger's Z-test (Steiger, 1980) via the quantpsy web implementation (Lee & Preacher, 2013a; 2013b). No significant difference in reliability was found between the IAS and the ICQ ($Z = 1.449$, $p = .15$), or the IAS and the BPQ ($Z = 1.099$, $p = .27$). However, the ICQ showed greater test re-test reliability than the BPQ ($Z = 2.282$, $p = .022$).

The relationship between scores on these questionnaires was also examined. A moderate correlation was found between the ICQ and the IAS at the first stage of testing ($r(115) = -.631$, $p<.001$), whereby greater self-reported interoceptive accuracy was associated with lower interoceptive confusion (also an accuracy measure). In contrast, no relationship was found between the IAS and the BPQ ($r(115) = .066$, $p>.250$) or the ICQ and the BPQ ($r(115) = .013$, $p>.250$), suggesting no relationship between self-reported interoceptive awareness (BPQ) and accuracy (IAS and ICQ) measures (Figure 2, panel a). This pattern remained after removal of ESL and individuals reporting diagnoses. The size of the correlation between the ICQ and the IAS was significantly larger than the correlation between the IAS and BPQ, and BPQ and ICQ (all $p<.01$). The same pattern of results was observed at the second stage of testing; the IAS and ICQ were highly correlated ($r(115) = -.712$, $p<.001$), whereas the BPQ did not correlate with the IAS ($r(115) = .022$, $p>.250$) or the

ICQ ($r(115) = .091$, $p>.250$). This pattern of significance remained after removal of ESL participants and individuals reporting diagnoses. As before, the correlation between the IAS and ICQ was significantly larger than the correlations between the IAS and BPQ or the ICQ and BPQ (all $p<.01$). As shown in Figure 2a, one individual had an extremely low IAS score, but analyses removing this individual produced the same pattern of results.

5. Study 3: Replication and Convergent validity

Study Three was conducted in order to provide an additional opportunity to examine the relationship between the IAS and BPQ and, as a measure of convergent validity, to examine the relationship between the IAS and another measure previously associated with interoception, alexithymia (difficulties identifying and describing one's own emotions; e.g., Bagby, Parker, & Taylor, 1994; Sifneos, 1973), measured using the Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994). Whilst alexithymia is often associated with poor objective and self-reported interoceptive accuracy, and reduced interoceptive awareness as measured by objective tests (e.g., Herbert, Herbert, & Pollatos, 2011; Murphy et al., 2018a; Murphy et al., 2018b; Brewer et al., 2016; Fiene et al., 2018; Zamariola, Vlemincx, & Luminet, 2018), some studies report a positive relationship between alexithymia and self-reported awareness of internal sensations (e.g., Ernst et al., 2014; Longarzo et al., 2015; Betka et al., 2018). Based on this evidence, it was expected that the TAS-20 would correlate negatively with the IAS, but positively with the BPQ.

5.1 Method

76 participants completed questionnaires ($M_{age} = 39.25$, $SD_{age} = 22.09$, Range 18-91, 51 Female, 25 Male, 12 ESL, 89.5% had no previous or current psychiatric conditions): the IAS and BPQ along with the TAS-20 and the Rosenberg self-esteem scale (Rosenberg, 1965), a 10-item measure assessing both positive and negative views about oneself with

scores rated on a 4-point Likert scale (strongly agree to strongly disagree) with higher scores indicating higher self-esteem. The self-esteem measure was included to ensure that any observed relationship between alexithymia and measures of interoception was not driven by a tendency to report oneself as impaired, regardless of the specific ability being assessed. Questionnaires were completed online via Qualtrics in a randomised order. Participants were recruited via pre-existing databases. A prize draw (£25 Amazon voucher) was offered as an incentive.

5.2 Results

Replicating results from Study Two, no relationship was observed between the BPQ and the IAS ($r(74) = .040$, $p>.250$) (Figure 2, Panel b). The TAS-20 was negatively correlated with the IAS ($r(74) = -.430$, $p<.001$) but not the BPQ ($r(74) = .079$, $p>.250$). This pattern remained after removing ESL and individuals with diagnoses. The size of the correlation between the IAS and TAS-20 was significantly larger than the correlation between the BPQ and TAS-20 ($Z = -3.354$, $p<.001$). Self-esteem scores were negatively correlated with TAS-20 scores ($r(74) = -.528$, $p<.001$), and positively with IAS scores ($r(74) = .448$, $p<.001$), but no relationship was found between self-esteem and BPQ scores ($r(74) = -.098$, $p>.250$). However, even after controlling for self-esteem using partial correlations, the relationship between the TAS-20 and IAS remained significant ($r(73) = -.255$, $p=.027$), and the relationship between the TAS-20 and BPQ remained not significant ($r(73) =.032$, $p>.250$). As shown in Figure 2 (panel b), one individual had an extreme score for the IAS (<50) that did not meet formal criteria to be considered an outlier. However, removing this individual did not change the pattern of results.

6. Study 4: Preliminary evidence of the relationship with objective interoceptive accuracy

Previous work examining the relationship between self-reported moment-by-moment judgements of interoceptive accuracy, and objective measures of interoceptive accuracy, often reveal small, but significant, correlations between the two measures (e.g., Garfinkel et al., 2015; though this may depend on the measure of interoceptive accuracy utilised; see Schulz et al., 2013; Forkmann et al., 2016). In contrast, there tends to be no relationship between questionnaire measures of trait interoceptive awareness (e.g., the BPQ) and objectively measured interoceptive accuracy (e.g., Garfinkel et al., 2015). Here we present initial data examining the relationship between trait-based subjective judgements of interoceptive accuracy obtained using the IAS, and objectively measured cardiac interoceptive accuracy.

In addition, the inclusion of confidence judgements to measure moment-by-moment beliefs concerning interoceptive accuracy allows an individual's 'interoceptive insight' to be calculated (Khalsa et al., 2017; referred to as interoceptive awareness by Garfinkel et al., 2015). This is a metacognitive measure calculated during objective tests of interoceptive accuracy, which indexes the degree to which an individual's subjective belief about the accuracy of their interoceptive perception accurately reflects their objective interoceptive accuracy (Khalsa et al., 2008). Interoceptive insight has been argued to be an important predictor of mental health, particularly levels of anxiety (e.g., Garfinkel et al., 2015; Khalsa et al., 2017). The IAS is not a subjective measure of interoceptive insight as it asks respondents to report their interoceptive accuracy, not their belief as to how well their perceived degree of interoceptive accuracy predicts their actual interoceptive accuracy. However, by comparing an individual's perception of their interoceptive accuracy as indexed by the IAS with their interoceptive accuracy as measured by objective tests, one could calculate an additional measure of interoceptive insight. However, such a measure would be less specific than the measure of interoceptive insight calculated using confidence judgements (unlike confidence measures the IAS measures perceived interoceptive accuracy across domains and over an extended period of time), and noisier (due to the reduced number of measurements), and therefore interoceptive insight was calculated using confidence judgements in Study 4.

