The Mint Scale: A Fresh Validation of the Multimodal Interoception Questionnaire and Comparison to the MAIA, BPQ and IAS

Dominique Makowski[1,2], Ana Neves[1], Emma Benn[1], Maisie Bennett[1], and Giulia Poreiro[1]
[1]School of Psychology, University of Sussex
[2]Sussex Centre for Consciousness Science, University of Sussex

Interoception, the sense of the body's internal state, is crucial for mental and physical health, yet its measurement is hindered by tools with conceptual and psychometric issues. To address these limitations, we developed and validated the Multimodal Interoception Questionnaire (Mint). In Study 1 (N=559), we generated a comprehensive item pool using a novel "modality-by-context" framework and employed the recently developed psychometric network analysis for data-driven item reduction. In Study 2 (N=737), we validated the resulting instrument in an independent sample, establishing its structure and comparing its predictive power against established interoception questionnaires (MAIA-2, BPQ, IAS) across a wide range of clinical and cognitive outcomes. This rigorous process yielded a 33-item scale with a stable hierarchical structure comprising three metaclusters (Interoceptive Deficit, Interoceptive Awareness, Visceroception) and 11 distinct facets. The Mint demonstrated superior criterion validity, consistently outperforming existing scales in predicting outcomes tied to altered bodily perception, including alexithymia, ADHD, autism, and somatic symptom clusters. The Mint offers a psychometrically robust and nuanced measure of self-reported interoception that integrates and extends existing scales, providing researchers and clinicians with a practical and readily usable tool to investigate the role of bodily sensations in health and disease.

Keywords: Interoception questionnaire, interoceptive accuracy scale, MAIA, Mint Validation, Body Awareness, Health

Introduction

Interoception - the sensing and interpretation of the body's internal physiological signals - plays a fundamental role in processes ranging from homeostatic regulation to emotion, decision-making, and mental health (Craig, 2002; Critchley & Garfinkel, 2017). Despite its growing recognition, the field is hindered by inconsistent definitions and fragmented measurement approaches (Desmedt et al., 2022, 2025; Khalsa et al., 2018). Available tools include "objective" physiological tasks (e.g., heartbeat counting or tracking) and "subjective" self-reports, and encompass both "explicit" paradigms that direct participants' attention and consciousness to internal sensations and "implicit" paradigms that do not (e.g., heartbeat evoked potentials recorded during rest). These assessments are often targeted at only one of several interoceptive modalities (e.g., cardioception, respiroception, gastroception), and are mapped onto theoretical dimensions (e.g., accuracy, sensitivity, awareness, Garfinkel et al., 2015) within a rapidly evolving theoretical scaffolding. However, no consensus exists on a gold-standard measure, and each approach has distinct limitations: physiological tasks often require specialised equipment and expertise, are time-consuming, and are challenging to administer at scale, whereas questionnaires are constrained by their inherent subjectivity. Crucially, the psychometric quality of many of these measures is questionable (e.g., see criticism about the heartbeat counting task, Zamariola et al., 2018), yielding low inter-task correlations (Murphy et al., 2018), raising serious concerns about their reliability and validity (e.g., measures presented as measuring one theoretical dimension but likely measuring something else). This measurement conundrum hinders progress in comparing, replicating, and interpreting interoception-related findings, underscoring the need to develop improved tools that align with modern psychometric standards.

The present work focuses on self-report questionnaires, which offer a scalable and accessible tool complementing physiological tasks. Among the numerous questionnaires available, a handful seem to dominate the literature (Desmedt et al., 2022), including the Body Perception Questionnaire (BPQ, Kolacz et al., 2018), the Multidimensional Assessment of Interoceptive Awareness (MAIA, Mehling et al., 2018), and the Interoceptive Accuracy Scale (IAS, Murphy et al., 2018). These tools were developed in different eras of a rapidly evolving field, each with its own objectives and philosophies. The BPQ, one of the earliest measures, emphasizes the assessment of body awareness and autonomic nervous system (ANS) reactivity, focusing on subjective experiences of physiological stress reactions in organs innervated by the ANS, such as those triggered by stress or trauma (Kolacz et al., 2018). The MAIA incorporated ideas from the then-booming emotion regulation and mindfulness fields, including subscales assessing adaptive coping strategies and metacognitive insights about the body. By contrast, the IAS was recently crafted to sidestep such high-level - possibly tangential - constructs and instead capture subjective reports of concrete bodily events (e.g., coughing, sneezing) as proxies for interoceptive accuracy.

These differing philosophies have practical consequences, leading to poor convergence between scales that are assumed to measure the same construct (Vig et al., 2022) - a classic example of the "jingle fallacy" where the shared label of "interoception" can mask underlying measurand differences, which in turn hinders replication and theoretical integration. Recent work has been focused on aligning subjective measures and facets with theoretical models, debating whether they measure interoceptive "sensitivity" or other dimensions (accuracy, attention, deficits). However, this focus on a theory-driven approach might be premature when theoretical models are in flux, and when a topographic mapping to self-report dimensions is not guaranteed (e.g. distinctions that are relevant at a neuroanatomical or cognitive level might not necessarily translate into the same structural landscape of subjective experiences). More fundamentally, existing questionnaires suffer from two major structural blind spots. First, they are often narrow or unsystematic in their coverage of bodily *modalities*, with a heavy bias towards cardio-respiratory sensations and a neglect of others like gastric, urogenital, or thermoregulatory signals. Second, and perhaps most critically, they fail to control for the *context* in which an interoceptive experience occurs, introducing significant variability in how participants interpret and respond to items. For example, an item about "feeling your heart beat" is ambiguous; its frequency, experience and its interpretation differ vastly depending on whether it occurs during physical exercise, a moment of anxiety, or quiet rest (e.g., someone who exercises regularly, or is anxious, might more easily recall an instance of occurrence of feeling their heart beating and respond based on that contextual saliency).

To address these gaps, we introduce the Multimodal Interoception Questionnaire (Mint). The overarching aim of this work is to develop and validate a new questionnaire that systematically addresses these limitations. In Study 1, we detail the bottom-up development of the Mint, using a "modality-by-context" framework to generate a comprehensive item pool, which is then refined using cutting-edge psychometric network analysis. In Study 2, we validate the final version of the Mint against the established interoception questionnaires, demonstrating its robust structure and predictive validity for a range of clinically-relevant outcomes.

Study 1: Item Selection

The primary goal of Study 1 was to develop a new interoception questionnaire by generating a comprehensive item pool and refining it into a psychometrically robust scale. To overcome the limitations of existing measures, we adopted a bottom-up, theoretically agnostic approach. Instead of imposing a pre-defined theoretical structure, we focused on two tangible and systematically controllable aspects of interoceptive experience: the *modality* of the sensation and the *context* in which it occurs. The objective was not to create *yet another scale* measuring an alleged distinct construct or variation, but rather to synthesize existing ideas and frameworks into a comprehensive and modern tool that captures the richness of interoceptive experiences and encompasses the most relevant facets of interoception.

Our item generation process was guided by a "modality-by-context" grid designed to overcome the limitations of existing scales. For *modalities*, we expanded beyond the typical cardio-respiratory focus to include a broader range of bodily systems: cardiac, respiratory, gastric, genital, bladder & colon, and skin & temperature (the latter of which being increasingly recognized as a key interoceptive channel, Crucianelli & Ehrsson, 2023). We also included a "general state" category to capture the integrated sense of the body's overall physiological condition (Craig, 2002; Craig, 2003). For *contexts*, we designed facets to capture specific situational contexts like negative (anxious) and positive (sexual) arousal, which reflect distinct bodily states with differing interoceptive signatures, as well as other experiential qualities tapping into elaboration and threshold of perception, such as

---

Dominique Makowski
Ana Neves
Giulia Poreiro

This preprint is a non-peer-reviewed work from the **Reality Bending Lab**.

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Dominique Makowski: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft; Ana Neves: Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing; Emma Benn: Investigation; Maisie Bennett: Investigation; Giulia Poreiro: Investigation, Methodology, Writing – original draft, Writing – review & editing

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

sensitivity, accuracy, and confusion. To maintain continuity with established work while making targeted improvements, we purposefully adapted items from existing questionnaires (e.g., IAS-like items related to observable behaviors). This systematic framework yielded a diverse pool of 120 initial items, ensuring comprehensive coverage while minimizing the ambiguity inherent in context-free questions.

The final step involved a data-driven reduction of this item pool. We employed cutting-edge psychometric network-based techniques to identify and remove redundant items and to uncover the most stable and coherent dimensional structure. Additionally, we designed the study so that we could explore an often overlooked psychometric quality: the invariance of the structure to the presentation format (how items are grouped together on the screen, whether by modality, facet, or randomly) influences the resulting psychometric structure. This rigorous, multi-stage process was designed to produce a subset of items with a clear, stable, and theoretically rich hierarchical structure, ready for formal validation.

Methods

Participants

We recruited 760 English-speaking participants using Prolific©. We excluded 191 for failing at least one attention check, and 10 based on measures significantly related to the probability of failing attention checks (namely, the multivariate distance obtained with the OPTICS algorithm, Thériault et al., 2024). The final sample includes 559 participants (age = 37.0 ± 12.2 [18, 77]; 50.8% women; country of residence: 63.86% UK, 26.65% USA). This study was approved by the University of Sussex' Ethics Committee (ER/MB2021/1).

Item Generation

Based on the two goals outlined for this scale, namely to include different interoceptive modalities, and to explicitly state the context of the interoceptive experience (e.g., whether negative or positive), we generated items in a systematic way following a combinatory approach, where each item's category corresponds to the union of a specific modality and context (Figure 1).

