Mega-analysis of the Interoceptive Accuracy Scale (IAS) Structure and its Subjective Correlates

Ana Neves[1], Magdalena Pfaff[1], and Dominique Makowski[1,2]
[1]School of Psychology, University of Sussex
[2]Sussex Centre for Consciousness Science, University of Sussex

Blabla the abstract blabla.

Keywords: keyword1, keyword2, keyword3

Introduction

Interoception is referred to the process of sensing, interpreting and integrating information pertaining to internal organs, such as the heart, the lungs or the gut (Khalsa et al., 2018). While recent research emphasizes a key role of interoception in a variety of processes (e.g., emotion regulation, decision making) and of outcomes (physical and psychological well being), the field remains clouded by concerns about how interoception is assessed.

The Interoceptive Assessment Puzzle

Various measures of interoception have been developed (see Figure 1), forming a combination of "objective" and "subjective" assessments (i.e., physiological tasks such as the heart beat counting or tracking *vs.* questionnaires and subjective scales involving metacognitive judgments), "explicit" and "implicit" paradigms (i.e., directing participants' awareness and attention to interoceptive processes *vs.* measuring interoception unbeknownst to them), various interoceptive modalities (e.g., cardioception, respiroception, gastroception) and theoretical dimensions (e.g., accuracy, sensitivity, awareness). While there is no consensus as to which particular approach provides the most accurate and "pure" measure of interoception and interoceptive abilities (assuming it is a unidimensional construct), it is instead plausible that each measure has strengths and limitations, and a utility dependent on the context and goal at hand (Desmedt et al., 2023; Jahedi & Méndez, 2014).

Although the use of subjective self-report questionnaires to measure deeply embodied functions might seem paradoxical at first, recent redefinitions of interoception emphasize the role of high-level and metacognitive elaboration of interoceptive information. These redefinitions provide theoretical grounding to support the idea that some facets of interoception, including participants' metacognitive beliefs, can be assessed subjectively (Khalsa et al., 2018; Suksasilp & Garfinkel, 2022). Moreover, the notion that self-reports might not reflect the same processes as other interoception tasks might be important to contextualize the apparent lack of convergence between measures in the field (Desmedt et al., 2022). For instance, existing findings typically show weak or no correlations between questionnaires and objective measures, such as the Heartbeat Counting Task (HCT, Schandry, 1981) and the Heartbeat Detection Task (HDT, Kleckner et al., 2015), including for measures of the same theoretical dimensions (Arslanova et al., 2022; Brand et al., 2023; e.g., task-based accuracy vs. self-reported accuracy, Murphy et al., 2019). Additionally, even various objective measures assessing in theory the same interoceptive dimension, such as accuracy, either show no or weak correlation (respectively, Brand et al., 2023; Hickman et al., 2020). Perhaps more surprisingly, low correlations have been observed even among questionnaires, suggesting (in parallel to major validity concerns)

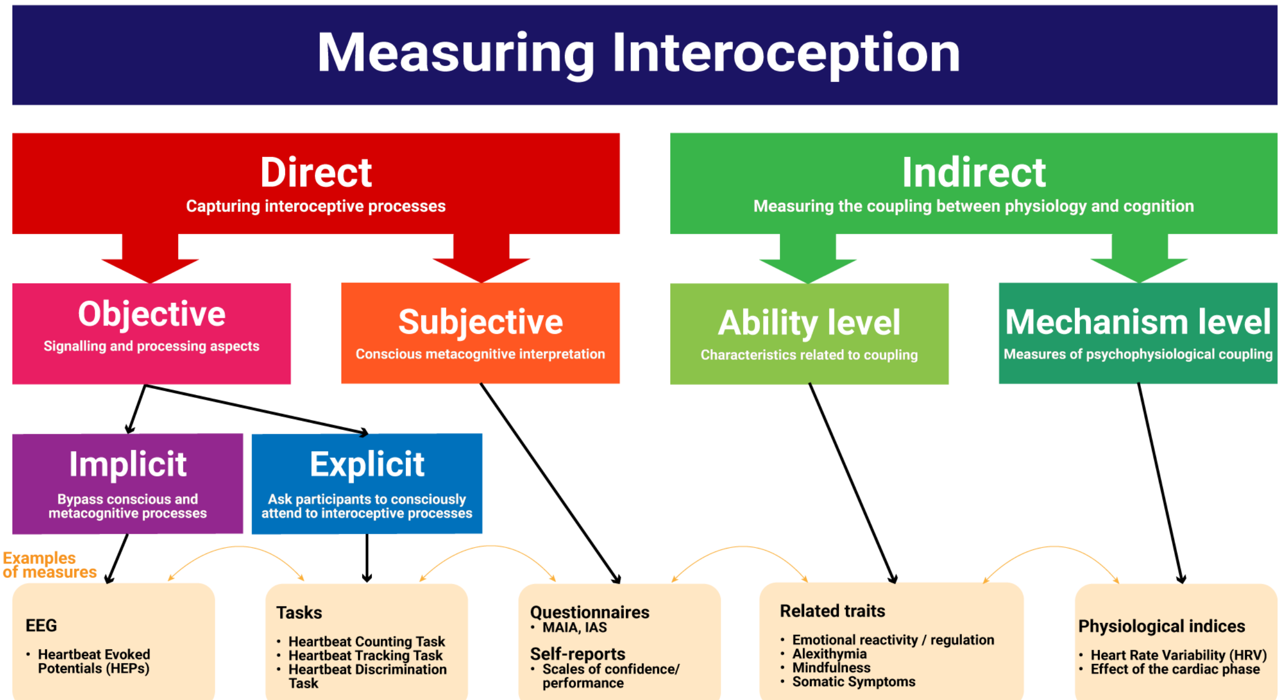
Ana Neves
Magdalena Pfaff
Dominique Makowski

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Ana Neves: Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing; Magdalena Pfaff: Data curation, Writing – original draft; Dominique Makowski: Project administration, Data curation, Formal Analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing

Correspondence concerning this article should be addressed to Dominique Makowski, Email: D.Makowski@sussex.ac.uk

Figure 1

The Interoceptive Assessment Puzzle. The different modalities of interoception (e.g., cardioception) can be assessed directly or indirectly. Direct assessments can further be subjective or objective, depending on whether they involve conscious metacognitive appraisals or more performance-based indices. Interoceptive tasks can be explicit (the participant is aware of the interoceptive nature of the task and must consciously attend to interoceptive signals; e.g., the heartbeat counting task) or implicit (measurements of interoception done unbeknowst to the participant; e.g., heartbeat evoked potentials measured during resting state). Indirect assessments evaluate constructs typically related (and ideally dependent on) to interoceptive processes or ability (or its deficit).



the potential targeting of different facets related to interoception.

One striking example concerns the assessment of interoceptive sensibility, which is broadly defined as the self-reported tendency to focus on and detect internal sensations (Garfinkel et al., 2015), but more narrowly as the subjective tendency to focus on interoceptive signals, without necessarily implying detection ability (Khalsa et al., 2018). A recent systematic review suggested that various questionnaires designed to assess interoceptive sensibility may, in fact, measure distinct constructs, with the risk of researchers treating them as equivalent despite overall low convergence (Desmedt et al., 2022). Notably, this review adopted a broad definition of sensibility, incorporating both interoceptive sensibility and interoceptive self-report scales, following the eight-facet model by Khalsa et al. (2018). Several widely used questionnaires were included in the review, such as the Multidimensional Assessment of Interoceptive Awareness (MAIA, Mehling et al., 2012; MAIA-2, Mehling et al., 2018), the Body Perception Questionnaire (BPQ, Porges, 1993), the Private subscale of the Body Consciousness Questionnaire (PBCS, Miller et al., 1981), the Body Awareness Questionnaire (BAQ, Shields et al., 1989), and the Eating Disorder Inventory (Garner et al., 1983; EDI, Garner, 1991). The lack of correlations to moderate correlations among these questionnaires highlight the need for greater conceptual clarity regarding what each measure captures, how they relate to different dimensions of interoception, and their potential overlaps with other constructs, such as alexithymia and body awareness.

The Interoceptive Accuracy Scale (IAS)

Focusing on another dimension of interoception, a recently developed scale with a rapidly growing popularity is the Interoceptive Accuracy Scale (IAS, Murphy et al., 2019). The IAS consists of 21 Likert-scale items that query how accurately one can perceive different bodily signals, with one item per physiological modality such as respiration (*"I can always accurately perceive when I am breathing fast"*), heart

(*"I can always accurately perceive when my heart is beating fast"*), skin (*"I can always accurately perceive when something is going to be ticklish"*), arousal or bodily functions like coughing (*"I can always accurately perceive when I am going to cough"*) or urinating (*"I can always accurately perceive when I need to urinate"*). Appealingly, the IAS' statements are about specific interoceptive behaviours, which is a distinct difference with other popular interoception questionnaires, such as the MAIA-2, which contains more general and metacognitive items (e.g., *"I trust my body sensations"*, *"I can notice an unpleasant body sensation without worrying about it"*), as well as dimensions related to attention regulation (e.g., Not-distracting) or emotion regulation (e.g., Not-worrying).