6.1 Method

6.1.1 Participants

67 participants took part in this online study facilitated by a custom-built mobile phone application in exchange for Amazon Mechanical Turk credits worth $30. Participants who provided no response on the interoception task (e.g., for every trial the default values of 0 heartbeats and total confidence were given) were classified as incomplete. 59 participants (24 males) completed both the interoception task and questionnaires and were thus included in the final dataset. Age information was gathered in bands (under 18, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84 and 85 or older). The standard deviation for age was equivalent to one band (SD = 1.208). In this sample, 18-24 and 65-74 were the lowest and highest age bands recorded, respectively.

6.1.2 Questionnaire measures

Participants completed the IAS as well as the Depression, Anxiety and Stress Scale (DASS-21; (Lovibond & Lovibond, 1995)) prior to the interoception task. The DASS-21 is a 21-item questionnaire assessing mood over the past week that has excellent psychometric properties and demonstrated convergent validity (Lovibond & Lovibond, 1995). Separate subscales measure depression, anxiety and stress, with higher scores indicating greater depression, anxiety, and stress, respectively.

6.1.3 Interoceptive accuracy and confidence

The heartbeat counting task (HCT; Schandry, 1981; Dale & Anderson, 1978) was used to quantify interoceptive accuracy with a time estimation control task (TET) also employed (Ainley, Brass, & Tsakiris, 2014; Murphy et al., 2017b; Shah et al., 2016). Although the HCT is the most commonly used measure of interoceptive accuracy, it has been the subject of a number of recent critical papers (e.g., Khalsa et al., 2009; Ring et al., 2015; Murphy et al., 2018b). Most of these criticisms, however, have been made using HCT data which was obtained without the full range of control measures. As such, the validity of the HCT when properly controlled as a measure of interoceptive accuracy is currently unclear (see Murphy et al. 2018b for further discussion).

Objective heartbeat was quantified using photoplethysmography (PPG) implemented via a custom built mobile application (for details of the technology see Morelli, Bartoloni, Colombo, Plans & Clifton, 2017; Cropley et al., 2017). Participants first completed 4 intervals of the HCT (22, 32, 42, 98 or 25, 35, 45 100 or 28, 38, 48, 103 or 31, 41, 51, 106 seconds) followed by 3 intervals of the TET (23, 40, 56 seconds). For both the HCT and TET, intervals were presented in a randomised order.

After a baseline heart rate measurement of 120 seconds was taken, participants began the HCT task. Participants were shown the following instructions on the smartphone screen and were played an audio file of a voice reading the same: ‘In this task, you'll be asked to silently count your heartbeats, without physically monitoring your heartbeat using your hands (such as feeling your pulse by putting your fingers on your wrist). You'll be given a countdown, and then you will count heartbeats you can perceive from an initial sound to a final sound. After, you'll be asked to enter the number of heartbeats you perceived. It's important you don't guess or count seconds instead. If you didn't feel any heartbeats, enter 0.' A countdown timer of three seconds was used to indicate the start of a trial with a beep indicating the commencement and completion of each trial. Upon completion, participants were asked to report the number of heartbeats they perceived by inputting the number into a box provided. This immediately triggered the start of the next trial, that always began with a reiteration of the instructions. If the heartrate signal was interrupted during a trial, detected by the absence of their finger on the smartphone's camera (through which the PPG algorithm gathers heartrate data) a prompt was given to the participant asking them to readjust their grip.

After each trial participants provided confidence ratings on a 1-10 scale for the HCT (1 = *Total Guess/No Heartbeat Awareness*; 10 = *Complete Confidence/Full perception of Heartbeat*) using a visual analogue scale adapted from Garfinkel et al., (2015). As well as providing a specific measure of beliefs regarding interoceptive accuracy on the HCT, inclusion of confidence ratings allows calculation of interoceptive insight (see section 6.1.5).

6.1.4 Additional control measures

As well as TET performance, body-mass index (BMI) was also calculated using self-reported height and weight. Resting heart rate and heart rate variability were also calculated and used as control variables (see supplementary materials [S1] for details).

6.1.5 Scoring and data analysis

Interoceptive accuracy on the heartbeat tracking task was estimated on a scale from 0 – 400 using the following equation: Σ(1 – (|Actual number of heartbeats – participant's estimate|/Actual number of heartbeats)) x 100. Higher scores indicate better perception of one's heartbeat (Murphy et al., 2017b; Shah, Hall, Catmur, & Bird, 2016). Timing scores were estimated similarly: Σ(1 – (|Actual number of seconds – participant's estimate|/Actual number of seconds)) x 100. Again, higher scores indicate better time estimation. Whilst this scoring method is appropriate in cases where participants do not over estimate more than 2x the amount of actual recorded heartbeats, for cases where sizeable overestimation occurs the following equation is more appropriate $1 - (|nbeatsreal - nbeatsreported|) /((nbeatsreal + nbeatsreported)/2)$ (see Garfinkel et al., 2015). For completeness, and ease of comparison with other studies, the results using the first (hereafter 'standard') formula excluding over-estimators for the HCT (N=5) and TET (N=2) are reported below in addition to the whole sample using the second (hereafter 'alternative') formula that controls for overestimation. However, the overall pattern of results was unchanged regardless of the scoring method utilised.

For the HCT, confidence ratings provided for each trial on a 1-10 scale were averaged. Higher scores indicate greater confidence. To calculate confidence-accuracy discrepancies (interoceptive insight), we first standardised the confidence scale from a 1-10 to a 0-10 scale and converted to a percentage (((confidence rating – 1)/9)*100) to enable comparison with the HCT as scored using the standard method. Then, for each trial, an absolute difference score was taken between participants' accuracy on the HCT for that trial as a percentage and participants' confidence [ABS (Objective Accuracy – Confidence)]. These scores were then averaged across all trials with higher scores indicating a greater discrepancy between accuracy and confidence (poor insight). As the alternative scoring method includes negative values it was not possible to calculate confidence-accuracy discrepancies. Instead, like Garfinkel et al., (2015) the within-participant Pearson correlation between confidence and accuracy served as a measure of interoceptive insight. For this measure, a high score represents good insight.

6. 1.6 Results

In the full sample, a significant correlation was observed between the IAS and the HCT both when over estimators were removed and the standard scoring scheme was utilised ($r(52) = .271$, p=.047), and when over estimators were retained using the alternative scoring scheme ($r(57) = .266$, p=.042). Additional analyses with successively more stringent data quality checks revealed increasing effect sizes (see supplementary materials [S2]). Furthermore, regression analyses controlling for age-group, gender, BMI, TET performance, HRV and resting heartrate had little influence on this relationship; for both the standard and the alternative scoring schemes the IAS remained a significant predictor of HCT performance (both $p<.015$; for details see supplementary materials [S2]).

Confidence ratings from the whole sample were not correlated with the IAS ($r(57) = -.026$, $p>.250$) but confidence ratings were significantly correlated with HCT performance using both scoring systems (standard: $r(52) = .507$, $p<.001$; alternative $r(57) = .454$, $p<.001$). Additional analyses with successively more stringent data quality checks did not change the pattern of results and regression analyses controlling for various confounds did not change the pattern of results (see supplementary materials [S2]).