We identified 7 "modalities" (cardiac, respiratory, gastric, genital, skin & temperature (thermoregulation), bladder & colon, and a "general state" category corresponding to a holistic and general awareness of an interoceptive state or dimension). Through iterative refinement (e.g., splitting or merging different categories together), we settled on 6 "facets", which encompass both *contexts* of experience (negative and positive arousal, e.g., anxious and sexual states), and potential distinct *mechanisms* (nociception & pleasure, sensitivity, accuracy, and confusion).

Using this orthogonal 7x6 modality/facet grid as a conceptual scaffolding, we generated 120 initial items, striving for a balanced number of items with consistent phrasing within modalities and facets[1]. We additionally crafted 8 "attention check" items blending in (and distributed across) each category.

Procedure

To avoid presenting all the 120 items on a single long and discouraging page, we split them into different pages. Participants were randomly assigned to one of three conditions, driving how items were grouped on the same page: 1) items grouped by modality (i.e., all cardiac items on the first page, all colon & bladder items on the second, etc.), 2) items grouped by facet, or 3) items presented fully randomly (but with their number balanced across 6 pages). The order of the item on any given page and the order of the modalities/facets was randomized. Each participant completed the full set of 120 items, with 8 attention check items interspersed throughout. The online experiment was implemented using JsPsych (De Leeuw, 2015), and item responses were recorded using 7-points Likert scales (0 = Disagree, 6 = Agree).

Data Analysis

In order to test whether the grouping condition had an effect on the structure (i.e., how items relate to one-another), we compared the correlation matrix obtained in the random condition to the ones obtained in the modality and facet conditions, focusing on 3 indices of correlation matrix similarity - the Procrustes Similarity Index (PSI, Sibson, 1978), the Adjusted RV (Rvadj, Mayer et al., 2011), and the Similarity of Matrices Index (SMI, Indahl et al., 2018). For each index, we bootstrapped the difference between the similarity with the facet and modality conditions to test whether the correlation matrix in the random-grouping condition is significantly more similar to any of the two other conditions.

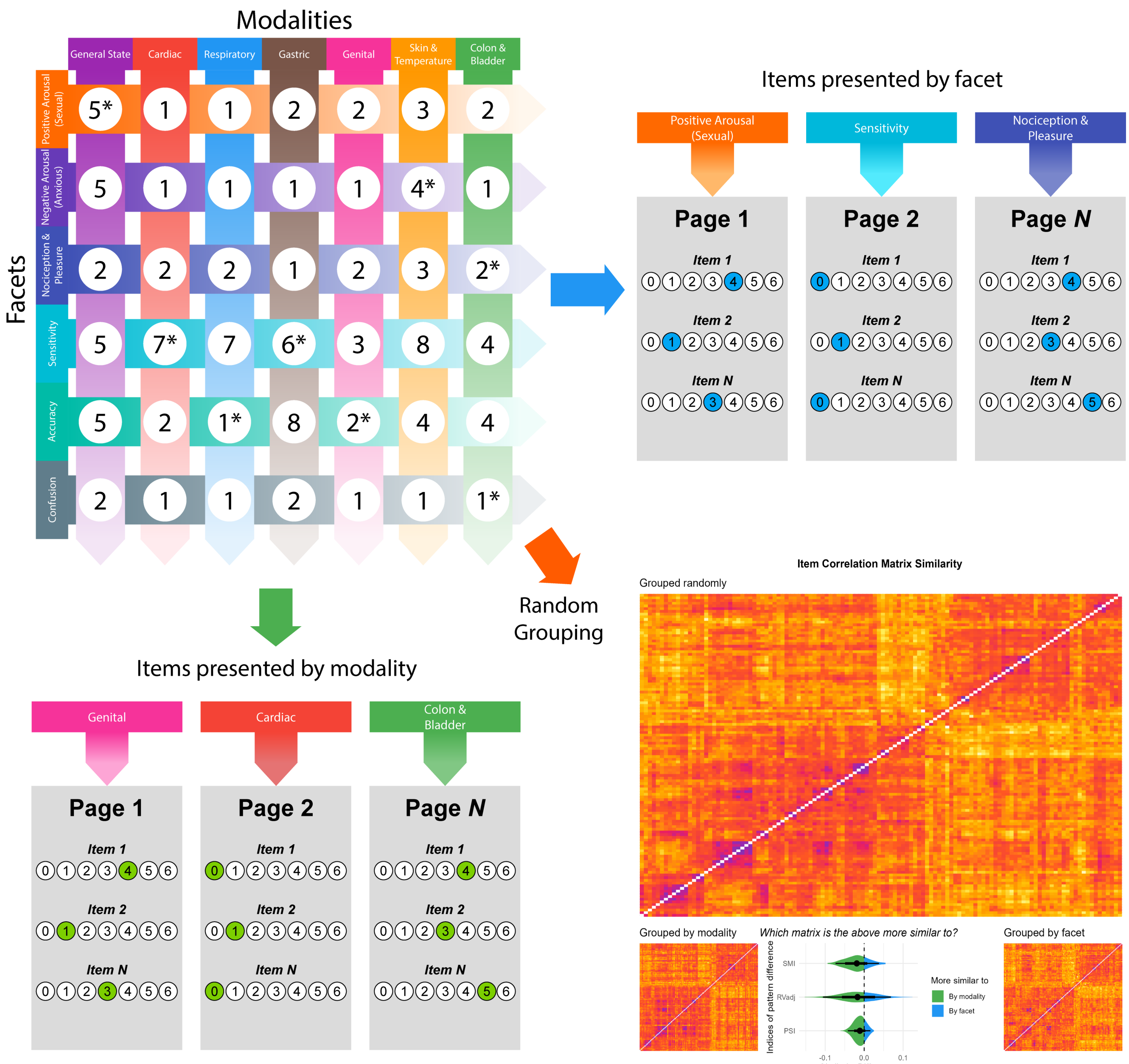
Items deemed "redundant" (which can distort the item structure estimation by introducing multicollinearity or local dependencies) were identified (using the recommended threshold of 0.25) using Unique Variable Analysis (UVA, Christensen et al., 2023), a novel and principled method derived from network psychometrics.

The structure of the items was analyzed using the recently-developed Exploratory Graph Analysis (EGA, H. F. Golino & Epskamp, 2017) framework, which allows to jointly estimate the number of dimensions (i.e., clusters of items), the structure, as well as its stability using bootstrapping (H. Golino et al., 2020). At a fundamental level, EGA conceptualizes variables as nodes in a network, with connections (edges) reflecting associations between them. Evidence has underlined its suitability as an alternative to traditional factor analysis, addressing some of its limitations such as the assumption of

[1]The initial item list is available at realitybending.github.io/InteroceptionScale/study1/analysis/2_analysis.html

Figure 1

The conceptual grid used to generate the 120 initial items (top-left). Each item belongs both to an interoceptive modality (vertical) and a facet (horizontal), with the number of each item per category indicated in the circles. The asterisk denotes the additional presence of an attention check item in that category. In the experiment, these items were presented on different pages grouped either by modality (bottom-left), by facet (top-right), or randomly. The Correlation Similarity (bottom-right) analysis suggested that the correlation matrix obtained from the participants assigned to the random-grouping condition was slightly more similar (but non-significantly) to the one obtained in the modality-grouping condition, suggesting that the scale's structure is robust to different presentation conditions, and that the modality-grouping might potentially tend to facilitate the emergence of the underlying item structure (and thus be interpreted as being more "natural").



a "latent" source of variability, issues in estimation of the optimal factor numbers, and poor performance in complex population structures, while remaining comparable and interpretable (Christensen & Golino, 2021; Jiménez et al., 2023). In particular, nodes communities (i.e., clusters of items) can be in practice interpreted as distinct "dimensions", similarly to traditional latent factors - but without explicitly assuming their existence (Christensen & Golino, 2021).

After removing redundant items using UVA, we iteratively fitted hierarchical EGA models (which additionally estimate higher-order "meta" clusters) using "glasso" (Friedman et al., 2008) and the "Leiden" algorithm (Traag et al., 2019) for community detection, refining the item pool at each step. We started by removing items with a low (< 80%) cluster stability (i.e., volatile items which jump between clusters across bootstrapped samples), followed by odd items belonging to no clusters or pairs of items (i.e., we keep items belonging to clusters of more than 2 items). Finally, for each lower-level cluster, we selected the 3 items with the highest node centrality (i.e., the highest loading in the cluster).

Results

The correlation matrix similarity analysis yielded no significant differences between the similarity of the random-grouping condition with the modality-grouping and facet-grouping conditions ($PSI_{\text{Random vs. Facet}} = 0.81$, $PSI_{\text{Random vs. Modality}} = 0.82$, $p = .45$; $RVadj_{\text{Random vs. Facet}} = 0.77$, $RVadj_{\text{Random vs. Modality}} = 0.78$, $p = .74$; $SMI_{\text{Random vs. Facet}} = 0.49$, $SMI_{\text{Random vs. Modality}} = 0.51$, $p = .52$), despite a consistent bias in favour of the modality-grouping condition.

From the 120 initial items, 4 redundant items were flagged by UVA. We then removed 40 items that showed low cluster stability, and 9 items that were part of clusters with less than 3 items. Finally, We kept the 3 items with the highest loading in their lower-level structure (removing 13 items in the process), resulting in 54 items in the final item pool.