The original validation study suggested a two-factor structure for the IAS, one reflecting the perception of general interoceptive signals (urinate, hungry, defecate, thirsty, pain, heart, taste, breathing, temperature, muscles, affective touch, vomit, sexual arousal), and other relating to signals that may be difficult to perceive solely through interoceptive information (itch, tickle, cough, burp, bruise, blood sugar, sneeze, wind). The authors however underlined its acceptable but imperfect fit (Murphy et al., 2019, p. 127), and several follow-up studies have indeed identified different optimal solutions. For instance, Brand et al. (2023) reported a 1-factor solution, while Lin et al. (2023) - using Exploratory Graph Analysis (EGA, H. F. Golino & Epskamp, 2017) - and Campos et al. (2021) found bifactor solutions (i.e., one general factor above a set of lower-level factors, Rodriguez et al., 2016) to be the best fit. Using a 2-factors Exploratory Factor Analysis (EFA), Koike and Nomura (2023) suggested that the items could be grouped into cutaneous (itching, tickling, coughing, burping, affective touch, bruising, passing gas, sneezing, muscle sensations, sexual arousal, and taste) and visceral sensations (urination, defecation, hunger, thirst, pain, breathing, fatigue/blood sugar, temperature, vomiting, and heartbeat).

Discussions have also been focused on specific items. For instance, Murphy et al. (2019) notes that some items might measure direct interoceptive signals such as cardioception, while others might capture phenomena not perceivable through interoceptive signals alone (e.g., "bruising"; p. 119). Lin et al. (2023) additionally highlights five locally dependent pairs and three items (touch, blood sugar, bruise) with exceptionally high difficulty and low discrimination, and Campos et al. (2021) reported "tickle" to be the only item that reflected more specific factors than the general factor. Furthermore, localization issues also arose, with both "itch" and "tickle" corresponding to the same Chinese character, leading to their collapse into a single item (Lin et al., 2023).

Regarding its validity, the IAS has naturally been compared to other interoception-related measures, and shows a positive correlations with most facets of the MAIA (Mehling et al., 2018), except for the Not-Distracting and Not-Worrying subscales (Brand et al., 2023) - which were previously highlighted as related to non-interoceptive abilities (Ferentzi et al., 2021). Interestingly, findings on the correlation between the IAS and the body awareness dimension of the BPQ (i.e., BPQ-A) have been mixed: some studies report small positive correlations (Brand et al., 2023; Campos et al., 2021; Koike & Nomura, 2023), while others find small negative correlations (Lin et al., 2023) or no correlation at all (Murphy et al., 2019). Small positive correlations have also been observed with the "observation" and "description" subscales of the Five Facet Mindfulness Questionnaire (FFMQ, Baer et al., 2006; Brand et al., 2023; Koike & Nomura, 2023), as well as with the "non-reactivity" and "acting with awareness" subscales (Koike & Nomura, 2023). Additionally, the IAS has shown a positive correlation with the interoceptive awareness subscale of the EDI [**ANA Write full name**] (EDI, Lin et al., 2023) and a negative correlation with the Interoceptive Confusion Questionnaire (ICQ, Brewer et al., 2016), as reported by Brand et al. (2023) and Murphy et al. (2019). Lastly, the correlation with the Interoceptive Attention Scale (IATS, Gabriele et al., 2022) [**ANA is that the ref for the scale validation?**] appears rather small (Koike & Nomura, 2023; Lin et al., 2023).

While assessing the predictive validity of an interoception scale can be conceived as theoretically challenging, expected negative associations were observed between the IAS and alexithymia (Brand et al., 2023; Campos et al., 2021; Koike & Nomura, 2023; Lin et al., 2023; Murphy et al., 2019), somatic symptoms (Brand et al., 2023; Koike & Nomura, 2023; Lin et al., 2023), depressive symptoms (Brand et al., 2023; Koike & Nomura, 2023; Lin et al., 2023), anxiety (Brand et al., 2023), neuroticism (Brand et al., 2023) and self-esteem (Murphy et al., 2019). Taken together, these findings support the IAS as measuring an adaptive aspect of interoception, although its pattern of associations with other interoception (or interoception-related) questionnaires points towards some overlap across various theoretical dimensions, casting some doubt on the orthogonal models of interoception and the possibility of its faithful capture by questionnaires.

The current study aims at 1) clarifying the structure of the IAS with a mega-analytic (which involves a re-analysis at the raw data level by aggregating datasets) approach that leverages existing data and contrast the traditional CFA/SEM factor-based analyses with network-based ones (Exploratory Graph Analysis); 2) provide an overview of the dispositional correlates of the IAS, clarifying its general pattern of associations, which is key to better understand the nature, place and role of interoception questionnaires within a larger context.

Study 1

Study 1 will re-analyse and assess the factor structure of the IAS by taking advantage of the large number of open-access datasets (Arslanova et al., 2022; Brand et al., 2022;

Brand et al., 2023; Campos et al., 2021; Gaggero et al., 2021; Lin et al., 2023; Murphy et al., 2019; Todd et al., 2022; Von Mohr et al., 2023). While combining these studies might provide a more robust and generalizable understanding of the IAS' factor structure, we also additionally provide an individual analysis (i.e., applying the same method on all samples separately) to add nuance to the general picture, as all studies differ in their sample size, demographic characteristics, language, and procedure.

Methods

Datasets

Our search focused on studies citing the original IAS validation paper (Murphy et al., 2019), identifying 136 papers (as of 01/05/2024). To qualify for inclusion, papers needed to (1) provide accessible data in open-access, (2) employ the IAS as a measure, and (3) report individual IAS items scores. We also included the data of four unpublished studies. A total of 16 datasets was included (see **Table 1**).

The total number of participants was 32,214 participants (Mean = 48.6 ± 13.1, 71.6% Female).

Data Analysis

Psychometrically good items should exhibit various qualities, such as validity, reliability and discrimination, to which one of the contributing factors is the amount of inter-individual variability captured by an item. Items to which all participants' answers are concentrated around one option - i.e., exhibiting a narrow distribution - will be flagged as potentially problematic.

After examining the distributions and correlations of all IAS items, we will test for "redundant" items (e.g., due to multicollinearity or local dependency) using Unique Variable Analysis (UVA, Christensen et al., 2023), a novel method derived from network psychometric designed to identify and merge items that share substantial variance (which can distort the structure estimation). We will use a threshold of 0.30 that detects large to very large overlap.

Following the analysis of items, we will analyze the factor structure of the IAS using three different approaches, each with particular trade-offs and assumptions, to provide a multi-verse picture of likely solutions. Namely, we will apply traditional exploratory and confirmatory Factor Analysis (EFA/CFA), hierarchical clustering (HCA), and Exploratory Graph Analysis (EGA), to the whole sample, as well as to each dataset separately (details being available in the analysis document at https://github.com/RealityBending/InteroceptionIAS), and our decisions and conclusions will try to take into account both levels of analysis.

By combining network analysis with psychometric methods, the recently-developed EGA framework allows to jointly estimate the number of dimensions (i.e., groups of items), the structure as well as its stability (H. Golino et al., 2020; H. F. Golino & Epskamp, 2017). Evidence has underlined its suitability as an alternative to traditional factor analysis, addressing some of its limitations such as the assumption of a "latent" source of variability, possible biasing in the estimation of the optimal factor numbers depending on sample size, and the poor performance of other methods in complex population structures, while remaining comparable and interpretable (Christensen & Golino, 2021b; Jiménez et al., 2023). At a fundamental level, EGA conceptualizes variables as nodes in a network, with connections (edges) reflecting associations between them. Clustering these nodes reveals distinct communities of related items, in practice akin to traditional latent factors - but without explicitly assuming their presence (Christensen & Golino, 2021b). We used the EGAnet package (Christensen & Golino, 2021a) to fit a hierachical EGA with the Leiden community detection algorithm.

While EGA offers a robust alternative to traditional factor analysis, factor analysis remains a widely used method for dimensionality assessment. As our goal is to provide a general - yet nuanced - picture, with room to show potential discrepancies emerging from the methods used, we will also include it in the present study. Unlike EGA, factor analysis assumes that a latent source of variability — a common latent variable — underlies the observed set of manifest variables (Cosemans et al., 2022). A critical step in factor analysis is determining the optimal number of factors, for which we will used the Method Agreement Procedure (Lüdecke et al., 2021), which involves a consensus-based decision based on multiple factor estimation methods applied concurrently.

Finally, we will also apply Hierarchical Clustering Analysis (HCA, Murtagh & Legendre, 2014), which differs from factor analysis in that it does not assume any latent source of variability, but instead iteratively groups items based on their similarity (e.g., correlation) into a hierarchy of clusters. The benefits of HCA include its interpretability and ability to capture complex relationships among items without relying on strict assumptions about data distribution or latent variables.

In a typical 2-step fashion, this first analysis run will be followed by a structure refinement with a further selection of items, and the final pool of items will be tested again. Additionally, various solutions (e.g., adding general factors) will be statistically compared using Confirmatory Factor Analysis (CFA).

Results

The distribution of the items across samples suggests the presence of a consistent modal value (Figure 2). In other words, participants are most likely to answer 4/5 (i.e., agree) on most items, with the exception of "blood sugar" and "bruise", which exhibit a different distributional pattern with a lower mode (~2/5). While it is not a problem *per se*, the con-

Table 1

Description of the samples used in the study.

