Biological Psychology 184 (2023) 108693

Contents lists available at ScienceDirect

Biological Psychology

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

The validity issues of the heartbeat counting task are not ruled out by Schulz et al. (2021): A commentary

Olivier Desmedt [a,b,c,d,*], Olivier Luminet [c,d], Pierre Maurage [c,d], Olivier Corneille [c]

[a] Institute of Psychology, University of Lausanne, Lausanne, Switzerland
[b] The Swiss National Science Foundation (SNSF), Switzerland
[c] Psychological Science Research Institute, UCLouvain, Louvain-la-Neuve, Belgium
[d] Fund for Scientific Research – Belgium (FRS-FNRS), Belgium

Interoception is the processing of internal bodily states by the nervous system. One of its most studied dimensions is interoceptive accuracy (IAcc), the objective capacity to detect internal bodily signals. Although the Heartbeat Counting Task (HBCT; Schandry, 1981) is the most frequently used measure of this construct, the interoception community has increasingly acknowledged its lack of construct validity to measure IAcc, following a large body of evidence (e.g., Brener & Ring, 2016; Desmedt, Luminet, & Corneille, 2018). However, Schulz et al. (2021) recently made a completely opposite conclusion in a paper entitled "On the construct validity of interoceptive accuracy based on heartbeat counting: Cardiovascular determinants of absolute and tilt-induced change scores". They indeed concluded that their "findings support the convergent and discriminant validity" of the HBCT.

In this comment, we argue that we disagree with their conclusion based on three arguments: (1) their definition of IAcc makes the HBCT a valid measure of this construct because it accepts any contributor of the correspondence between actual and reported heartbeats; this, however, does not correspond to the commonly used definition (Garfinkel et al., 2015); (2) contrary to their claim, we believe the correlation size (i.e., $r = .42$) between $IAcc_{HBCT}$ scores and Heartbeat Discrimination Task (HBDT; Whitehead et al., 1977) scores is likely over-estimated and does not reach the minimum threshold required to support convergent validity; (3) their statistical analyses that test the contribution of time estimation in HBCT performance are not valid, which does not allow to conclude that the task discriminant validity is supported.

Conceptual issue: Measuring "interoceptive accuracy" in light of theoretical views

Schulz et al. (2021) define interoceptive accuracy as "the correspondence between actual and perceived bodily signals". We see value in this definition as it allows us to match construct and measurement. Nevertheless, researchers should be aware that the said correspondence (i.e., task performance) may be driven by a host of mechanisms, some of which are irrelevant to "the process of accurately detecting and tracking internal bodily sensations" (i.e., the commonly used definition of interoceptive accuracy; Garfinkel et al., 2015, p. 66). In other words, the definition proposed by Schulz et al. is at risk of conflating mechanisms that are relevant but also irrelevant to interoceptive accuracy as currently understood by most researchers.

The distinction between relevant and irrelevant contributors is sometimes easy. For instance, there is a consensus that if a participant achieves good performance by monitoring a smartwatch or by taking their pulse, their performance does not indicate high interoceptive abilities. Researchers are therefore controlling these potential contributors by asking participants not to take their pulse or use a smartwatch. Perhaps less clear, however, is the theoretical status of estimation processes (such as relying on cardiac knowledge) as contributors to task performance. If we assume that these estimation processes are undesirable (i.e., do not index how participants *detect* their bodily signals), then new instructions can be designed to reduce their influence (see Desmedt et al., 2018). In their paper, Schulz et al. specify that bodily signals should be *perceived* and thus state that the absence of correlation between $IAcc_{HBCT}$ scores[1] and time estimation (i.e., a guessing strategy) would support the HBCT discriminant validity.