The final hierarchical EGA model (Generalized Total Entropy Fit Index = -119.18) - in which all 54 items yielded a high cluster stability (> 90%) suggested 3 metaclusters and 15 lower-level clusters (each containing 3 items): "Interoceptive Deficits" (containing 5 clusters: *Urointestinal Inaccuracy* - UrIn; *Cardiorespiratory Confusion* - CaCo; *Cardiorespiratory Noticing* - CaNo; *Olfactory Compensation* - Olfa; *Satiety Noticing* - Sati), "Interoceptive Awareness" (containing 7 clusters: *Sexual Arousal Awareness* - SexA; *Sexual Arousal Sensitivity* - SexS; *Sexual Organs Sensitivity* - SexO; *Urointestinal Sensitivity* - UrSe; *Relaxation Awareness* - RelA; *State Specificity* - StaS; *Expulsion Accuracy* - ExAc), and "Interoceptive Sensitivity" (containing 6 clusters: *Cardioception* - Card; *Respiroception* - Resp; *Signalling* - Sign; *Gastroception* - Gast; *Dermal Hypersensitivity* - Derm; *Sexual Arousal Changes* - SexC).

To further reduce and balance the remaining items, we collectively decided on removing the lower-level clusters *Sexual Arousal Awareness* (too general and overlapping with the other more specific sex-related items), *Signalling* (which items started with "when something important is happening in my life", which meaning we deemed too much open to interpretation), and *Sexual Arousal Changes* (low fit with the other modality-focused clusters of its group). The final set included 45 items.

Discussion

The primary aim of Study 1 was to develop and refine a large, systematically generated pool of interoceptive items into a psychometrically sound and structurally coherent questionnaire. Starting with 120 items derived from a modality-by-context grid, we employed a multi-step, data-driven reduction process using cutting-edge structure analysis methods. This rigorous procedure yielded a final, balanced set of 45 items, organized within a stable hierarchical structure comprising three higher-order dimensions (Interoceptive Deficits, Interoceptive Awareness, Interoceptive Sensitivity) and 12 distinct lower-level facets.

One innovation of this study was to test the effect of item presentation by comparing structures across different grouping conditions. While the analysis revealed no significant effect, a consistent (though non-significant) trend suggested that grouping items by modality may be slightly more conducive to revealing the underlying psychometric structure than grouping by facet or randomly. This finding, while preliminary due to the sample size per condition ($N \pm 200$), highlights a potentially important and under-investigated aspect of questionnaire design that warrants future research.

The resulting hierarchical structure offers novel yet interpretable insights into self-reported interoception, with flexible degrees of granularity. Notably, we categorised several facets (typically related to objective external behaviours, like coughing or sneezing) under the label "Awareness", that might traditionally be named "Accuracy" in other scales. Our rationale is that subjective self-report, by its very nature, cannot be a pure measure of objective accuracy, but at best a proxy of an individual's subjective certainty or awareness of their interoceptive processes. Furthermore, our bottom-up approach revealed intriguing facets, such as "Olfactory Compensation". This finding provide further evidence for a pseudo-interoceptive pathway, whereby individuals with potentially poorer access to internal signals may compensate by relying on external cues (e.g., odours) to infer their bodily state. The "Sensitivity" metacluster also grouped core visceroceptive facets (e.g., "Cardioception", "Gastroception") with "Dermal Hypersensitivity", pointing to a potential common dimension of heightened perceptual sensitivity to bodily sensations.

In summary, Study 1 successfully leveraged a systematic item generation process and modern analytical techniques to produce a 45-item questionnaire with a clear, stable, and theoretically rich hierarchical structure. The next crucial step is to formally validate this new instrument by examining its psychometric properties and its relationship with existing measures in a new, independent sample, which is the objective of Study 2.

Study 2: Validation

Following the generation and selection of a set of 45 interoceptive items, the next objective is to formally validate this new instrument in a large, independent sample, a crucial step for establishing the psychometric soundness and unique contribution of the Mint scale to the field of interoception research. To this end, we sought to 1) re-assess the structure of the shorter version and 2) establish its convergent validity by examining the correlations with other established and popular interoception questionnaires (namely the MAIA-2, BPQ, and IAS), as well as other related measures (e.g., alexithymia). Finally, we assessed its criterion validity by directly comparing the predictive power of the interoceptive scales across a broad range of outcomes, including emotional, cognitive, and clinical variables.

Methods

Participants

We recruited 921 English-speaking participants via SONA and Prolific©, from which 118 were excluded for failing at least one attention check and 6 based on multivariate distance (using the same procedure as for study 1). 60 participants were further excluded due to missing data following a technical issue. The final sample includes 737 participants (age = 36.8 ± 14.7 [18, 87]; 57.3% women; Country of residence: 75.17% United Kingdom, 24.83% other). This study was approved by the University of Sussex' Ethics Committee (ER/EB672/2).

Measures

Interoception Questionnaires. The 45 items of the Multimodal Interoception Questionnaire (Mint) scale were presented in a random order, with the same 7-point Likert scale as in study 1 (0 = *Disagree*, 6 = *Agree*).

The Multidimensional Assessment of Interoceptive Awareness (MAIA-2, Mehling et al., 2018) measures 8 dimensions with 37 items presented on a 7-point Likert scale (0 = *Never*, 6 = *Always*). It includes bodily dimensions such as Body Listening, Noticing, Emotional Awareness, dimensions related to emotion regulation, such as Trust, Not-worrying, and Self-Regulation, as well as dimensions related to attention, such as Attention Regulation and Not-Distracting. The relationship of some of these dimensions to interoception remains debated, particularly Not-Distracting and Not-Worrying, which tend to show weak or non-existent correlations with the other MAIA subscales (Ferentzi et al., 2021).

The Body Perception Questionnaire - Very Short Form (BPQ-VSF, Cabrera et al., 2018) measures a general score of bodily awareness with 12 items presented on 5-point Likert scale (0 = *Never*, 5 = *Always*).

The Interoceptive Accuracy Scale (IAS, Murphy et al., 2020) measures a general score of interoceptive accuracy to physical sensations with 21 items presented on a 5-point Likert scale (1 = *Disagree Strongly*, 5 = *Strongly Agree*).

Emotions and Cognition. The Toronto Alexithymia Scale TAS (TAS-20, Leising et al., 2009) measures 3 Alexithymia dimensions with 20 items 5-point Likert scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*: Difficulty Identifying Feelings (DIF), Difficulty Describing Feelings (DDF), and Externally Oriented Thinking (EOR).

The Emotion Reactivity Scale - Brief Version (B-ERS, Veilleux et al., 2024) measures 3 dimensions with 6 items on a 5-point Likert scale (0 = *Not like me at all*, 4 = *Extremely like me*): Arousal (*"My moods are very strong and powerful"*), Sensitivity (*"I tend to get very emotional very easily"*), and Persistence (*"When I am angry/upset, it takes me much longer than most people to calm down"*).

The Cognitive Emotion Regulation Questionnaire (CERQ-short, Garnefski & Kraaij, 2006) measures 9 adaptive and maladaptive emotion regulation strategies with 18 items (1 = *Almost Never*, 5 = *Almost Always*). The strategies include Self-blame, Other-blame, Rumination, Catastrophizing, Putting into Perspective, Positive Refocusing, Positive Reappraisal, Acceptance, and Planning.

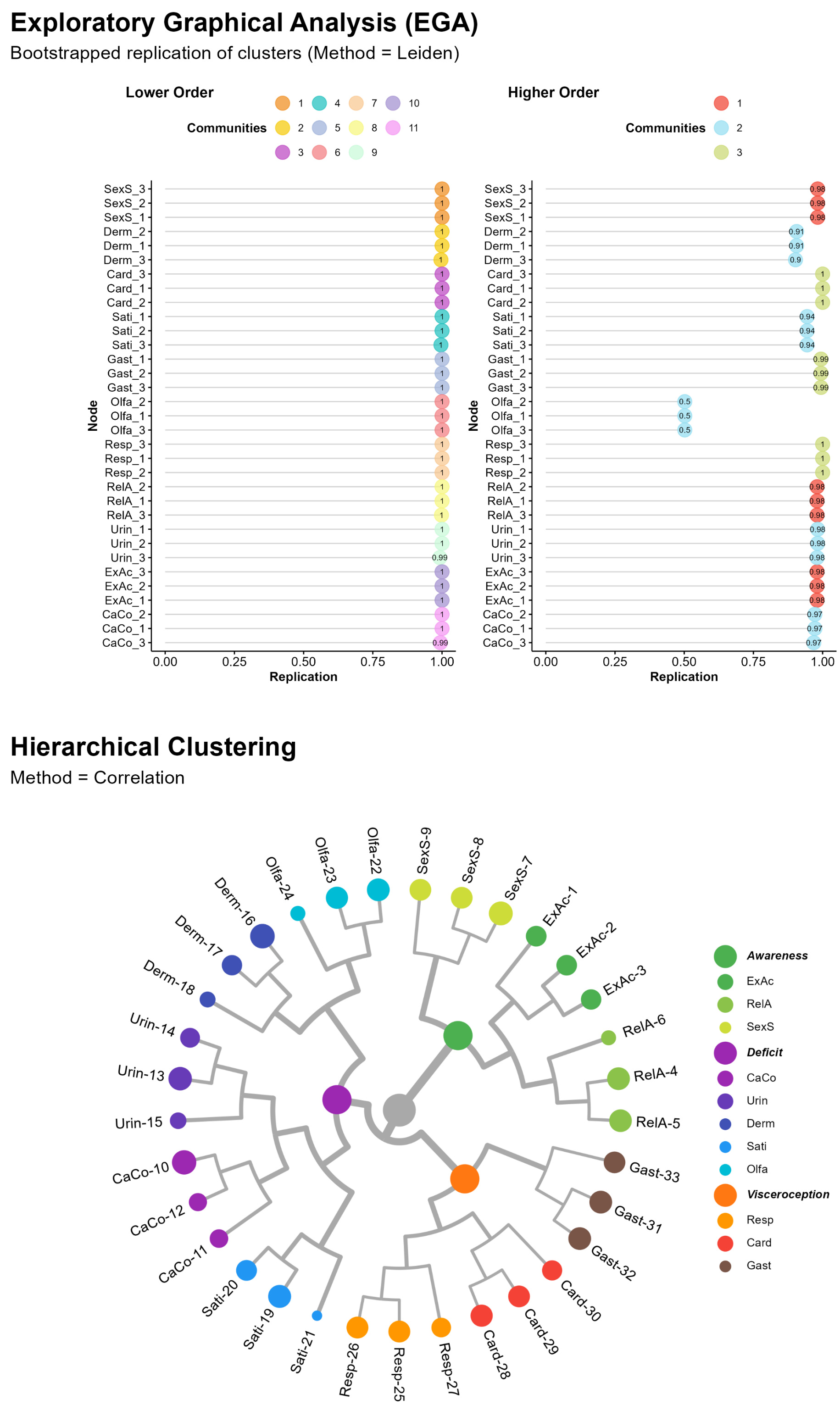
The Černis Felt Sense of Anomaly scale - short form (CEFSA-14, Černis et al., 2024) measures 7 dissociative experiences with 14 items on a 5-point Likert scale (0 = *Never*, 4 = *Always*). This scale is designed to measure dissociative experiences in adolescence and adulthood, specifically focusing on the felt sense of anomaly-type dissociation. It includes Anomalous Experience of the Self, Anomalous Experience of the Body, Anomalous Experiences of Emotion, Altered Sense of Familiarity, Altered Sense of Connection, Altered Sense of Agency, and Altered Sense of Reality.

