Biological Psychology 169 (2022) 108289

Contents lists available at ScienceDirect

Biological Psychology

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

What do measures of self-report interoception measure? Insights from a systematic review, latent factor analysis, and network approach

Olivier Desmedt [a,b,*], Alexandre Heeren [a,b,c], Olivier Corneille [a], Olivier Luminet [a,b]

[a] Psychological Science Research Institute, UCLouvain, Louvain-la-Neuve, Belgium
[b] Fund for Scientific Research – Belgium (FRS-FNRS), Belgium
[c] Institute of Neuroscience, UCLouvain, Brussels, Belgium

A R T I C L E I N F O

Keywords:
Interoception
Interoceptive sensibility
Exploratory factor analysis
Network analysis

A B S T R A C T

Recent conceptualizations of interoception suggest several facets to this construct, including "interoceptive sensibility" and "self-report interoceptive scales", both of which are assessed with questionnaires. Although these conceptual efforts have helped move the field forward, uncertainty remains regarding whether current measures converge on their measurement of a common construct. To address this question, we first identified -via a systematic review- the most cited questionnaires of interoceptive sensibility. Then, we examined their correlations, their overall factorial structure, and their network structure in a large community sample ($n = 1003$). The results indicate that these questionnaires tap onto distinct constructs, with low overall convergence and interrelationships between questionnaire items. This observation mitigates the interpretation and replicability of findings in self-report interoception research. We call for a better match between constructs and measures.

Interoception is the processing of internal bodily states by the nervous system (Khalsa et al., 2017). It is essential for survival as it allows the nervous system to be informed about physiological needs and maintain homeostasis (Craig, 2015). Interoception is thought to play a crucial role in emotional identification (Craig, 2004) and regulation (Füstös, Gramann, Herbert, & Pollatos, 2013) as well as in various other psychological phenomena (e.g., decision-making and body ownership; Tsakiris, Tajadura-Jiménez, & Costantini, 2011), therefore potentially explaining associations between interoceptive abilities and mental health (anxiety and depression; e.g., Paulus & Stein, 2010; Pollatos et al., 2008). As a result, interoception is also seen as a relevant candidate for clinical interventions.

At the conscious level, different dimensions of interoception have been proposed. Until 2015, the taxonomy for these dimensions varied a lot between researchers. Based on this observation, Garfinkel and colleagues (2015; 2013) proposed a three-dimensional model of interoception. This model, which has quickly been broadly endorsed by the scientific community (i.e., 388 citations of Garfinkel, 2015, in the Scopus database on January 26, 2021), encompasses three dimensions that can be defined as follows.

Interoceptive accuracy is the capacity to detect accurately and track internal sensations; it is assessed by behavioral performance measures. *Interoceptive sensibility* is the self-reported tendency to focus on internal sensations and the capacity to detect them. *Interoceptive awareness* is the correspondence between interoceptive accuracy and sensibility; it represents the degree to which interoceptive accuracy is predicted by self-report confidence in one's behavioral performance. With this conceptualization, Garfinkel and colleagues' main goal was to distinguish between objective, self-report, and "metacognitive" interoceptive processes.

Recently, a panel of interoception experts has formulated another taxonomy of interoception comprising eight features (Khalsa et al., 2017). This new taxonomy proposes additional features of interoception and, therefore, includes complementary measures (e.g., the perceived intensity of internal signals) that were excluded from the previous conceptualization. Among the eight features, two refer to self-report interoceptive processes. *Interoceptive sensibility* is defined as the self-perceived tendency to focus on interoceptive stimuli. According to Khalsa and colleagues (2017), this construct is well captured by the Multidimensional Assessment of Interoceptive Awareness (MAIA; Mehling et al. 2012; Mehling, Acree, Stewart, Silas, & Jones, 2018). *Interoceptive self-report scales* are defined as "the ability to reflect upon one's autobiographical experiences of interoceptive states, make judgments about their outcomes, and describe them through verbal or motor

* Correspondence to: UCLouvain-IPSY, 10 Place du Cardinal Mercier, B-1348 Louvain-la-Neuve, Belgium.
E-mail address: olivier.desmedt@uclouvain.be (O. Desmedt).

https://doi.org/10.1016/j.biopsycho.2022.108289
Received 30 January 2021; Received in revised form 25 January 2022; Accepted 7 February 2022
Available online 9 February 2022
0301-0511/

responses" (Khalsa et al., 2017). According to these authors, many measures may underly this dimension, such as the Visceral Sensitivity Index (VIS; Labus, Mayer, Chang, Bolus, & Naliboff, 2007), the Body Awareness Questionnaire (BAQ; Shields, Mallory, & Simon, 1989), the Body Perception Questionnaire (BPQ; Porges, 1993), and the MAIA (Mehling et al., 2012; Mehling etal., 2018).

Garfinkel, Seth, Barrett, Suzuki, and Critchley (2015) and Khalsa et al. (2017) should be commended for their conceptual efforts that allowed for more conceptual clarity and better communication between researchers. However, three important comments are in order here. First, these conceptualizations do not exactly ascribe the same meaning to "interoceptive sensibility". In Khalsa et al. (2017), the self-reported capacity to detect internal signals is removed from the "interceptive sensibility" construct. Hence, it is important to note that the same label relates to different understandings in the two taxonomies, preventing good communication between researchers.

Second, there has been a recent tendency to assign an ever-growing number of questionnaires to the "interoceptive sensibility" (as defined by Garfinkel et al., 2015) construct. Specifically, while "interoceptive sensibility" was initially restricted to the self-perceived ability to detect internal signals and the tendency to focus on them, it now additionally covers many other interoceptive-related (but perhaps distinct) features of interoception (e.g., the trust given to internal sensations, tendency to focus on internal sensations, awareness of symptoms, capacity to predict disease from symptoms, emotional awareness). Prominent self-report questionnaires of interoceptive sensibility include the MAIA (Mehling et al., 2012, 2018), the BPQ (Porges, 1993), the BAQ (Shields et al., 1989), or the Eating Disorder Inventory (EDI; Gardner, 1991; Garner, Olmsted, & Polivy, 1985).

Third, and directly relevant to the current research, no study has yet empirically investigated whether the construct(s) captured by these questionnaires (when considered altogether) do tap onto a common construct or different constructs. To address this question, we first performed a systematic review to identify the most frequently cited questionnaires of *interoceptive sensibility* or *interoceptive self-report scales*. Second, we ran an item-level exploratory factor analysis and network analysis on these questionnaires after completion by a large sample of respondents. The latent factor analysis allowed us to explore the factors underlying these questionnaires and to probe if it is empirically justified to assume one common factor. The network analysis examined interrelationships between the different questionnaires' items and tested whether the distinct questionnaires' items cohere as a unitary network or whether they constitute distinct communities (or subnetworks) of nodes. We used an exploratory approach since we had no strong theoretical expectations about the potential number of factors/communities covered by these questionnaires.

1. Systematic review

1.1. Method

This systematic review was conducted with CADIMA and in accordance with PRISMA guidelines (Moher, Liberati, Tetzlaff, Altman, & Group, 2009).

1.2. Preregistration and data sharing

The systematic review was preregistered at https://osf.io/fzreh/.

1.3. Eligibility criteria

We only included articles that identified interoception questionnaire (s) as measure(s) of interoceptive sensibility (Garfinkel et al., 2015; Khalsa et al., 2017) and interoceptive self-report scales (Khalsa et al., 2017). This means that included articles should not necessarily have administered the questionnaire (e.g., they cited a given questionnaire in their introduction), but instead, have merely named (and possibly also administered) it. No other restriction was made.[1]

1.4. Search strategy

A systematic literature search (see Table S1 in Supplementary Materials for the search strategy) was performed by the first author on PubMed, Scopus, PsychINFO, and ScienceDirect from April 2015—the date of publication of Garfinkel et al.'s (2015) study— to September 12, 2020, by restricting to peer-reviewed English papers. The following keywords were entered: ("interoceptive sensibility" or "interoceptive self-report scales") and ("questionnaire" or "inventory" or "scale" or "rating" or "instrument"). The term "interoceptive self-reported scales", proposed by Khalsa et al. (2017), was never used. We restricted our search strategy to these terms and these publication dates, as our goal was not to identify *all* questionnaires of interoceptive-related phenomena, but rather to identify the questionnaires that are acknowledged to assess the dimensions proposed by Garfinkel et al. (2015) and Khalsa et al. (2017).

1.5. Study selection

Titles, abstracts, and full texts were screened by the first author.

1.6. Data collection process

The first author extracted the title, the reference, the questionnaire (s) identified as measure(s) of interoceptive sensibility or interoceptive self-report scales, and the endorsed definition (Garfinkel et al.'s one vs Khalsa et al.'s one).[2] The outcome of interest was the frequency with which each questionnaire had been cited in the selected studies. Fig. 1.