Using the standard scoring method, interoceptive insight was correlated with HCT confidence ($r(52) = -.424$, $p<.001$), whereby higher confidence was related to lower discrepancies in interoceptive insight, but not with the IAS ($r(52) = .090$, $p>.250$) or HCT accuracy ($r(52) = -.171$, $p>.20$). Using the alternative scoring method for insight and HCT accuracy, insight was uncorrelated with confidence, the IAS or HCT accuracy (all $p>.250$). Additional data controls did not change the pattern of results (see supplementary materials [S2]).

Finally, in the total sample the IAS was negatively correlated with depression scores ($r(57) = -.295$, $p=.023$) but not anxiety ($r(57) = -.186$, $p>.15$). Neither depression, nor anxiety, was correlated with HCT performance (accuracy, confidence, or insight) using the standard or alternative scoring systems (all $p>.05$).

7. Study 5: Further evidence of the relationship with objective interoceptive accuracy

Although preliminary evidence of the relationship between the IAS and interoceptive accuracy was provided in Study 4 it is important to acknowledge that the full range of control variables advocated by Murphy et al., (2018b) were not included. The TET counting intervals were also not adequately matched to the HCT time intervals and these tasks were not counterbalanced. Moreover, as only the relationship between the IAS and the HCT was examined, it is not possible to conclude that only measures of accuracy, and not awareness, are predictive of performance. Study 5 therefore sought to replicate the association between the IAS and HCT including the full range of control variables, and including additional questionnaires (the ICQ and BPQ), to determine whether the relationship between the IAS and performance on the HCT was specific to measures of accuracy.

7.1 Method

7.1.1 Participants

40 participants were recruited via pre-existing databases and local advertisements. Participants were recruited on the basis that they had no current psychiatric diagnoses and had English as their first language (this was deemed crucial for the present study given that objective accuracy has been found to differ as a function of diagnostic status (see Khalsa & Lapidus, 2016; Murphy et al., 2017a). Participants were given a small honorarium to cover their travel expenses. Five participants' data were excluded (four due to reporting a posteriori

not meeting one of the recruitment criteria; one due to equipment failure). This resulted in 35 valid cases (M_{age} = 28.51, SD_{age} = 9.84, Range 20-56, 25 females, 10 males).

7.1.2 Questionnaire measures

Approximately 24 hours before the study, participants completed a number of questionnaires online via Qualtrics in a randomised order. These questionnaires included the TAS-20, BPQ, ICQ and IAS, and the DASS-21.

7.1.3 Interoceptive accuracy and confidence

As before, the HCT and TET were used to quantify interoceptive accuracy and as a control measure, respectively. Objective heartbeat was measured using a pulse oximeter. To ensure these tasks were well matched, each participant completed both the HCT and the TET over four durations with two duration sets employed (either 25, 35, 45, 100 seconds or 28, 38, 48, 103 seconds). Duration set (e.g., 25 vs. 28 seconds) was counterbalanced across the TET and HCT. Additionally, the order of individual durations was counterbalanced according to a Latin-square across participants. No significant differences were observed as a function of task order, for the HCT ($t(33) = .665$, $p>.250$) or the TET ($t(33) = -.557$, $p>.250$).

Throughout the task, participants were seated with hands on the table and feet flat on the floor. Participants were instructed to silently count their heartbeats from when the experimenter said “start” until they heard a beep, at which point they should report the number counted. Participants were instructed not to physically monitor their heartbeat and only count heartbeats they felt (not to count seconds or guess). They were explicitly informed that if they felt no heartbeats they should report zero as their answer. Participants were given two minutes to practice prior to the first HCT trial, and no feedback was provided. The TET was identical to the HCT except participants were asked to count seconds instead of heartbeats.

As in Study 4, confidence ratings were taken after each trial. In contrast to Study 4, a 0-10 rather than 1-10 scale was used for ease of comparison with the HCT and TET data. Additionally, given the possibility that there may have been confusion regarding confidence ratings (e.g., it is possible that participants misinterpreted the instructions and reported complete confidence in feeling no heartbeats, for example, rather than their confidence that they counted the correct amount of heartbeats), participants were explicitly informed that an answer of 0 would indicate low confidence in the accuracy of their answer, whereas an answer of 10 would indicate complete confidence. As before, the inclusion of confidence allows interoceptive insight to be calculated (see section 7.1.5).

7.1.4 Additional control measures

As the HCT has been found to be influenced by various physiological and psychological factors, a number of controls were recorded during/post task including resting heart rate (HR), HR variability, accuracy of participants' beliefs about the average resting HR (high scores indicate greater error), systolic blood pressure and BMI. For details see Murphy et al., (2017b); Murphy et al., (2018b); supplementary materials [S3].

7.1.5 Scoring and data analysis

Scoring of the HCT and TET data was identical to Study 4. However, as no participants in this dataset overestimated 2x the number of actual heartbeats, and when no overestimation occurs the scoring methods yield highly similar results (Study 4: r = .917), for brevity only scores using the standard scoring system are reported. Scoring of both confidence ratings and interoceptive insight was in accordance with Study 4, though confidence ratings were made on a 0-10 scale and only the standard scoring system was utilised. For scoring of additional control measures (e.g., beliefs, heartrate parameters) please see Murphy et al., (2017b); supplementary materials [S3].

7.2 Results

7.2.1 Relationship between questionnaire measures

Replicating the results of Study 2 and Study 3, a moderate relationship was observed between the IAS and ICQ ($r(33) = -.574$, $p<.001$) whilst the BPQ did not correlate with either the IAS ($r(33) = .091$, $p>.250$) or ICQ ($r(33) = .291$, $p=.09$). The correlation between the IAS and ICQ was significantly larger than the correlation between the IAS and BPQ and the correlation between the ICQ and BPQ (both $p<.05$). Similarly, the TAS-20 correlated with both the IAS ($r(33) = -.572$, $p<.001$) and the ICQ ($r(33) = .648$, $p<.001$) but not the BPQ ($r(33) = .067$, $p>.250$; Table 2). The correlation between the TAS-20 and IAS was significantly larger than the correlation between the BPQ and TAS-20 ($Z = -3.042$, $p=.002$). Likewise, the correlation between the ICQ and TAS-20 was significantly larger than the correlation between the BPQ and TAS-20 ($Z = 3.303$, $p<.001$). To determine whether the relationship between the TAS-20 and self-reported interoceptive accuracy (IAS and ICQ) would remain after controlling for anxiety and depression, a series of entry method regression analyses was conducted with either the IAS, ICQ or BPQ score as the dependent variable, and age (years), gender (0 = female, 1 = male), depression, anxiety, and alexithymia as predictor variables. Results from these analyses suggested that IAS scores were predicted by age ($b = -.359$, $t = -2.510$, $p=.018$) and alexithymia ($b = -.606$, $t = -3.445$, $p=.002$) only. All other predictors were non-significant (all $p>.250$). ICQ scores were predicted by both depression ($b = .374$, $t = -2.512$, $p=.018$) and alexithymia ($b = .367$, $t = 2.518$, $p=.018$). All other predictors were non-significant (all ps$>.190$). In contrast, BPQ scores were predicted only by anxiety ($b = .588$, $t = 2.424$, $p=.022$); all other predictors were non-significant (all *p*s>.150). These regression analyses were confirmed using robust regressions conducted in Matlab with the default tuning function employed. The same pattern of significance was obtained for all regression analyses, except for BPQ scores; when robust statistics were utilised none of the examined factors predicted BPQ scores.

