Peer Review File

Interoceptive ability is uncorrelated across respiratory and cardiac axes in a large scale psychophysical study

Corresponding Author: Dr Leah Banellis

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Version 0:

Decision Letter:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors **

Dear Dr Banellis,

Thank you for your patience during the peer-review process. Your manuscript titled "Interoceptive ability is uncorrelated across respiratory and cardiac axes: a large scale psychophysical study" has now been seen by 2 reviewers, and I include their comments at the end of this message. They find your work of interest but raised some important points. We are interested in the possibility of publishing your study in Communications Psychology, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response to the reviewers. Please highlight all changes in the manuscript text file.

Editorially, we consider it crucial that the robustness and recoverability of the parameter estimates are demonstrated in the revised manuscript. Please also address reviewers' concerns regarding the two tasks (HRDT & RRST), such as the difference in task demand between the two tasks.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

I am attaching an Editorial Requests Table that details critical reporting requirements for the revised manuscript. Please attend to each item and ensure your manuscript is fully compliant. If your revised manuscript is not aligned with these requests on major issues, such as those concerning statistics, it may be returned to you for further revisions without re-review.

Please submit the following items:

- Revised manuscript
- Point-by-point response to the referees' comments
- Cover letter (as a separate document)
- Nature Research Reporting Summary
- Completed Editorial Request Table (attached).

via this link: Link Redacted .

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Additional guidance is available in our style and formatting guide Communications Psychology formatting guide.

We hope to receive your revised paper within 8 weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, provided it still presents a significant contribution to the literature at that stage.

We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Troby Lui

Troby Lui, PhD
Associate Editor
Communications Psychology

REVIEWER EXPERTISE:
Reviewer #1: interoception
Reviewer #2: interoception, metacognition/confidence

REVIEWER REPORTS:

Reviewer #1 (Remarks to the Author):

This is a well-written, methodologically rigorous, and conceptually important paper that I believe makes a valuable contribution to the field of interoception research.

The psychophysical approach is a clear addition to the traditional interoceptive measures, and the use of a large sample adds credibility to the conclusions. The study is methodologically sound and transparent, with open data, and thorough reporting. The Bayesian framework used to test for null correlations is appropriately applied and strengthens the authors' claims regarding modality specificity.

I have some suggestions that could strengthen introduction/discussion:

I feel a broader contextualisation of the novel tasks (HRDT & RRST) presented in this paper is necessary:

First, while the authors refer to HRDT as an improvement over traditional heartbeat counting/discrimination tasks, it would be helpful to give a brief overview of whether and how HRDT is related to these more traditional tasks. If HRDT is un-correlated to other cardioception tasks, then that seems to complicate the interpretation further of interoception as modality-specific rather than task-specific.

Second, regarding RRST, the paper assumes a degree of familiarity with the respiratory interoception literature that some readers may not have. A brief review of other commonly used respiratory tasks and how the novel RRST compares and relates to them would be helpful.

Another suggestion I have is to include a more critical discussion of task demands:

HRDT seems to require integration between internal (cardiac) and external (auditory) signals, potentially engaging multisensory or crossmodal comparison processes. In contrast, RRST appears to be intra-modal (respiratory alone). Could the observed lack of correlation be partially due to these differences in task demands, rather than reflecting a clean separation of interoceptive axes?

As a related point, would a tactile–auditory rate discrimination task (rather than purely auditory) had served as a better exteroceptive control, given its closer alignment to the multisensory nature of HRDT?

Lastly, I feel there could be more clarity on the next steps for future research.
The authors suggest broad directions for future research (larger samples, longitudinal designs, neuroimaging). These are

important, but it would be helpful to tie this more directly to the current findings. So do we, as a field, have the definite tasks then, how should we address the difference in task demands. I feel these questions align more the aim of the current work and I would have liked to see specific ideas for follow-up studies in the context of those.

As a minor point, maybe I misunderstood, but I thought that correlation was only in metacognitive bias and not metacognitive efficiency? There is a paragraph in the discussion, where the authors say: “Comparably, we observe a domain-general relationship of overall confidence but modality- specific effects of metacognitive efficiency across cardiac and respiratory domains. We utilised M-Ratio as our estimate of metacognitive efficiency which attempts to correct for differences in underlying task performance, although recent studies highlight a residual potential for biases from confidence criterion and decision accuracy (Rahnev, 2025; Rahnev & Fleming, 2019; Xue et al., 2021b, 2021a).” I got a bit lost here. Perhaps a more intuitive explanation of what M-Ratio stands for and a bit more on how to reconcile that the correlation was only between the bias but not efficiency.

Thank you for the excellent work!
Irena

Reviewer #2 (Remarks to the Author):

Summary and Overall Assessment
This is a well-written and highly relevant manuscript that makes a valuable contribution to the interoception literature. The study addresses a fundamental question—whether interoceptive abilities are domain-general or modality-specific—using a large sample and consistent psychophysical methods across cardiac, respiratory, and auditory domains. The authors find that interoceptive sensitivity, precision, and metacognitive measures do not significantly correlate across modalities, with the exception of subjective confidence. Bayesian analyses support the absence of correlation across most dimensions, pointing toward modality-specific interoceptive processing, while confidence may reflect a domain-general metacognitive process.

This is an important and timely study that addresses an important question in the field. The paper is clearly written, methodologically careful, and theoretically grounded. Given that many of the key conclusions rest on model-derived parameters, my main suggestions relate to confirming the robustness and recoverability of these estimates, particularly the psychometric slopes.

Major Comments
HRDT Task. The HRDT-Intero task is described as involving auditory feedback tones presented at varying tempos above or below the participant’s heart rate. One point where additional detail would be helpful is in how the auditory feedback tones were generated in relation to participants’ real-time heart rate. Specifically, could the authors clarify how heart rate was estimated during the task? Was it based on a moving average of the last five interbeat intervals, or was a longer time window used? The reason I ask is that using only five IBIs may introduce variability due to natural heart rate fluctuations (HRV), potentially affecting the stability of the reference rate. A slightly longer sliding window could help smooth out these fluctuations and yield a more robust estimate for the comparison tones.

In addition, it would be helpful to know whether the auditory tone sequences included any variability or jitter to simulate the natural fluctuations in actual heart rhythms. If such variability was incorporated, was it individualized based on each participant’s heart rate variability (HRV), or was a standardized approach used across participants?

Use of Pulse Oximeter for Heart Rate Measurement
One methodological detail that may warrant further clarification is the use of the pulse oximeter for measuring heart rate. Pulse oximeters can sometimes produce subtle but perceptible pressure or tactile sensations on the finger, especially when worn for longer durations. This opens the possibility that some participants may have received incidental feedback about their heartbeat from the device itself—potentially influencing performance on tasks intended to assess internal interoceptive awareness. Did the authors assess whether participants were aware of such feedback, for instance through a post-task questionnaire or debriefing?

Cardiac Results From Figure 1, it appears that participants exhibit a systematic bias in the cardiac task, with threshold estimates clustering around a delta BPM of approximately -10 and the algorithm iterates in step sized of almost 30 BPM. I found this striking, as a 10 BPM or even 30 BPM difference is relatively large in the context of typical resting heart rates, which usually fall between 50 and 80 BPM. Given this, I would have expected participants to more easily detect such a discrepancy. I’m curious whether this kind of bias is a known phenomenon in heartbeat discrimination tasks, or whether there might be other explanations—for example, anchoring biases, or task-specific perceptual confounds (realtively short time window to estimate HR over 5 beats).