The Primal World Beliefs Inventory (PI-18, Clifton & Yaden, 2021) measures higher-order beliefs about the basic character of the world: Good, Safe, Enticing, and Alive. The 18 items are presented on a 6-point Likert scale (0 = *Strongly Disagree*, 5 = *Strongly Agree*).

Health and Wellbeing. The Single Item Life Satisfaction scale (SILS, Jovanović & Lazić, 2020) was followed by the PHQ-4 (Kroenke et al., 2009) assessing depression and anxiety symptoms with 4 items. We used the refined version of the PHQ-4 (Makowski et al., 2025), which adds an additional response option ("Once or twice"), increasing sensitivity to subclinical fluctuations.

Figure 2

Structure analysis of the final set of 33 Mint items. Hierarchical Exploratory Graph Analysis (EGA) was applied to jointly identify clusters of items, as well as higher-order metaclusters. The reliability of each item can be quantified by the proportion of structure replication across bootstrapped samples, with high values indicative of high structural stability. The Mint scale displayed an excellent structural consistency, with the exception of the Olfa *cluster which belonging to the* Deficit *metaclusters warrants further research. Bottom plot shows the result of hierarchical clustering, providing evidence for structure statbility across methods, and allowing for more granular understanding of the relationships between variables.*



Participants were asked to self-report any current, medically diagnosed psychiatric disorders using a checklist based on DSM-5 categories. If one or more conditions were endorsed, participants were asked to indicate any current treatments, including pharmacological (e.g., antidepressants, anxiolytics, antipsychotics, mood stabilizers), psychological (e.g., psychotherapy, mindfulness), or lifestyle interventions. Binary variables (0 = absent, 1 = present) were created to identify participants reporting mood disorders (MDD, GAD, Bipolar Disorder; with a stricter subgroup of participants undergoing a pharma- or psychological treatment), anxiety-centred disorders (GAD, Panic Disorder, Social Anxiety Disorder, Specific Phobias), eating disorders, addiction-related disorders, borderline personality disorder, autism spectrum disorder (ASD), and ADHD.

Participants were asked to select somatic symptoms or conditions they experienced from a list of 36 options. To facilitate mechanistic interpretation and reduce redundancy, answers were grouped into four non-overlapping clusters based on shared physiological pathways and known etiological mechanisms. The *Afferent Sensitivity* cluster included conditions associated with heightened interoceptive awareness and neurogenic excitability, such as migraine, neuropathy, dizziness, nausea, muscle tension, epilepsy, and frequent urination. The *Central Sensitization* cluster comprised syndromes characterized by chronic pain and fatigue, likely reflecting central amplification of sensory signals and HPA-axis dysregulation, such as fibromyalgia, chronic fatigue syndrome, chronic pain, back pain, pelvic pain, irritable bowel syndrome (IBS). The *Autonomic Dysfunction* cluster captured disorders linked to dysregulation of the autonomic and cardiopulmonary systems, including joint hypermobility, cardiac arrhythmia, chest pain, shortness of breath, hypo-/hypertension, sleep apnea, chronic obstructive pulmonary disease (COPD), and chronic bronchitis. Finally, the *Immune-Inflammatory* cluster encompassed conditions associated with immune dysregulation, barrier dysfunction, and gut-brain axis disturbance, such as eczema, psoriasis, skin rashes, asthma, celiac disease, gluten and lactose sensitivity, inflammatory bowel diseases (Crohn's disease and ulcerative colitis), gastroesophageal reflux disease (GERD), multiple sclerosis, and Sjögren's syndrome. We scored each cluster as a binary variable based on whether the participant selected at least one symptom from that cluster.

Lifestyle. Participants reported about owning any wearable devices to monitor health indices such as heart rate, number of steps, calories burned, sleep quality, and weight. For each selected device, a follow-up question inquired about the frequency of usage and subjective importance of that measure.

Participants were asked to rate how physically active they consider themselves and how many hours of active workout they engage in per week. Participant's BMI was computed using height and weight.

Procedure

In order to avoid the repetition of similar types of questions and and balance longer and shorter questionnaires, we partitioned the measures into three groups (and randomized the order within them): 1) interoception questionnaires (MAIA, IAS, BPQ), 2) emotions (TAS, CERQ, ERS, PI-18), and 3) health (somatic symptoms, mental health, LS + PHQ-4, CEFSA). After completing demographic questions, participants always started with the Mint scale, and each following interoception questionnaire was interspersed with two questionnaires from the emotions and pathology groups. 7 attention checks were embedded throughout (one within each major questionnaire). In order to make the experiment more enjoyable, the experiment ended with a radar chart summarizing the participants' responses to the Mint scale[2].

Data Analysis

We started by confirming and further refining the structure of the Mint scale (see Figure 2) using the same EGA model as in Study 1. We then computed the lower-level dimensions and the higher-level metaclusters' scores by averaging their corresponding items. The convergent validity of the final set of items was assessed by computing the correlations between the Mint scale and the other interoception questionnaires (MAIA, BPQ, IAS).

Next, we tested the predictive power of the Mint scale relative to other interoception questionnaires. We will assessed the importance of each interoceptive dimension (from all the scales) as a unique predictor by fitting regression models (linear for continuous measures - e.g., depression score - and logistic for binary variables - e.g., presence *vs.* absence of mood disorders) to predict different outcomes, and compare the best between the 4 interoception scales (based on the R2). We then evaluated the predictive performance of each scale as a whole by comparing regression models with all the dimensions entered as predictors (note that the IAS and BPQ only have one total score variable). We assessed the models based on their total explained variance (R2), as well as on the Bayesian Information Criterion (BIC), which penalizes for predictor number, thus offering a balance between model parsimony and predictive power. For the logistic models, we quantified the discriminative power by computing the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve, which assesses the model's discriminative power (the combination of sensitivity and specificity).

In order to evaluate the potential for a short version of our scale, we compared 4 variants of the Mint: the (full) *Mint*

[2]The experiment can be tested by following the link on https://github.com/RealityBending/InteroceptionScale

Figure 3

Item loadings. The table shows each item of the Mint with its cluster centrality, which is the EGA equivalent (in terms of interpretation) of factor loadings. It also shows how each lower-level cluster is related to higher-order metaclusters. The lower-level clusters are Expulsion Accuracy (ExAc), Relaxation Awareness (RelA), Sexual Arousal Sensitivity (SexS), Cardiorespiratory Confusion (CaCo), Urointestinal Inaccuracy (UrIn), Dermal Hypersensitivity (Derm), Satiety Noticing (Sati), Olfactory Compensation (Olfa), Respiroception (Resp), Cardioception (Card), Gastroception (Gast).