2. Results

2.1. Selection process

2.1.1. Synthesis of results

The results of individuals studies have been summarized in Table S2 in the Supplementary Materials. Sixty-eight studies met the inclusion criteria. Included articles cited one to five relevant questionnaires. Fourteen questionnaires were identified: BPQ (34%), MAIA (32%), BAQ (9%), Private subscale of the Body Consciousness Questionnaire (PBCS; 8%; Miller, Murphy, & Buss, 1981), Self-Awareness Questionnaire (SAQ; 3.5%; Longarzo et al., 2015), EDI (3%), Body Sensations Questionnaire (2%; Chambless, Caputo, Bright, & Gallagher, 1984), Five Facet Mindfulness Questionnaire (1%; Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006), Emotional Susceptibility Scale (1%; Caprara et al., 1985), Autonomic Perception Questionnaire (1%; Mandler & Uviller, 1958), Visceral Sensitivity Index (1%; Labus et al., 2007), Body Vigilance Scale (1%; Schmidt, Lerew, & Trakowski, 1997), Somatic Absorption Scale (1%; Köteles, Simor, & Tolnai, 2012), and Interoceptive Awareness Questionnaire (1%; Bogaerts, Walentynowicz, Van Den Houte, Constantinou, & Van den Bergh, 2020). Fig. 2 shows the citation frequency of questionnaires.

[1] Systematic reviews and meta-analyses were not excluded, as they could cite interoceptive sensibility questionnaires in the introduction and the discussion. However, we did not count questionnaires that were cited while describing included studies of the systematic review/meta-analysis to avoid double-counting. This criterion was not pre-registered, as we did not anticipate this problem.

[2] This variable was not pre-registered. However, during the data extraction, we noticed that interoceptive sensibility was either conceptualized as Khalsa et al. (2017) or Garfinkel et al. (2015) definition.

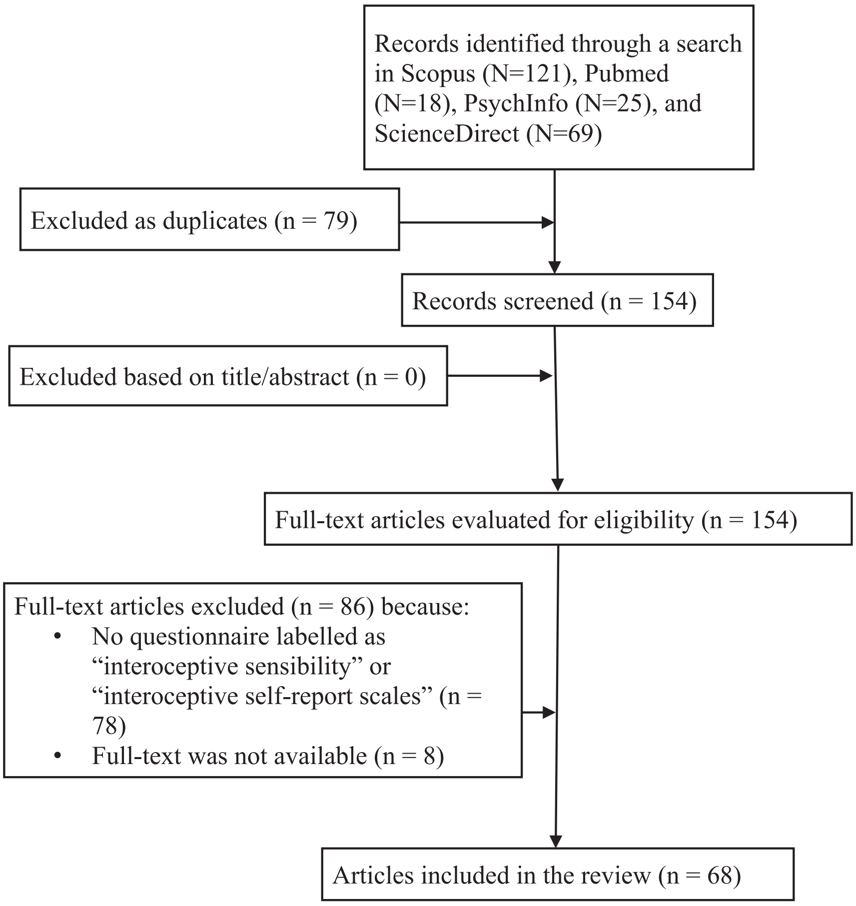

Fig. 1. Flow diagram.

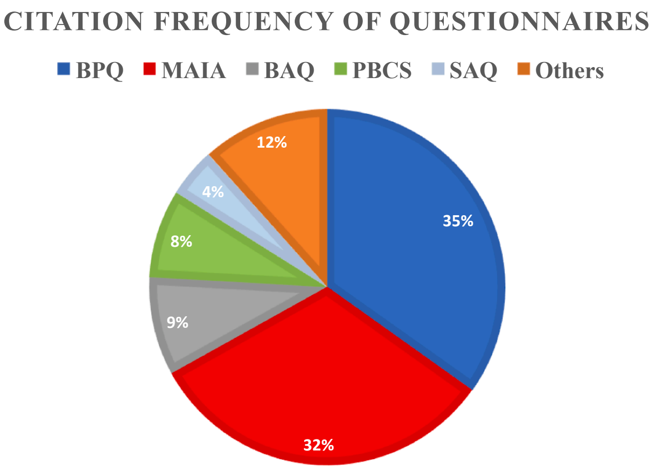

Fig. 2. ***Legend***. Others: EDI, Body Sensations Questionnaire, Five Facet Mindfulness Questionnaire, Emotional Susceptibility Scale, Autonomic Perception Questionnaire, Visceral Sensitivity Index, Body Vigilance Scale, Somatic Absorption Scale, Interoceptive Awareness Questionnaire.

3. Factor and network analyses

3.1. Method

3.1.1. Preregistration and data sharing

The study design, data collection, and analysis plan were preregistered at https://osf.io/fzreh/. Our R code and de-identified data are available at https://osf.io/e2ax7/files/.

3.2. Participants

We recruited 1003 participants (M_{age} = 35.57, SD_{age} = 12.77) on Prolific. There were 60.3% of women (N = 605), 39.6% of men (N = 397) and 0.1% of other (N = 1). Participants were from different ethnicity: Asian or Pacific Islander (7.78%), Black or African American (4.09%), Hispanic or Latino (1%), Native American or Alaskan Native (0%), White or Caucasian (81.65%), multiracial or biracial (4.29%), and others (1.49%). Participants had to (1) be between 18 and 70 years old, (2) have English as their first language, (3) live in the UK, Ireland, USA, Canada, or Australia,[3] (4) be free of current mental conditions, and (5) be free of any chronic diseases (e.g., diabetes, heart disease, stroke).

3.3. Materials and procedure

We expected that participants may experience fatigue and would be at risk of routine questionnaire completion in case the task was too long. Therefore, we selected only the most frequently used questionnaires of interoceptive sensibility or interoceptive self-report scales (as identified in the systematic review). The task was pre-tested for the duration. The presentation order of questionnaires, and the presentation order of items within questionnaires, were randomized.

[3] This criterion was not pre-registered. Given that Prolific includes, by default, many countries around the world, we thought it was important to restrict our study to English-speaking countries and ensure that participants had a sufficient level of English to correctly understand the items.

3.4. Statistical analyses

3.4.1. Outliers detection and handling

Multivariate outliers[4] were detected via Mahalanobis Distance larger than the critical chi-square value for df=the number of DVs at alpha = 0.001 (Mahalanobis, 1936). 126 multivariate outliers were detected. The final sample consisted of 877 participants.

3.4.2. Descriptive statistics

Descriptive statistics (M, SD, minimum, maximum, and range) for the items and scale scores can be found in the Table S3 et S4 of the Supplementary Materials.

3.4.3. Correlations between questionnaires

We performed Pearson correlations between each pair of questionnaires. We expected at least $r = 0.70$ to consider that two measures assess the same construct (Chmielewski, Sala, Tang, & Baldwin, 2016; Ruscio & Roche, 2012) and $r = 0.80$ to consider that they are interchangeable (Chmielewski et al., 2016).

3.4.4. Factor analysis

The item-level Exploratory Factor Analysis (EFA) was performed with *psych* (Revelle, 2020) and *GPArotation* (Jennrich, 2014) R packages. We first performed Bartlett's Test of Sphericity (Bartlett, 1954) and Kayser-Meyer-Olkin's (which should be above 0.7; Kaiser, 1970, 1974) test to verify correlation and sample adequacy with factor analysis. An oblique rotation (i.e., oblimin) was applied as we expected correlations between factors. The model was estimated with weighted least square (WLS) and polychoric correlations, as our variables were categorical (Costello & Osborne, 2005; Goretzko, Pham, & Bühner, 2019). To decide on the number of factors extracted,[5] we used the Scree Plot, parallel analysis (Horn, 1965),[6] comparison data (CD) approach (Ruscio & Roche, 2012) with Spearman correlations, and interpretability of factors. Results coming from these different methods were compared to reach a final decision. We also achieved a simple structure by running several rounds of analyses after having removed items that do not load ($r > 0.30$) on any factor and items that load on more than one factor. Finally, we also computed reliability statistics (i.e., Cronbach's alpha, Omega coefficients, and average inter-item correlation) and fit indices (i.e., the goodness of fit and residual statistics) including the Tucker-Lewis index (TLI), comparative fit index (CFI), root mean square error of approximation (RMSEA), and root means square of the residual (RMSR).