Sample	Reference	Language	N	Difference	Age (Mean ± SD)	Range	Female %	Availability
	Murphy et al., (2020)							osf.io/3m5nh
Sample 1a		English	451		25.8 ± 8.4	18-69	69.4%	
Sample 1b		English	375		35.3 ± 16.9	18-91	70.1%	
Sample 2	Gaggero et al., (2021)	English and Italian	814		24.9 ± 5.3	18-58	60.3%	osf.io/5x9sg
Sample 3	Campos et al., (2022)	Portuguese	515		30.7 ± 10.5	18-72	59.6%	osf.io/j6ef3
Sample 4	Todd et al., (2022)	English	802		48.6.6 ± 14.1*	18-92*	50%*	osf.io/ms354
Sample 5	Arslanova et al., (2022)	English	143		28.5 ± 7.6	18-73	46.8%	osf.io/mp3cy
Sample 6	Brand et al., (2022)	German	619		43.9 ± 14.5	18-78	78.7%	osf.io/xwz6g
	Brand et al., (2023)							osf.io/3f2h6
Sample 7a		German	522		23.4 ± 6.7	18-79	79.5%	
Sample 7b		German	1993		32.0 ± 12.6	16-81	77.7%	
Sample 7c		German	802		27.3 ± 9.3	18-72	68.9%	
	Lin et al., (2023)							osf.io/3eztd
Sample 8a		Chinese	1166	Collapsed ”Itch” and ”Tickling”	32.5 ± 8.4	16-60	57.0%	
Sample 8b		Chinese	500	Collapsed ”Itch” and ”Tickling”	37.4 ± 7.4	20-60	56.2%	
Sample 9	VonMohr et al., (2023)	English	21843		56.5 ± 14.4	18-93	73.2%	osf.io/7p9u5
Sample 10	Makowski et al., (2023a)	English	485	Analog scales. No Temp, Blood sugar and Cough	30.1 ± 10.1	18-73	50.3%	github.com/RealityBending/IllusionGameReliability
Sample 11	Makowski et al., (2023b)	English	836	Analog scales	25.1 ± 11.3	17-76	53.0%	github.com/DominiqueMakowski/PHQ4R
Sample 12	Makowski et al., (2023c)	English	146	Analog scales	21.1 ± 4.3	18-50	76%	github.com/RealityBending/InteroceptionPrimals
Sample 13	Makowski et al., (2024)	English	737		36.8 ± 14.9	17, 87	57.3%	github.com/RealityBending/InteroceptionScale
Sample 14	Poerio et al., (2024)	English	107		26.8 ± 9.2	18-57	74.8%	osf.io/49wbv
Sample 15	Poerio et al., unpublished	English	131		30.9 ± 12.0	18-60	75.9%	
Sample 16	Total		32993		48.6 ± 13.1	17-93	71.6%	

Note. *Information taken from the sample description of relevant paper rather than recomputed.

trasting distribution might be indicative of items with non-homogeneous psychometric “difficulty”. This is also the case for “affective touch” in samples 8a and 8b (the Chinese validation samples), which might indicate localisation issues. Additionally, one can note the low occurrence of extreme values (1 and 5), meaning that the bulk of answers varies between 3 values (assuming the IAS is implemented as a 5-point Likert scale following its validation). The samples using an analogue scale (samples 10a, 10b and 10c in the figure) display a more continuous and progressive spread of answers, seemingly improving the interindividual variability, although potentially displaying a secondary lower mode at ~2 (which might be suggesting the existence of potential clusters of participants). The correlation matrix between all items shows an overall positive correlation pattern, with highly correlated pairs of items (e.g., Tickle - Itch, Urinate - Defecate, Pain - Wind, Hungry - Thursty) or triplets (Vomit - Sneeze - Cough, Temperature - Muscles - Pain).

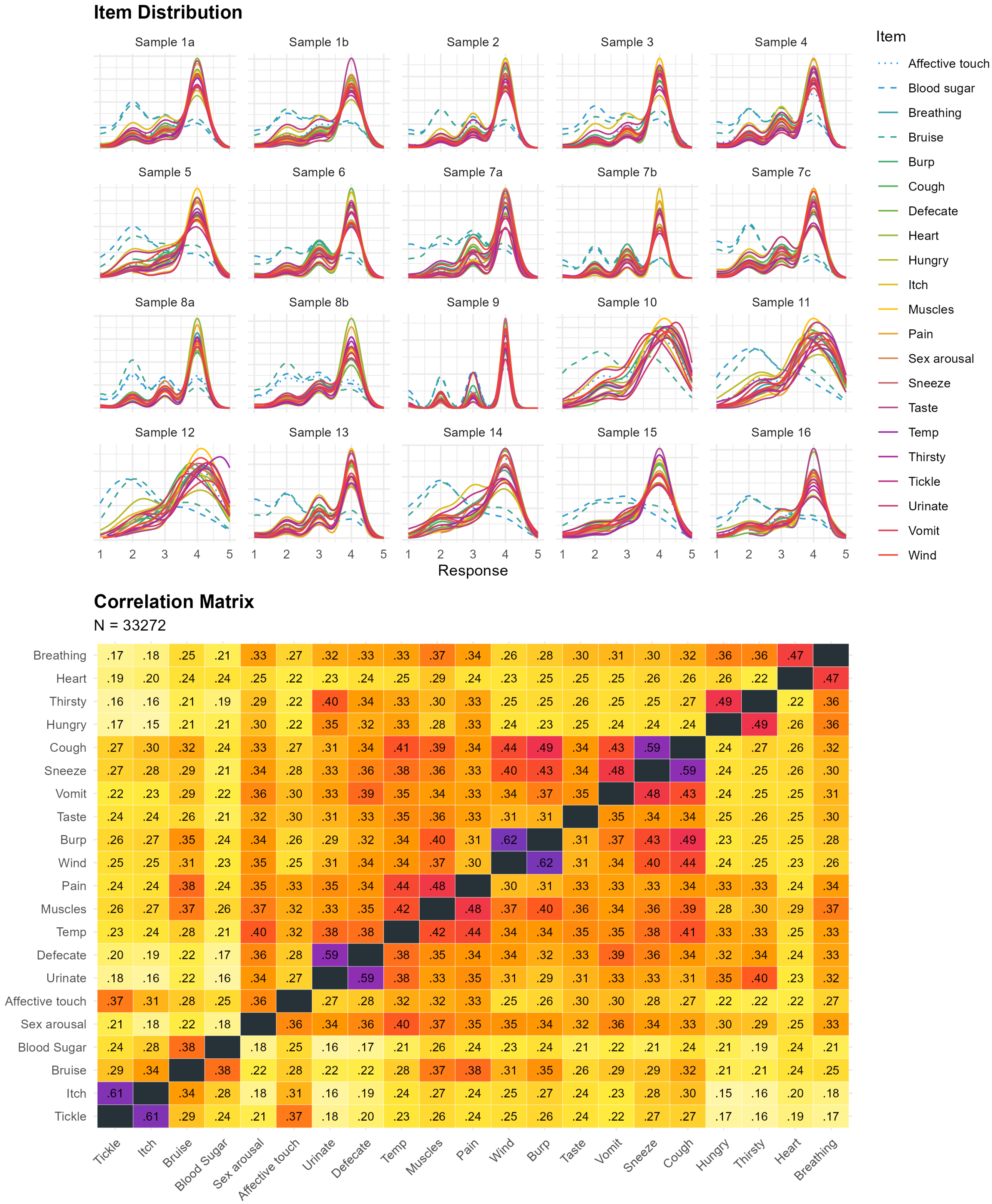
UVA flagged two strongly redundant variables, “itch” and “tickle” - suggesting to remove the latter. Several more pairs of items were flagged as moderately redundant (“wind” and “burp”; “urinate” and “defecate”) and mildly redundant (“sneeze” and “cough”; “heart” and “breathing”; “hungry” and “thirsty”). These patterns consistently appeared in most samples when considered individually. We removed “tickle” from further analysis due to its high redundancy (and because it is absent from some datasets due to translation issues).

The HCA highlighted pairs and triplets of items consistently grouped together across samples, such as “wind” and “burp”, “sneeze” and “cough”, “itch” and “bruise”, “urinate” and “defecate”, and “pain”, “muscles”, and “temperature”. This pattern was largely replicated by the EGA, with the additional presence of a unique cluster comprising “Sex arousal”, “Affective touch”, “Temperature”, “Pain”, “Muscles”, and “Taste”. EFA suggested the optimal number of factors to be 3, yielding one dimension with expulsion-related items (“burp”, “wind”, “cough”, “sneeze”, and “vomit”), a second dimension with visceroceptive items (“heart”, “breathing”, “hungry”, “thirsty”, “urinate”, and “defecate”), and a third dimension with skin-related items (“bruise” and “blood sugar”).

Importantly, this initial structure analysis run highlighted some problematic items: Taste typically displayed a lone or unstable pattern of associations, Affective touch exhibited cross-loadings and instability, Vomit was less strongly asso-

Figure 2

Top: Distribution of responses across datasets reveals a consistent modal value, typically around 4 or 5 (indicating agreement), except for 'blood sugar' and 'bruise' - in dashed lines - and 'affective touch' (dotted lines) in the Chinese validation sample, which have lower modes. Most responses cluster around the middle values, with few extreme scores (1 and 5). Samples using an analogue scale (10a, 10b, 10c) show a more continuous distribution and increased interindividual variability. Since most samples use Likert scales (discrete), density plots may not be the most accurate representation but were chosen to clearly highlight variability patterns in the data. Bottom: The correlation matrix between all items shows an overall positive correlation pattern, with correlated pairs (e.g., Wind, Burp) of items or triplets (e.g., Vomit, Sneeze and Cough).



ciated with other items, and Itch did not form a consistent cluster. Finally, Temperature and Sexual arousal showed redundant patterns of associations but were less reliable. These 6 items were thus removed, and a second run of structure analysis was performed on the remaining 14 items.