7.2.3 Relationship with objective interoceptive accuracy and interoceptive insight

HCT performance was highly correlated with confidence ratings ($r(33) = .806$, $p<.001$) (Figure 3, left), even when participants who felt zero heartbeats for every trial of the HCT ($r(28) = .681$, $p<.001$), or zero heartbeats for at least one trial of the HCT ($r(23) = .572$, $p=.003$), were excluded. In contrast, time estimation scores were not predicted by time confidence ratings ($r(33) = .011$, $p>.250$). Confidence in time estimation was not correlated with confidence in HCT performance ($r(33) = .149$, $p>.250$), though a trend emerged when participants who felt zero heartbeats were removed ($r(23) = .361$, $p=.076$). Interoceptive insight (confidence-accuracy discrepancy; high scores indicate a greater discrepancy between accuracy and confidence) for the HCT was correlated with HCT accuracy ($r(33) = .580$, $p<.001$), suggesting worse interoceptive insight with respect to interoceptive accuracy was associated with better HCT performance, but HCT confidence ratings were not correlated with interoceptive insight ($r(33) = .050$, $p>.250$). In contrast, for time estimation, insight of time estimation accuracy was only correlated with confidence ($r(33) = -.644$, $p<.001$; see Table 2).

In terms of the relationship between questionnaire measures (IAS, BPQ, ICQ and TAS-20) and HCT performance, HCT performance was correlated with the IAS ($r(33) = .336$, $p=.049$; Figure 3) but not with any other questionnaire (all *p*s>.13). In contrast, time perception scores were not correlated with any questionnaire (all *p*s>.12). Confidence ratings for the HCT were correlated with the IAS ($r(33) = .409$, $p=.015$; Figure 3), but not with any other questionnaire (all *p*s>.09). No questionnaire was associated with interoceptive insight (all *p*s>.250). For time estimation, no questionnaire was associated with confidence ratings (all *p*>.250), or insight (all *p*>.250) (Table 2). Controlling for additional confounds had little effect on the results obtained (see supplementary materials [S4]).

8. Study 6: Confirmatory factor analysis

8.1 Method & Results

To confirm the 2-factor structure of the IAS, confirmatory factor analysis (CFA) was conducted on all available data. For comparison, CFA was also conducted on the available BPQ data (see below). Duplicate participants (based on IP addresses) were removed so that only the first completion of the questionnaire was retained, and individuals who contributed data for the initial factor analysis were also removed. This resulted in 375 unique cases (M_{age} = 35.23, SD_{age} = 16.89, Age Range = 18-91, Males =107, Females = 263, Other = 5, 309 English first language).

CFA was conducted in R (Team, 2013) using the Laavan package (Rosseel, 2012). Given that the data for all IAS items (all *D*(375) >.190, all *ps*<.001) were skewed, robust CFA was utilised. To evaluate the fit of the model, the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) are reported. Whilst an exact cut-off for the RMSEA has not been agreed, it is suggested that a cut-off value between 0.06-0.10 indicates acceptable fit (see Hooper, Coughlan, & Mullen, 2008). Likewise, clear cut-offs for the CFI and TLI have not been decided, though values >.90-.95 are generally considered to be indicators of good fit (Hooper et al., 2008). Results from the CFA indicated the two-factor solution provided an acceptable, but imperfect, explanation of the data (RMSEA = .070 [.062, .079], CFI = .865, TLI = .849). A comparable analysis on a 1-factor solution yielded similar results (RMSEA = .084 [.076, .092], CFI = .807, TLI = .786), with the 2-factor solution slightly superior.

The same analysis was conducted on the available BPQ data to provide a comparison with the IAS data (N = 350, M_{age} = 36.29, SD_{age} = 17.31, Age Range, 18-91, Males =115, Females = 230, Other = 5, 307 English first language). As before, duplicate values (based on IP addresses) were removed. Likewise, due to substantial skew (all $D(350) = >.20$; all $ps<.001$), robust CFA was utilised. Although a factor solution for the long BPQ body awareness scale has not been published, body awareness for the short version of the BPQ (26 items) has been reported to be described by a single factor (Cabrera et al., 2018) and can be computed by retaining specific items from the long form used in the present studies (Kolacz, Holmes & Porges, 2018). Therefore, CFA in the current study was utilised on the items that comprise the short BPQ. Results from the CFA indicated that the 1-factor solution was an acceptable, though imperfect explanation of these data (RMSEA = .085 [.078, .092], CFI = .877, TLI = .866).

9. General discussion

The aim of this set of studies was to develop and validate a new measure of self-reported trait interoceptive accuracy and evaluate the distinction between interoceptive accuracy and awareness. Results from these studies indicate that the interoceptive accuracy scale (IAS) has good internal consistency (Study 1), and its reliability is comparable to that of existing interoception questionnaires (BPQ and ICQ; Study 2). However, whilst the initial factor analysis indicated a 2-factor solution, the reliability of this factor structure requires further scrutiny (Study 6). Nevertheless, using total IAS scores, results support the proposals by Murphy et al., (2018a); as predicted, scores on both trait interoceptive accuracy questionnaires (the IAS and ICQ) were highly correlated, whilst neither of the interoceptive accuracy questionnaires was associated with scores on the interoceptive awareness questionnaire (BPQ; Study 2). This pattern was replicated in Studies 3 and 5. Furthermore, increasing alexithymia was associated with worse self-reported trait interoceptive accuracy, but not trait awareness, a pattern which remained after controlling for self-esteem (Study 3). This was replicated in Study 5, with additional analyses confirming that the relationship between poor self-reported trait interoceptive accuracy (ICQ and IAS) and alexithymia was independent of depression and anxiety. As predicted, preliminary data from Study 4 confirmed that objectively-measured interoceptive accuracy was predicted by self-reported interoceptive accuracy (both trait and moment-by-moment judgements). This was replicated and extended in Study 5 where it was found that objectively measured interoceptive accuracy was only predicted by self-reported interoceptive accuracy (both trait and moment-by-moment judgements), and not by self-reported trait interoceptive awareness. Furthermore, interoceptive insight concerning one's interoceptive accuracy was found to be independent of self-reported trait interoceptive accuracy (Study 5). Finally, moment-by-moment judgements of cardiac interoceptive accuracy were predicted by self-reported trait interoceptive accuracy, but not by self-reported trait interoceptive awareness (Study 5).

The results from this series of studies indicate that the IAS may be a useful tool for assessing self-reported trait interoceptive accuracy across a number of interoceptive domains. Initial validation attests to both the reliability (internal, test-retest) and convergent validity of the IAS, which were at least comparable to the other interoception questionnaires examined in this study (measured here, and in previous studies e.g., Brewer et al., 2016; Ferentzi, Drew, Tihanyi, & Köteles, 2018). The initial exploratory PCA of the IAS indicated a 2-factor solution, possibly reflecting a distinction between solely interoceptive signals and interoceptive signals that may be difficult to perceive using interoceptive information alone, or perturbations of bodily functions that are socially unacceptable. The results of the confirmatory factor analysis, however, indicated that this 2-factor solution was an imperfect explanation of the data. As such, caution is required when using this factor solution, though it is notable that a comparably imperfect fit was found for the most popular measure of interoceptive awareness (the BPQ). These preliminary data therefore suggest that further cross-validation studies are required to confirm the factor structure of self-report measures of interoception, both in typical and clinical populations.