3.5. Network analysis

3.5.1. Network estimation

We present a graphical Gaussian model (GGM) that was regularized via the graphical LASSO (Least Absolute Shrinkage and Selection Operator) algorithm (Friedman, Hastie, Tibshirani, & Tibshirani, 2015), which has two main goals. First, it estimated regularized partial correlations between pairs of nodes, thereby excluding spurious associations (or edges) resulting from the influence of other nodes in the network. Second, it shrunk trivially small associations to zero, thus removing possibly "false positive" edges from the model and returning a sparser network including only the strongest edges (Epskamp, Borsboom, & Fried, 2018). We did so via the R package *qgraph* (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012; Epskamp et al., 2018), which automatically implements the graphical LASSO regularization in combination with the Extended Bayesian Information Criterion (EBIC) model selection (Foygel & Drton, 2011). In this procedure, 100 models with varying degrees of sparsity are estimated; a final model is selected according to the lowest EBIC value, given a specific hyperparameter gamma (γ), which regulates the trade-off between including false-positive edges and removing true edges. The hyperparameter γ can be set between 0 (favoring a model with more edges) and 0.5 (promoting a simpler model with fewer edges). Following recent recommendations (e.g., Epskamp et al., 2018), we set γ to 0.5 to be confident that our edges are true. To assess the accuracy of the edge weights, we implemented a nonparametric bootstrapping procedure (with 1000 bootstrapped samples with replacement) to bootstrap the edge weights' confidence regions using the R package bootnet (Epskamp et al., 2018). Using a bootstrapped difference test (Epskamp et al., 2018), we also examined whether the edge weights significantly differed from one another. Given the ordinal nature of the data, we based the network estimation on polychoric correlations to relax the assumption of normality. We did so via the implementation of the R built-in *cor_auto* function that automatically computes polychoric correlations when dealing with ordinal data.[7]

3.5.2. Centrality estimates

To quantify each node's importance in the regularized GGM, we computed the expected influence centrality indices (Robinaugh, Millner, & McNally, 2016). This centrality index quantifies the cumulative importance of each node and describes the sum of the edge weights attached to this node, considering both positive and negative values (Robinaugh et al., 2016). Hence, higher values indicate greater centrality in the network and so greater importance (Meessen et al., 2016). The plot represents the raw expected influence value of each node. The stability of this metric's estimates was assessed using the case-dropping subset bootstrap procedure (Costenbader & Valente, 2003), with 1000 bootstrapped samples. In this procedure, the correlation between the original centrality indices and the centrality indices as obtained from smaller subsets, with up to 75% of participants dropped, is assessed. To quantify the stability of the indices, we also calculated the centrality stability correlation coefficient (CS-coefficient). The CS-coefficient represents the maximum proportion of participants that can be dropped to maintain, with a 95%-probability, a correlation with the original centrality indices of at least.70. A minimum CS-coefficient of.25 (and preferably of at least.50) is recommended for interpreting centrality indices (Epskamp et al., 2018). Capitalizing on this case-dropping subset bootstrap procedure, we performed a bootstrapped difference test (Epskamp et al., 2018) to examine whether nodes significantly differ from one another in terms of expected influence.

3.5.3. Community detection

We investigated the GGM's community structure—that is, whether nodes (i.e., items) cohere as a unitary network structure or whether they cluster into distinct communities of nodes by implementing the Spinglass modularity-based community detection algorithm (Reichardt & Bornholdt, 2006). As in previous studies (Heeren, Bernstein, & McNally, 2018; Robinaugh, LeBlanc, Vuletich, & McNally, 2014), we chose this algorithm given its suitability for revealing the community structure of signed networks, i.e., networks including both positive and negative edge weight values (Traag & Bruggeman, 2009; Yang, Algesheimer, &

[4] Although we initially planned to identify univariate outliers (i.e., values greater/smaller than three times the interquartile), we dispensed with it since this approach is not optimal for ordinal variables as their values are constrained from 1 to n and thus, no answer can generally be considered as an outlier (Riani, Torti, & Zani, 2011).

[5] Although we initially planned to use the Kaiser criterion, we finally decided to remove this criterion following the comments of a reviewer pointing out that this method should not be used for EFA anymore, which is consistent with recent guidelines (Goretzko et al., 2019).

[6] We performed the parallel analysis based on the original dataset and with the eigenvalues of principal factor analysis.

[7] Note that we also estimated the GGM using Pearson Correlations and that the two networks were almost identical.

Tessone, 2016). We implemented this algorithm using the *spinglass.community* function of the R package *igraph* (Csardi & Nepusz, 2006).

Following previous studies (e.g., Bernstein, Heeren, & McNally, 2019; Heeren, Bernstein, & McNally, 2020), we also identified important nodes that may serve as bridges between the resultant communities by computing the bridge expected influence of each node. Bridge expected influence reflects the sum of all edges that exist between a given node and the nodes in the other communities. To do so, we relied on the bridge function of the R package *networktools* (Jones, Ma, & McNally, 2019). The plot represents the raw expected influence value of each node. As with the expected influence, the stability of the bridge centrality indices was assessed using a case-dropping subset bootstrap procedure and the computation of the related CS-coefficient. We likewise performed a bootstrapped difference test (Epskamp et al., 2018) to examine whether nodes significantly differ from one another in terms of bridge expected influence.

4. Results

4.1. Internal consistency of questionnaires

Internal consistency was excellent for BPQ ($\alpha = 0.96$ and $\omega = 0.97$) and SAQ ($\alpha = 0.93$ and $\omega = 0.94$), good for MAIA ($\alpha = 0.88$ and $\omega = 0.95$) and BAQ ($\alpha = 0.88$ and $\omega = 0.9$), and acceptable for PBCS ($\alpha = 0.69$ and $\omega = 0.73$), as indicated by Cronbach's alpha and omega coefficients. Average inter-item correlations were acceptable for MAIA ($r = 0.16$), SAQ ($r = 0.27$), PBCS ($r = 0.28$), BAQ ($r = 0.29$), and BPQ ($r = 0.47$), as they fall within the range 0.15–0.50 (Clark & Watson, 2016; Clark & Watson, 2019). Overall, the internal consistency of questionnaires suggests that our data are reliable.

4.2. Correlation between questionnaires

We first tested the bivariate correlation for each pair of questionnaires. We found correlations ranging from $r = 0.03$–0.55 (see Table 1). These correlations were surprisingly low for questionnaires that should measure the same construct (see Discussion).

4.3. Factor analysis

An item-level exploratory factor analysis (EFA) was used to investigate the underlying factors in these questionnaires. One hundred and twenty-six multivariate outliers were detected and removed based on the Mahalanobis distance.[8] Multivariate assumptions (i.e., additivity, normality, linearity, homogeneity, and homoscedasticity) were met. Regarding normality, skewness values were within the acceptable range -2 and $+2$ for all items (except for SAQ 35) (George & Mallery, 2010). Kurtosis values were mostly within the range -7 and 7 (expect for SAQ 16 and 29) which is also considered to be acceptable (Byrne, 2010; Hair, Black, Babin, & Anderson, 2010). However, we performed estimation methods that do not assume normality of the data (e.g., WLS estimation with polychoric correlations and Spearman correlations).[9] No missing data was found. KMO test ($MSA = 0.95$) and Barlett's test ($X^2(7260) = 54{,}387.39$, $p < .001$) indicated a good sampling and correlation adequacy, respectively.

The number of factors to extract depended on the factor retention criterion. The comparison data approach, Scree Plot, and the parallel analysis suggested 5, 5 or 6, and 13 factors, respectively. We thus tried to extract 5–13 factors and selected the most interpretable one. The 7- to 13-factor solutions were particularly difficult to interpret. In particular, the seventh to thirteenth factors had very few items (and sometimes no item) loading on only one factor, rendering the factor solution particularly difficult to interpret. For instance, the eleventh factor comprised items related to the trustworthiness of body sensations and the awareness of palms sweating. The 5 and 6-factor solutions were much more interpretable. The five-factor and six-factor solutions were very similar for the first four factors. In the five-factor solution, the fifth factor includes items assessing the tendency not to distract from uncomfortable sensations, but the variable describing the tendency not to worry about these sensations is removed from the solution. On the contrary, the six-factor solution teases apart these two tendencies with the fifth and sixth factors.

Next, we tried to achieve a Very Simple Structure (VSS) for the five- and six-factor solutions by removing items that did not load ($r < 0.30$) on any factor or that loaded on several factors. The VSS allows identifying a factorial structure that is theoretically meaningful (Revelle & Rocklin, 1979). When applying this procedure, the last factor of the six-factor solution was not composed of any marker item (i.e., items loading sufficiently ($r > 0.30$) on one factor and not splitting across factors) anymore. We thus proceeded with the five-factor solution. The VSS procedure led us to remove 14 items, the remaining items being marker items.