Item Loadings

Node centrality

Item	Label	ExAc	RelA	SexS	CaCo	Urin	Derm	Sati	Olfa	Resp	Card	Gast
Metaclusters												
M1	Interoceptive Awareness	0.42	0.41	0.30	−0.20	−0.06	0.06	−0.07	0.08	0.13	0.02	0.12
M2	Interoceptive Deficit	0.00	−0.17	0.01	0.49	0.44	0.24	0.23	0.18	0.21	0.20	0.12
M3	Visceroception	0.02	0.15	0.02	0.20	0.01	0.11	0.00	0.13	0.60	0.58	0.33
Items												
1	*I can always accurately feel when I am about to fart*	0.48	0.02	0.05	0.00	0.00	0.04	−0.01	0.10	0.00	0.00	0.00
2	*I can always accurately feel when I am about to sneeze*	0.48	0.09	0.01	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00
3	*I can always accurately feel when I am about to burp*	0.46	0.06	0.01	−0.02	−0.04	0.05	−0.06	0.00	0.03	0.00	0.00
4	*I always feel in my body if I am relaxed*	0.02	0.59	0.02	−0.04	0.00	0.00	−0.01	0.00	0.04	0.00	0.03
5	*I always know when I am relaxed*	0.10	0.58	0.03	−0.10	−0.07	0.00	−0.03	−0.01	0.03	0.03	0.02
6	*My body is always in the same specific state when I am relaxed*	0.04	0.28	0.00	−0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.00
7	*During sex or masturbation, I often feel very strong sensations coming from my genital areas*	0.01	0.03	0.65	−0.05	−0.04	0.00	0.00	0.00	0.00	0.00	0.00
8	*My genital organs are very sensitive to pleasant stimulations*	0.00	0.02	0.53	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00
9	*When I am sexually aroused, I often notice specific sensations in my genital area (e.g., tingling, warmth, wetness, stiffness, pulsations)*	0.08	0.02	0.53	0.00	0.01	0.03	0.00	0.02	0.00	0.01	0.02
10	*Sometimes my breathing becomes erratic or shallow and I often don't know why*	−0.02	−0.01	0.00	0.68	0.11	0.00	0.09	0.02	0.07	0.00	0.00
11	*I often feel like I can't get enough oxygen by breathing normally*	0.00	−0.07	−0.04	0.39	0.04	0.04	0.02	0.02	0.06	0.04	0.00
12	*Sometimes my heart starts racing and I often don't know why*	0.00	−0.07	0.00	0.37	0.06	0.05	0.03	0.00	0.00	0.16	0.00
13	*I sometimes feel like I need to urinate or defecate but when I go to the bathroom I produce less than I expected*	0.00	0.00	0.01	0.08	0.63	0.03	0.01	0.07	0.00	0.00	0.00
14	*I often feel the need to urinate even when my bladder is not full*	0.00	−0.01	0.00	0.04	0.44	0.09	0.00	0.02	0.00	0.00	0.01
15	*Sometimes I am not sure whether I need to go to the toilet or not (to urinate or defecate)*	−0.04	−0.06	−0.03	0.09	0.32	0.00	0.12	0.00	0.00	0.00	0.00
16	*In general, my skin is very sensitive*	0.00	0.00	0.00	0.01	0.00	0.70	0.02	0.00	0.04	0.00	0.01
17	*My skin is susceptible to itchy fabrics and materials*	0.00	0.00	0.00	0.07	0.07	0.46	0.01	0.01	0.00	0.00	0.01
18	*I can notice even very subtle stimulations to my skin (e.g., very light touches)*	0.13	0.00	0.04	0.00	0.05	0.30	0.00	0.00	0.02	0.03	0.04
19	*I don't always feel the need to eat until I am really hungry*	−0.01	0.00	0.00	0.00	0.00	0.01	0.60	0.00	0.00	0.00	0.00
20	*Sometimes I don't realise I was hungry until I ate something*	−0.05	−0.02	0.00	0.09	0.10	0.01	0.49	0.02	0.01	0.00	0.00
21	*I don't always feel the need to drink until I am really thirsty*	0.00	−0.03	0.00	0.04	0.03	0.00	0.23	0.03	0.00	−0.01	0.00
22	*I often check the smell of my armpits*	0.00	−0.01	0.00	0.02	0.03	0.00	0.04	0.59	0.04	0.03	0.00
23	*I often check the smell of my own breath*	0.00	0.00	0.01	0.02	0.00	0.00	0.02	0.55	0.04	0.00	0.04
24	*I often check the smell of my farts*	0.10	0.00	0.01	0.00	0.06	0.01	0.00	0.28	0.00	0.00	0.01
25	*In general, I am very sensitive to changes in my breathing*	0.00	0.04	0.00	0.06	0.00	0.07	0.00	0.02	0.54	0.15	0.02
26	*I can notice even very subtle changes in my breathing*	0.03	0.03	0.00	0.03	0.00	0.00	0.00	0.01	0.54	0.14	0.02
27	*I am always very aware of how I am breathing, even when I am calm*	0.00	0.02	0.00	0.04	0.00	0.00	0.01	0.04	0.43	0.05	0.06
28	*In general, I am very sensitive to changes in my heart rate*	0.00	0.00	0.00	0.07	0.00	0.03	−0.01	0.00	0.11	0.55	0.03
29	*I often notice changes in my heart rate*	0.00	0.00	0.01	0.14	0.00	0.00	0.00	0.03	0.08	0.53	0.01
30	*I can notice even very subtle changes in the way my heart beats*	0.00	0.07	0.00	0.01	0.00	0.00	0.00	0.00	0.17	0.45	0.05
31	*I can notice even very subtle changes in what my stomach is doing*	0.00	0.02	0.00	0.00	0.01	0.04	0.00	0.03	0.05	0.04	0.59
32	*In general, I am very sensitive to what my stomach is doing*	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.02	0.00	0.02	0.58
33	*I am always very aware of what my stomach is doing, even when I am calm*	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.07	0.04	0.53

(including all lower-level dimensions), the *metaMint* (only including the metaclusters), the *miniMint* (including the metaclusters computed from a reduced number of items), and the *microMint* (including only the most representative dimension of each metacluster). Moreover, we also included an alternative "interoception-focused" version of the MAIA (*iMAIA*) that only contains the 3 most interoceptive dimensions (Body Listening, Noticing, and Emotional Awareness)[3].

Results

The application of EGA to the initial set of 42 items reproduced the expected 14 lower-level clusters of triplets and 3 higher-order metaclusters, with the exception of the *UrSe* (Urointestinal Sensitivity), for which one item moved to the *Olfa* (Olfactory Compensation) cluster. In order to further balance and reduce the items, we removed the *UrSe* cluster, as well as *StaS* (State Specificity), which in comparison to the other items stood out as containing vague and overlapping items ; *SexO* (Sexual Organs Sensitivity), which overlapped with the *SexS* (Sexual Arousal Sensitivity) cluster; and *CaNo* (Cardiorespiratory Noticing), which overlapped with the *CaCo* (Cardiorespiratory Confusion) cluster.

The final set of 33 items yielded a good fit (Generalized Total Entropy Fit Index = -77.23), with all items showing high cluster stability (> 90%), with the exception of *Olfa*. The final structure included 11 lower-level clusters, grouped into 3 higher-order metaclusters: "Interoceptive Deficit" (containing *CaCo*, *UrIn*, *Derm*, *Sati*, *Olfa*), "Interoceptive Awareness" (containing *ExAc*, *RelA*, *SexS*), and "Visceroception" (containing *Card*, *Resp*, *Gast*). Item centrality (intepretable as loadings in traditional factor analysis frameworks) are shown in Figure 3. Additionally, we applied hierarchical clustering analysis which replicated this structure (see Figure 2), suggesting consistency across methods.

The correlation matrix of the Mint dimensions revealed an interesting and rich tapestry of relationships (see Figure 4), with contrasting patterns of associations with dimensions from the same group. For instance, *Derm* and *Olfa*, despite being positively correlated with the other dimensions from the *Deficit* cluster, were not negatively correlated to *Awareness* and its dimensions. Similarly, *Sati*, unlike the other *Deficit* dimensions, was not positively correlated to *Visceroception*. This complex structure suggests that the dimensions indeed capture distinct aspects of interoception, and that the metaclusters, rather than being simple aggregates of rather-redundant elements, might actually capture unique combinations (greater than the sum of their parts). It also provides evidence against a simplistic adaptive *vs.* maladaptive dichotomy, as *Deficit* was not negatively bit positively correlated with *Visceroception*. Interestingly, while *Awareness* was also positively correlated with *Visceroception*, it yielded an insignificant negative correlation with *Deficit* (this finding was also aligned with the hierarchical clustering results, showing a greater proximity between *Deficit* and *Visceroception* than with *Awareness*), underlining again the complex web of relationships captured by the Mint.

The correlation matrix between the Mint and the other interoception scales revealed high levels of overlap, as well as some unique contributions (see Figure 4). The BPQ was positively correlated with most Mint dimensions (the highest with *Visceroception*, $r = .46$). The IAS was positively correlated with *Visceroception* and *Awareness* dimensions (the highest with *Awareness*, $r = .63$), but negatively with most *Deficit* dimensions (with the exception of *Olfa* and *Derm*). In turn, the *Visceroception* metacluster most strongly correlated with MAIA's Noticing ($r = .55$). Interestingly, MAIA's TTrusting correlated selectively with *Awareness*, and negatively with *Deficit* dimensions, but not with *Visceroceptive* dimensions (underlining its high-level metacognitive nature). MAIA's Emotional Awareness and Body Listening displayed a similar pattern to Noticing, and Attention Regulation and Self Regulation positively correlated with *Awareness* and *Visceroception* dimensions, but negatively with some *Deficit* dimensions (*CaCo* and *UrIn*). Not-distracting only yielded mild negative associations with *Sati* and *Olfa*. Overall, the Mint dimensions were able to capture most of the (relevant) variance and intricacies present in the other interoception scales.

Exploratory correlations with the emotion regulation (CERQ) dimensions revealed stronger associations with most of the MAIA dimensions (supporting its proximity with emotion regulation). Interestingly, Rumination (and Self-Blame) stood out as selectively related to the Mint's *Visceroception* and *Deficit* dimensions (note that Rumination was also related to the MAIA's interoceptive dimensions, namely Noticing, Emotional Awareness and Body Listening). The only primal world belief that correlated particularly with the Mint's *Visceroception* and the MAIA's interoceptive dimension was the belief that the world was alive (i.e., the belief that the universe is conscious, purposeful, and in active relationship with oneself, Clifton & Yaden, 2021).

Most of the target outcomes measured in the study to assess validity were best predicted by one of the Mint dimension (see Figure 5). This included Alexithymia (best predicted by *Deficit*); ERS' Arousal (best predicted by *Deficit*); CEFSA's anomalous experiences of the body (best predicted by *CaCo*), Self and Reality (best predicted by *Deficit*); ADHD and Autism (best predicted by *Deficit*), Somatic symptoms (best predicted by *Deficit*); BMI (best predicted by *Sati*). Most of the exceptions showed an advantage for MAIA dimensions, in particular Not-worrying (which best predicted ERS' Sensitivity and Persistence) and Trusting (which best predicted LS, Depression and Anxiety, CEFSA's Emotion and Connection, and self-reported physical activity).

Model comparison included 4 variants of the Mint scale,

[3] See correlation results to further justify this selection.