As a follow-up, it would be informative to examine whether threshold or slope estimates in the cardiac task are systematically related to an individual’s resting heart rate. For example, is there a correlation between absolute heart rate and either the estimated threshold or precision? Including such an analysis would not only clarify whether baseline physiology influences task performance, but also help interpret the observed group-level bias in a more nuanced way.

Breathing Resistance. A related point concerns the breathing resistance task. From the reported results, it appears that participants approached the upper end of the resistance scale, with many settling around 90% resistance. I was curious whether 100% resistance in this context corresponds to full occlusion of the breathing tube. If so, that would suggest that participants were experiencing very strong resistance levels, and the data may be approaching a ceiling effect. Is it known that individuals are generally quite insensitive to large inspiratory resistances, or might this reflect a task-specific limitation?

Additionally, if a subset of participants converged at or near the maximum resistance level, this could potentially skew group-level estimates of both threshold and precision. For instance, it might compress variability across individuals or artificially inflate slope estimates for those whose responses saturated near ceiling. If such a pattern was observed, it would be helpful to know how it was addressed—e.g., whether any exclusions were applied or whether the modeling approach was adjusted to account for this.

Modeling Considerations and Parameter Robustness
Given that the central findings of the study rely on psychometric parameters—particularly threshold and slope estimates derived from the Psi adaptive staircase method—it would be helpful to include additional validation analyses to confirm the robustness and recoverability of these estimates. While the Psi method is a powerful and efficient approach for estimating both the point of subjective equality (PSE) and slope, it is also known that slope estimates can be less stable than thresholds, especially when the number of trials is limited or when the stimulus range does not adequately sample the steep region of the psychometric function. Since these potential limitations apply to both tasks in the present study, it would be useful to reassure readers that the model parameters are reliable under the conditions used. I would recommend adding the following checks:

1. Parameter Recovery Analysis: One straightforward and informative approach is to conduct a parameter recovery analysis. This involve generating synthetic observers with known psychometric parameters (e.g., using a cumulative Gaussian with specified threshold and slope values) and simulating responses over the same number of trials as used in the actual experiment. Including realistic noise and lapse rates would increase the ecological validity of the simulation. The degree to which Psi recovers the original parameters—particularly slope—is then evaluated by plotting recovered versus true values for slope and threshold.

2. Sensitivity to the prior: In addition, it would be valuable to assess the sensitivity of parameter estimates to the choice of prior. Since Psi uses a Bayesian updating framework, the width and shape of the prior distribution can meaningfully influence the resulting estimates. I would recommend repeating the recovery simulations with a range of priors—wide vs. narrow, centered vs. offset—to determine how robust the slope estimates are to potential prior mis-specification.

3. Stimulus Grid: A third important factor to consider is the initial stimulus grid used in the Psi algorithm. The range and placement of stimuli directly affect the algorithm's ability to sample the psychometric function effectively. It would therefore be helpful to simulate Psi runs using different grid configurations—for example, one that samples the entire function well, one that is too narrow to capture the tails, and one that is misaligned with the true threshold. This analysis could reveal how grid setup influences the reliability of both slope and threshold estimates. In the heartbeat task, in particular, it appears that some individuals may not have had dense sampling in the steep region of the psychometric function, which could compromise slope estimation for those participants. Similarly in the breathing task the grid based on possible breathing resistances could explain the correlation between slope and threshold. See comment below.

Interindividual Differences and Model Transparency
Given the focus on individual differences in interoceptive ability, it would be extremely valuable to provide access to individual-level model fits—ideally through the GitHub repository mentioned in the manuscript. This would allow interested readers and researchers to assess model convergence and fit quality across participants and to verify that the reported group-level inferences are grounded in robust individual-level estimates. Even a representative subset of psychometric curves or posterior distributions would help illustrate the reliability of the modeling procedure across the sample.

Correlation Between Threshold and Slope in the Breathing Task
One aspect that stood out to me was the relatively strong correlation between threshold and slope in the breathing task. Typically, one would not expect a substantial correlation between these two parameters, as they capture distinct psychophysical properties: threshold reflects bias, while slope reflects precision. While it is certainly possible that this relationship reflects a meaningful psychological trait—such as individuals with lower interoceptive sensitivity also demonstrating more precise discrimination—it may also be a byproduct of the modeling approach.

Specifically, the correlation could be an artifact of over-shrinkage in the hierarchical model, where the prior structure induces correlation in the group-level posteriors when individual-level data are noisy or sparse. One way to assess this possibility would be to simulate hierarchical datasets in which threshold and slope are uncorrelated at the individual level, then refit the same model and examine whether the recovered posteriors show spurious correlation. If a correlation emerges in the absence of one in the generative process, this would suggest the observed effect may be model-induced.

Another reason could be the above mentioned ceiling effects close to a 100% resistance.

Discussion: The authors conclude that interoceptive sensitivity, precision, and metacognitive efficiency are largely domain-specific, challenging the notion of a general interoceptive dysfunction. While the data are compelling, one caveat worth mentioning is that the respiratory task specifically targets detection of an artificial inspiratory resistance, rather than more naturalistic dimensions of respiration such as breath frequency or tidal volume. This distinction is important, as it limits the generalizability of the findings to other aspects of respiratory interoception. For example, respiratory sinus arrhythmia (RSA) reflects the strong coupling between breathing and heart rate. Overall we knwo that there are significant changes in heart rate occurring across a single breathing cycle. Given this, it remains possible that naturalistic aspects of breath perception may show different interrelationships with cardiac interoception than those observed in this resistance detection task.

As a related side note—though not something that needs to be addressed within the scope of this paper—it would be interesting in future work to examine whether cardiac interoceptive accuracy varies across different phases of the respiratory cycle. If the authors collected respiratory data during the heartbeat task, it could open the door to exploratory analyses on this front, though I mention this more as a speculative thought than as a required addition.

Frederike Petzschner

Version 1:

Decision Letter:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors **

Dear Dr Banellis,

Your manuscript titled "Interoceptive ability is uncorrelated across respiratory and cardiac axes: a large scale psychophysical study" has now been seen by our reviewer, whose comments appear below. Reviewer #2 could not provide a report in this round; we have therefore elicited feedback from an editorial board member who is an expert in the domain. In light of their advice I am delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Psychology.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers and a list of editorial requests. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Although the returning reviewer is satisfied with the revisions, editorially, we require that you clarify whether this analysis is part of several other analyses derived from the same dataset. Please report thresholds rather than Bayes Factors in Figure 4. Last but not least, please nuance the claims about domain specificity and address the possibility that one measure could be more biased and less sensitive than the other.

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table". Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed here https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Psychology is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit https://www.nature.com/commspsychol/open-access

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

* TRANSPARENT PEER REVIEW: Communications Psychology uses a transparent peer review system. On author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons.

* CODE AVAILABILITY: All Communications Psychology manuscripts must include a section titled "Code Availability" at the end of the methods section. We require that the custom analysis code supporting your conclusions is made available in a publicly accessible repository at this stage; please choose a repository that generates a digital object identifier (DOI) for the code; the link to the repository and the DOI must be included in the Code Availability statement. Publication as Supplementary Information will not suffice.

* DATA AVAILABILITY:
All Communications Psychology manuscripts must include a section titled "Data Availability" at the end of the Methods section. More information on this policy, is available in the Editorial Requests Table and at http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf.

Please use the following link to submit the above items:
Link Redacted
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Troby Lui

Troby Lui, PhD
Associate Editor
Communications Psychology

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have addressed my comments.