The final solution was the following (see Table 2 for the standardized loadings and Table S5, in the Supplementary Materials, for actual items)[10]:

- Factor 1 – *Neutral and Negative Body Sensations Awareness*: It included all items (i.e., 26) of the BPQ that assesses awareness of neutral and uncomfortable bodily sensations (e.g., "During most situations I am aware of watering or tearing of my eyes"). It has also been proposed that the BPQ assesses the tendency to be aware of sensations felt during anxiety, explaining its association with anxiety questionnaires (Trevisan, Mehling, & McPartland, 2021).
- Factor 2 – *Adaptive Interoception*: It included 25 (out of 37) items from the MAIA assessing the adaptive relationship with body sensations. In particular, (1) the capacity to notice and focus on uncomfortable, comfortable, and neutral body sensations (e.g., "I notice where in my body I am comfortable" and "I can pay attention to my breath without being distracted by things happening around me"), (2) the capacity to regulate distress by attention to body sensations, (3) active listening to the body for insight (e.g., "when I feel overwhelmed I can find a calm place inside"), and (4) experience of one's body as safe and trustworthy (e.g., "I feel my body is a safe place").
- Factor 3 – *Negative feelings propensity*: It included 34 (out of 35) items from the SAQ assessing the frequency with which one feels uncomfortable, painful, or symptomatic bodily sensations (e.g., "I feel a pain that seems to migrate around the body")
- Factor 4 – *Extero-Interoceptive Awareness*: It included 2 (out of 5) items from the PBCS assessing hunger and temperature sensations, as well as 17 (out of 18) items from the BAQ assessing the capacity to

Table 1
Pearson correlations between total scores of questionnaires.

	BPQ	MAIA	BAQ	PBCS	SAQ
BPQ	–				
MAIA	0.21[a]	–			
BAQ	0.26[a]	0.56[a]	–		
PBCS	0.35[a]	0.34[a]	0.41[a]	–	
SAQ	0.48[a]	0.03	0.14[a]	0.32[a]	–

[a] $< .001$

[8] Analyses were performed with and without outliers. Similar conclusions were reached with outliers. Data and R codes for analyses with outliers can be found at https://osf.io/e2ax7/files/. Outputs are available upon request.

[9] We also performed the analyses with Pearson correlations that suggested six (vs. five) factors. We discuss it in the discussion.

[10] Factor labels were chosen based on our content analysis.

Table 2
5-Factor Model Standardized Loadings.

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
BPQ_1	**0.74**	0.07	0.01	0.03	0.03
BPQ_2	**0.75**	-0.04	-0.01	-0.07	-0.08
BPQ_3	**0.77**	0.01	-0.01	0.02	-0.02
BPQ_4	**0.67**	0.11	0.11	0.01	0.11
BPQ_5	**0.84**	-0.07	-0.03	-0.02	-0.04
BPQ_6	**0.68**	0.02	-0.05	0.07	-0.08
BPQ_7	**0.77**	-0.01	-0.01	0.04	0.02
BPQ_8	**0.85**	-0.12	-0.04	-0.03	-0.07
BPQ_9	**0.69**	0.11	-0.01	0.06	0.03
BPQ_10	**0.65**	0.05	0.02	0.07	0.02
BPQ_11	**0.57**	0.15	0.18	-0.01	0.06
BPQ_12	**0.79**	-0.07	0.03	-0.03	-0.01
BPQ_13	**0.79**	0.05	-0.06	0.02	-0.09
BPQ_14	**0.74**	0.02	-0.12	0.10	-0.07
BPQ_15	**0.70**	0.05	0.07	-0.02	0.09
BPQ_16	**0.80**	-0.09	0.02	-0.05	-0.01
BPQ_17	**0.72**	0.09	0.10	-0.06	0.13
BPQ_18	**0.69**	-0.04	0.00	-0.03	-0.02
BPQ_19	**0.62**	0.07	0.07	0.03	0.01
BPQ_20	**0.54**	0.03	0.08	0.00	0.06
BPQ_21	**0.71**	0.09	-0.03	0.03	0.05
BPQ_22	**0.71**	-0.02	0.12	-0.03	0.02
BPQ_23	**0.63**	0.14	-0.04	0.03	0.00
BPQ_24	**0.78**	0.03	0.02	-0.01	-0.02
BPQ_25	**0.66**	0.13	0.08	-0.01	0.18
BPQ_26	**0.78**	-0.06	-0.09	0.03	0.00
MAIA_1	0.08	0.01	**0.47**	0.18	-0.01
MAIA_2	0.18	0.05	**0.36**	0.22	0.00
MAIA_3	0.09	-0.04	**0.55**	0.16	-0.01
MAIA_4	0.15	0.12	**0.43**	0.08	0.06
MAIA_5	0.01	-0.07	-0.05	-0.03	**0.67**
MAIA_7	-0.04	-0.01	-0.24	-0.06	**0.65**
MAIA_8	0.04	-0.08	-0.09	0.04	**0.70**
MAIA_16	-0.04	-0.01	**0.69**	-0.05	-0.04
MAIA_17	0.08	0.07	**0.65**	0.04	0.03
MAIA_18	-0.01	0.00	**0.51**	0.02	-0.05
MAIA_19	0.04	0.06	**0.77**	-0.10	-0.11
MAIA_20	0.00	0.05	**0.73**	0.03	-0.02
MAIA_21	-0.01	-0.02	**0.62**	0.07	-0.17
MAIA_22	0.01	0.01	**0.75**	0.02	-0.07
MAIA_23	0.07	0.09	**0.40**	0.22	0.04
MAIA_25	0.04	0.04	**0.53**	0.20	0.03
MAIA_26	0.08	-0.03	**0.51**	0.22	0.01
MAIA_27	0.02	0.07	**0.53**	0.28	0.07
MAIA_28	-0.11	-0.04	**0.76**	-0.07	-0.04
MAIA_29	-0.02	0.01	**0.65**	0.11	-0.02
MAIA_30	0.00	0.01	**0.68**	-0.04	-0.02
MAIA_31	0.01	0.00	**0.75**	-0.08	0.03
MAIA_32	0.02	0.06	**0.54**	0.26	0.10
MAIA_33	0.01	0.15	**0.62**	0.08	0.14
MAIA_34	0.05	-0.01	**0.61**	0.19	0.03
MAIA_35	-0.03	-0.26	**0.54**	0.03	-0.05
MAIA_36	0.02	-0.20	**0.62**	0.00	-0.08
MAIA_37	0.09	-0.19	**0.60**	0.15	-0.08
BAQ_1	0.06	0.09	0.06	**0.57**	-0.04
BAQ_2	0.04	0.00	-0.04	**0.50**	0.00
BAQ_3	0.02	-0.01	0.04	**0.54**	-0.10
BAQ_4	0.02	0.10	0.07	**0.60**	0.00
BAQ_5	-0.03	0.11	0.01	**0.51**	0.02
BAQ_6	-0.02	-0.01	0.08	**0.48**	-0.04
BAQ_7	0.00	-0.12	0.03	**0.59**	-0.02
BAQ_8	-0.08	0.00	0.00	**0.66**	-0.02
BAQ_9	-0.02	-0.01	0.15	**0.59**	-0.01
BAQ_11	-0.02	0.01	0.05	**0.62**	0.02
BAQ_12	0.03	0.00	0.01	**0.48**	0.07
BAQ_13	0.03	0.04	-0.01	**0.60**	-0.06
BAQ_14	0.03	0.10	0.05	**0.47**	0.00
BAQ_15	0.01	-0.16	0.08	**0.57**	0.06
BAQ_16	0.02	-0.07	0.12	**0.64**	0.01
BAQ_17	0.01	-0.11	0.00	**0.54**	0.04
BAQ_18	0.06	0.04	-0.05	**0.57**	-0.01
PBCS_4	0.06	0.05	-0.02	**0.40**	0.00
PBCS_5	0.05	0.15	0.10	**0.33**	0.04
SAQ_1	0.03	**0.52**	0.02	-0.01	-0.08

Table 2 (*continued*)

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
SAQ_2	0.10	**0.54**	-0.01	0.05	0.15
SAQ_3	-0.01	**0.58**	0.13	-0.08	0.09
SAQ_4	0.02	**0.51**	0.00	0.04	-0.02
SAQ_5	-0.02	**0.62**	0.10	-0.02	0.02
SAQ_6	0.02	**0.69**	-0.12	0.09	0.04
SAQ_7	0.03	**0.66**	0.08	-0.03	0.01
SAQ_8	0.00	**0.55**	-0.01	-0.06	-0.15
SAQ_9	0.00	**0.54**	-0.09	0.20	-0.01
SAQ_10	0.01	**0.49**	-0.06	0.05	-0.12
SAQ_11	0.06	**0.54**	-0.06	-0.07	-0.15
SAQ_12	0.06	**0.66**	-0.06	-0.02	-0.06
SAQ_13	-0.02	**0.64**	0.14	-0.08	0.13
SAQ_14	0.17	**0.44**	-0.14	0.09	-0.14
SAQ_15	0.14	**0.48**	-0.09	0.09	-0.13
SAQ_16	-0.03	**0.68**	0.08	-0.01	0.05
SAQ_17	0.01	**0.67**	0.07	-0.01	-0.01
SAQ_18	0.04	**0.66**	-0.01	-0.03	-0.09
SAQ_20	0.20	**0.42**	-0.10	0.02	-0.11
SAQ_21	0.05	**0.64**	0.00	-0.07	0.03
SAQ_22	0.19	**0.45**	-0.11	0.00	-0.13
SAQ_23	0.11	**0.31**	0.08	-0.05	-0.06
SAQ_24	0.01	**0.64**	-0.10	0.06	-0.09
SAQ_25	0.11	**0.56**	-0.06	0.06	-0.07
SAQ_26	-0.02	**0.75**	0.05	-0.05	0.03
SAQ_27	0.04	**0.66**	0.08	-0.06	0.15
SAQ_28	0.07	**0.55**	-0.03	0.12	-0.07
SAQ_29	-0.13	**0.75**	0.02	0.07	-0.01
SAQ_30	-0.04	**0.73**	-0.08	0.07	-0.03
SAQ_31	0.08	**0.55**	0.09	-0.02	0.04
SAQ_32	0.07	**0.63**	0.05	-0.08	-0.01
SAQ_33	0.01	**0.61**	-0.06	0.03	-0.02
SAQ_34	0.22	**0.42**	0.06	-0.04	0.01
SAQ_35	0.04	**0.68**	-0.06	-0.02	0.02

Items' labels are available in Table S5 in Supplementary Materials.

notice and predict body reactions to internal and external factors such as weather, seasons, foods, blows, diseases (e.g., the flu), and energy level (e.g., "I can accurately predict what time of day lack of sleep will catch up with me").

