Submitted to: Neuroscience and Biobehavioral Reviews
Word count: 996 words
Submission date: 19th December 2023; Revision date: 15/01/2024

Auditing measures of interoceptive accuracy: important clarifications

Jennifer Murphy[1,*] & Geoffrey Bird[2,3]

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

[2]Department of Experimental Psychology, University of Oxford

[3]School of Psychology, University of Birmingham

*Corresponding author

Jennifer Murphy

Department of Psychology

Royal Holloway, University of London

Egham, Surrey

TW20 0EX

Jennifer.murphy@rhul.ac.uk

Desmedt, Luminet, Walentynowicz, & Corneille, (2023) audit measures of interoceptive accuracy, the perception of internal bodily signals. Whilst a useful overview of tasks and limitations, as well as practical challenges, the authors 1) do not differentiate tasks of interoceptive propensity and accuracy, and 2) present critiques that we believe overlook the nuance of promising tasks.

Interoceptive accuracy is defined as accuracy of the perception of internal bodily states. It is assessed by comparing participant reports to an objective measure (Murphy, 2023). Whilst individuals can differ with respect to interoception in many ways (Murphy, 2023), one neglected area is propensity to use internal signals (Murphy, 2022).

As outlined by Murphy (2022), individuals differ in the cues they use to gauge internal states (e.g., hunger); some rely more on internal information (e.g., internal feelings of hunger), others on external information (e.g., time of day). Evidence that propensity is distinct from accuracy comes from examining gender (Murphy, 2022); although males typically outperform females on lab-based interoceptive accuracy tasks, women have a greater propensity to use external cues, resulting in comparable performance in real-life.

Whilst few tasks assess propensity to use internal signals, one was developed by Murphy, Catmur, & Bird, (2018). In this task, participants perform an exhalation into a device that measures respiratory output. Participants then perform a second exhalation aiming for a specific target percentage of the first. After performing this exhalation, they are asked to estimate how close their second exhalation was to the target percentage. The difference between their actual second exhalation and their estimate is taken as a perceptual measure. Importantly, this task was developed to be administered under two conditions – where participants do and do not have access to external auditory cues caused by their exhalations. Performance differences across conditions are taken as a measure of interoceptive propensity;

if performance does not change when exteroceptive cues are removed, then participants rely solely on interoceptive cues even when both internal and external cues are available to gauge their respiratory output (under certain assumptions about relative performance using each cue).

Although Desmedt et al., (2023) classify this task as a measure of interoceptive accuracy (acknowledging this as debatable), it was not designed to assess accuracy. Indeed, neither the internal condition alone is a measure of interoceptive accuracy – scores would be influenced by a multitude of factors (e.g., self-esteem, size of first exhalation etc.) – nor the difference score between internal and external conditions, as the subjective estimate is not compared with an objective measure. Given the multitude of signals which may be used in respiratory tasks (skin/muscle sensations as the chest expands and contracts, temperature/pressure changes in the mouth/throat etc.), we agree with the authors that perhaps *no* respiratory task is a pure measure of interoceptive accuracy. This is one reason why a more practical definition of interoceptive accuracy might be whether internal signals can be perceived unaided, rather than perceived via pathways over which there is debate as to which qualify as 'truly' interoceptive.

The Phase Adjustment Task (PAT) is also discussed as a promising new measure (Plans et al., 2020). In the PAT, participants rotate a dial to change the phase relationship between their heartbeats and a tone until they believe they are synchronous. As the starting phase of tones are random across trials, accuracy is inferred from the consistency of participants' selected delays. Whilst the authors outline constructive next steps for examining psychometric properties (e.g., internal consistency), several criticisms of the PAT are made. First, the authors speculate that integrating both internal and external cues may prevent detection of heartbeats. Whilst possible, to our knowledge there is currently no empirical evidence for this, and in the absence of a valid purely perceptual task no way to address this

question. While the proportion of participants judged as interoceptive on the PAT matches previous estimates, again, in the absence of purely perceptual tests it is impossible to test whether the PAT is difficult enough to underestimate the proportion of interoceptive individuals. As noted in Plans et al., (2021), rather than relying on single assessments, it is perhaps more useful to consider accuracy across multiple occasions, situations (at rest, perturbed) and domains (cardiac, gastric etc.). A third criticism relates to the low-to-moderate test-retest reliability. However, given state effects on interoception (Murphy, 2023), it is debatable whether we should expect higher reliability; without further research, we cannot determine whether lower test-retest reliability reflects task unreliability or expected state effects. Finally, the authors point to the ability of the PAT to be administered remotely (though it can also be administered in the lab) as a cause of the PAT's 'low discrimination ability', classifying participants as interoceptive, unclassified and non-interoceptive. Importantly, the authors miss that this was a deliberate choice (continuous consistency scores are available). It was argued that classification provides a more interpretable metric than a continuous score; if two individuals score below chance, it is not informative to interpret differences as meaningful as they likely reflect noise and not differences in interoceptive accuracy. Whilst differences between individuals scoring above chance may be informative (potentially reflecting perceptual differences), without an adequately matched control task it is unclear whether differences are perceptual or due to another factor (attention/motivation). Indeed, one strength is quantification of the degree of uncertainty with the use of varying thresholds (Bayes Factors of 3-30) rather than cut-off scores, given that any threshold is arbitrary. Crucially, this approach was developed as an improvement on standard approaches and not driven by the remote testing strategy – when a test measures ability to perceive heartbeats at rest, on a single occasion, over a short period of time, it is unclear how much scores can differ from either 'able to perceive heartbeats' or not.

Whilst we agree work is required to improve measurement of interoceptive accuracy, we argue for the importance of considering *what* is assessed by measures (separating accuracy from propensity), the influence of state effects, and a need for multiple assessments under varying conditions using multiple (valid) tests.

Bibliography

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

Murphy, J. (2022). Propensity to use interoceptive signals: An important individual difference. *Biological Psychology*, *171*, 108326. https://doi.org/10.1016/j.biopsycho.2022.108326

Murphy, J. (2023). Interoception: Where do we go from here? *Quarterly Journal of Experimental Psychology*, 17470218231172725. https://doi.org/10.1177/17470218231172725

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

Plans, D., Ponzo, S., Morelli, D., Cairo, M., Ring, C., Keating, C. T., ... & Bird, G. (2021). Measuring interoception: The phase adjustment task. *Biological psychology*, *165*, 108171.