The EGA analysis employed a sparse Gaussian graphical model (GGM) with the graphical lasso (glasso) method. We applied the Walktrap (**REF**) and the Louvain (**REF**) algorithms for clusters identification, which yielded a five and four clusters solution, respectively. The four-cluster solution appeared as being more stable (based on the bootstrapped stability analysis: it had a higher proportion of items consistently assigned to the same cluster across bootstrap samples). In this solution, item stability for all variables exceeded 0.90, except for "affective touch". The first cluster included skin-related items "itch", "tickle", "bruise", and "blood sugar"; the second cluster contained expulsion-related items "burp", "wind", "cough", "sneeze", and "vomit"; the third cluster (**NAME or interpretation?**) comprised "affective touch", "sexual arousal", "muscles", "temperature", "pain", and "taste"; and the fourth cluster included visceroceptive items like "heart", "breathing", "hungry", "thirsty", "urinate", and "defecate". This solution was generally consistent when applied to samples separately.

Regarding Factor Analysis, the Method Agreement Procedure recommended a four-factor solution for the whole sample. A four-factor solution was extracted using an *oblimin* rotation, accounting for 41.58% of the total variance. The first factor (14.45%) included items "burp", "cough", "wind", "sneeze", "vomit", "temperature", "sexual arousal" and "taste". The second factor (11.76%) included "breathing", "hungry", "heart", "thirsty", "pain", "muscles", "blood sugar", and "bruise". The third factor (8.09%) contained skin-related items "tickle", "itch" and "affective touch", and the final factor (7.28%) included "urinate" and "defecate". Notably, this structure differed from that suggested by EGA.

UVA, then re-run EGA and EFA. Then CFA.

We then attempted at testing various candidate structures using confirmatory factor analysis (CFA).

It also demonstrated the lowest uniqueness value in the factor analysis, leading to its exclusion from further analysis. Similarly, several "ambiguous" items — such as "temp", "vomit", "affective touch", "sexual arousal", and "taste" — were removed due to their context-dependent nature as well as weak loadings on multiple factors, which may compromise measurement consistency and reduce the clarity of factor structure.

CFA was performed to identify the best-fitting model for the data. A total of five models were computed: Model 1 assumed a general interoception factor (1-factor solution), Model 2 proposed four factors, Model 3 assumed five factors, Model 4 proposed six factors, and Model 5, the most refined model, separated the factors into smaller, more specific groups, resulting in a 7-factor solution.

The results of model comparison indicate that the Model 5 - with 7 factors - had the best fit among all the models (AIC = -171638, BIC = -171228, χ^2 (56) = 2195.20, p < .001). Parameter estimates for the model with 7 factors were also analyzed, with all factor loadings found to be significant (p < .001) and standardized coefficients ranging from 0.50 to 0.82. Furthermore, the correlations between latent factors were all significant, ranging from 0.43 to 0.80, with the strongest correlation observed between ItchBruise and MusclesPain (0.80).

To assess whether the inclusion of higher-order factors was justified, three models were evaluated. Model 1 assumed one higher-order factor. Model 2 introduced two higher-order factors, while Model 3 included three higher-order factors. The baseline model, which consisted of 7 factors, was compared with these alternative models. The results indicated that the baseline model provided the best fit (AIC = -171638, BIC = -171228, $\chi^2(56)$ = 2195.20, p < .001), suggesting no evidence to support the inclusion of higher-order factors.

When taking acount all the samples, the fit statistics indicated a very good model fit ($\chi^2(56)$ = 2195.20, p < .001, RMSEA = 0.03, CFI= 0.98 SRMR = 0.92) for the model with 7-factors. These fit statistics collectively suggest that the model adequately represents the data structure. Overall, the CFA results suggest that Model 7 provides the best fit for the data, with significant factor loadings and strong correlations between the factors.

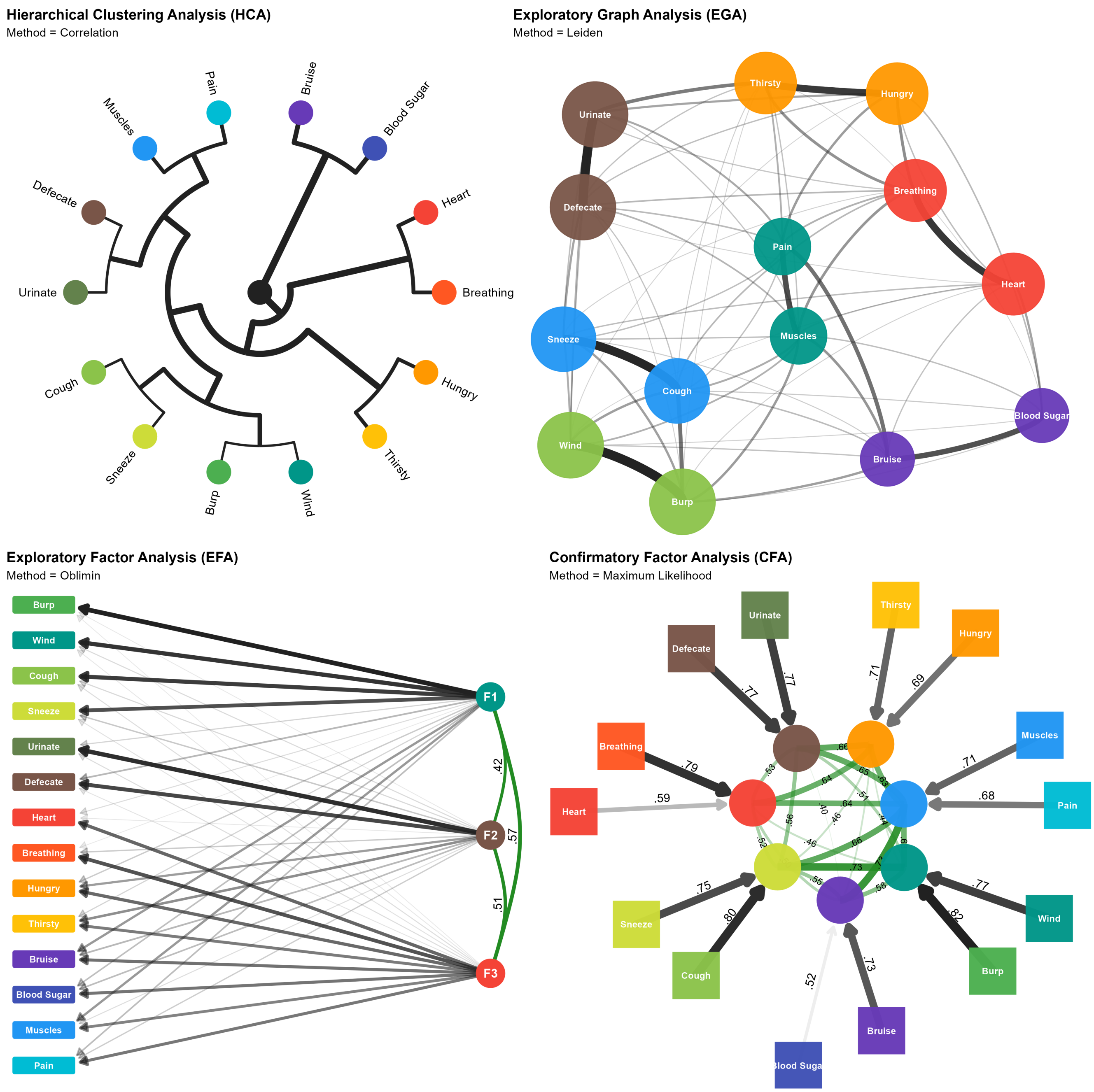
Discussion

The comprehensive structural analysis of a large body of IAS item-level datasets revealed a 4-factor model as the most appropriate solution when taking into account all the items in the scale. These findings contrast with previous research which all found that 2-factor model (Koike & Nomura, 2023; Murphy et al., 2019), 1-factor model (Brand et al., 2023) and bifactor model (Campos et al., 2021; Lin et al., 2023) fits the data best. While this analysis also revealed an okay fit for the 1-factor model, the 4-factor model was ultimately superior. Notably, this 4-factor solution was found in both the EGA and EFA analysis, altough groupings differed.

The EGA analysis revealed different 'hubs' of items that are related, not only in this structure analysis, but also in underlying mechanisms. The 'wind-burp-cough-sneeze-vomit' category, for example, only entails items that are linked to excretion through the mouth. The 'heart-breathing-hungry-urinate-defecate' category includes items linked to basic physiological functions essential for homeostasis and survival. The 'affective-touch-sexual arousal-temperature-pain-taste-muscles' category represents sensations closely tied to both affective experiences and bodily awareness. Lastly, the 'tickle-itch-bruise-blood-sugar' category includes items that reflect more localized or transient bodily sensations, often linked to surface-level or metabolic responses.

Figure 3

TODO.



The EFA groupings, in contrast, revealed a less balanced division of items into clusters. One factor consisted of just two items—'urinate' and 'defecate'—while another contained three items—'tickle', 'itch', and 'affective touch'. Of notice is the loading for 'affective touch' as 0.33, barely meeting the acceptable threshold of >0.32 (Costello & Osborne, 2005), while the other items in this factor had loadings above 0.75.