Figure 4

Correlation Matrices between the Mint dimensions (upper-left), and between the Mint and other interoception questionnaires (MAIA, BPQ and IAS; upper-right). The bottom matrix shows the relationships between interoception and other measures included in the study, such as alexithymia (TAS-20), depression and anxiety (PHQ-4), emotion reactivity (ERS), abnormal and dissociative experiences (CEFSA), and Primal World Beliefs. Stars indicate dimensions that have been score-reversed for better in-context interpretation. Correlation coefficients are shown only for significant correlations (p < .001).

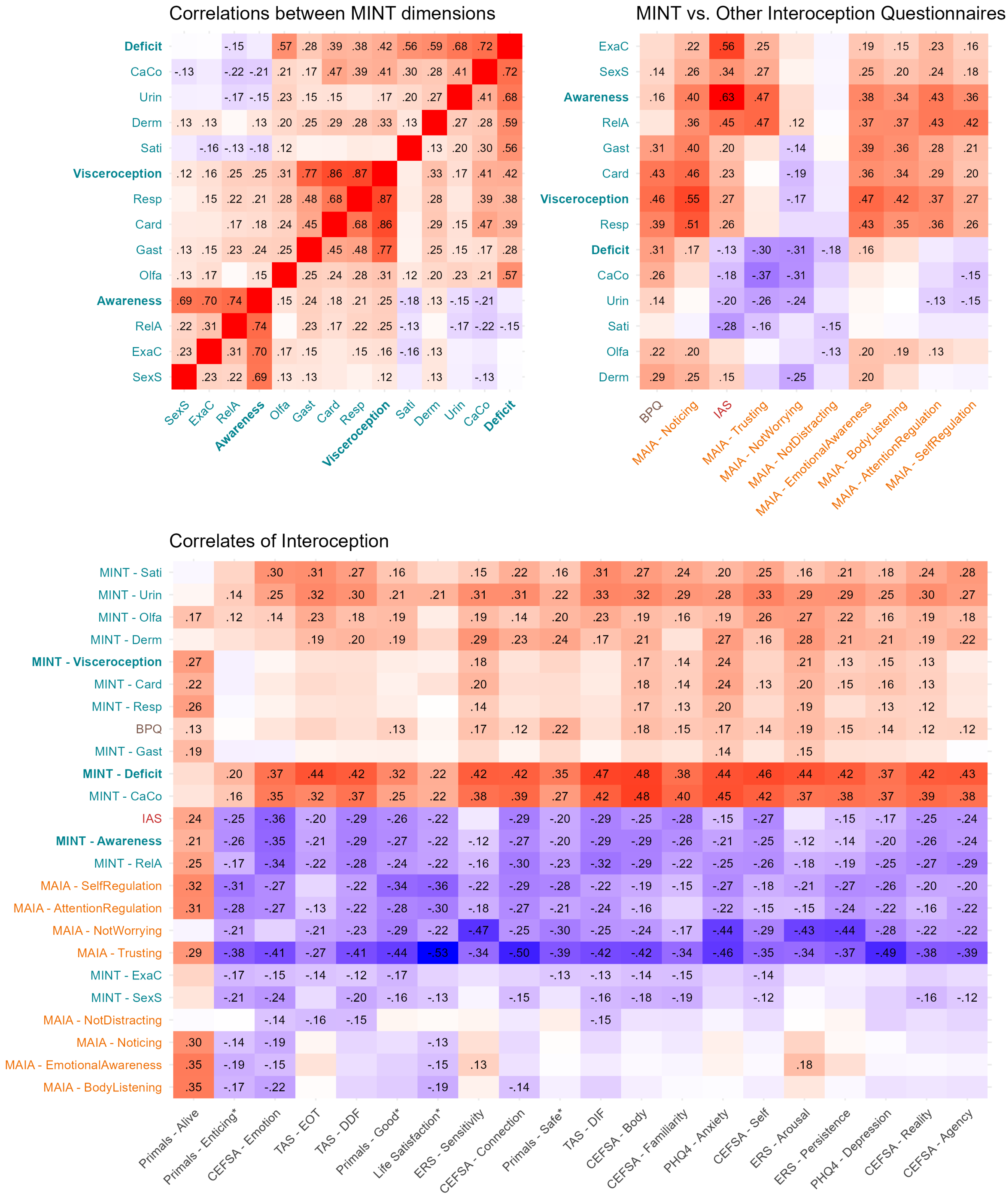


Figure 5

Summary table of the comparison between interoception questionnaires (Mint, MAIA, BPQ and IAS). For various outcomes included the study, we tested the Mint and Non-Mint dimension that had the strongest link (as a unique predictor). Values in parenthesis represents correlation coefficient for continuous variables and log-odds ratios for binary variables (the occurrence of mental and somatic health conditions). We also compared the predictive performance of regression models including multiple predictors, and present their ranking based on raw predictive power (R2) and BIC, which favours more parsimonious models (with less predictors). Green background indicates an advantage for the Mint, while red backgrounds indicate an advantage for another interoception scale. Orange backgrounds indicate an advantage for the MAIA driven by its emotion-regulation dimensions (Trusting and Not-Worrying).

Interoception Questionnaires Comparison

MINT vs. MAIA, IAS, BPQ

Outcome	Best Predictor (MINT)	Best Predictor (Non-MINT)	Best Models (R2)	Best Models (BIC)
Alexithymia				
Difficulty Identifying Feelings (DIF) *TAS-20*	Deficit *(0.47)*	MAIA - Trusting *(-0.42)*	**mint** > **metamint** > **minimint** > maia > **micromint** > ias > imaia > bpq	**metamint** > **mint** > **minimint** > maia > **micromint** > ias > bpq > imaia
Difficulty Describing Feelings (DDF) *TAS-20*	Deficit *(0.42)*	MAIA - Trusting *(-0.41)*	**mint** > **metamint** > maia > **minimint** > **micromint** > ias > imaia > bpq	**metamint** > **mint** > **minimint** > maia > **micromint** > ias > bpq > imaia
Externally Oriented Thinking (EOT) *TAS-20*	Deficit *(0.44)*	MAIA - Trusting *(-0.27)*	**mint** > **metamint** > **minimint** > maia > **micromint** > ias > bpq > imaia	**metamint** > **mint** > **minimint** > **micromint** > maia > ias > bpq > imaia
Emotional Reactivity				
Arousal *ERS*	Deficit *(0.44)*	MAIA - NotWorrying *(-0.43)*	maia > **mint** > **metamint** > **minimint** > **micromint** > imaia > bpq > ias	maia > **metamint** > **mint** > **minimint** > **micromint** > bpq > imaia > ias
Sensitivity *ERS*	Deficit *(0.42)*	MAIA - NotWorrying *(-0.47)*	maia > **mint** > **metamint** > **minimint** > **micromint** > bpq > imaia > ias	maia > **metamint** > **minimint** > **mint** > **micromint** > bpq > imaia > ias
Persistence *ERS*	Deficit *(0.42)*	MAIA - NotWorrying *(-0.44)*	maia > **mint** > **metamint** > **minimint** > **micromint** > ias > bpq > imaia	maia > **metamint** > **minimint** > **mint** > **micromint** > ias > bpq > imaia
Mood				
Life Satisfaction	Deficit *(-0.22)*	MAIA - Trusting *(0.53)*	maia > **mint** > **minimint** > **metamint** > **micromint** > ias > imaia > bpq	maia > **minimint** > **metamint** > ias > **micromint** > **mint** > imaia > bpq
Anxiety *PHQ-4*	CaCo *(0.45)*	MAIA - Trusting *(-0.46)*	maia > **mint** > **metamint** > **minimint** > **micromint** > bpq > ias > imaia	maia > **metamint** > **minimint** > **mint** > **micromint** > bpq > ias > imaia
Depression *PHQ-4*	Deficit *(0.37)*	MAIA - Trusting *(-0.49)*	maia > **mint** > **metamint** > **minimint** > **micromint** > ias > bpq > imaia	maia > **metamint** > **minimint** > **micromint** > **mint** > ias > bpq > imaia
Mental Health				
Mood Disorder	Deficit *(0.84)*	MAIA - Trusting *(-0.90)*	maia > **mint** > **metamint** > **micromint** > **minimint** > ias > bpq > imaia	maia > **metamint** > **micromint** > **minimint** > ias > bpq > **mint** > imaia
ADHD	Deficit *(0.60)*	IAS *(-0.48)*	**mint** > **metamint** > **minimint** > maia > **micromint** > ias > imaia > bpq	**metamint** > **minimint** > ias > **micromint** > bp > imaia > **mint** > maia
Autism	Deficit *(0.71)*	MAIA - Trusting *(-0.58)*	**mint** > **metamint** > **minimint** > maia > **micromint** > ias > imaia > bpq	**metamint** > **minimint** > ias > bpq > **micromin** > imaia > **mint** > maia
Somatic Health				
Afferent Sensitivity *Migraine, neuropathy, muscle tension, dizziness, ...*	Deficit *(0.50)*	MAIA - Trusting *(-0.36)*	**mint** > maia > **metamint** > **minimint** > **micromint** > bpq > ias > imaia	**metamint** > **minimint** > **micromint** > bpq > ia > imaia > maia > **mint**
Central Sensitization *Fibromyalgia, chronic fatigue, back pain, IBS, ...*	Deficit *(0.38)*	MAIA - Trusting *(-0.36)*	maia > **mint** > **metamint** > **minimint** > **micromint** > imaia > bpq > ias	**metamint** > **minimint** > bpq > ias > maia > **micromint** > imaia > **mint**
Autonomic Dysfunction *Hypermobility, chest pain, hypo/hypertension, ...*	Card *(0.46)*	MAIA - Trusting *(-0.20)*	**mint** > **minimint** > maia > **micromint** > **metamint** > bpq > imaia > ias	**minimint** > bpq > **micromint** > ias > **metamin** > imaia > **mint** > maia
Immune-Inflammatory *Allergies, eczema, autoimmune, ...*	Derm *(0.60)*	MAIA - Trusting *(-0.36)*	**mint** > maia > **metamint** > **minimint** > **micromint** > bpq > imaia > ias	**metamint** > **minimint** > bpq > **micromint** > **mint** > ias > maia > imaia
Dissociative Symptoms				
Body *CEFSA*	CaCo *(0.48)*	MAIA - Trusting *(-0.42)*	**mint** > **metamint** > **minimint** > **micromint** > maia > ias > bpq > imaia	**metamint** > **minimint** > **mint** > **micromint** > maia > ias > bpq > imaia
Self *CEFSA*	Deficit *(0.46)*	MAIA - Trusting *(-0.35)*	**mint** > **metamint** > **minimint** > **micromint** > maia > ias > bpq > imaia	**metamint** > **minimint** > **mint** > **micromint** > maia > ias > bpq > imaia
Emotions *CEFSA*	Deficit *(0.37)*	MAIA - Trusting *(-0.41)*	**mint** > **metamint** > maia > **minimint** > **micromint** > ias > imaia > bpq	**metamint** > **mint** > **minimint** > maia > **micromint** > ias > imaia > bpq
Reality *CEFSA*	Deficit *(0.42)*	MAIA - Trusting *(-0.38)*	**mint** > **metamint** > **minimint** > maia > **micromint** > ias > bpq > imaia	**metamint** > **minimint** > **mint** > **micromint** > maia > ias > bpq > imaia
Lifestyle				
BMI	Sati *(-0.17)*	IAS *(0.08)*	**mint** > **micromint** > **metamint** > **minimint** > maia > ias > imaia > bpq	ias > **micromint** > bpq > **metamint** > **minimin** > imaia > **mint** > maia
Physical Activity	RelA *(0.22)*	MAIA - Trusting *(0.36)*	maia > **mint** > **minimint** > **micromint** > imaia > **metamint** > ias > bpq	maia > **minimint** > **micromint** > imaia > **metamint** > ias > **mint** > bpq
Cardiac Monitoring	Card *(0.31)*	MAIA - BodyListening *(0.36)*	**mint** > maia > imaia > **metamint** > **minimint** > **micromint** > ias > bpq	imaia > **metamint** > **minimint** > ias > maia > bpq > **micromint** > **mint**
Sleep Monitoring	RelA *(0.22)*	MAIA - NotWorrying *(-0.22)*	**mint** > maia > **metamint** > imaia > **minimint** > **micromint** > bpq > ias	bpq > ias > **metamint** > imaia > **minimint** > **micromint** > maia > **mint**
Steps Monitoring	Awareness *(0.26)*	MAIA - Noticing *(0.28)*	**mint** > maia > **metamint** > imaia > **minimint** > **micromint** > ias > bpq	**metamint** > imaia > **minimint** > ias > **micromint** > bpq > **mint** > maia