- Factor 5 – *Interoceptive Not-Distracting*: It included 3 (out of 37) items from the MAIA assessing the tendency not to ignore or distract oneself from sensations of pain or discomfort (e.g., "I try to ignore pain" [reversed]).

Then, we tested the internal reliability of each factor and computed the fit indices of the 5-factor solution. The reliability was excellent for Factors 1 ($\alpha = 0.96$), 2 ($\alpha = 0.94$), and 3 ($\alpha = 0.94$), good for Factor 4 ($\alpha = 0.89$), and acceptable for Factor 5 ($\alpha = 0.75$). The 5-factor model had a moderate fit. The RMSR and RMSEA indicated good fit at 0.04 and 0.056 (90% *CI*[0.055, 0.057]), respectively. However, CFI (0.75) and TLI (0.72) were not good enough. Finally, we computed the inter-factor correlations (see Table 3).

4.4. Network analysis

4.4.1. GGM

Fig. 3 shows the estimated GGM network wherein edges represent

Table 3
5-Factor Model Standardized Loadings.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	–				
Factor 2	0.27[a]	–			
Factor 3	0.48[a]	0.09[b]	–		
Factor 4	0.27[a]	0.59[a]	0.15[a]	–	
Factor 5	-0.07[a]	-0.22[a]	-0.10[a]	-0.14[a]	–

[a] < .001
[b] < .05

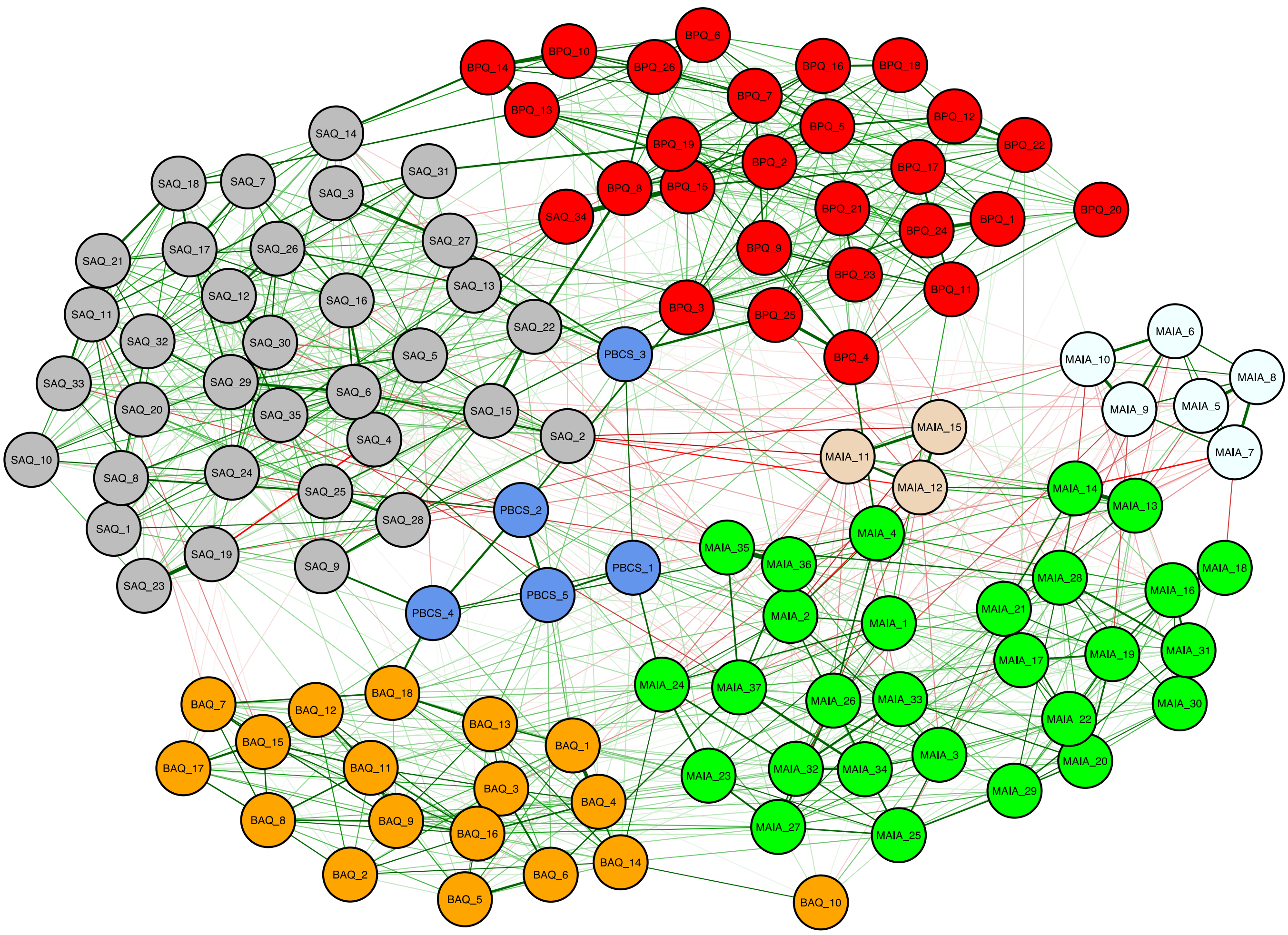

Fig. 3. Graphical Gaussian model and community detection, ***Note.*** The thickness of an edge reflects the magnitude of the association (the thickest edge representing a value of.46). Items ' labels are detailed in Table 3 in Appendices. Node's color reflects the community structure (see the Community detection analysis).

regularized partial correlations between nodes (i.e., the items). The thickness of the edge denotes the strength of the association, with a thicker edge indicating a larger association. Green edges represent positive regularized partial correlations, whereas red ones represent negative regularized partial correlations. The force-directed Fruchterman and Reingold's (1991) layout algorithm was used to visualize the network, which pulls nodes with the strongest associations nearer to the center of the graph.

Several features were immediately apparent. First, the thickest edges were noticeable between items coming from the same questionnaire. For instance, there were many strong associations between nodes denoting SAQ's items or BPQ's items. The only exception was BAQ_10 (the only reversed item of the BAQ), which was barely connected to the other BAQ's items but showed a few thin associations with MAIA's items. Second, most of the thinnest edges were between items of different questionnaires. For instance, only a few thin edges connected SAQ's nodes to BAQ's nodes and MAIA's nodes to BPQ's nodes. The only thick edge denoting a cross-questionnaire association was between SAQ_34 ("I feel my palms sweaty") and BPQ_15 ("During most situations, I am aware of palms sweating"). Third, six MAIA's items (i.e., MAIA_5; MAIA_6; MAIA_7; MAIA_8; MAIA_9; MAIA_10) emerged as a distinct cluster of nodes that were connected via negative edges to the other items of the MAIA. These items refer to the tendency not to ignore or distract oneself from sensations of pain or discomfort (e.g., "I try to ignore pain" [reversed]). Finally, three MAIA's items (i.e., MAIA_11, MAIA_12, and MAIA_15) formed another distinct cluster of nodes that all allude to not worrying about uncomfortable body sensations (e.g., "I start to worry that something is wrong if I feel any discomfort" [reversed]).

The nonparametric bootstrapping procedure showed that the bootstrapped CIs for the edge-weights were small (see Fig. S1 in the Supplementary Materials), thus indicating that they were fairly accurate.

4.4.2. Centrality estimates

BPQ_13 ("During most situations, I am aware of stomach and gut pains"), BPQ_14 ("During most situations, I am aware of stomach distension or bloatedness"), BPQ_15 ("During most situations, I am aware of palms sweating"), and SAQ_27 ("I feel my heart thudding") were the nodes yielding the highest expected influence values (values are provided in Fig. S2 in the Supplementary Materials). In contrast, BAQ_10 ("I don't notice seasonal rhythms and cycles in the way my body functions" [Reversed]), MAIA_11 ("When I feel physical pain, I become upset" [Reversed]), and MAIA_12 ("I start to worry that something is wrong if I feel any discomfort" [Reversed]) were the nodes with the lowest expected influence values.

Case-dropping subset bootstrap indicated the stability of this centrality metric in the present sample (see Fig. S3 in the Supplementary materials). The CS-coefficient was.52. The bootstrapped different test confirmed that SAQ_27, BPQ_13, BPQ_14, and BPQ_15 were significantly more central than the remaining nodes.