While the 4-factor structure repeatedly showed up reagrdless of the dimensionality method, some items were flagged as redudant and in particular 'tickle' was suggested to be removed. Additionally, some items appear to be ambiguous, as their interpretation and grouping may be contextually dependent. For instance, sensations such as feeling one's heart beat faster and experiencing nausea-related urges to vomit often co-occur. This overlap is not merely coincidental; these sensations may reflect interconnected physiological processes, such as autonomic arousal or the body's stress response. Such interdependencies challenge the assumption that IAS items neatly map onto distinct bodily sensations. Empirical evidence further highlights item ambiguity. Items such as 'taste' showed similarly weak loadings across multiple factors (0.24 for Factor 1 and 0.20 for Factor 3), falling below the commonly accepted threshold for meaningful loadings. Likewise, 'temperature' loaded weakly on both Factor 1 (0.26) and Factor 3 (0.25), while 'sexual arousal' demonstrated similarly low and split loadings (0.25 to Factor 1 and 0.23 to Factor 3). Although 'vomit' showed a slightly stronger loading on Factor 1, its context-dependent nature further complicates its conceptual clarity. Consequently, refining the scale by removing items with overlapping or ambiguous content could enhance its precision, ensuring that each item captures a unique and meaningful aspect of interoceptive awareness. Based on this, the CFA was conducted with 14 items out of the 21, suggetsing that 7-factors (i.e., pairs of items) is the model with the best fit for the data.

There are couple recomendations that follow from this mega-analysis on the IAS. The findings indicate a high proportion of answers at 4 (see Figure 2), especially when using a 5-step scale. The analogue scale shows a more dispersed distribution, with some answers indicating the highest 5/5, which was not the case in Likert-scales. Therefore, we recommend using an analog scale for the IAS, which provides a novel approach to improving the scale in a simple manner.

Secondly, there are several items that show redundancy suggesting that adapting the IAS would be beneficial for validity. Based on the given results, we suggest removing the tickle, while keeping the itch item [wTO = 0.364]. Other items with moderate redundancy include 'wind' and 'burp', as well as 'urinate' and 'defecate'. Additionally, there are pairs with slight redundancy, such as 'hungry' and 'thirsty', 'sneeze' and 'cough', and 'heart' and 'breathing'. Interestingly, Lin et al. (2023) also found redundancy between 'tickle' and 'itch', ultimately deciding to exclude the former due to the fact that the characters for both terms are identical in the Chinese language. In contrast, Campos et al. (2021) also found 'tickle' to be redudant but did not suggest excluding it from the analysis. Lin et al. (2023) also developed a shortened version of the IAS, removing additional items to create a 12-item scale, including the following items:'hunger', 'breath', 'urinate', 'taste', 'vomit', 'cough', 'temperature', 'sexual arousal', 'wind', 'muscle', 'pain', and 'itch'. As demonstrated by Lin et al. (2023), reducing redundancy and streamlining the scale through a short-form IAS can improve the model fit and enhance the scale's unidimensionality, ultimately making it a more efficient tool for assessing interoceptive accuracy.

The IAS aims to measure subjective interoceptive accuracy by assessing individuals' self-reported ability to detect internal bodily sensations. However, the interpretation of these sensations is highly dependent on their physiological context. Sensations such as heart rate acceleration, breathing changes, or muscle soreness can arise in diverse contexts — for instance, during exercise, emotional arousal, or illness. Without considering the underlying physiological context, individuals may misattribute these sensations to different causes, potentially confounding their self-assessments. For example, a fast heartbeat may be accurately detected but interpreted as anxiety rather than exertion. Similarly, sensations like nausea or sexual arousal may fluctuate in intensity depending on contextual factors such as stress, fatigue, or hormonal cycles. In fact, previous research has highlighted the need for interoceptive questionnaires to prioritize neutral bodily sensations that are less susceptible to cognitive or emotional reinterpretation (Vlemincx et al., 2023).

In conclusion this study provided valuable insights into the structure and content of the IAS. While the original factor structure was not replicated, our findings revealed a more stable 4-factor model that emerged across both EGA and EFA approaches. This structure highlights meaningful item clusters that align with underlying physiological processes, improving our understanding of subjective interoceptive accuracy. However, our analysis also uncovered several limitations that warrant further attention. The presence of redundant items (e.g., 'tickle') and items with ambiguous or weak factor loadings (e.g., 'taste' and 'temperature') suggests a need for scale refinement. Adopting an analog scale format may further enhance variability and improve response accuracy. Additionally, the observed overlap between context-dependent sensations underscores the importance of developing items that account for physiological contexts to minimize misattributions.

Study 2

The second study focuses on the dispositional correlates of the IAS.

Correlations of the IAS will be computed to assess the relationship between subjective interoceptive accuracy and other subjective measures of interoception, mood, psychopathology, personality, and beliefs. Investigating correlates will help validate the IAS, as well as other interoceptive measures in the future.

Methods

Materials

The questionnaires used for the IAS correlates are listed in **Table 2** (**TODO: add the rest of the questionnaires, sample items and references**).

Questionnaire	Number of Dimensions	Assessment	Number of Items	Scoring
Interoceptive Related				
MAIA-2	8	Interoception	37	6-point Likert scale
BPQ	2	Body awareness and autonomic reactivity	49	5-point Likert scale
TAS-20	3	Alexithymia	20	5-point Likert scale
BVAQ	2	Alexithymia	40	5-point Likert scale
Mood				
BDI-II	1	Severity of depressive symptoms	21	0-to-3-point values
PHQ-4	2	Anxiety and depressive symptoms	4	5-point Likert scale
STAI-T	1	Trait anxiety	20	4-point Likert scale
GAD-2	1	General Anxiety	2	4-point Likert scale
Personality				
NEO-FFI	1	Neuroticism	12	5-point Likert scale
Mini IPIP6	6	Personality	24	Analogue scales
BFI	5	Personality	44	5-point Likert scale
PID-5-SF	5	Dysfunctional personality traits	25	4-point Likert scale
Psychopathology				
SPQ-BRU	4	Schizotypy	32	5-point Likert scale
MSI-BPD	1	Borderline personality disorder	10	Analogue scales
ASQ - Short	5	Autistic Traits	28	4-point Likert scale
Beliefs and Misbeliefs				
GCB	5	Conspiracy beliefs	15	5-point Likert scale
PI-18	1	Beliefs about the world	99	6-point Likert scale
LIE scale	4	Lying tendencies	16	Visual analogue scales

Data Analysis

Correlations will be computed using the correlation package under a Bayesian framework (Ben-Shachar et al., 2020).

Results

Average correlations

The EGA components captured grouping of items such as ‘wind’ and ‘burp, ’cough’ and ‘sneeze’, ‘muscle’ and ‘pain’. **TODO: but that’s not true, we used only pairs, not EGA clusters?J**

These groupings were used in correlational analysis to analyse how much each pairing is associated with other factor such as Alexithymia and with Mood disorders (see figure 2).

TODO: Massively streamline. It can be all summarize in one or two paragraphs max.

Correlations with body measures. Alexithymia was assessed in the samples with the Bermond–Vorst Alexithymia Questionnaire (BVAQ, Vorst & Bermond, 2001) and the Toronto Alexithymia Scale (TAS, Bagby et al., 1994).

The BVAQ consists of 5 subscales - fantasising, idenitying, analysing; verbalising and emotionalising - assessed with 40 items on a 5-point Likert scale, from ‘defnitely applies to me’ to ‘in no way applies to me’. Additionally, the BVAQ reduces these subscales into two high order factors, an affective component and a cognitive one, with high scores being indicative of high proneness to alexithymia.

On average, the cognitive component of the BVAQ was weakly and negatively correlated with all IAS pairs of items with the biggest correlation being with the Itch/Bruise pair (r = -0.112) and the lowest correlation beeing with the Muscle/-Pain pair (r = -0.244). The affective component of the BVAQ was positively, but very weak, correlated with all pairs, with the biggest correlation being with the Itch/Bruise pair (r = 0.107). The only exception was a negative correlation with the Urinate/Defecate pair (r = -0.036).

The TAS contains 20-items rated on a 5-point forced scale, from ‘strongly disagree’ to ‘strongly afree’, divided into 3 dimensions - difficulty identiying feelings, difficulty describing feelings, and externally thinking. High scores on this scale also reflect higher alexithymia.

All the three dimensions assessed with the TAS were on average negatively correlated with all pairs of IAS items. The difficulty describing feelings had its strongest correlation with Hungry/Thirsty (r = -0.179) and weakest with the Wind/Burp (r = -0.117). while, the difficulty describing feelings had its strongest correlation with Muscle/Pain (r = -0.247) and weakest with Itch/Bruise (r = -0.157). Lastly, the external thinking dimension was more correlated with the Cough/Sneeze pair (r = -0.138) and less correlated with the Hungry/Thirsty (r = -0.018).

The studies within our sample used the Body Perception Questionnaire short-form (BPQ-SF) and the very-short form (BPQ-VSF) to assess interoception (Cabrera et al., 2018). The BPQ-SF comprised of 46 items on a 5-point Likert scale assessing body awareness (26 items) and autonomic reactivity (21 items). The BPQ-VSF comprises of 12 items from the body awareness subscale of the BPQ-SF. In this study, all scores assessing these two dimensions were grouped together, hence no disitintion is made between awareness measured with the BPQ-SF and the BPQ-VSF, or eith scores obtained only using the awareness subscale.

note to add: discuss later

In general, all pairs of the IAS were positively,and weakily, associated with the body awareness subscales, while negative and weakily correlated with the autonomic reactivity subscale. The strongest correlation identified between the IAS pairs and the body awereness subscale was with the Heart/Breathing pair (r = 0.151) whilst the strongest correlation with the autonomic reactivity was with the Urinate/Defecate pair (r = -0.235). The weakest correlation between the body awareness and the IAS was with the Hungry/Thirsty pair (r = 0.055) and between the autonomic reactivity and the IAS was with the Heart/Breathing pair (r = -0.106).