Reviewers Response

We sincerely thank the editor and reviewers for their thoughtful efforts and valuable feedback on our manuscript. The referee's comments are presented in black, our responses are in blue, and quoted changes to the manuscript are formatted in quoted blue italics.

REVIEWER EXPERTISE:
Reviewer #1: interoception
Reviewer #2: interoception, metacognition/confidence

REVIEWER REPORTS:

Reviewer #1 (Remarks to the Author):

This is a well-written, methodologically rigorous, and conceptually important paper that I believe makes a valuable contribution to the field of interoception research.

The psychophysical approach is a clear addition to the traditional interoceptive measures, and the use of a large sample adds credibility to the conclusions. The study is methodologically sound and transparent, with open data, and thorough reporting. The Bayesian framework used to test for null correlations is appropriately applied and strengthens the authors' claims regarding modality specificity.

Thank you so much for helping us improve our manuscript for publication. We are deeply appreciative of your positive comments regarding the manuscript!

I have some suggestions that could strengthen introduction/discussion:

I feel a broader contextualisation of the novel tasks (HRDT & RRST) presented in this paper is necessary:

First, while the authors refer to HRDT as an improvement over traditional heartbeat counting/discrimination tasks, it would be helpful to give a brief overview of whether and how HRDT is related to these more traditional tasks. If HRDT is un-correlated to other cardioception tasks, then that seems to complicate the interpretation further of interoception as modality-specific rather than task-specific.

We appreciate your comment highlighting the importance of understanding relationships among interoceptive tasks within each organ modality. In the original publication of these paradigms (Legrand et al., 2022; Nikolova et al., 2022), each task was directly compared with traditional measures targeting the same modality, specifically, the HRDT with the traditional heartbeat counting task, and the RRST with the respiratory filter detection task. To clarify this, we have now added the following text to the manuscript:

> *"To address these limitations, more recent psychophysical tasks such as the Heart Rate Discrimination Task (HRDT) and the Respiratory Resistance Sensitivity Task (RRST) measure interoceptive sensitivity, precision, and metacognitive ability across cardiac and respiratory modalities, while also being validated with respect to previous interoception tasks (i.e., the heartbeat counting task, or the respiratory filter detection task) (Legrand et al., 2022; Nikolova et al., 2022)."*

Second, regarding RRST, the paper assumes a degree of familiarity with the respiratory interoception literature that some readers may not have. A brief review of other commonly used respiratory tasks and how the novel RRST compares and relates to them would be helpful.

Thank you for directing our attention to areas in our introduction which can be improved. We have added more information on previous respiratory interoception tasks, including methodological details in which the RRST improves:

"Respiratory interoception tasks such as those based on the detection or discrimination of breathing resistance through a circuit via the manual addition of filters (Garfinkel et al., 2016; Harrison et al., 2021; Zhao et al., 2002), also suffer from similar issues including low stimulus granularity, unreliable estimates of performance, and an absence of information regarding precision (i.e., slope)."

Another suggestion I have is to include a more critical discussion of task demands:

HRDT seems to require integration between internal (cardiac) and external (auditory) signals, potentially engaging multisensory or crossmodal comparison processes. In contrast, RRST appears to be intra-modal (respiratory alone). Could the observed lack of correlation be partially due to these differences in task demands, rather than reflecting a clean separation of interoceptive axes?

Thank you for providing the opportunity for us to address the crossmodal task differences. While indeed the HRDT involves multisensory processes of cardiac and auditory integration, the RRST is primarily intra-modally focused on respiration. However, respiroception itself is fundamentally more influenced by a combination of interoceptive and exteroceptive sensory information as it involves both tactile and thermosensation of air in the mouth and nose, stretch sensations of the chest during the expansion/contraction of the lungs, and chemosensation of blood gas changes (oxygen/carbon dioxide). Despite this, we agree with your suggestions for future research addressing this by increasing the similarity of the cardiac and respiratory tasks, as well as control for exteroceptive influences by adding a mutli-sensory bodily-related tactile component to the exteroceptive control task. We added details of this below:

"Third, the cardiac and respiratory tasks differ in their multisensory (HRDT is cardiac and auditory) or unimodal (RRST is respiratory alone) nature. This is an overall issue with advanced cardiac interoception tasks as it is difficult to manipulate cardiac signals, however a more similar bodily-related exteroceptive control involving multisensory tactile-auditory integration could partially serve as a method for controlling for exteroceptive auditory influences."

As a related point, would a tactile–auditory rate discrimination task (rather than purely auditory) had served as a better exteroceptive control, given its closer alignment to the multisensory nature of HRDT?

We agree that a multisensory bodily-related discrimination task could serve as a more similar exteroceptive control. Therefore, as suggested, we have added details of this in our discussion (see points directly above and below).

Lastly, I feel there could be more clarity on the next steps for future research.
The authors suggest broad directions for future research (larger samples, longitudinal designs, neuroimaging). These are important, but it would be helpful to tie this more directly to the current findings. So do we, as a field, have the definite tasks then, how should we address the difference in task demands. I feel these questions align more the aim of the current work and I would have liked to see specific ideas for follow-up studies in the context of those.

Thank you so much for noticing that our future directions could be more specific. Considering this we expanded upon this section in the Discussion:

"Longitudinal and intervention-based designs that manipulate physiological states, such as exercise or pharmacological stimulation, may help elucidate whether the absence of correlation between cardiac and respiratory interoception is specific to resting conditions or persists under altered physiological states (Khalsa et al., 2009; Pollatos et al., 2007). Future research should minimise design asymmetries between the cardiac and respiratory tasks, for example by pairing the HRDT with a tactile–auditory exteroceptive control or by adding an exteroceptive reference to the RRST. This would render both tasks comparably multisensory and provide a bodily–exteroceptive control that accounts for multimodal auditory influences."

As a minor point, maybe I misunderstood, but I thought that correlation was only in metacognitive bias and not metacognitive efficiency? There is a paragraph in the discussion, where the authors say: "Comparably, we observe a domain-general relationship of overall confidence but modality- specific effects of metacognitive efficiency across cardiac and respiratory domains. We utilised M-Ratio as our estimate of metacognitive efficiency which attempts to correct for differences in underlying task performance, although recent studies highlight a residual potential for biases from confidence criterion and decision accuracy (Rahnev, 2025; Rahnev & Fleming, 2019; Xue et al., 2021b, 2021a)." I got a bit lost here. Perhaps a more intuitive explanation of what M-Ratio stands for and a bit more on how to reconcile that the correlation was only between the bias but not efficiency.

We can see how that section of the Discussion could be confusing. To increase clarity, we emphasise that the only crossmodal relationship is with overall confidence (bias), not with metacognitive efficiency (M-Ratio):

> *"Importantly, the crossmodal relationship was confined to overall confidence only, while metacognitive efficiency did not correlate across cardiac and respiratory domains. Bias reflects overall confidence reflecting how liberally they rate high confidence independent of performance, and thus often generalises across tasks (Ais et al., 2016; Hoogervorst et al., 2025; Lund et al., 2023; Mazancieux et al., 2020). In contrast, M-Ratio indexes metacognitive calibration: how tightly confidence tracks trial-by-trial performance after accounting for perceptual sensitivity, it therefore depends on modality-specific sensory noise and priors."*

Thank you for the excellent work!
Irena

Reviewer #2 (Remarks to the Author):

Summary and Overall Assessment
This is a well-written and highly relevant manuscript that makes a valuable contribution to the interoception literature. The study addresses a fundamental question—whether interoceptive abilities are domain-general or modality-specific—using a large sample and consistent psychophysical methods across cardiac, respiratory, and auditory domains. The authors find that interoceptive sensitivity, precision, and metacognitive measures do not significantly correlate across modalities, with the exception of subjective confidence. Bayesian analyses support the absence of correlation across most dimensions, pointing toward modality-specific interoceptive processing, while confidence may reflect a domain-general metacognitive process.