4.4.3. Community detection

The spinglass algorithm detected 7 communities ("subnetworks") of nodes. The communities are represented by distinct colors in Fig. 3. The first community (i.e., red nodes in Fig. 3) comprised all the BPQ's items and the item SAQ_34. The second community (i.e., orange nodes in Fig. 3) comprised all the BAQ items. The third community (i.e., blue nodes in Fig. 3) comprised all the PCBS items. The fourth community (i. e., grey nodes in Fig. 3) consisted of all the SAQ items except the SAQ_34. The fifth community consisted of the MAIA_5, MAIA_6, MAI_7, MAIA_8, MAIA_9, and MAIA_10 (i.e., light blue nodes in Fig. 3). The sixth community included the MAIA_11, MAIA_12, and MAI_15 (i.e., sandy tan nodes in Fig. 3). The seventh community included the remaining MAIA's items: from MAIA_1 to MAIA_4 and from MAIA_16 to MAIA_37 (i.e., green nodes in Fig. 3).

The bootstrapped different test revealed that PBCS_3, SAQ_34, BPQ_25, PBCS_2, BPQ_3, and SAQ_14 yielded significantly higher bridge expected value than the remaining nodes (see Fig. S5 and Fig. S6 in the Supplementary materials). The case-dropping subset bootstrap procedure confirmed the stability of this metric (see Fig. S3 in the Supplementary materials) and the CS-coefficient was.51.

5. Discussion

The main question we sought to address in the present research was whether various questionnaires thought to measure interoceptive sensibility (and "interoceptive self-report scales") do indeed measure a common construct (i.e., have acceptable convergent validity). To do so, we (1) identified in a systematic review the most frequently cited questionnaires of interoceptive sensibility, and we (2) examined their correlations, (3) their overall factorial structure, and (4) their network structure. In the general discussion, we summarize the main results and discuss their implications for the interpretation and replicability of the findings. We also discuss future directions for interoception research.

5.1. What are the most frequently mentioned questionnaires and their inter-relations?

The first aim was to identify the most frequently mentioned interoceptive sensibility questionnaires. Our systematic review indicated that no less than 14 different questionnaires had been identified to measure interoceptive sensibility. Among them, five (BPQ, MAIA, BAQ, SAQ, and PBCS) cover 86.5% of all citations of interoceptive sensibility questionnaires. Correlational analyses revealed inexistent to moderate associations between these questionnaires (range from $r = 0.03$–0.55). This finding casts preliminary doubts on the assumption that the most frequently cited questionnaires of interoceptive sensibility measure a common construct. Even a correlation of 0.55 does not reach satisfactory convergent validity (that should be above 0.70; Carlson & Herdman, 2012), and this may be consequential (see below).

It should be noted that our search terms (e.g., "interoceptive sensibility") were quite literal, as we used the terms proposed by the recent conceptualizations of interoception. Therefore, we may have missed some papers containing alternative terminologies for these constructs (e. g., subjective interoception). However, these alternative terms have been seldom used, suggesting that their inclusion/exclusion should have a low influence on the results. Moreover, including these terms would have increased the probability of including measures designed to assess (slightly) different constructs. Finally, the use of literal terms should have increased the chance of finding convergence, not divergence, between the selected questionnaires.

5.2. Self-report interoception: a single dimension?

Another central question was to test whether the items from the different "interoceptive sensibility" questionnaires assess a single dimension. Both the factor and the network approaches supported the existence of one factor/community per questionnaire except for the PBCS which was partly included in the factor associated with the BAQ and the MAIA, which appeared to be underpinned by 2 factors or 3 communities. One subscale from the MAIA (i.e., Not-worrying; the tendency not to worry about sensations of pain or discomfort) was removed from the final factor solution because its items loaded on several factors. However, the network approach identified this MAIA subscale and the PBCS as subnetworks, although some items (e.g., "I am sensitive to internal bodily tensions" [PBCS_1]; "I can often feel my heart beating" [PBCS_3]) show strong associations with items belonging to a different questionnaire (see the results of the bridge centrality analysis). These results, therefore, suggest that the various questionnaires measure different constructs that cannot be subsumed under a common conceptual umbrella. This is further supported by the inter-factor correlations that range from $r = -0.07$ to $r = 0.59$, which suggests that these questionnaires measure related, but distinct, constructs. In short, interoception researchers wrongly assume that they work on the same construct when they use or compare responses on different questionnaires (i.e., a jingle fallacy).

This conclusion is consistent with the fact that all these questionnaires have been developed in different contexts with different purposes, although they are now reported to measure an identical construct. The BPQ was aimed at assessing the level of body awareness, and more specifically the awareness of the autonomic nervous system reactivity (Porges, 1993). The MAIA was the result of a more complex and multidimensional perspective on "body awareness", which includes both interoceptive and proprioceptive signals and distinguishes modes of attention such as mindfulness (Mehling et al., 2012). This questionnaire can be used in experimental interoception research and for the assessment of mind-body therapies. The BAQ was developed to assess the self-reported tendency to attend to normal, non-emotive, body processes (vs. somatic complaints or emotion-related sensations; Shields et al., 1989). In particular, it aimed to evaluate the sensitivity to body cycles and rhythms, the ability to detect small changes in normal functioning, and the ability to anticipate bodily reactions. The BCS was developed to distinguish between the private (what cannot be observed by others; e.g., stomach gurgles) and public (what can be observed by others; e.g., posture) aspects of body awareness. Finally, the SAQ was built to measure the perception of a wide range of bodily sensations among non-clinical individuals (Longarzo et al., 2015). More specifically, it evaluates how and how frequently individuals feel signals from their own body.

Finally, one may argue that confirmatory factor analysis is better suited for testing whether different questionnaires assess a single dimension. However, we had no alternative model to compare with, as there are no strong theoretical expectations about the number of factors underlying self-report measures of interoception. Moreover, we were genuinely interested in identifying the potential constructs measured by these questionnaires using a bottom-up approach. For these reasons, the exploratory approach was the most relevant. In future studies, confirmatory factor analyses (CFA) should be performed to test if a five-factor solution should be favored over a single-factor solution or other solutions. This is particularly important as some fit indices (i.e., CFI and TLI) suggested that the five-factor solution may not be optimal.

5.3. What are the consequences for the reliability and validity of results?

This is most concerning for the interpretation and the replicability of findings in interoception research. The current analysis suggests that results found with one specific measure (e.g., BPQ) are unlikely to be replicated with another measure (e.g., BAQ), given the low correlations between questionnaires and the presence of almost one factor/community by questionnaire. Carlson and Herdman (2012) have indeed mathematically demonstrated that, if two measures (a and b) correlate to $r_{a,b} = 0.50$ and the first measure (a) correlates with an outcome (y) to $r_{a,y} = 0.30$, the correlation between the second measure (b) and this outcome (y) can range from $r_{b,y} = -0.68$ to 0.98. Hence, assuming that two measures assess the same construct while evidence indicates only a moderate association between the two can lead to huge between-studies heterogeneity and, as a result, to low robustness/replicability of the results.

This is also particularly concerning for meta-analyses that aggregate "interoceptive sensibility" measures. For instance, a recent meta-analysis by Trevisan et al. (2019) tested the relationship between alexithymia and interoceptive sensibility (among other dimensions) by aggregating several loosely related questionnaires of interoception (Desmedt, Luminet, & Corneille, 2021; Trevisan et al., 2021). Given that these questionnaires measure different constructs, the results cannot be meaningfully interpreted (Trevisan et al., 2021), which questions the validity of the meta-analytical conclusions.

This situation is also problematic with regards to the construct validity of self-report measures of interoception. An important gap exists between the definition of interoceptive sensibility (Garfinkel et al., 2015; Khalsa et al., 2017) and the constructs that are measured by questionnaires meant to indicate it. As a reminder, interoceptive sensibility is (1) the self-perceived tendency to focus on internal sensations and the capacity to detect them (Garfinkel et al., 2015) or (2) only the self-perceived tendency to focus on internal sensations (Khalsa et al., 2017). Our systematic review indicates that the first definition is the most largely shared across authors. However, a content analysis of the items used in the questionnaires included in this study indicates that the questionnaires measure distinct—and often more specific—constructs. Here are a few examples. The BPQ measures awareness of neutral and uncomfortable bodily sensations. The BAQ measures the capacity to notice and predict body reactions to internal and external factors such as diseases (e.g., the flu) and food. Finally, the SAQ measures the frequency with which one feels uncomfortable, painful, or symptomatic bodily sensations. This discrepancy between construct definition and measurement might lead to invalid interpretations, as researchers might draw conclusions on the role of interoceptive sensibility without considering what the specific questionnaire is actually measuring.

5.4. *What are the consistencies and discrepancies between the factor and network analyses?*

Results from the factor and network approaches were generally very consistent, which is not surprising given the assumed mathematical equivalence between them (Christensen, Golino, & Silvia, 2020; but see, Bringmann & Eronen, 2018 for a discussion on the ontological differences). However, a few minor discrepancies arose between the two approaches. First, the network analysis identified two additional communities covered by PBCS items and one subscale of the MAIA (Not-worrying). On the contrary, the factor analysis showed that some PBCS items loaded on the same factor as the BAQ and the MAIA subscale were removed from the final solution following the VSS procedure. It should, however, be noted that when we estimated the factor structure with Pearson (vs. polychoric) correlations, the subscale Not-worrying emerged, leading to a six-factor solution. This is consistent with previous studies that repeatedly supported the existence of this construct by confirmatory factor analyses (Ferentzi et al., 2020; Mehling et al., 2012, 2018). One explanation for these inconsistencies is the low discriminant validity of items from the subscale Not-worrying, as the VSS procedure removed items having cross-loadings. This is also true for the subscale Not-distracting that only contained three marker items. This low discriminant validity could suggest either that these constructs are sub-dimensions being subsumed by a broader construct or that they are not well captured by current subscales. Future studies should clarify this question.