The MAIA was one of the most commonly used measures of interoception in our study, with nine samples reporting its use. This 37-item questionnaire assesses eight state-trait dimensions of interoception: Noticing, Not-Distracting, Not-Worrying, Attention Regulation, Emotional Awareness, Self-Regulation, Body Listening, and Trust. Responses are rated on a scale from 0 (Never) to 5 (Always).

On average, all MAIA dimensions were positively and weakly to moderately correlated with IAS pairings. Notably, the strongest correlations were observed between the Noticing dimension and the Heart/Breathing pairing (r = 0.394), Trusting and Hungry/Thirsty (r = 0.347), and Attention Regulation and Heart/Breathing (r = 0.334). The Not-

Distracting and Not-Worrying subscales were generally positively correlated with IAS pairings, with a few exceptions: Not-Distracting showed minimal correlation with Cough/Sneeze (r = 0.0206) and Heart/Breathing (r = -0.007), while Not-Worrying had a low correlation with Itch/Bruise (r = 0.031).

The Interoceptive Confusion Questionnaire was used to assess individuals' difficulties in interpreting non-affective physiological states, such as pain and hunger. The ICQ consists of 20 items rated on a scale from 1 ("Does not describe me") to 5 ("Describes me very well"), with higher scores indicating greater interoceptive confusion.

The ICQ showed weak to moderate negative correlations with all IAS pairings. The strongest correlation was observed with the Hunger/Thirsty pairing (r = 0.348), while the weakest was with the Itch/Bruise pairing (r = 0.207).

Correlations with mood measures. Mood disorders were assessed using several standardized measures, including the General Anxiety Disorder-2 (GAD-2, Kroenke et al., 2007), the State-Trait Anxiety Inventory (STAI, Spielberger, 1970) and its shorter version, the STAI-5 (Zsido et al., 2020), Beck's Depression Inventory (BDI, Beck et al., 1996), and the Mood and Feelings Questionnaire [MFQ; Messer et al. (1995)]. Additionally, the Patient Health Questionnaire (PHQ) was administered in its 2-item [PHQ-2; Kroenke et al. (2003)], 9-item [PHQ-9; Kroenke et al. (2001)], and 15-item (PHQ-15, Kroenke et al., 2002) versions. Finally, borderline personality traits were assessed using the McLean Screening Instrument for Borderline Personality Disorder [MSI-BPD; Zanarini (2003)].

The GAD-2, a brief screening tool for generalized anxiety disorder, consists of two items rated on a scale from 0 (not at all) to 3 (nearly every day). The STAI, a 40-item questionnaire rated on a 4-point Likert scale (0 to 3), measures both state and trait anxiety. However, in our study, most participants primarily completed the trait anxiety subscale. In some samples, a shorter 5-item version (STAI-5) was used to assess both state and trait anxiety.

On average, anxiety measures showed weak negative correlations with all IAS pairs. Notably, the strongest correlations between the IAS pairings and the GAD-2, STAI-T and STAIT-5 were observed with the Hungry/Thirsty pair (r = -0.168, r = -0.270 and r = -0.248, respectively).

The BDI consists of 21 items measuring the severity of depressive symptoms on a scale from 0 to 3. The total score is calculated by summing the highest responses, which are then compared to six depression severity levels, ranging from 1–10 (normal fluctuations in mood) to over 40 (extreme depression). The PHQ-2 includes two items assessing the frequency of depressive symptoms and anhedonia. The PHQ-2 is derived from the PHQ-9, a nine-item screening tool used to assess depression severity and monitor treatment response. Both questionnaires are measured on on a scale from 0 (not at all) to 3 (nearly every day)

Depression measures showed weak to moderate negative correlations with IAS pairings. The BDI (r = -0.372), PHQ-2 (r = -0.148), and PHQ-9 (r = -0.241) correlated most with the Hungry/Thirsty pair, while the MFQ correlated most with Heart/Breathing (r = -0.345) pair.

The PHQ-15 is a 15-item questionnaire that assesses somatic symptoms on a 3-point scale (e.g., back pain). It exhibited its strongest correlation with the Hungry/Thirsty pair (r = -0.241) and, on average, showed weak negative correlations with all other IAS pairings.

Lastly, the MSI-BPD is a 10-item questionnaire used to assess personality disorder, where items are rated on a dichotomous scale of 1 (present) and 0 (absent). The MSI-BPD also showed its strongest negative correlation with the Hungry/Thirsty pair (r = -0.140) and was negatively correlated with all other pairings, except for Cough/Sneeze, which showed a slight positive correlation (r = 0.0219).

Correlations with psychopathology measures. Maladaptive personality traits were assessed using the Personality Inventory for DSM-5 Short Form [PID-5-SF; Thimm et al. (2016)], which measures five domains: disinhibition, antagonism, detachment, negative affect, and psychoticism. The scale consists of 25 items rated on a 4-point Likert scale, ranging from 0 (very false or often false) to 3 (very true or often true).

On average, all maladaptive personality traits assessed by the PID-5-SF were weakly and negatively correlated with IAS pairings. The strongest correlation was observed between the psychoticism dimension and the Muscle/Pain pairing (r = -0.173).

Schizotypy was assessed using the Schizotypal Personality Questionnaire – Brief Revised Updated (SPQ-BRU; Davidson et al. (2016)), which consists of 32 items rated on a 5-point Likert scale ranging from strongly agree to strongly disagree. This questionnaire evaluates four primary dimensions: cognitive-perceptual (positive), interpersonal (negative), disorganized, and social anxiety. These dimensions are further divided into nine secondary factors: constricted affect, eccentricity, magical thinking, lack of close friends, odd speech, referential thinking, social anxiety, suspiciousness, and unusual perceptions.

On average, all nine factors were weakly and negatively correlated with IAS pairings, with correlations ranging from r = -0.170 (between lack of close friends and Muscles/Pain) to r = 0.102 (between magical thinking and Itch/Bruise).

The short version of the Autism-Spectrum Quotient (ASQ-Short; Hoekstra et al., 2011) was used to assess five autistic traits: social skills, adherence to routines, cognitive flexibility (switching), imagination, and patterns/numbers. The questionnaire consists of 28 items rated on a 4-point Likert scale, ranging from 1 (definitely agree) to 4 (definitely disagree).

Overall, all pairings were weakly and negatively correlated with the ASQ dimensions, except for the Itch/Bruise and Heart/Breathing pairings, which showed weak positive correlations with the patterns/numbers trait (r = 0.184 and r = 0.038, respectively). The strongest correlation was observed between the imagination trait and the Wind/Burp pairing (r = -0.218).

Correlations with personality measures. The Big Five Inventory-Short Form [BFI-S; Lang et al. (2011)] and the Mini International Personality Item Pool [Mini-IPIP6; Sibley et al. (2011)] were used to assess general personality traits. The BFI-S consists of 15 items rated on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), measuring five personality factors: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The Mini-IPIP6 assesses six personality traits—Extraversion, Agreeableness, Conscientiousness, Neuroticism, Openness, and Honesty-Humility—using 24 items. While this questionnaire is typically scored on a 7-point Likert scale from 1 (very inaccurate) to 7 (very accurate), an analogous scale was used in the respective sample. Lastly, the Neuroticism subscale of the NEO Five-Factor Inventory [Neo-FFI; Costa and McCrae (1992)] was used to assess Neuroticism, consisting of 12 items rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

To assess correlations with the IAS pairings, scores were grouped across personality dimensions due to the overlap among these traits (with the exception of Honesty-Humility). On average, IAS pairings were positively associated with most personality dimensions, though these correlations were generally weak. The strongest correlation was observed between Conscientiousness and the Hungry/Thirsty pairing (r = 0.164). Both Honesty-Humility and Neuroticism were weakly and negatively correlated with the IAS pairings, with the highest correlations observed for the Hungry/Thirsty pairing (r = -0.217 and r = -0.200, respectively).

Correlations with other measures. The IAS was also correlated with primal world beliefs, as measured by the Primal Inventory [PI-18; J. D. Clifton and Yaden (2021)], which assesses beliefs about the world being alive, good, safe, and enticing. Items that evaluate neutral beliefs about the hierarchical order of importance in the world (i.e., hierarchical), as well as beliefs about the comprehensibility of most things and situations (i.e., understandable), and the belief that the world is characterized by flux (i.e., changing) were added as well. The scale contains 18 items ranging from 5 (Strongly agree) to 0 (strongly disagree).

Overall, most primal beliefs show weak positive correlations with all pairings of the IAS. The strongest correlation is between the hierarchical belief and the Hungry/Thirsty pairing (r = 0.181). Some beliefs, however, exhibit negative correlations with certain pairings. These negative correlations range from r = -0.0940 between the changing belief and the Hungry/Thirsty pairing, to r = -0.00490 between the Enticing belief and the Itch/Bruise pairing.

The Generic Conspiracist Beliefs Scale [GCBS; Brotherton et al. (2013)] was used to assess five facets of conspiracy beliefs: Extraterrestrial, Global Conspiracies, Government Malfeasance, Information Control, and Personal Wellbeing. The scale comprises 15 items rated on a 5-point Likert scale, ranging from definitely not true (1) to definitely true (5).

Overall, the GCBS showed a weak but positive correlation with all facets of the IAS, with the strongest correlation observed between Global Conspiracies and Hungry/Thirsty (r = 0.140). Negative correlations were found within the Global Conspiracies, Extraterrestrial, and Information Control facets, though these were small, ranging from r = -0.0101 to r = -0.0167.