In terms of the relationships between self-report measures, as predicted, no relationship was observed between the self-report measure of interoceptive awareness (BPQ) and the self-report measures of accuracy (IAS and ICQ), though moderate correlations were observed between the two interoceptive accuracy questionnaires, a result that replicated across studies. Likewise, self-reported confidence in one's accuracy during the HCT (Study 5) was only correlated with the IAS and not the BPQ, attesting to the conceptualisation of the IAS as a measure of accuracy, not awareness. Although self-reported accuracy was not correlated with the IAS in Study 4, this discrepancy is likely due to confusion regarding the instructions (see section 7.1.3). Taken together, these results are consistent with proposals by Murphy et al. (2018a) that self-reported interoceptive awareness and accuracy may dissociate. These results highlight the need for caution when interpreting the results of questionnaire measures of interoception that include items assessing both accuracy and awareness (e.g., the Body Awareness Questionnaire or the Interoceptive Accuracy Questionnaire; Shields et al., 1989; Van den Bergh et al., in prep).

Of interest is the relationship between the self-report measures and objectively measured interoceptive accuracy reported in Studies 4 and 5. Although from small samples, the preliminary data provided here support the proposal by Murphy et al. (2018a) that subjective and objective measures of interoception are more likely to be correlated if both measures are assessing interoceptive accuracy or awareness, than if one measure assesses awareness and the other accuracy (Study 5). Indeed, the lack of a relationship between scores on the BPQ and performance on the HCT is consistent with previous reports noting the absence of a relationship between objectively measured interoceptive accuracy and the BPQ (e.g., Critchley et al., 2004; Garfinkel et al., 2015; Ferentzi et al., 2018), and the proposed distinction between interoceptive accuracy and awareness speaks to discrepancies in the literature regarding the relationship between objectively measured interoceptive accuracy and other questionnaires that conflate accuracy and awareness (e.g., Chentsova-Dutton & Dzokoto, 2014; Zamariola, et al., 2018; Emanuelsen & Drew, 2015).

It should be noted, however, that subjective and objective measures of interoception may be differentially related depending on the specific interoceptive domain being tested; it is currently unclear whether individual differences in one interoceptive domain (e.g., respiratory) are associated with individual differences in another (e.g., cardiac; Garfinkel et al., 2016; Herbert, Muth, Pollatos, & Herbert, 2012; Steptoe & Vögele, 1992; Ferentzi, et al., 2018; see Khalsa et al., 2017 for a discussion). Likewise, within interoceptive domains (e.g., cardiac) it remains a possibility that differential associations may be obtained across different tasks of interoceptive accuracy (e.g., Garfinkel et al., 2015; Schulz et al., 2013; Forkmann et al., 2016). In particular, the measure of objective interoceptive accuracy employed in the present study, the HCT (Dale & Anderson, 1978; Schandry, 1981), has recently come under intense scrutiny (e.g., Khalsa et al., 2009; Murphy et al., 2018b), with a body of research indicating that performance may be a product of knowledge about heartrate rather than the ability to perceive heartbeats (e.g., Ring et al., 2015). In both Studies 4 and 5, the relationship between the IAS and HCT performance remained after the inclusion of multiple control variables (see supplementary materials), including estimated heartrate (Study 5), meaning the IAS-HCT relationship is unlikely to be driven by this factor. It is clear that further research is required to understand the processes indexed by the heartbeat counting task however (see

Murphy et al., 2018b), and the relationship between objective and subjective measures of interoception.

In conclusion, the results of this set of studies support a distinction between self-reported interoceptive accuracy and self-reported interoceptive awareness. They suggest that objective and trait measures of self-reported interoception may be related if quantified along the same dimension of accuracy and awareness. These data also speak to the possible utility of the IAS as a self-report measure of trait interoceptive accuracy.

References

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale--I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. doi:10.1016/0022-3999(94)90005-1

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

Betka, S., Pfeifer, G., Garfinkel, S., Prins, H., Bond, R., Sequeira, H., ... & Critchley, H. (2018). How Do Self-Assessment of Alexithymia and Sensitivity to Bodily Sensations Relate to Alcohol Consumption?. *Alcoholism: Clinical and Experimental Research*, *42*(1), 81-88.

Brener, J., & Ring, C. (2016). Towards a psychophysics of interoceptive processes: the measurement of heartbeat detection. *Phil. Trans. R. Soc. B*, *371*(1708), 20160015. DOI: 10.1098/rstb.2016.0015.

Brewer, R., Cook, R., & Bird, G. (2016). Alexithymia: a general deficit of interoception. *Royal Society Open Science*, *3*(10), 150664. doi:10.1098/rsos.150664

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. doi:10.1002/mpr.1596

Ceunen, E., Van Diest, I., & Vlaeyen, J. (2013). Accuracy and awareness of perception: related, yet distinct (commentary on Herbert et al., 2012). *Biological psychology*, *92*(2), 423-427. 10.1016/j.biopsycho.2012.09.012

Chentsova-Dutton, Y. E., & Dzokoto, V. (2014). Listen to your heart: the cultural shaping of interoceptive awareness and accuracy. *Emotion, 14*(4), 666–678. http://dx.doi.org/10.1037/a0036193

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. doi:10.1038/nrn894

Critchley, H. D., & Harrison, N. A. (2013). Visceral influences on brain and behavior. *Neuron*, *77*(4), 624–638. doi:10.1016/j.neuron.2013.02.008

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.

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

Damasio, A. R. (1994). *Descartes' error: Emotion, rationality and the human brain.* New York: Putnam

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

Emanuelsen, L., Drew, R., & Köteles, F. (2015). Interoceptive sensitivity, body image dissatisfaction, and body awareness in healthy individuals. *Scandinavian journal of psychology*, *56*(2), 167-174. https://doi.org/10.1111/sjop.12183

Ernst, J., Böker, H., Hättenschwiler, J., Schüpbach, D., Northoff, G., Seifritz, E., & Grimm, S. (2014). The association of interoceptive awareness and alexithymia with neurotransmitter concentrations in insula and anterior cingulate. *Social Cognitive and Affective Neuroscience*, *9*(6), 857–863. doi:10.1093/scan/nst058

Ferentzi, E., Bogdány, T., Szabolcs, Z., Csala, B., Horváth, Á., & Köteles, F. (2018). Multichannel investigation of interoception: Sensitivity is not a generalizable feature. *Frontiers in human neuroscience*, *12*, 223.

Ferentzi, E., Drew, R., Tihanyi, B. T., & Köteles, F. (2018). Interoceptive accuracy and body awareness - Temporal and longitudinal associations in a non-clinical sample. *Physiology & Behavior*, *184*, 100–107. doi:10.1016/j.physbeh.2017.11.015

Field, A. (2005). *Discovering statistics using SPSS*. Sage publications.