Second, the network analysis assigned one item from the SAQ to the same community as the BPQ items. This can easily be explained by a strong association ($r = 0.63$) between BPQ_15 ("During most situations, I am aware of palms sweating") and SAQ_34 ("I feel my palms sweaty") that both evaluate the self-perceived awareness of palms sweating. These discrepancies could be partly explained by the VSS procedure (that only keeps marker items) and the arbitrary decisions applied in the factor analysis. In particular, several steps of the factor analysis are subject to the subjectivity of the researchers, as they have to choose, e.g., the (1) rotation method, (2) factor retention criteria, and (3) the most interpretable solution. However, even though subjective choices were made, the pre-registration and transparency behind the method should maximize the replicability of EFA findings.

Of note, some commentators have recently cast doubts about the suitability of network analyses for uncovering psychological data structure, and particularly because some centrality indices have appeared unstable in cross-sectional and temporal networks (Bringmann et al., 2019; Forbes, Wright, Markon, & Krueger, 2017; Hallquist, Wright, & Molenaar, 2019; but see Borsboom et al., 2017). However, most of these concerns focused on closeness and betweenness centrality metrics and not expected influence, which was the metric used here. Indeed, this latter metric has been typically reported as more stable (Robinaugh et al., 2016), an observation confirmed in this study, as expected influence estimates were highly stable (see Supplementary Materials).

5.5. *What are the future directions?*

The present research questions the pragmatic value of the "interoceptive sensibility" construct, as it is dissociated from its measurement (various questionnaires used to capture the construct do not converge). We have discussed above that this is not just a mere terminological issue, but one that has the potential to greatly threaten interpretations, replications, and more generally a sound communication between researchers in interoception research.

Together with other empirical studies, and more generally, the present research also questions the *pragmatic* value of current conceptualizations of interoception. This is because, besides interoceptive sensibility, low convergence also applies to other dimensions of interoception. Notably, recent evidence shows that measures of interoceptive accuracy also loosely correlate within (as for HCT and HDT outcomes; Ring & Brener, 2018) and between bodily domains (see a review and empirical evidence in Ferentzi et al., 2018). This is not to say that the current conceptualizations are faulty. On the contrary, we see merits in those. Rather, we note that current conceptualizations are disconnected from current measurement (as various measures do not converge).

We call for the development of measures that better fit with current conceptualizations (or vice-versa). Ideally, we should endorse the construct validity approach to scale development (Chmielewski et al., 2016; Clark & Watson, 2016, 2019; Loevinger, 1957; Strauss & Smith, 2009; Watson, 2012), which will both clarify the constructs (i.e., self-report dimensions of interoception) and create psychometrically sound measures of these constructs. In their guidelines, Clark and Watson (2019) propose 13 steps for the validation process that goes from the clear conceptualization of target constructs to the use of cross-method analyses (e.g., questionnaires and interviews). Although this validation process is time-consuming, we believe it is essential to test validly the theory and the efficacy of interventions (e.g., mindfulness) on interoceptive dimensions. A helpful development in interoception research might consist in differentiating self-report measures of interoception according to the phenomena of interest (e.g., attention vs accuracy; Gabriele, Spooner, Brewer, & Murphy, 2020; Jennifer Murphy et al., 2019), bodily domain (e.g., cardiac), or system (e.g., cardiorespiratory; Wang, Tan, Van den Bergh, von Leupoldt, & Qiu, 2020), and type of physiological activation (e.g., activation vs deactivation; Vlemincx, Walentynowicz, Zamariola, Van Oudenhove, & Luminet, 2020). Recent promising questionnaires have been recently developed that pursue this goal: e.g., The Three-domain Interoceptive Sensations Questionnaire (THISQ; Vlemincx et al., 2020) and the Interoceptive Accuracy Scale (IAS; Jennifer Murphy et al., 2019).

In the meantime, researchers should select questionnaires according to the specific construct they intend to measure. For instance, if researchers seek to measure the awareness of neutral and negative bodily sensations, they should administer the BPQ, but if they seek to measure the tendency not to worry about internal sensations, they should administer the MAIA subscale. This will allow a better correspondence between researchers' intentions, hypotheses, results, and interpretations (i.e., more replicable and valid research).

6. Conclusion

In conclusion, the present research indicates that current interoception questionnaires do not measure a common construct. Instead, these questionnaires inform distinct, insufficiently related, entities. This lack of empirical convergence between questionnaires threatens the validity of interpretations and the replicability and generalization of the findings. This sort of limitations is not specific to the interoception research, but also applies to a diversity of constructs, including "depression" (Santor, Gregus, & Welch, 2006), "fear extinction", "implicit attitudes" (e.g., Corneille & Hütter, 2020), "impulsivity" (e.g., Stahl et al., 2014), "physical activity" (Chmielewski et al., 2016), and "experiential avoidance" (Rochefort, Baldwin, & Chmielewski, 2018). Future research should better match conceptualization with measurement (for a discussion, see Flake & Fried, 2020). This can be achieved by adapting either the conceptualization or the measurement in interoception research.

Acknowledgments

Olivier Desmedt (Ph.D. student; grant number: 34226579), Olivier Luminet (Research Director), and Alexandre Heeren (Research Associate; Grant "1.C.059.18F") are funded by the Fund for Scientific Research – Belgium (FRS-FNRS). We would like to thank Jérémy Bena for his assistance in running simulations for the polychoric analyses.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopsycho.2022.108289.

References[11]

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. https://doi.org/10.1177/1073191105283504

Bartlett, M. S. (1954). A note on the multiplying factors for various χ 2 approximations. *Journal of the Royal Statistical Society. Series B Methodological*, 296–298.

Bernstein, E. E., Heeren, A., & McNally, R. J. (2019). Reexamining trait rumination as a system of repetitive negative thoughts: A network analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 63*, 21–27. https://doi.org/10.1016/j.jbtep.2018.12.005

Bogaerts, K., Walentynowicz, M., Van Den Houte, M., Constantinou, E., & Van den Bergh, O. (2020). The interoceptive awareness questionnaire (IAQ) differentiates between and within groups with stress-related bodily complaints versus healthy controls. *Manuscript in Preparation*.

Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., … Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology, 128*(8), 892–903. https://doi.org/10.1037/abn0000446.

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. *Psychological Review, 125*(4), 606.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.

Caprara, G. V., Cinanni, V., D'Imperio, G., Passerini, S., Renzi, P., & Travaglia, G. (1985). Indicators of impulsive aggression: Present status of research on irritability and emotional susceptibility scales. *Personality and Individual Differences, 6*(6), 665–674. https://doi.org/10.1016/0191-8869(85)90077-7

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods, 15*(1), 17–32. https://doi.org/10.1177/1094428110392383

Chambless, D. L., Caputo, G. C., Bright, P., & Gallagher, R. (1984). Assessment of fear of fear in agoraphobics: The body sensations questionnaire and the agoraphobic cognitions questionnaire. *Journal of Consulting and Clinical Psychology, 52*(6), 1090–1097. https://doi.org/10.1037/0022-006X.52.6.1090

Chmielewski, M., Sala, M., Tang, R., & Baldwin, A. (2016). Examining the construct validity of affective judgments of physical activity measures. *Psychological Assessment, 28*(9), 1128. https://doi.org/10.1037/pas0000322

Christensen, A. P., Golino, H., & Silvia, P. J. (2020). A psychometric network perspective on the validity and validation of personality trait questionnaires. *European Journal of Personality*.

Clark, L. A., & Watson, D. (2016). *Constructing validity: Basic issues in objective scale development*. https://doi.org/10.1037/14805-012

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Corneille, O, & Hütter, M. Implicit? (2020). What Do You Mean? A Comprehensive Review of the Delusive Implicitness Construct in Attitude Research. *Personality and Social Psychology Review, 24*(3), 212–232. https://doi.org/10.1177/1088868320911325

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), 7. https://doi.org/10.7275/jyj1-4868

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks, 25*(4), 283–307. https://doi.org/10.1016/S0378-8733(03)00012-1

Craig, A. D. (2004). Human feelings: Why are some more aware than others? *Trends in Cognitive Sciences, 8*(6), 239–241. https://doi.org/10.1016/j.tics.2004.04.004

Craig, A. D. (2015). *How do you feel?: An interoceptive moment with your neurobiological self*. Princeton University Press.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *International Journal of Complex Systems, 1695*(5), 1–9.

Desmedt, O., Luminet, O., & Corneille, O. (2021). More convergence is needed in the measurement of interoception. *Manuscript in Preparation*.

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(1), 1–18. https://doi.org/10.18637/jss.v048.i04

Ferentzi, E, Bogdány, T, Szabolcs, Z, Csala, B, Horváth, Á, & Köteles, F (2018). Multichannel Investigation of Interoception: Sensitivity Is Not a Generalizable Feature. *Front. Hum. Neurosci., 12*, 223. https://doi.org/10.3389/fnhum.2018.00223

Ferentzi, Eszter, Olaru, G., Geiger, M., Vig, L., Köteles, F., & Wilhelm, O. (2020). Examining the factor structure and validity of the multidimensional assessment of interoceptive awareness. *Journal of Personality Assessment*, 1–10. https://doi.org/10.1080/00223891.2020.1813147

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017). Evidence that psychopathology symptom networks have limited replicability. *Journal of Abnormal Psychology, 126*(7), 969–988. https://doi.org/10.1037/abn0000276.