Lastly, the Lying Profile Questionnaire [LIE; Makowski, Pham, et al. (2023)] a 16 item visual analog scale was used to assess 4 dispositional lying simensions: ability; negativity, contextuality, and frequency.

Overall, most lying profile dimensions show weak correlations with IAS pairings. Ability exhibits primarily weak positive correlations, with the strongest observed for Wind/Burp (r = 0.082). In contrast, Frequency tends to show weak negative correlations, ranging from Wind/Burp (r = -0.062) to Muscles/Pain (r = -0.088). Contextuality displays mixed correlations, with Itch/Bruise showing the strongest negative association (r = -0.127), while Urinate/Defecate has a small positive correlation (r = 0.045). Finally, Negativity is consistently positively correlated with all pairings, with the strongest relationship found for Hungry/Thirsty (r = 0.090).

Discussion

Our findings underline how subjective measures of interoception exist within a complex network of correlates.

TODO:: Firstly, talk about the link with other interoceptive measures. Then, discuss other correlates by order of importance

TODO: this discussion is too much descriptive for now. Discuss that from a larger perspective. What does it mean in general.

Among these, alexithymia exhibits the strongest negative correlation with the IAS, whereas the MAIA questionnaire shows the strongest positive correlation. These correlates not only help explain different aspects of interoception but also serve as valuable tools for validating interoceptive measures.

TODO: integrate the following Note to dom: Ferentzi et al. (2021) suggests that all MAIA dimensions, except for Not-Worrying and Not-Distracting, relate to a general interoceptive awareness factor. The low correlations of NW and ND with this general factor indicate that these dimensions do not contribute significantly to the measurement of general interoceptive awareness. Instead, NW appears to measure low emotionality or neuroticism, while ND does not capture a coherent underlying structure.

While our results reveal various correlations with the IAS, they are limited to the scope of the given questionnaire. Nonetheless, they provide valuable insights into how interoception may relate to different psychological and personality traits. The results show a consistent pattern of correlations with other measures and highlight interesting exploratory results, such as correlations between primal world beliefs with the IAS.

Our analysis found a strong negative correlation between alexithymia and IAS scores, aligning with previous research (Brand et al., 2023; Herbert et al., 2011; Murphy et al., 2019). Similarly, a negative correlation between autism and interoceptive awareness was observed in our sample, consistent with prior findings (DuBois et al., 2016).

Conspiracy beliefs did not strongly correlate with IAS scores, though a slight positive correlation was present. To our knowledge, this relationship has not been previously explored. However, prior studies have suggested connections between interoception and (political) beliefs, potentially pointing to shared underlying mechanisms (Ruisch et al., 2022a).

The relationship between interoception and lying profiles was also weak. This contrasts with previous research suggesting associations between interoception and deception (Makowski, Lau, et al., 2023), warranting further investigation.

Mood and IAS scores exhibited a strong negative correlation, consistent with prior studies that have documented similar findings (Solano López & Moore, 2018). Additionally, personality traits correlated with interoceptive accuracy scores, reinforcing existing research linking personality and interoception (Erle et al., 2021).

We also observed negative correlations between schizotypy and interoception, in line with previous studies that identified a similar relationship with interoceptive awareness, particularly in individuals at risk for psychosis (Torregrossa et al., 2022).

Interestingly, world beliefs demonstrated significant positive correlations with interoception. While this relationship has not been previously documented, other forms of belief, such as political ideology, have been linked to interoception (Ruisch et al., 2022b). Further research is needed to determine whether world beliefs, which shape our perception of reality (J. D. W. Clifton, 2020), are meaningfully connected to interoception.

Overall, our findings highlight the broad relevance of interoception across various cognitive and affective traits, underscoring its significance in both research and clinical contexts. By identifying numerous correlates of the IAS, we contribute not only to a deeper understanding of interoception's role in daily life but also to the ongoing validation of the IAS and other interoceptive measures. This analysis lays an important foundation for the development of new interoceptive assessment tools, further advancing our comprehension of interoception and its impact on human experience.

General Discussion

The present study aimed… [always start with a description of the study].

Our analyses revealed that the IAS follows a four-factor structure with an uneven distribution. While the findings indicate that the IAS measures interoception adequately, there is room for improvement. Additionally, different correlation measures with the IAS suggest opportunities for further exploration of how interoception is assessed. In the following section, we discuss the strengths and shortcomings of the IAS, followed by proposed steps to enhance interoception measurement.

Overall, the IAS is straightforward in its sensation-centered items. However, several areas for improvement emerge from this study. Firstly, redundant items should be removed, such as the "itch" item, as highlighted in our analysis. Previous research also suggests redundancy between itch and tickle items Campos et al. (2021). Interestingly, while Campos et al. (2021) does not recommend the removal of either, Lin et al. (2023) argues for removing the itch item due to their overlapping character representation.

Furthermore, this study recommends using analog scales instead of 5-point scales. The limited variability of the 5-point scale often results in most responses clustering around 3 or 4. As shown in Figure 2, adopting an analog scale significantly increases variability. However, even with an analog scale, IAS variability remains constrained. Greater variability allows for better differentiation among participants, making dispersion an essential factor for obtaining meaningful results. Enhancing variability would therefore be beneficial for the IAS.

Despite these improvements, certain limitations persist in the IAS that affect its accuracy. Notably, some modalities are underrepresented—for instance, heart perception is measured by only one item. Expanding modality coverage would enhance variability within each category, leading to more nuanced results. Moreover, the IAS lacks a clear theoretical or empirical structure, with only small item groupings. Ideally, a scale should allow for clear groupings that support meaningful data analysis. In this study, each group contained only two items, resulting in low scores and limited variability. Additionally, some IAS items are ambiguous, with their interpretation depending on context. For example, an item about perceiving heartbeats and another about vomiting could both relate to anxiety, leading to results that may differ from initial expectations. Thus, the grouping and structure of the IAS require refinement.

Another concern is that all IAS items are phrased positively, which may influence participant responses. While positive phrasing has advantages, it can also introduce response bias, leading to unidimensional results. A more balanced phrasing approach, incorporating both positively and negatively framed items, could yield more accurate responses.

Given these considerations, it is clear that context-specific, cross-modal items—such as integrating cardioception and respiroception—are needed. Recognizing the necessity for a refined interoception scale, this study proposes the development of the Multidimensional Interoceptive Inventory (MInt). This new scale will be designed to align with recent findings on the IAS and interoception research while allowing for direct comparison with IAS correlates.

[**TO DO**: add - previous work suggests the importance of physiological contexts (Vlemincx et al., 2021)] **I would rather put that in the discussion in the suggestions for better scales**

Limitations and Future Directions

There are several limitations to the IAS. There are some redundant items, the 5-point scale does not provide great variability, and the structure could be improved. Therefore, improving the IAS, or creating a new questionnaire investigating interoception could be useful to achieving reliable and accurate indication of interoceptive awareness.

Conclusion

The IAS is a valuable tool for measuring interoception compared to existing questionnaires and methods. However, refining or even redesigning the questionnaire could lead to a more precise and comprehensive assessment. This study highlights the need for a new interoception scale to advance research in the field. By identifying various correlates of the IAS, this work paves the way for future investigations into optimal interoceptive measures, ultimately laying the foundation for the development of a more effective interoception survey.

References

Arslanova, I., Galvez-Pol, A., Kilner, J., Finotti, G., & Tsakiris, M. (2022). Seeing through each other's hearts: Inferring others' heart rate as a function of own heart rate perception and perceived social intelligence. *Affective Science*, *3*(4), 862–877.

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13*(1), 27–45.

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.

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. *Psychological Assessment*.

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815.

Brand, S., Meis, A. C., Tünte, M. R., Murphy, J., Woller, J. P., Jungmann, S. M., Witthöft, M., Hoehl, S., Weymar, M., Hermann, C., & Ventura-Bort, C. (2023). A multi-site German validation of the Interoceptive Accuracy Scale and its relation to psychopathological symptom burden. *Communications Psychology*, *1*(1). https://doi.org/10.1038/s44271-023-00016-x

Brand, S., Petzke, T. M., & Witthöft, M. (2022). The differential relationship between self-reported interoceptive accuracy and attention with psychopathology. *Zeitschrift für Klinische Psychologie Und Psychotherapie*.

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

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology*, *4*, 279.

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.

Campos, C., Rocha, N. B., & Barbosa, F. (2021). *Untangling self-reported interoceptive attention and accuracy: Evidence from the european portuguese validation of the body perception questionnaire and the interoceptive accuracy scale*. http://dx.doi.org/10.31234/osf.io/a7wdj

Christensen, A. P., Garrido, L. E., & Golino, H. (2023). Unique variable analysis: A network psychometrics method to detect local dependence. *Multivariate Behavioral Research*, *58*(6), 1165–1182.

Christensen, A. P., & Golino, H. (2021a). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A monte carlo simulation and tutorial. *Psych*, *3*(3), 479–500.

Christensen, A. P., & Golino, H. (2021b). On the equivalency of factor and network loadings. *Behavior Research Methods*, *53*(4), 1563–1580.

Clifton, J. D. W. (2020). Testing if primal world beliefs reflect experiences—or at least some experiences identified ad hoc. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01145

Clifton, J. D., & Yaden, D. B. (2021). Brief measures of the four highest-order primal world beliefs. *Psychological Assessment*, *33*(12), 1267.

Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory graph analysis for factor retention: Simulation results for continuous and binary data. *Educational and Psychological Measurement*, *82*(5), 880–910.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO personality inventory. *Psychological Assessment*, *4*(1), 5.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(1).