Fiene, L., Ireland, M. J., & Brownlow, C. (2018). The interoception sensory questionnaire (ISQ): A scale to measure interoceptive challenges in adults. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-018-3600-3

Forkmann, T., Scherer, A., Meessen, J., Michal, M., Schächinger, H., Vögele, C., & Schulz, A. (2016). Making sense of what you sense: Disentangling interoceptive awareness, sensibility and accuracy. *International Journal of Psychophysiology*, *109*, 71–80.

Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: new insights link internal physiology to social behaviour. Commentary on:: "Anterior insular cortex mediates bodily sensibility and social anxiety" by Terasawa et al. (2012). *Social Cognitive and Affective Neuroscience*, *8*(3), 231–234. doi:10.1093/scan/nss140

Garfinkel, S. N., Barrett, A. B., Minati, L., Dolan, R. J., Seth, A. K., & Critchley, H. D. (2013). What the heart forgets: Cardiac timing influences memory for words and is modulated by metacognition and interoceptive sensitivity. *Psychophysiology*, *50*(6), 505–512. doi:10.1111/psyp.12039

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). doi:10.1098/rstb.2016.0014

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. doi:10.1016/j.biopsycho.2014.11.004

Herbert, B. M., Herbert, C., & Pollatos, O. (2011). On the relationship between interoceptive awareness and alexithymia: is interoceptive awareness related to emotional awareness? *Journal of Personality*, *79*(5), 1149–1175. doi:10.1111/j.1467-6494.2011.00717.x

Herbert, B. M., Muth, E. R., Pollatos, O., & Herbert, C. (2012). Interoception across modalities: on the relationship between cardiac awareness and the sensitivity for gastric functions. *Plos One*, *7*(5), e36646. doi:10.1371/journal.pone.0036646

Heyes, C., & Bird, G. (2008). Mirroring, association, and the correspondence problem. *Sensorimot. Found. High. Cogn*, *22*(461), 10-1093.

Hooper, D., Coughlan, J., Mullen, M.: Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

Khalsa, S. S., & Lapidus, R. C. (2016). Can Interoception Improve the Pragmatic Search for Biomarkers in Psychiatry? *Frontiers in Psychiatry*, *7*, 121. doi:10.3389/fpsyt.2016.00121

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., ... & Meuret, A. E. (2017). Interoception and mental health: a roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*, 501-513.

Khalsa, S. S., Rudrauf, D., Damasio, A. R., Davidson, R. J., Lutz, A., & Tranel, D. (2008). Interoceptive awareness in experienced meditators. *Psychophysiology*, *45*(4), 671-677.

Khalsa, S. S., Rudrauf, D., Sandesara, C., Olshansky, B., & Tranel, D. (2009). Bolus isoproterenol infusions provide a reliable method for assessing interoceptive awareness. *International Journal of Psychophysiology*, *72*(1), 34-45.

Kolacz, J., Holmes, L., & Porges, S, W., (April, 2018). *Manual and Scoring Document for Body Perception Questionnaire (BPQ), Body Perception Questionnaire – Short Form (BPQ-SF), Body Perception Questionnaire – Very Short Form (BPQ-VSF)*. Retrieved from: http://stephenporges.com/images/bpqv2.pdf

Langer, N., Beeli, G., & Jäncke, L. (2010). When the sun prickles your nose: an EEG study identifying neural bases of photic sneezing. *Plos One*, *5*(2), e9208. doi:10.1371/journal.pone.0009208

Lee, I. A., & Preacher, K. J. (2013a, October). Calculation for the test of the difference between two dependent correlations with no variable in common [Computer software]. Available from http://quantpsy.org.

Lee, I. A., & Preacher, K. J. (2013b, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

Longarzo, M., D'Olimpio, F., Chiavazzo, A., Santangelo, G., Trojano, L., & Grossi, D. (2015). The relationships between interoception and alexithymic trait. The Self-Awareness Questionnaire in healthy subjects. *Frontiers in Psychology*, *6*, 1149. doi:10.3389/fpsyg.2015.01149

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343. doi:10.1016/0005-7967(94)00075-U

Mazzone, S. B., McLennan, L., McGovern, A. E., Egan, G. F., & Farrell, M. J. (2007). Representation of capsaicin-evoked urge-to-cough in the human brain using functional magnetic resonance imaging. *American Journal of Respiratory and Critical Care Medicine*, *176*(4), 327–332. doi:10.1164/rccm.200612-1856OC

Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The multidimensional assessment of interoceptive awareness (MAIA). *Plos One, 7*(11), e48230. doi:10.1371/journal.pone.0048230

Morelli, D., Bartoloni, L., Colombo, M., Plans, D., & Clifton, D. (2017). Profiling the Propagation of Error from PPG to HRV Features in a Wearable Physiological-Monitoring Device. *Healthcare Technology Letters*. https://doi.org/10.1049/htl.2017.0039

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

Murphy, J, Geary, H., Millgate, E., Catmur, C., & Bird, G. (2017b). Direct and indirect effects of age on interoceptive accuracy and awareness across the adult lifespan. *Psychonomic Bulletin & Review*, 1–10. doi:10.3758/s13423-017-1339-z

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

Murphy, J., Catmur, C., & Bird, G. (2018a). Alexithymia is associated with a multidomain, multidimensional failure of interoception: Evidence from novel tests. *Journal of Experimental Psychology: General*, *147*(3), 398-408. http://dx.doi.org/10.1037/xge0000366

Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*. Routledge.

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

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

Ring, C, Brener, J., Knapp, K., & Mailloux, J. (2015). Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: a cautionary tale about interoceptive awareness. *Biological Psychology*, *104*, 193–198. doi:10.1016/j.biopsycho.2014.12.010

Ring, C., & Brener, J. (1996). Influence of beliefs about heart rate and actual heart rate on heartbeat counting. *Psychophysiology*, *33*(5), 541–546. doi:10.1111/j.1469-8986.1996.tb02430.x

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton
S., Porges (1993) *Body perception questionnaire*: Laboratory of development assessment University of Maryland.

Schandry R, & Specht G. 1981. The influence of psychological and physical stress on the perception of heartbeats. Psychophysiology 18, 154.

Schandry, R. (1981). Heart Beat Perception and Emotional Experience. *Psychophysiology*, *18*(4), 483–488. doi:10.1111/j.1469-8986.1981.tb02486.x

Schulz, A., Lass-Hennemann, J., Sütterlin, S., Schächinger, H., & Vögele, C. (2013). Cold pressor stress induces opposite effects on cardioceptive accuracy dependent on assessment paradigm. *Biological Psychology, 93*(1), 167–174.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. doi:10.1016/j.tics.2013.09.007

Shah, P, Catmur, C., & Bird, G. (2017). From heart to mind: linking interoception, emotion, and theory of mind. *Cortex;* 220–223. doi: 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*, *81*, 215–220. doi:10.1016/j.cortex.2016.03.021

Shields, S. A., Mallory, M. E., & Simon, A. (1989). The body awareness questionnaire: reliability and validity. *Journal of personality Assessment*, *53*(4), 802-815. https://doi.org/10.1207/s15327752jpa5304_16

Sifneos, P. E. (1973). The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, *22*(2), 255–262. doi:10.1159/000286529

Sokol-Hessner, P., Hartley, C. A., Hamilton, J. R., & Phelps, E. A. (2015). Interoceptive ability predicts aversion to losses. *Cognition & Emotion*, *29*(4), 695–701. doi:10.1080/02699931.2014.925426

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251. doi:10.1037/0033-2909.87.2.245

Steptoe, A., & Vögele, C. (1992). Individual differences in the perception of bodily sensations: the role of trait anxiety and coping style. *Behaviour Research and Therapy*, *30*(6), 597–607. doi:10.1016/0005-7967(92)90005-2

Team, R. C. (2013). R: A language and environment for statistical computing.