This is an important and timely study that addresses an important question in the field. The paper is clearly written, methodologically careful, and theoretically grounded. Given that many of the key conclusions rest on model-derived parameters, my main suggestions relate to confirming the robustness and recoverability of these estimates, particularly the psychometric slopes.

We are deeply appreciative of your recognition of the strengths and impact of our study! Thank you for helping us improve the manuscript, we have responded to your concerns below.

Major Comments
HRDT Task. The HRDT-Intero task is described as involving auditory feedback tones presented at varying tempos above or below the participant's heart rate. One point where additional detail would be helpful is in how the auditory feedback tones were generated in relation to participants' real-time heart rate. Specifically, could the authors clarify how heart rate was estimated during the task? Was it based on a moving average of the last five interbeat intervals, or was a longer time window used? The reason I ask is that using only five IBIs may introduce variability due to natural heart rate fluctuations (HRV), potentially affecting the stability of the reference rate. A slightly longer sliding window could help smooth out these fluctuations and yield a more robust estimate for the comparison tones.

We apologise that the details of the auditory feedback tones generation was not in the manuscript, and appreciate the inquiry as to the role of the listening window. We settled on a 5 second window in the original task design as longer windows would actually be subject to a greater possible influence of heart-rate variability, for example of inhalation, rendering the comparison less clear (i.e., a longer window could contain discrete faster and slower phases, and it would be unclear what the participant should compare the reference stimuli to). We do agree however that addressing this empirically is an interesting topic for future study.

To clarify the recording of heart rate, this was estimated during the 5 second 'heart listening' phase via a finger pulse oximeter (PPG) on the non-dominant hand. Details of the online systolic peak detection algorithm has been added to the methods:

> *"During this phase, average heart rate was measured in real time using photoplethysmography, with signal peaks enhanced by squaring and normalising the signal over a 750 ms rolling window. If implausible heart rates (<40 or >120 BPM) were detected, an error message prompted the participant to keep still, and the trial was restarted."*

We have also added a note with a suggestion for future research to the Discussion:

> *"As in previous studies using the HRDT (Bodart et al., 2024; Jeganathan et al., 2024; Leganes-Fonteneau, 2024; Legrand et al., 2022; Tyrer et al., 2025), we here utilised a five second sliding window for the listening interval. This is an intentional design choice which aims to limit the role of heart rate variability in the discrimination window. However, an open question concerns whether longer intervals might impact the psychometric function, and indeed, if participants maintain a rolling average estimate of their heart, at what interval."*

In addition, it would be helpful to know whether the auditory tone sequences included any variability or jitter to simulate the natural fluctuations in actual heart rhythms. If such variability was incorporated, was it individualized based on each participant's heart rate variability (HRV), or was a standardized approach used across participants?

Thank you for your question. The heart rate feedback is generated directly from the psychophysical staircasing procedure, where the offset interval is determined by the adaptive procedure. Introducing jitter into these test values would reduce the efficiency of the estimation procedure and require more trials to reach stable convergence. But the stimuli will reflect the natural HRV of the participant inherently, as they are presented at test intervals above and below their true heart rate.

Use of Pulse Oximeter for Heart Rate Measurement
One methodological detail that may warrant further clarification is the use of the pulse oximeter for measuring heart rate. Pulse oximeters can sometimes produce subtle but perceptible pressure or tactile sensations on the finger, especially when worn for longer durations. This opens the possibility that some participants may have received incidental feedback about their heartbeat from the device itself—potentially influencing performance on tasks intended to assess internal interoceptive awareness. Did the authors assess whether participants were aware of such feedback, for instance through a post-task questionnaire or debriefing?

This indeed can be an issue with the use of certain fingertip pulse oximeters during cardiac interoception tasks. However, to mitigate such effects for the heart rate discrimination task, we used a specific 'softtouch' pulse oximeter product, and instructed participants to reposition to another finger if they felt heartbeat sensations from the pressure of the pulse oximeter itself. To ensure this is clear we added the following to the methods:

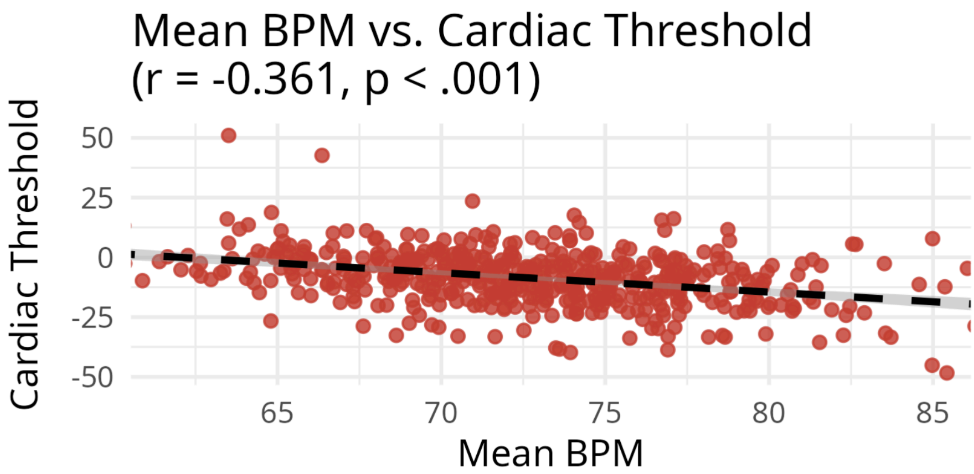
> *"To minimise tactile feedback, we used a soft-clip pulse oximeter, previously shown not to induce fingertip sensations (Murphy et al., 2019), and participants were instructed to reposition the device to another fingertip if sensations occurred."*

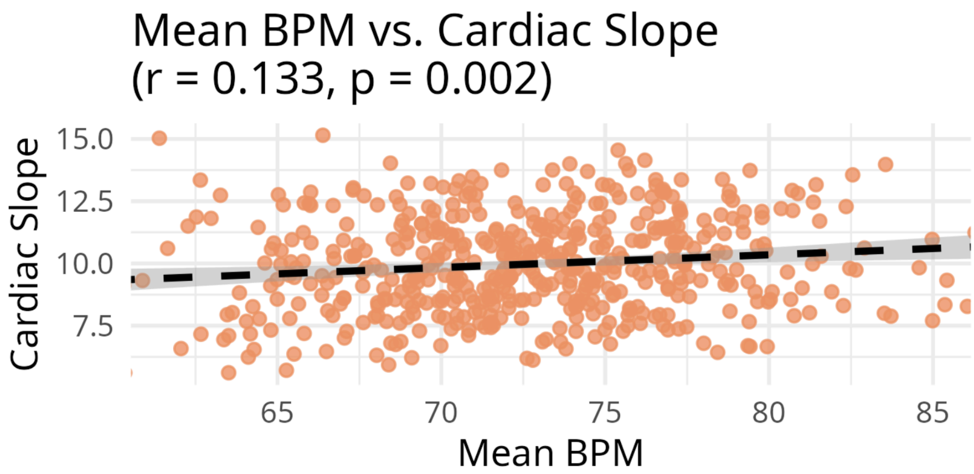
Cardiac Results From Figure 1, it appears that participants exhibit a systematic bias in the cardiac task, with threshold estimates clustering around a delta BPM of approximately -10 and the algorithm iterates in step sized of almost 30 BPM. I found this striking, as a 10 BPM or even 30 BPM difference is relatively large in the context of typical resting heart rates, which usually fall between 50 and 80 BPM. Given this, I would have expected participants to more easily detect such a discrepancy. I'm curious whether this kind of bias is a known phenomenon in heartbeat discrimination tasks, or whether there might be other explanations—for example, anchoring biases, or task-specific perceptual confounds (relatively short time window to estimate HR over 5 beats).