the full *Mint* (including all lower-level clusters - 33 items), the *metaMint* (including only the 3 metaclusters based on all items), the *miniMint* (including the 3 metaclusters computed from 2 of its most loading triplets - 18 items), and the *microMint* (including only the most representative dimension of each metacluster - 9 items). This analysis confirmed the clear advantage for the Mint over the other interoception scales. The full *Mint* model was typically the best model based on R2, and the *metaMint* was the best model when parsimony was taken into account (i.e., based on BIC). Moreover, most of the instances were the *MAIA* was the best model were explainable by the inclusion of Trusting, and the interoception-only version *iMAIA* typically yielded worse performance than the Mint. In many instances, the *miniMint* yielded reasonable performance, although the *microMint* was typically less promising.

The predictive performance for the mental health outcomes (see Figure 6) displayed a consistent pattern, with an advantage for the *MAIA* for depression and anxiety which dropped below the *Mint* versions for its *iMAIA* variant. The *Mint*, however, displayed a clear advantage for the prediction of autism, ADHD, and all somatic groups of symptoms (with the exception of Central Sensitization, which was best predicted by the *MAIA*).

Finally, the importance of heart monitoring (for owners of such wearable) was best predicted by MAIA's Body listening (r = .36), followed the *Card* dimension of the Mint (r = .31). The importance of sleep monitoring importance was best predicted by *RelA* (r = .22), followed by MAIA's not worrying (r = −.219). Daily activity via steps monitoring was best predicted by MAIA's Noticing (r = .28), followed by *Awareness* (r = .26).

Discussion

The primary goal of Study 2 was to validate the items selected in Study 1. This involved re-assessing its structure in a new, large sample, and establishing its convergent and criterion validity against other established interoception questionnaires. The analysis led to a final, refined 33-item Mint scale with a stable and theoretically coherent hierarchical structure, which demonstrated substantial predictive power across a range of outcomes.

Re-validating the structure after item selection is a crucial step in questionnaire development. When a large item pool is reduced, the context changes, which can alter the perceived meaning and relationships between the remaining items. For instance, a questionnaire with many items suggesting abnormality or pathology might colour other neutral items with a negative connotation. While the original structure was mostly replicated, the re-validation allowed for a clarification of some of the higher-order dimensions (for instance, the "Sensitivity" metacluster became more clearly defined around core interoceptive clusters). The final structure, comprising 11 facets and 3 metaclusters, showed excellent structural consistency and stability, providing a solid foundation for the scale. However, the *Olfa* (Olfactory Compensation) cluster showed lower stability in its placement within the *Deficit* metacluster. This suggests its role may be more complex, and future studies are needed to investigate its place and invariance across different populations and contexts.

The criterion validity analysis revealed a clear and compelling pattern of results. The Mint scale, particularly its *Deficit* metacluster, consistently outperformed other scales in predicting outcomes directly related to alterations in bodily experience and self-regulation, such as alexithymia, dissociative experiences, ADHD, autism, and various somatic symptoms. The main contender, the MAIA-2, showed an advantage primarily for outcomes related to general well-being, such as life satisfaction, depression, and anxiety. However, this advantage was almost entirely driven by dimensions highly related to emotion regulation abilities, namely *Trusting* and *Not-Worrying*, whose inclusion as core interoceptive facets has been questioned (Ferentzi et al., 2021). When these dimensions were removed (in the *iMAIA* version), the MAIA-2's predictive power for these outcomes dropped considerably, often below that of the Mint.

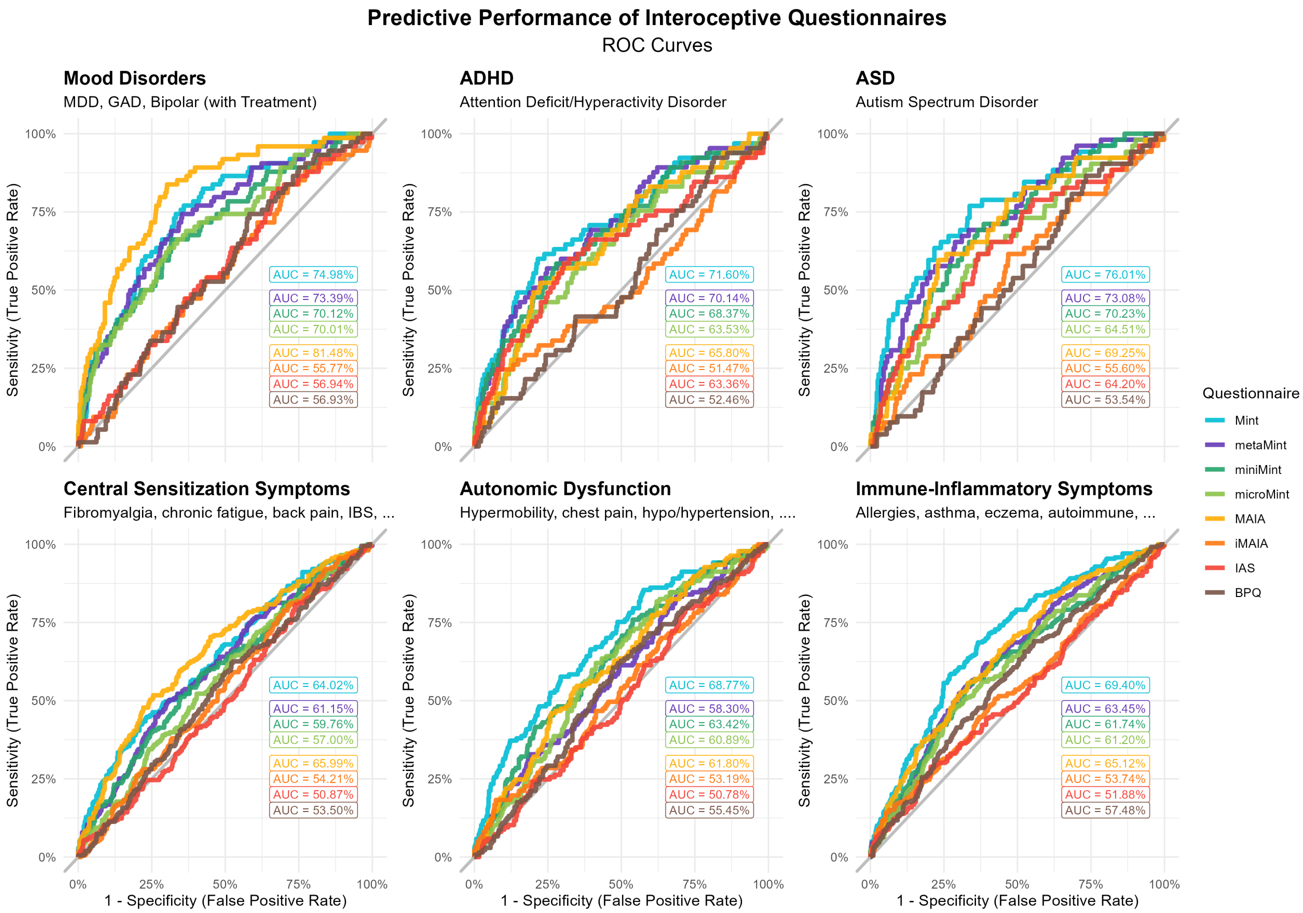
This highlights a crucial issue in the field: the potential for misattributing findings to "interoception" when they may be driven by broader emotion regulation capacities (Murphy et al., 2020). While valuable, the MAIA-2's most predictive subscales appear less related to core interoceptive processing and more to metacognitive beliefs about one's body and emotional coping. The Mint, in contrast, appears to tap more directly into distinct facets of interoceptive experience, from visceral sensitivity to perceived deficits. The pattern is clear: for high-level, metacognitive aspects of affective state and well-being, the MAIA-2's emotion regulation facets are highly relevant. For outcomes tied more specifically to bodily processes, clinical conditions characterized by sensory dysregulation, and nuanced interoceptive experiences, the Mint offers superior predictive validity and specificity.