Terasawa, Y., Moriguchi, Y., Tochizawa, S., & Umeda, S. (2014). Interoceptive sensitivity predicts sensitivity to the emotions of others. *Cognition & Emotion*, *28*(8), 1435–1448. doi:10.1080/02699931.2014.888988

Van den Bergh, O., Bogaerts, K., Walentynowicz, M., & Van Diest, I., (in preparation). The Interoceptive Awareness Questionnaire: Unraveling the distinction between awareness of neutral and negative bodily sensations.

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. doi: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. doi:10.1007/BF00998623

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, *8*(3).

Yves Rosseel (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Zamariola, G., Vlemincx, E., Luminet, O., & Corneille, O. (2018). Relationship between interoceptive accuracy, interoceptive sensibility, and alexithymia. *Personality and Individual Differences*, *125*, 14-20. https://doi.org/10.1016/j.paid.2017.12.024

Author Notes

The data pertaining to the relationship between alexithymia and heartbeat counting (Study 5) was previously included as part of a larger study reported in Murphy et al., (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.

JM was supported by a doctoral studentship from the Economic and Social Research Council [1599941; ES/J500057/1]. GB was supported by the Baily Thomas Trust. We gratefully acknowledge the help of Edward Millgate and Sharmin Khan in data collection.

Footnotes

[1] Note that our use of the term awareness differs from that of Garfinkel et al., (2015) who use the term in a metacognitive sense to refer to the degree to which beliefs about interoceptive ability map onto objective measures of ability. Our use of the term also differs from Khalsa et al., (2017) who use awareness to refer to all conscious aspects of interoception.

[2] Principal components analysis (PCA) rather than exploratory factor analysis (EFA) was utilised for examining the underlying structure of the IAS as Field (2005; page 638) argues that PCA is a psychometrically sound procedure which often yields similar results to a EFA. However, it is notable that a 2-factor solution for the IAS was observed when EFA was utilised.

Tables

Table 1. Rotated factor solution for the IAS.

	Factor 1	Factor 2
Urinate (Item 5)	.688	
Hungry (Item 2)	.624	
Defecate (Item 6)	.611	
Thirsty (Item 4)	.593	
Pain (Item 17)	.540	
Heart (Item 1)	.512	
Taste (Item 7)	.504	
Breathing (Item 3)	.502	
Temperature (Item 11)	.496	
Muscles (Item 15)	.466	
Affective touch (Item 19)	.448	
Vomit (Item 8)	.412	
Sexual arousal (Item 12)	.404	
Itch (Item 21)		.726
Tickle (Item 20)		.649
Cough (Item 10)		.631
Burp (Item 14)		.622
Bruise (Item 16)		.597
Blood Sugar (Item 18)		.563
Sneeze (Item 9)		.549
Wind (Item 13)		.512

Table 2. Correlations between questionnaire and HCT measures in Study 5

Measure	Mean	SD	1	2	3	4	5	6	7	8	9
1. BPQ	2.10	0.69	-								
2. IAS	82.86	8.82	.091	-							
3. ICQ	46.14	9.54	.291	-.574**	-						
4. TAS-20	44.89	15.83	.067	-.572**	.648**	-					
5. HCT ACC	191.27	113.91	.233	.336*	-.252	-.256	-				
6. HCT Confidence	35.61	24.86	.207	.409*	-.289	-.287	.806**	-			
7. HCT Insight	16.22	14.72	.098	.000	-.080	-.043	.580**	.050	-		
8. TET ACC	296.71	61.42	-.265	.174	-.164	.109	.042	-.095	.235	-	
9. TET Confidence	64.61	16.20	.008	-.018	.016	.075	-.191	.149	-.451**	.011	-
10. TET Insight	18.55	16.91	.064	-.102	.128	.106	.097	-.221	.420*	.260	-.644**

*Denotes significant at *p*<.05 **Denotes significant at *p*<.001

BPQ = Porges body perception questionnaire. IAS = Interoceptive accuracy scale. ICQ = Interoceptive confusion questionnaire. TAS-20 = Toronto Alexithymia Scale. HCT = Heartbeat counting task. TET = Time Estimation Task

Figure captions

Figure 2. Panel a) depicts the relationship between the three questionnaires for which test re-test reliability was assessed in Study 2. A moderate relationship was observed between the IAS and ICQ (panel a, left) whereby higher self-reported interoceptive accuracy was associated with reduced interoceptive confusion. In contrast, no relationship was observed between the BPQ and IAS (panel a, middle) or the BPQ and ICQ (panel a, right), suggesting no relationship between accuracy and awareness. Panel b) depicts the relationship between the questionnaires in Study 3. A negative relationship was observed between the TAS-20 and the IAS (panel b, left). Replicating the results from Study 2, no relationship between the IAS and BPQ was observed (panel b, middle). A negative relationship was observed between the TAS-20 and the IAS (panel b, middle) whereby increasing rates of alexithymia were associated with lower self-reported interoceptive accuracy. In contrast, no relationship between the TAS-20 and the BPQ was observed. IAS = interoceptive accuracy scale. BPQ = Porges body perception questionnaire. ICQ = Interoceptive confusion questionnaire. TAS-20 = Toronto Alexithymia Scale.

Figure 3. Panel a) The relationship between the IAS, accuracy and confidence ratings for the heartbeat tracking task as reported in Study 4 using the standard scoring system and removal of individuals who overestimated (see text). A significant relationship was found between the IAS and HCT (left) but not the IAS and HCT confidence ratings (middle). A moderate relationship was observed between HCT confidence and accuracy. Panel b) The relationship between the IAS, accuracy and confidence ratings for the heartbeat counting task as reported in Study 5. A significant relationship was found between the IAS and HCT accuracy (left) and the IAS and HCT confidence (middle). A strong relationship was also observed between HCT confidence and accuracy. IAS = Interoceptive accuracy scale. HCT ACC = Heartbeat counting task accuracy. HCT Confidence = average confidence ratings following the heartbeat counting task.

Figures

Figure 1. Interoceptive accuracy scale

Below are several statements regarding how accurately you can perceive specific bodily sensations. Please rate on the scale how well you believe you can perceive each specific signal. For example, if you often feel you need to urinate and then realise you do not need to when you go to the toilet you would rate your accuracy perceiving this bodily signal as low.

Please only rate how well you can perceive these signals without using external cues, for example, if you can only perceive how fast your heart is beating when you measure it by taking your pulse this would not count as accurate internal perception.