Thank you for your interesting point. The psychometric traces in Figure 1 are only from a single example participant. Inter-individual differences in threshold values can be seen in Figure 2 (mean raw HRDT threshold = -8.610, SD = 11.218). Indeed, we and others have found that participants at rest tend to underestimate their BPM by ~8 BPM. This aligns well with previous studies which show consistent heart rate underestimation in >95% of individuals at rest using traditional cardiac interoception tasks (Zamariola et al., 2018). The algorithm step sizes are not however constant, as they are adjusted continuously by the psi algorithm and will generally be near 1BPM steps as model confidence increases. While we do not currently know precise psychophysiological mechanisms drive heart-rate estimation, we agree it is an interesting question for future research, one which is being actively investigated by our own and other labs around the world.

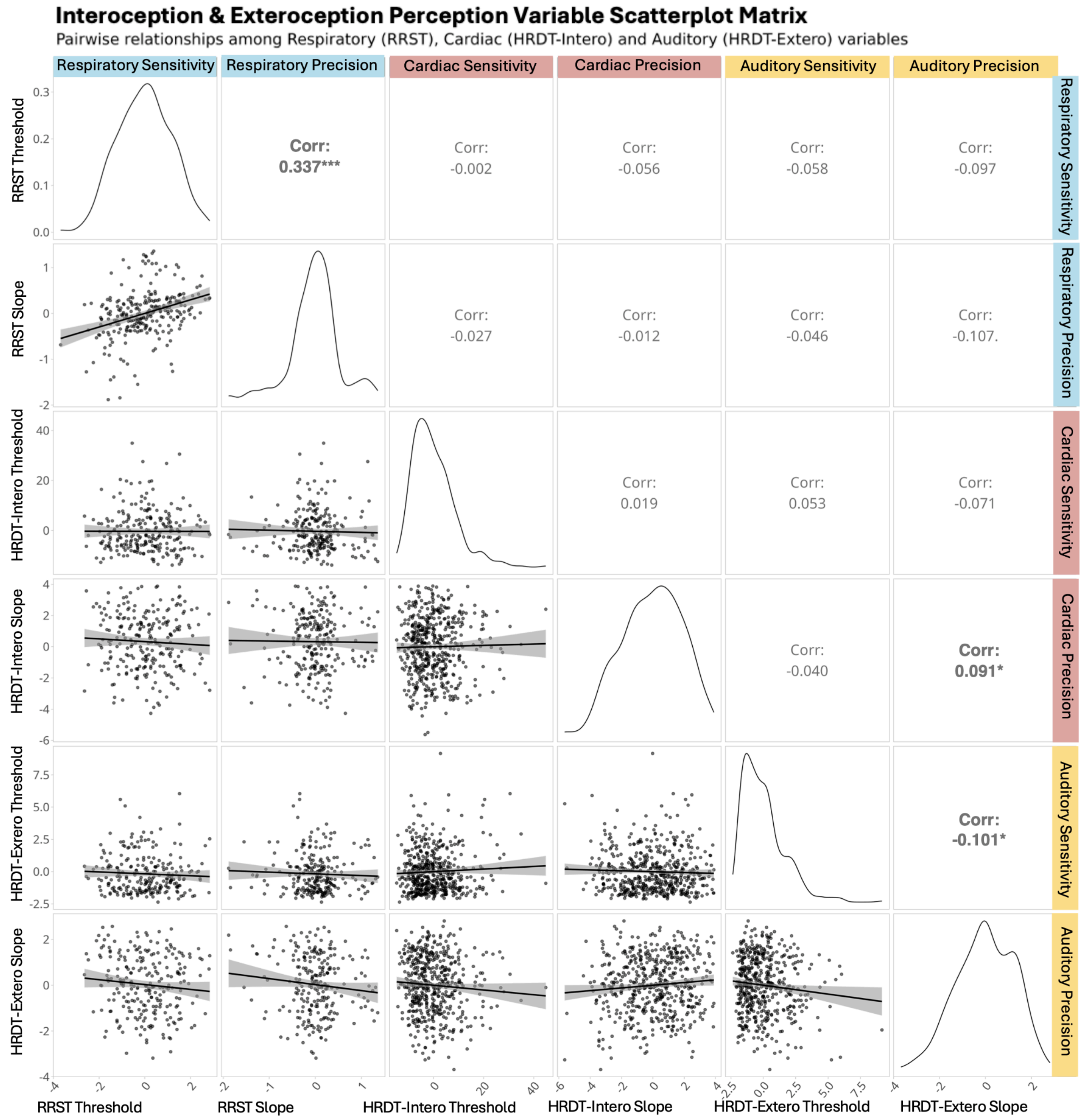
As a follow-up, it would be informative to examine whether threshold or slope estimates in the cardiac task are systematically related to an individual's resting heart rate. For example, is there a correlation between absolute heart rate and either the estimated threshold or precision? Including such an analysis would not only clarify whether baseline physiology influences task performance, but also help interpret the observed group-level bias in a more nuanced way.

This is an important point, heart rate can relate to threshold and slope estimated by the HRDT. We illustrate this here:



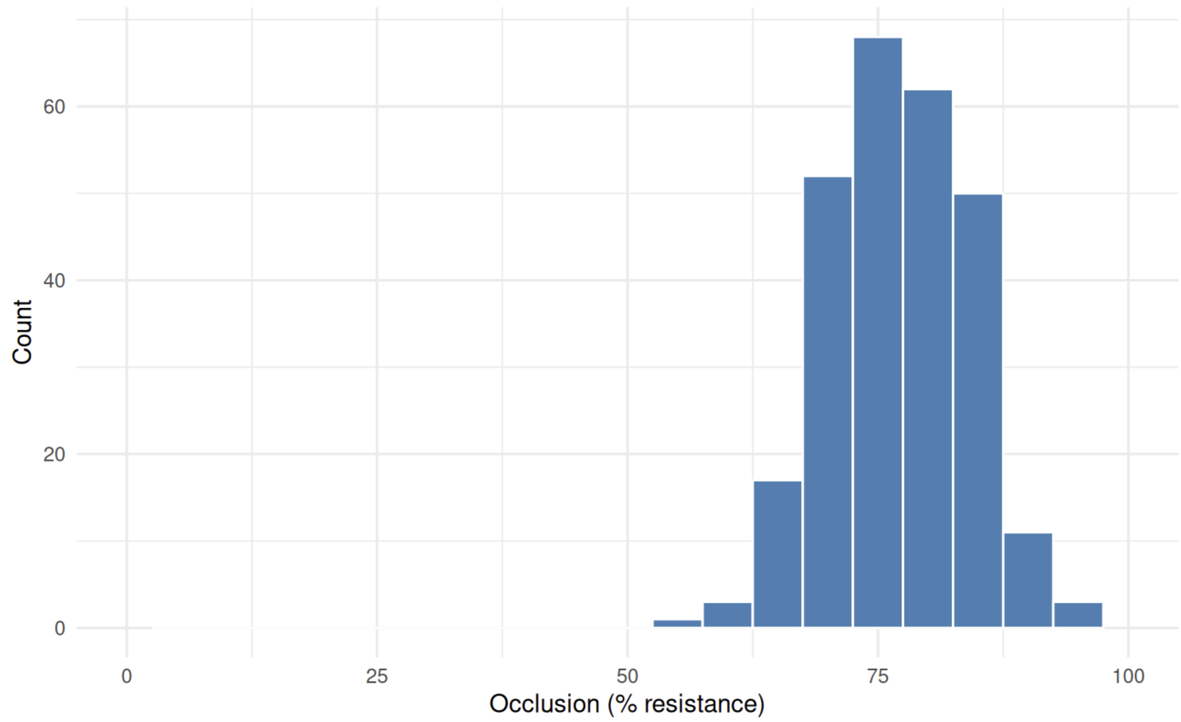


In general, this is why in HRDT studies where we directly manipulate heart rate (Tyrer et al., 2025) or when studying groups with clear baseline differences in heart rate (Jeganathan et al., 2024) we typically control for intra and/or interindividual differences in heart rate to be sure that any effects on threshold or slope are above and beyond these effects. To clarify this point here, we now add this partial correlation analysis controlling for heart rate to the Supplementary Material, which shows highly similar results (see Supplementary Figure 7):



Breathing Resistance. A related point concerns the breathing resistance task. From the reported results, it appears that participants approached the upper end of the resistance scale, with many settling around 90% resistance. I was curious whether 100% resistance in this context corresponds to full occlusion of the breathing tube. If so, that would suggest that participants were experiencing very strong resistance levels, and the data may be approaching a ceiling effect. Is it known that individuals are generally quite insensitive to large inspiratory resistances, or might this reflect a task-specific limitation?

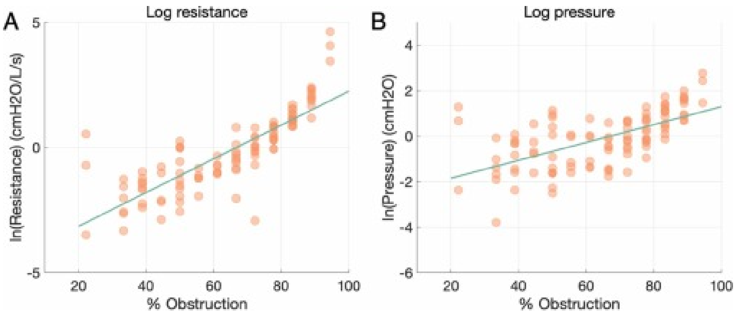
We appreciate this thoughtful question. When we plot a histogram of each individual's RRST threshold (in percentage occlusion units) we see that most individuals are around 75% occlusion, with only 3% of participants threshold settling above 90% occlusion (in Figure 1 is plotted one individual example). This mirrors group thresholds from recent studies (Stoffel et al., 2025; Tyrer et al., 2025).



This is also demonstrated in our recent preprint on the hierarchical modelling details (Courtin et al., 2025):

The figure shows the fitted psychometric functions (PF) for the RRST (N = 267). The solid line indicates the group-level posterior mean PF, while the dashed line represents the 95% credible interval. Shaded regions show the posterior predictive range within which a new participant's psychometric function is expected to fall, visualised across 60%, 80%, 90%, and 95% intervals.

Moreover, as you can see in the original psychophysiological validation paper (Nikolova, Harrison, et al., 2022), resistance and airway pressure increase log linearly with aperture closure, meaning that values above 90% are indeed exerting a highly noticeable effect on respiratory interoception:



Additionally, if a subset of participants converged at or near the maximum resistance level, this could potentially skew group-level estimates of both threshold and precision. For instance, it might compress variability across individuals or artificially inflate slope estimates for those whose responses saturated near ceiling. If such a pattern was observed, it would be helpful to know how it was addressed—e.g., whether any exclusions were applied or whether the modeling approach was adjusted to account for this.

We agree that ceiling effects can sometimes be a concern in psychophysical tasks. However, in our data this is unlikely, as only a small percentage of participants (3%) reached a threshold above 90% occlusion.

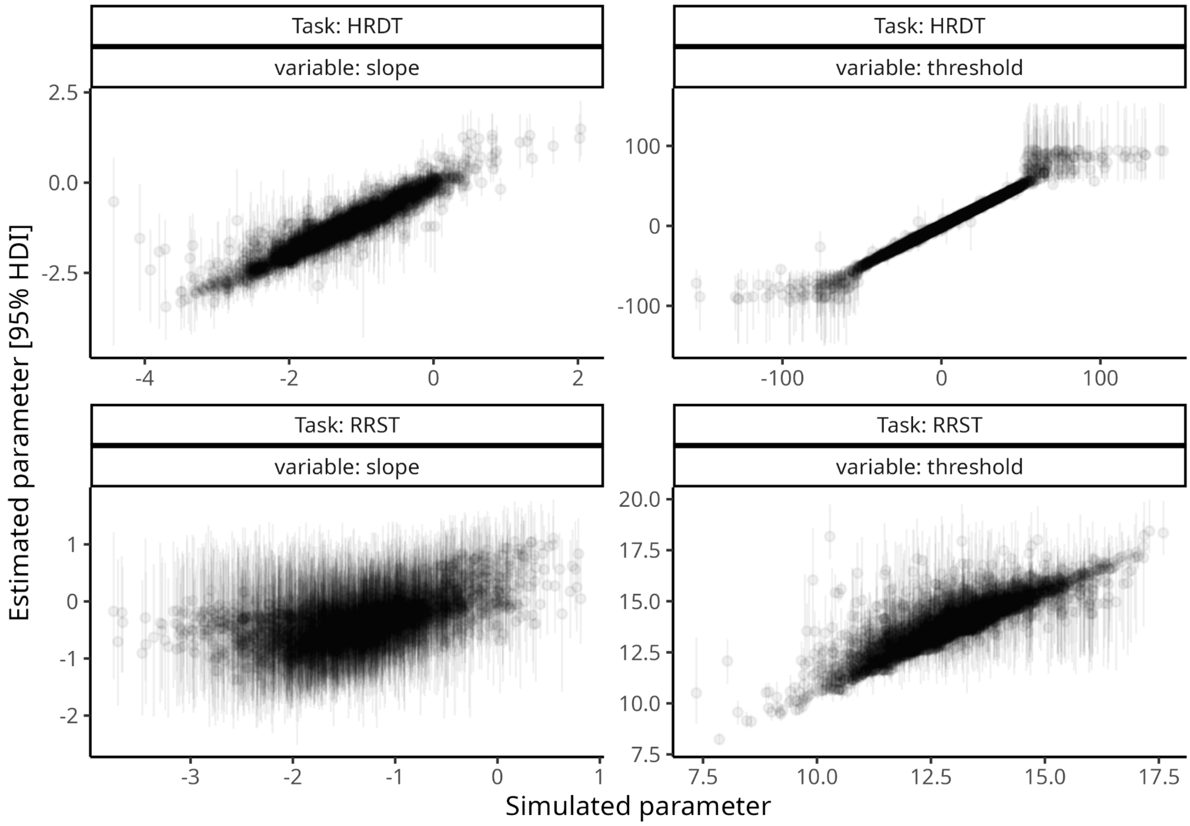
Modeling Considerations and Parameter Robustness
Given that the central findings of the study rely on psychometric parameters—particularly threshold and slope estimates derived from the Psi adaptive staircase method—it would be helpful to include additional validation analyses to confirm the robustness and recoverability of these estimates. While the Psi method is a powerful and efficient approach for estimating both the point of subjective equality (PSE) and slope, it is also known that slope estimates can be less stable than thresholds, especially when the number of trials is limited or when the stimulus range does not adequately sample the steep region of the psychometric function. Since these potential limitations apply to both tasks in the present study, it would be useful to reassure readers that the model parameters are reliable under the conditions used. I would recommend adding the following checks:

We agree with the Reviewer that the robustness of threshold and especially slope estimates can be a concern when estimated from limited trial numbers or noisy data. This is one of the reasons that led us to use estimates derived from a hierarchical model jointly estimating group- and individual-level parameters, rather than estimates extracted from individual PSI runs. Such an approach allows us to use PSI to optimize stimulus placement to the sensitivity of the participant during the experiment but without sacrificing statistical power and rigour. We have now made this clearer in the manuscript:

> *"Using this hierarchical model explicitly accounted for the nested structure of the data which allowed for principled pooling of information across participants, yielding more precise estimates with fewer trials (Courtin et al., 2025)."*

1. Parameter Recovery Analysis: One straightforward and informative approach is to conduct a parameter recovery analysis. This involve generating synthetic observers with known psychometric parameters (e.g., using a cumulative Gaussian with specified threshold and slope values) and simulating responses over the same number of trials as used in the actual experiment. Including realistic noise and lapse rates would increase the ecological validity of the simulation. The degree to which Psi recovers the original parameters—particularly slope—is then evaluated by plotting recovered versus true values for slope and threshold.

We recently preprinted a manuscript (Courtin et al., 2025: https://doi.org/10.1101/2025.08.27.672360) in which we explain our inference strategy, validate it and explore its practical implications. As part of this project, we ran the type of simulations the Reviewer describes to assess model and parameter recovery. As the current manuscript explores subject level correlations we have here extracted the simulated subject level parameters and the estimates of the hierarchical models for the HRDT (top) and RRST (bottom). Here we simulated 100-datasets for 50 subjects for 80 trials of the HRDT and 30 datasets for 200 subjects for 120 trials of the RRST: The convergence of the simulated hierarchical model resulted with 74 and 18 datasets for the two tasks respectively. Below is the parameter recovery plot showing the simulated subject level parameter value on the x-axis and the recovered estimate and their 95% credibility interval on the y-axis. The correlation between simulated and estimated parameters were 0.94 0.99 for the slope and threshold of the HRDT and 0.60 and 0.86 for the slope and threshold of the RRST.



The parameter recovery plot has been added to the Supplementary Materials and we have added a mention of the parameter recovery analysis in the manuscript:

> *"To ensure that our hierarchical models could satisfactorily recover these subject level parameters we show parameter recovery for the subject-level threshold and slope parameters for each model in the Supplementary Material (see Supplementary Figure 7). Further details of the hierarchical modelling and parameter recovery methods can be found in Courtin et al., 2025."*

2. Sensitivity to the prior: In addition, it would be valuable to assess the sensitivity of parameter estimates to the choice of prior. Since Psi uses a Bayesian updating framework, the width and shape of the prior distribution can meaningfully influence the resulting estimates. I would recommend repeating the recovery simulations with a range of priors—wide vs. narrow, centered vs. offset—to determine how robust the slope estimates are to potential prior mis-specification.

The choice of priors for the threshold and slope in the PSI-algorithm was uniform over a reference grid spanning -40.5 to 40.5 BPM for the threshold in increments of 1 and 0.1 to 20 for the slope in increments of 0.1 (following the paper introducing the task). This means that we chose the least informative prior making the trial-by-trial data determine the posterior belief over the parameters. The reviewer raises an interesting methodological point on whether stimulus intensities would have been more informative with a more realistic prior distribution for the parameters. Given that we only used the stimulus-response pairs and not the parameter estimates generated by the PSI method in the final analysis, we believe that such a prior-sensitivity analysis is beyond the scope of this paper.

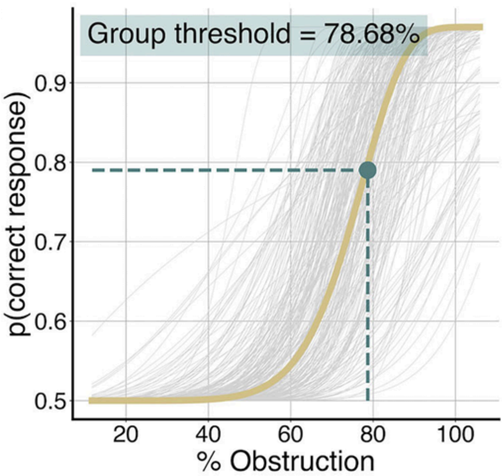
Additionally, with over 200 subjects, the prior distribution on the group level parameters in the hierarchical Bayesian model (which were only very weakly informative see preprint: Courtin et al., 2025) has little influence on the posterior distribution. As a confirmation of this we calculated the ratio between posterior and prior standard deviation of the group level means of the threshold and slope for both HRDT and RRST. As a rule of thumb, a prior is considered to be informative when this ratio is larger than or equal to 0.1 (https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations):

Task	Parameter	Posterior SD/Prior SD
HRDT (Intero)	Threshold	0.010
HRDT (Intero)	Slope	0.023
HRDT (Extero)	Threshold	0.030
HRDT (Extero)	Slope	0.015
RRST	Threshold	0.021
RRST	Slope	0.023

3. Stimulus Grid: A third important factor to consider is the initial stimulus grid used in the Psi algorithm. The range and placement of stimuli directly affect the algorithm's ability to sample the psychometric function effectively. It would therefore be helpful to simulate Psi runs using different grid configurations—for example, one that samples the entire function well, one that is too narrow to capture the tails, and one that is misaligned with the true threshold. This analysis could reveal how grid setup influences the reliability of both slope and threshold estimates. In the heartbeat task, in particular, it appears that some individuals may not have had dense sampling in the steep region of the psychometric function, which could compromise slope estimation for those participants. Similarly in the breathing task the grid based on possible breathing resistances could explain the correlation between slope and threshold. See comment below.

Thank you for providing the opportunity for us to address this point. For the cardiac task, PSI could select stimuli on a grid spanning -40.5 to 40.5 delta-BPM in 1 delta-BPM increments. Since the transition range of individual participant psychometric functions typically spans several dozen delta-BPM units, we are not concerned that PSI has access to insufficiently precise stimuli to effectively estimate the psychometric function slope.

The stimulus grid is much sparser for the respiratory task: 17 levels in total, going from no to near-complete compression of the breathing circuit. Still, our data indicate that the transition range of each participant's psychometric function spanned several stimulus intensities, suggesting that the granularity was sufficient for reliable estimation of the psychometric slope (see below figure from Nikolova et al., 2025):



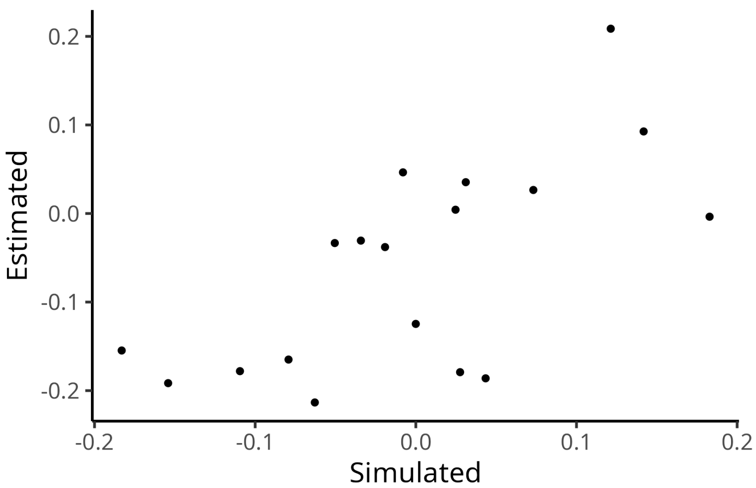
Additionally, parameter recovery is excellent for both tasks, when estimated using simulations (see plot above), confirming that the stimulus range the PSI method had access to was sufficiently granular to reliably estimate the thresholds and slopes. Whereas we agree with the Reviewer that further exploring the impact of the stimulus range on PSI performance might be interesting, we believe that this is beyond the scope of this paper.