General Discussion

Across two studies, we developed and validated the Multimodal Interoception Questionnaire (Mint), a new 33-item self-report measure of interoception. Our approach addressed key limitations of existing scales by starting with a systematically generated, broad item pool that explicitly controlled for context and modality, and then using a data-driven reduction process. The result is a psychometrically robust instrument with a stable, hierarchical structure comprising three higher-order metaclusters, namely *Interoceptive Deficit*, *Interoceptive Awareness*, and *Visceroception*, and 11 distinct lower-level facets. The validation study demonstrated that the Mint not only captures variance present in other popular scales (MAIA-2, BPQ and IAS) but also offers superior pre-

Figure 6

Predictive performance of interoception questionnaires for various self-reported mental health conditions (mood disorders with psychopharmacological treatment, ADHD and ASD) and somatic symptoms. The Receiver Operating Characteristic (ROC) curves are shown for the logistic regression models for various questionnaires and versions. A high Area Under the Curve (AUC) indicates a good discriminative power of the model (i.e., a strong combination of sensitivity and specificity). The Mint scale (in blue-green) consistently outperformed the other interoception scales, with the exception of the MAIA which, driven by its emotion-regulation dimensions (Trusting and Not-worrying), performed better for mood disorders and central sensitization symptoms. In several instances, the short versions of the Mint scale (miniMint and microMint) yielded reasonable performance, still outperforming other measures.



dictive power for a range of clinical and cognitive outcomes, particularly those tied to altered bodily perception.

A key contribution of the Mint is its potential ability to disentangle core interoceptive experiences from more general emotion regulation capacities. Our findings show that while the MAIA-2 is a potent predictor of general well-being, this is largely driven by its *Trusting* and *Not-Worrying* subscales, which heavily tap into emotion regulation skills and coping strategies. The Mint, conversely, excelled at predicting conditions like alexithymia, ADHD, autism, and somatic symptoms, where alterations in bodily sensing are a core feature. This suggests the Mint provides a more specific measure of interoceptive processing, offering researchers a finer-grained tool to probe the role of bodily awareness in health and disease.

Despite these strengths, the present work has limitations that open avenues for future research. First, as a self-report measure, the Mint captures subjective beliefs about interoceptive abilities, not objective accuracy. A crucial next step is to investigate the relationship between Mint scores and physiological or behavioural measures of interoception (e.g., heartbeat tracking tasks, heartbeat evoked potentials). This would help establish the extent to which subjective reports on the Mint correspond to objective interoceptive performance. Future research could also systematically examine the cost of shorter versions, by investigating the tradeoffs incurred by versions with only the one or two most central facets per metacluster and their impact on psychometric qualities and predictive power.

While the overall structure of the Mint is stable, some facets warrant deeper investigation. The *Olfa* (Olfactory Compensation) and *Derm* (Dermal Hypersensitivity) clusters, for instance, showed more complex patterns of association. Their placement within the *Deficit* metacluster is theoretically intriguing but requires further validation. The former aligns with predictive coding models suggesting that individuals may rely on exteroceptive cues (like odours) to infer internal states (Paulus & Stein, 2010), a form of pseudo-interoception (or extero-interoception). The latter points to a potential link between visceral and cutaneous sensitivity, possibly reflecting a general trait of heightened bodily awareness or a shared mechanism of central sensitization, which has been observed in conditions like chronic pain (Di Tella & Castelli, 2016). Future studies should examine the invariance of the Mint's structure across different populations (e.g., clinical vs. non-clinical groups) and contexts (e.g., heightened arousal vs. rest) to clarify the role of these more nuanced facets. Such work could reveal whether their relevance and pattern of associations differ in conditions like anxiety disorders, chronic pain, or neurodevelopmental disorders.

In conclusion, the Mint scale represents a significant methodological advancement for the field of interoception research. By providing a validated, structurally coherent, and nuanced tool, it allows for a more precise assessment of self-reported interoceptive experiences. Its various forms (full, meta, mini, and micro) offer flexibility for researchers and clinicians, paving the way for a deeper understanding of the multifaceted role of interoception in human psychology and pathophysiology.

Data Availability

Data, code, and all materials are available at https://github.com/RealityBending/InteroceptionScale.

Acknowledgements

We would like to thank the dissertation students from the University of Sussex for their help in data collection. DM would also like to thank Anastasia for the motivation provided to write this paper.

References

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.

Černis, E., Loe, B. S., Lofthouse, K., Waite, P., Molodynski, A., Ehlers, A., & Freeman, D. (2024). Measuring dissociation across adolescence and adulthood: Developing the short-form Černis felt sense of anomaly scale (ČEFSA-14). *Behavioural and Cognitive Psychotherapy*, *52*(2), 163–177.

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. (2021). On the equivalency of factor and network loadings. *Behavior Research Methods*, *53*(4), 1563–1580.

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

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

Craig, A. D. (2003). Interoception: The sense of the physiological condition of the body. *Current Opinion in Neurobiology*, *13*(4), 500–505.

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology*, *17*, 7–14. https://doi.org/10.1016/j.copsyc.2017.04.020

Crucianelli, L., & Ehrsson, H. H. (2023). The role of the skin in interoception: A neglected organ? *Perspectives on Psychological Science*, *18*(1), 224–238.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.

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., Maurage, P., & Corneille, O. (2025). Discrepancies in the definition and measurement of human interoception: A comprehensive discussion and suggested ways forward. *Perspectives on Psychological Science*, *20*(1), 76–98.

Di Tella, M., & Castelli, L. (2016). Alexithymia in chronic pain disorders. *Current Rheumatology Reports*, *18*(7), 41.

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.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

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

Garnefski, N., & Kraaij, V. (2006). Cognitive emotion regulation questionnaire–development of a short 18-item version (CERQ-short). *Personality and Individual Differences*, *41*(6), 1045–1053.

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.

Indahl, U. G., Næs, T., & Liland, K. H. (2018). A similarity index for comparing coupled matrices. *Journal of Chemometrics*, *32*(10), e3049.

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*.

Jovanović, V., & Lazić, M. (2020). Is longer always better? A comparison of the validity of single-item versus multiple-item measures of life satisfaction. *Applied Research in Quality of Life*, *15*, 675–692.

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

Kolacz, J., Holmes, L., & Porges, S. W. (2018). Body perception questionnaire (BPQ) manual. *Traumatic Stress Research Consortium*.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ–4. *Psychosomatics*, *50*(6), 613–621.

Leising, D., Grande, T., & Faber, R. (2009). The toronto alexithymia scale (TAS-20): A measure of general psychological distress. *Journal of Research in Personality*, *43*(4), 707–710.

Makowski, D., Te, A., Neves, A., & Chen, S. A. (2025). *Measuring depression and anxiety with 4 items? Adaptation of the PHQ-4 to increase its sensitivity to subclinical variability*. PsyArXiv. https://doi.org/10.31234/osf.io/436np_v2

Mayer, C.-D., Lorent, J., & Horgan, G. W. (2011). Exploratory analysis of multiple omics datasets using the adjusted RV coefficient. *Statistical Applications in Genetics & Molecular Biology*, *10*(1).

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.

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

Murphy, J., Catmur, C., & Bird, G. (2018). Classifying individual differences in interoception: Implications for the measurement of interoceptive accuracy. *Psychonomic Bulletin & Review*, *25*(4), 1424–1431. https://doi.org/10.3758/s13423-017-1339-z

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure and Function*, *214*(5), 451–463.

Sibson, R. (1978). Studies in the robustness of multidimensional scaling: Procrustes statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, *40*(2), 234–238.

Thériault, R., Ben-Shachar, M. S., Patil, I., Lüdecke, D., Wiernik, B. M., & Makowski, D. (2024). Check your outliers! An introduction to identifying statistical outliers in r with easystats. *Behavior Research Methods*, *56*(4), 4162–4172.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From louvain to leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 1–12.

Veilleux, J. C., Schreiber, R. E., Warner, E. A., & Brott, K. H. (2024). Development and validation of a brief version of the emotion reactivity scale: The b-ERS. *Current Psychology*, *43*(14), 12586–12600.

Vig, L., Köteles, F., & Ferentzi, E. (2022). Questionnaires of interoception do not assess the same construct. *PLoS One*, *17*(8), e0273299.

Zamariola, G., Maurage, P., Luminet, O., & Corneille, O. (2018). Interoceptive accuracy scores from the heartbeat counting task are problematic: Evidence from simple bivariate correlations. *Biological Psychology*, *137*, 12–17.