1. I can always accurately perceive when my heart is beating fast
2. I can always accurately perceive when I am hungry
3. I can always accurately perceive when I am breathing fast
4. I can always accurately perceive when I am thirsty
5. I can always accurately perceive when I need to urinate
6. I can always accurately perceive when I need to defecate
7. I can always accurately perceive when I encounter different tastes
8. I can always accurately perceive when I am going to vomit
9. I can always accurately perceive when I am going to sneeze
10. I can always accurately perceive when I am going to cough
11. I can always accurately perceive when I am hot/cold
12. I can always accurately perceive when I am sexually aroused
13. I can always accurately perceive when I am going to pass wind
14. I can always accurately perceive when I am going to burp
15. I can always accurately perceive when my muscles are tired/sore
16. I can always accurately perceive when I am going to get a bruise
17. I can always accurately perceive when I am in pain
18. I can always accurately perceive when my blood sugar is low
19. I can always accurately perceive when someone is touching me affectionately rather than non-affectionately
20. I can always accurately perceive when something is going to be ticklish
21. I can always accurately perceive when something is going to be itchy

Scale: Strongly Agree (5), Agree (4), Neither agree nor disagree (3), Disagree (2), Disagree Strongly (1).

Figure 2.

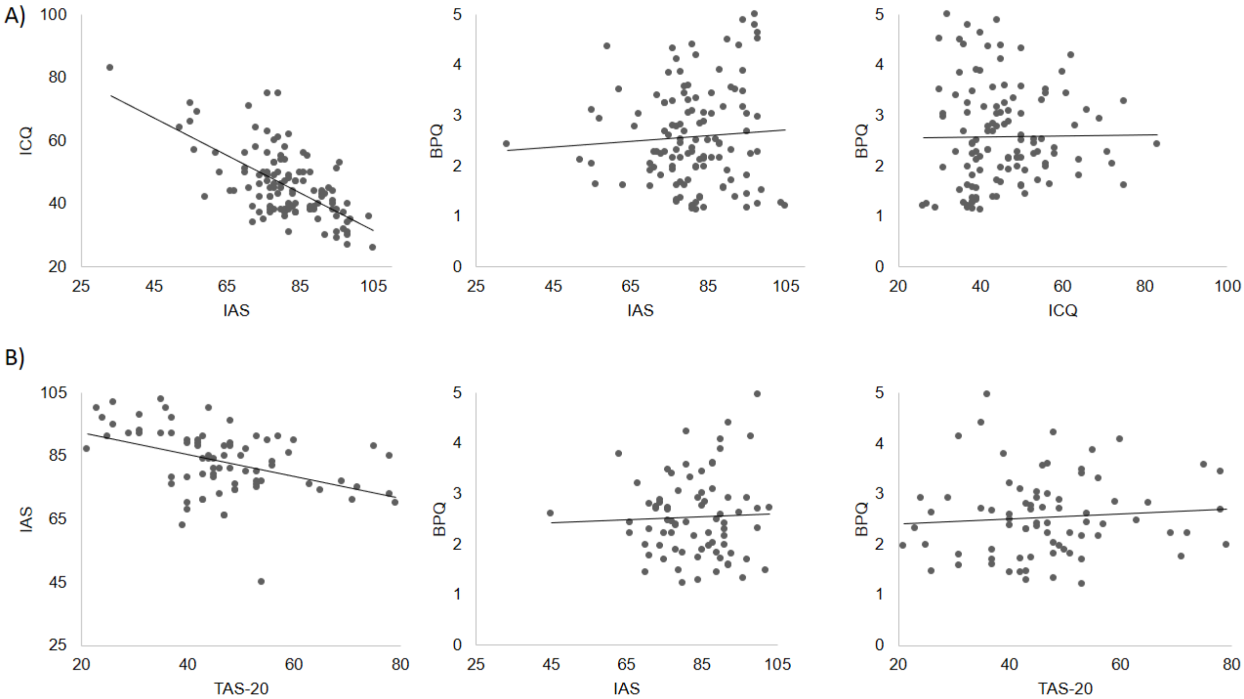
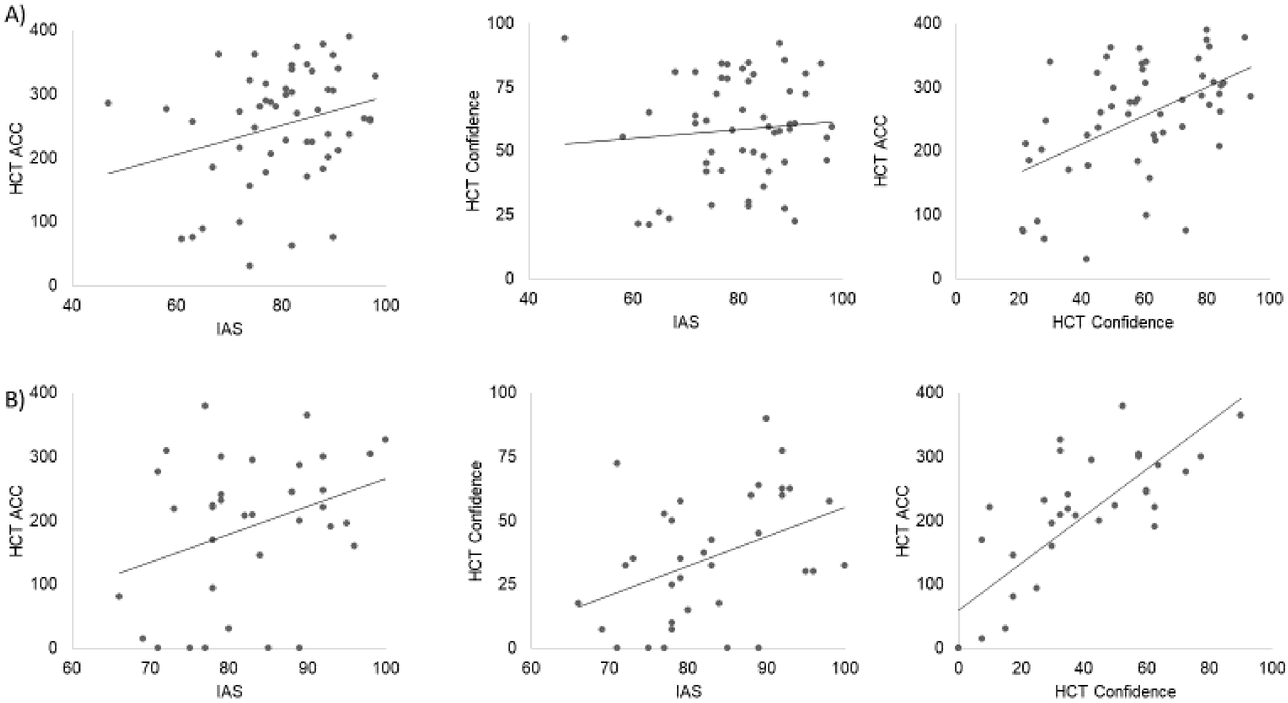


Figure 3


A)
B)
HCT ACC
HCT Confidence
IAS
HCT Confidence
0
100
200
300
400
25
50
75
100
40
60
80
100
0
20
70
90