Foygel R., Drton M. (2011). Bayesian model choice and information criteria in sparse generalized linear models. *ArXiv:1112.5635 [Math, Stat]*. ⟨http://arxiv.org/abs/1112.5635⟩.

Friedman, J., Hastie, T., Tibshirani, R., Tibshirani, M.R. (2015). Package 'glasso.'

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

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

Gardner, D. M. (1991). *Eating disorder inventory 2: Professional manual. Odessa, FL:* Psychological Assessment Resources.

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

Garner, D. M., Olmsted, M. P., & Polivy, J. (1985). Eating disorder inventory. *Psychopharmacology Bulletin, 21*(4), 1009–1010.

George D., Mallery M. (2010). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.) Boston: Pearson.

Goretzko, D., Pham, T. T. H., & Bühner, M. (2019). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 1–12.

Hair J., Black W.C., Babin B.J., Anderson R.E. (2010) Multivariate data analysis (7th ed.). Upper Saddle River, New Jersey: Pearson Educational International.

Hallquist, M. N., Wright, A. G. C., & Molenaar, P. C. M. (2019). Problems with Centrality Measures in Psychopathology Symptom Networks: Why Network Psychometrics Cannot Escape Psychometric Theory. *Multivariate Behavioral Research, 0*(0), 1–25. https://doi.org/10.1080/00273171.2019.1640103.

Heeren, A., Bernstein, E. E., & McNally, R. J. (2018). Deconstructing trait anxiety: A network perspective. *Anxiety, Stress, & Coping, 31*(3), 262–276. https://doi.org/10.1080/10615806.2018.1439263

Heeren, A., Bernstein, E. E., & McNally, R. J. (2020). Bridging maladaptive social self-beliefs and social anxiety: A network perspective. *Journal of Anxiety Disorders, 74*, Article 102267. https://doi.org/10.1016/j.janxdis.2020.102267

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

[11] References marked with an asterisk are included in the systematic review.

Jennrich, C.B.R(2014). GPArotation: GPA Factor Rotation (2014.11–1) [Computer software]. ⟨https://CRAN.R-project.org/package=GPArotation⟩.
Jones, P. J., Ma, R., & McNally, R. J. (2019). Bridge centrality: A network approach to understanding comorbidity. *Multivariate Behavioral Research, 0*(0), 1–15. https://doi.org/10.1080/00273171.2019.1614898
Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika, 35*(4), 401–415.
Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575
(*) Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., … Mehling, W. E. (2017). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*, 501–503. https://doi.org/10.1016/j.bpsc.2017.12.004.
Köteles, F., Simor, P., & Tolnai, N. (2012). Psychometric evaluation of the Hungarian version of the Somatic Absorption Scale. *Mentalhigiene Es Pszichoszomatika, 13*(4), 375–395. https://doi.org/10.1556/Mental.13.2012.4.2
Labus, J. S., Mayer, E. A., Chang, L., Bolus, R., & Naliboff, B. D. (2007). The central role of gastrointestinal-specific anxiety in irritable bowel syndrome: Further validation of the visceral sensitivity index. *Psychosomatic Medicine, 69*(1), 89–98.
Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological reports, 3*(3), 635–694.
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.
Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Science of India, 12*, 49–55.
Mandler, J. M., & Uviller, E. T. (1958). Autonomic feedback: The perception of autonomic activity. *The Journal of Abnormal and Social Psychology, 56*(3), 367–373.
(*) Meessen, J., Mainz, V., Gauggel, S., Volz-Sidiropoulou, E., Sütterlin, S., & Forkmann, T. (2016). The relationship between interoception and metacognition: A pilot study. *Journal of Psychophysiology, 30*(2), 76–86. https://doi.org/10.1027/0269-8803/a000157.
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). https://doi.org/10.1371/journal.pone.0208034
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), Article e48230. https://doi.org/10.1371/journal.pone.0048230
Miller, L. C., Murphy, R., & Buss, A. H. (1981). Consciousness of body: Private and public. *Journal of Personality and Social Psychology, 41*(2), 397.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine, 6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097
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*, 1747021819879826.
Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure and Function, 214*(5–6), 451–463. https://doi.org/10.1007/s00429-010-0258-9
Pollatos, O., Kurz, A.-L., Albrecht, J., Schreder, T., Kleemann, A. M., Schöpf, V., … Schandry, R. (2008). Reduced perception of bodily signals in anorexia nervosa. *Eating Behaviors, 9*(4), 381–388.
S. Porges , (1993). Body perception questionnaire. Laboratory of Developmental Assessment, University of Maryland.
Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E, 74*(1), Article 016110. https://doi.org/10.1103/PhysRevE.74.016110
W. Revelle , (2020). psych: Procedures for Psychological, Psychometric, and Personality Research (2.0.12) [Computer software]. ⟨https://CRAN.R-project.org/package=psych⟩.
Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research, 14*(4), 403–414. https://doi.org/10.1207/s15327906mbr1404_2
Riani, M., Torti, F., & Zani, S. (2011). *Outliers and robustness for ordinal data. In modern analysis of customer surveys* (pp. 155–169). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119961154.ch9
Ring, C., & Brener, J. (2018). Heartbeat counting is unrelated to heartbeat detection: A comparison of methods to quantify interoception. *Psychophysiology*, Article e13084.
Robinaugh, D. J., LeBlanc, N. J., Vuletich, H. A., & McNally, R. J. (2014). Network analysis of persistent complex bereavement disorder in conjugally bereaved adults. *Journal of Abnormal Psychology, 123*(3), 510–522. https://doi.org/10.1037/abn0000002
Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology, 125*(6), 747–757. https://doi.org/10.1037/abn0000181
Rochefort, C., Baldwin, A. S., & Chmielewski, M. (2018). Experiential avoidance: An examination of the construct validity of the AAQ-II and MEAQ. *Behavior Therapy, 49*(3), 435–449.
Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*(2), 282–292. https://doi.org/10.1037/a0025697
Santor, D. A., Gregus, M., & Welch, A. (2006). Focus article: Eight decades of measurement in depression. *Measurement: Interdisciplinary Research and Perspectives, 4*(3), 135–155. https://doi.org/10.1207/s15366359mea0403_1
Schmidt, N. B., Lerew, D. R., & Trakowski, J. H. (1997). Body vigilance in panic disorder: Evaluating attention to bodily perturbations. *Journal of Consulting and Clinical Psychology, 65*(2), 214–220. https://doi.org/10.1037/0022-006X.65.2.214
Shields, S. A., Mallory, M. E., & Simon, A. (1989). The body awareness questionnaire: Reliability and validity. *Journal of Personality Assessment, 53*(4), 802–815.
Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General, 143*(2), 850–886. https://doi.org/10.1037/a0033981
Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology, 5*, 1–25. https://doi.org/10.1146/annurev.clinpsy.032408.153639
Traag, V. A., & Bruggeman, J. (2009). Community detection in networks with positive and negative links. *Physical Review E, 80*(3), Article 036115. https://doi.org/10.1103/PhysRevE.80.036115
Trevisan, D. A., Altschuler, M. R., Bagdasarov, A., Carlos, C., Duan, S., Hamo, E., … Stahl, D. (2019). A meta-analysis on the relationship between interoceptive awareness and alexithymia: Distinguishing interoceptive accuracy and sensibility. *Journal of Abnormal Psychology, 128*(8), 765–776. https://doi.org/10.1037/abn0000454
Trevisan, D. A., Mehling, W. E., & McPartland, J. C. (2021). Adaptive and maladaptive bodily awareness: Distinguishing interoceptive sensibility and interoceptive attention from anxiety-induced somatization in autism and alexithymia. *Autism Research, 14*(2), 240–247. https://doi.org/10.1002/aur.2458
Tsakiris, M., Tajadura-Jiménez, A., & Costantini, M. (2011). Just a heartbeat away from one's body: Interoceptive sensitivity predicts malleability of body-representations. *Proceedings of the Royal Society of London B: Biological Sciences, 278*(1717), 2470–2476.
E. Vlemincx , M. Walentynowicz , G. Zamariola, , L. Van Oudenhove , O. Luminet, (2020). The THree-domain Interoceptive Sensations Questionnaire (THISQ). Manuscript in Preparation.
(*) Wang, X., Tan, Y., Van den Bergh, O., von Leupoldt, A., & Qiu, J. (2020). Intrinsic functional brain connectivity patterns underlying enhanced interoceptive sensibility. *Journal of Affective Disorders, 276*, 804–814. https://doi.org/10.1016/j.jad.2020.07.032.
Watson, D. (2012). Objective tests as instruments of psychological theory and research. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (Vol 1, pp. 349–369). Washington, DC: American Psychological Association (Foundations, planning, measures, and psychometrics).
Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports, 6*(1), 30750. https://doi.org/10.1038/srep30750
Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. J., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger (2017). Journal of Abnormal Psychology, 126 (7), 989–999. https://doi.org/10.1037/abn0000306.