Interindividual Differences and Model Transparency
Given the focus on individual differences in interoceptive ability, it would be extremely valuable to provide access to individual-level model fits—ideally through the GitHub repository mentioned in the manuscript. This would allow interested readers and researchers to assess model convergence and fit quality across participants and to verify that the reported group-level inferences are grounded in robust individual-level estimates. Even a representative subset of psychometric curves or posterior distributions would help illustrate the reliability of the modeling procedure across the sample.

We agree it's important for the individual-level model fits to be openly available. The STAN model files have been uploaded to OSF (the files were too large for GitHub): https://osf.io/tnxsh/overview
The link has been made available on the associated GitHub.

Correlation Between Threshold and Slope in the Breathing Task
One aspect that stood out to me was the relatively strong correlation between threshold and slope in the breathing task. Typically, one would not expect a substantial correlation between these two parameters, as they capture distinct psychophysical properties: threshold reflects bias, while slope reflects precision. While it is certainly possible that this relationship reflects a meaningful psychological trait—such as individuals with lower interoceptive sensitivity also demonstrating more precise discrimination—it may also be a byproduct of the modeling approach.
Specifically, the correlation could be an artifact of over-shrinkage in the hierarchical model, where the prior structure induces correlation in the group-level posteriors when individual-level data are noisy or sparse. One way to assess this possibility would be to simulate hierarchical datasets in which threshold and slope are uncorrelated at the individual level, then refit the same model and examine whether the recovered posteriors show spurious correlation. If a correlation emerges in the absence of one in the generative process, this would suggest the observed effect may be model-induced.

From our 18 simulated datasets of 200 subjects we investigated whether a correlation appeared between the threshold and the slope. Below we display the correlation coefficient from the simulated threshold and slope values versus the estimated correlation between the estimated threshold and slope values.



As can be seen the model does recover the simulated correlation coefficient and does not overestimate it. Making it highly unlikely that the correlation is due to an artifact of the hierarchical model fitting procedure, since the estimated and simulated parameter correlations converge suggesting the model is well specified.

Another reason could be the above mentioned ceiling effects close to a 100% resistance.

We note that only 3% of participants showed thresholds above 90%, so ceiling effects are unlikely to explain the observed results.

Discussion: The authors conclude that interoceptive sensitivity, precision, and metacognitive efficiency are largely domain-specific, challenging the notion of a general interoceptive dysfunction. While the data are compelling, one caveat worth mentioning is that the respiratory task specifically targets detection of an artificial inspiratory resistance, rather than more naturalistic dimensions of respiration such as breath frequency or tidal volume. This distinction is important, as it limits the generalizability of the findings to other aspects of respiratory interoception. For example, respiratory sinus arrhythmia (RSA) reflects the strong coupling between breathing and heart rate. Overall we know that there are significant changes in heart rate occurring across a single breathing cycle. Given this, it remains possible that naturalistic aspects of breath perception may show different interrelationships with cardiac interoception than those observed in this resistance detection task.

We appreciate the reviewer's insightful comment regarding the specificity of our respiratory task. We acknowledge that detecting inspiratory resistances captures only one aspect of respiratory interoception and future work should explore other respiratory features such as breath duration, frequency, or tidal volume. We have included this limitation in the Discussion and noted that future work should examine these dimensions. Notably, consistent with our findings, a respiratory task focusing on exhalation duration also showed no correlation with cardiac interoception (Bodart et al., 2024).

> *"Fourth, while our respiratory task assessed the ability to detect breathing resistances, future work could examine other features of respiration, such as breath duration, frequency, or volume. Notably, consistent with our findings, a respiratory task focusing on exhalation durations also showed no correlation with cardiac interoception (Bodart et al., 2024)."*

As a related side note—though not something that needs to be addressed within the scope of this paper—it would be interesting in future work to examine whether cardiac interoceptive accuracy varies across different phases of the respiratory cycle. If the authors collected respiratory data during the heartbeat task, it could open the door to exploratory analyses on this front, though I mention this more as a speculative thought than as a required addition.

We thank the reviewer for this thoughtful suggestion. We agree that examining whether cardiac interoceptive accuracy varies across different phases of the respiratory cycle would be an interesting direction for future research, though this is beyond the scope of the present study.

Frederike Petzschner

Reviewers Response

We sincerely thank the editor and reviewers for their thoughtful efforts and valuable feedback on our manuscript. The editor board members' comments are presented in black, our responses are in blue, and quoted changes to the manuscript are formatted in quoted blue italics.

Although the returning reviewer is satisfied with the revisions, editorially, we require that you clarify whether this analysis is part of several other analyses derived from the same dataset.

We clarify that the revision analyses use the same dataset as this manuscript. Courtin et al. (2025) is a methodological preprint describing the hierarchical modelling of these tasks, based on the same Visceral Mind Project data used here. This has been clarified in the main manuscript:

> *"Further details of the hierarchical modelling and parameter recovery methods of this dataset can be found in Courtin et al., 2025."*

All figures in this manuscript (including the supplementary parameter recovery figure) are original to this publication. The core correlation and Bayes factor analyses were performed exclusively for the present manuscript.

Please report thresholds rather than Bayes Factors in Figure 4.

Thank you for pointing our direction to missing this important detail. We have added clear Bayes Factor thresholds to the legend of Figure 4, and we edited the results text to be consistent with these thresholds.:

> *"BF_{01} values greater than three (blue colours) indicate moderate evidence in favor of the null hypothesis (no association), and values greater than one (lime green colour) show anecdotal evidence for the null, while values below one (yellow colours) suggest evidence against the null (Lee & Wagenmakers, 2014)."*

Last but not least, please nuance the claims about domain specificity and address the possibility that one measure could be more biased and less sensitive than the other.

We agree that this is an important point. We have nuanced our domain specificity claims in the manuscript. For example:

> *"our findings support that interoceptive sensitivity, precision, and metacognitive efficiency may operate in a domain-specific manner across cardiac and respiratory modalities."*

Also, the Discussion limitations section now features descriptions of task differences and suggestions for how to overcome this in future research. For example:

> *"While our hierarchical modelling approach yields robust psychometric estimates, differences in stimulus granularity and multisensory input across tasks may nonetheless contribute to asymmetries in sensitivity or criterion. Future research should minimise design asymmetries between the cardiac and respiratory tasks, for example by pairing the HRDT with a tactile–auditory exteroceptive control or by adding an exteroceptive reference to the RRST. This would render both tasks comparably multisensory and provide a bodily–exteroceptive control that accounts for multimodal auditory influences."*

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table". Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

Formatting and additional requests from "Editorial Requests Table" completed and attached.