Davidson, C. A., Hoffman, L., & Spaulding, W. D. (2016). Schizotypal personality questionnaire–brief revised (updated): An update of norms, factor structure, and item content in a large non-clinical young adult sample. *Psychiatry Research*, *238*, 345–355.

Desmedt, O., Heeren, A., Corneille, O., & Luminet, O. (2022). What do measures of self-report interoception measure? Insights from a systematic review, latent factor analysis, and network approach. *Biological Psychology*, *169*, 108289.

Desmedt, O., Luminet, O., Walentynowicz, M., & Corneille, O. (2023). The new measures of interoceptive accuracy: A systematic review and assessment. *Neuroscience & Biobehavioral Reviews*, *153*, 105388. https://doi.org/https://doi.org/10.1016/j.neubiorev.2023.105388

DuBois, D., Ameis, S. H., Lai, M.-C., Casanova, M. F., & Desarkar, P. (2016). Interoception in Autism Spectrum Disorder: A review. *International Journal of Developmental Neuroscience*, *52*(1), 104–111. https://doi.org/10.1016/j.ijdevneu.2016.05.001

Erle, T. M., Mitschke, V., & Schultchen, D. (2021). Did my heart just leap or sink? The role of personality for the relation between cardiac interoception and well-being. *Personality and Individual Differences*, *170*, 110493. https://doi.org/10.1016/j.paid.2020.110493

Ferentzi, E., Olaru, G., Geiger, M., Vig, L., Köteles, F., & Wilhelm, O. (2021). Examining the factor structure and validity of the multidimensional assessment of interoceptive awareness. *Journal of Personality Assessment*, *103*(5), 675–684.

Gabriele, E., Spooner, R., Brewer, R., & Murphy, J. (2022). Dissociations between self-reported interoceptive accuracy and attention: Evidence from the interoceptive attention scale. *Biological Psychology*, *168*, 108243.

Gaggero, G., Bizzego, A., Dellantonio, S., Pastore, L., Lim, M., & Esposito, G. (2021). Clarifying the relationship between alexithymia and subjective interoception. *PLoS One*, *16*(12), e0261126.

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.

Garner, D. M. (1991). Eating disorder inventory-2; professional manual. *Psychological Assessment Resources*.

Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders*, *2*(2), 15–34.

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS One*, *12*(6), e0174035.

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292.

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. https://doi.org/10.1111/j.1467-6494.2011.00717.x

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*, 107949.

Jahedi, S., & Méndez, F. (2014). On the advantages and disadvantages of subjective measures. *Journal of Economic Behavior & Organization*, *98*, 97–114. https://doi.org/10.1016/j.jebo.2013.12.016

Jiménez, M., Abad, F. J., Garcia-Garzon, E., Golino, H., Christensen, A. P., & Garrido, L. E. (2023). Dimensionality assessment in bifactor structures with multiple general factors: A network psychometrics approach. *Psychological Methods*.

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., Feusner, J. D., Garfinkel, S. N., Lane, R. D., Mehling, W. E., et al. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(6), 501–513.

Kleckner, I. R., Wormwood, J. B., Simmons, W. K., Barrett, L. F., & Quigley, K. S. (2015). Methodological recommendations for a heartbeat detection-based measure of interoceptive sensitivity. *Psychophysiology*, *52*(11), 1432–1440.

Koike, H., & Nomura, M. (2023). Development and validation of japanese versions of the interoceptive accuracy scale and interoceptive attention scale. *SAGE Open*, *13*(4), 21582440231214639.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine*, *64*(2), 258–266.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The patient health questionnaire-2: Validity of a two-item depression screener. *Medical Care*, *41*(11), 1284–1292.

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine*, *146*(5), 317–325.

Lang, F. R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. G. (2011). Short assessment of the big five: Robust across survey methods except telephone interviewing. *Behavior Research Methods*, *43*, 548–567.

Lin, X.-X., Shen, H.-R., Lin, J.-X., Zhang, Y.-H., Murphy, J., Wang, Y.-Z., Sun, Y.-B., Wang, N., Wang, J.-Y., Wei, G.-X., & Luo, F. (2023). Psychometric validation and refinement of the Chinese Interoceptive Accuracy Scale (IAS) in general population and patients with chronic pain. *Journal of Psychosomatic Research*, *175*, 111541. https://doi.org/10.1016/j.jpsychores.2023.111541

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An r package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60).

Makowski, D., Lau, Z. J., Pham, T., Te, A., Kirk, S., & Liauw Yong Tong, C. (2023). *The heart can lie: The role of interoception and theory of mind in deception*. http://dx.doi.org/10.31234/osf.io/p342w

Makowski, D., Pham, T., Lau, Z. J., Raine, A., & Chen, S. A. (2023). The structure of deception: Validation of the lying profile questionnaire. *Current Psychology*, *42*(5), 4001–4016.

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

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. https://doi.org/10.1371/journal.pone.0048230

Messer, S. C., Angold, A., Costello, E. J., Loeber, R., Van Kammen, W., & Stouthamer-Loeber, M. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents: Factor composition and structure across development. *International Journal of Methods in Psychiatric Research*, *5*, 251–262.

Miller, L. C., Murphy, R., & Buss, A. H. (1981). Consciousness of body: Private and public. *Journal of Personality and Social Psychology*, *41*(2), 397.

Murphy, J., Brewer, R., Plans, D., Khalsa, S. S., Catmur, C., & Bird, G. (2019). Testing the independence of self-reported interoceptive accuracy and attention. *Quarterly Journal of Experimental Psychology*, *73*(1), 115–133. https://doi.org/10.1177/1747021819879826

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, *31*(3), 274–295.

Porges, S. (1993). Body perception questionnaire. *Laboratory of Developmental Assessment, University of Maryland*, *10*, s15327752jpa5304_1.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137.

Ruisch, B. C., Von Mohr, M., Naber, M., Tsakiris, M., Fazio, R. H., & Scheepers, D. T. (2022a). Sensitive liberals and unfeeling conservatives? *Interoceptive sensitivity predicts political liberalism*. *Politics and the Life Sciences*, *41*(2), 256–275. https://doi.org/10.1017/pls.2022.18

Ruisch, B. C., Von Mohr, M., Naber, M., Tsakiris, M., Fazio, R. H., & Scheepers, D. T. (2022b). Sensitive liberals and unfeeling conservatives? *Interoceptive sensitivity predicts political liberalism*. *Politics and the Life Sciences*, *41*(2), 256–275. https://doi.org/10.1017/pls.2022.18

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology*, *18*(4), 483–488.

Shields, S. A., Mallory, M. E., & Simon, A. (1989). The body awareness questionnaire: Reliability and validity. *Journal of Personality Assessment*, *53*(4), 802–815.

Sibley, C. G., Luyten, N., Purnomo, M., Mobberley, A., Wootton, L. W., Hammond, M. D., Sengupta, N., Perry, R., West-Newman, T., Wilson, M. S., et al. (2011). The mini-IPIP6: Validation and extension of a short measure of the big-six factors of personality in new zealand. *New Zealand Journal of Psychology*, *40*(3).

Solano López, A. L., & Moore, S. (2018). Dimensions of Body-Awareness and Depressed Mood and Anxiety. *Western Journal of Nursing Research*, *41*(6), 834–853. https://doi.org/10.1177/0193945918798374

Spielberger, C. D. (1970). Manual for the state-trait anxiety inventory (self-evaluation questionnaire). *(No Title)*.

Suksasilp, C., & Garfinkel, S. N. (2022). Towards a comprehensive assessment of interoception in a multi-dimensional framework. *Biological Psychology*, *168*, 108262.

Thimm, J. C., Jordan, S., & Bach, B. (2016). The personality inventory for DSM-5 short form (PID-5-SF): Psychometric properties and association with big five traits and pathological beliefs in a norwegian population. *BMC Psychology*, *4*, 1–11.

Todd, J., Swami, V., Aspell, J. E., Furnham, A., Horne, G., & Stieger, S. (2022). Are some interoceptive sensibility components more central than others? Using item pool visualisation to understand the psychometric representation of interoception. *Plos One*, *17*(12), e0277894.

Torregrossa, L. J., Amedy, A., Roig, J., Prada, A., & Park, S. (2022). Interoceptive functioning in schizophrenia and schizotypy. *Schizophrenia Research*, *239*, 151–159. https://doi.org/10.1016/j.schres.2021.11.046

Vlemincx, E., Walentynowicz, M., Zamariola, G., Van Oudenhove, L., & Luminet, O. (2023). A novel self-report scale of interoception: The three-domain interoceptive sensations questionnaire (THISQ). *Psychology & Health*, *38*(9), 1234–1253.

Von Mohr, M., Silva, P. C., Vagnoni, E., Bracher, A., Bertoni, T., Serino, A., Banissy, M. J., Jenkinson, P. M., & Fotopoulou, A. (2023). My social comfort zone: Attachment anxiety shapes peripersonal and interpersonal space. *Iscience*, *26*(2).

Vorst, H. C., & Bermond, B. (2001). Validity and reliability of the bermond–vorst alexithymia questionnaire. *Personality and Individual Differences*, *30*(3), 413–434.

Zanarini, M. C. (2003). Zanarini rating scale for borderline personality disorder (ZAN-BPD): A continuous measure of DSM-IV borderline psychopathology. *Journal of Personality Disorders*, *17*(3), 233–242.

Zsido, A. N., Teleki, S. A., Csokasi, K., Rozsa, S., & Bandi, S. A. (2020). Development of the short version of the spielberger state—trait anxiety inventory. *Psychiatry Research*, *291*, 113223.