Finally, the influence of cardiac signal properties on performance is even more ambiguous. This is in contrast to exteroception research, where signal properties are standardized across participants to avoid invalid conclusions about individual perceptual abilities (Corneille et al., 2020). For example, in a visual chart test, the size of the letters and

* Corresponding author at: Institute of Psychology, University of Lausanne, Géopolis, CH-1015 Lausanne, Switzerland.
E-mail address: olivier.desmedt@unil.ch (O. Desmedt).

[1] We do not subscribe to the use of the terms "$IAcc_{HBCT}$" or "$IAcc_{HBDT}$" scores, as it suggests that these scores validly index IAcc. We, however, use these terms in the present comment to be consistent with the original paper.

https://doi.org/10.1016/j.biopsycho.2023.108693
Received 8 August 2023; Received in revised form 21 September 2023; Accepted 24 September 2023
Available online 28 September 2023
0301-0511/

the distance between them and the participant are similar from person to person. In interoception research, signal properties are often not standardized across participants. In heartbeat detection tasks, a stronger heartbeat signal - due to stable individual differences, but also due to context-dependent states - may lead to better performance because this signal would be easier to detect. Thus, task scores may capture not only detection ability but also a physical condition (e.g., body fat, which may interfere with the heart signal) or a physical state (e.g., physiological activation, which is associated with stronger heart contractions). Again, it is important to determine whether individual differences in signal properties can contribute to task performance and, if not, to find ways to limit their influence. The theoretical status of these properties remains ambiguous in Schulz et al. (2021). On the one hand, the authors state in the introduction that "…the contribution of […] cardiac signal properties support the convergent validity of the HBCT…" (p.2). On the other hand, in their conclusion, they seem to suggest that the lack of correlation observed between cardiac signal properties and $IAcc_{HBCT}$ scores supports the discriminant validity of the task (p.8). We call for clarification of whether or not interoceptive accuracy can be confounded with cardiac signal properties. If these properties are irrelevant to the construct, experimental (e.g., by manipulating physiological states) or statistical controls should be implemented to standardize them across participants.

More generally, we believe that it is important that interoception researchers state and control for whichever contributors they see as undesirable to task performance, *in light of* the theoretical views they endorse. We also encourage them to rely on different terminologies depending on their theoretical standpoint. For instance, researchers interested in the role of prior semantic knowledge or expectations, especially those working under a predictive coding framework, may want to refer to "interoceptive beliefs" rather than "interoceptive accuracy" (Legrand et al., 2022).

Coherence issue: Convergence between tasks (or lack thereof)

Another major issue in Schulz et al. (2021) is the interpretation of the association (i.e., $r = 0.42$) between $IAcc_{HBDT}$ and $IAcc_{HBCT}$ scores as an indication of the convergent validity of these tasks. In our view, 18 % of shared variance between two tasks meant to measure the exact same construct indicates low convergence. Even more worrisome, we believe the association reported by the authors is largely overestimated. A recent meta-analysis (Hickman et al., 2020) found a weak association between $IAcc_{HBDT}$ and $IAcc_{HBCT}$ scores ($r = 0.21$, IC 95 % [0.13; 0.29]). Only 2 out of the 22 included studies found an effect size equal or superior to the one found by Schulz et al. (2021). One could argue that this difference in effect size is explained by procedural differences in HBCT administration (e.g., instructions, time intervals, heart rate measurement devices). Nevertheless, when performing a meta-analysis on eight effect sizes extracted from previous studies of Schulz et al., we also found a weak association ($r = 0.24$, IC 95 % [0.13; 0.35]; Forkmann et al., 2016; Michal et al., 2014; Schulz et al., 2013, 2020, 2021; Wittkamp et al., 2018). The high effect size found by Schulz et al. (2021) could thus represent an overestimation of the true effect size partly due to a limited sample size ($N = 49$) giving rise to unstable estimates. Indeed, in personality (and social) psychology, correlations usually stabilize with sample sizes of $N > 200$ (Schönbrodt & Perugini, 2013). If we rely on Hickman et al.'s (2020) meta-analysis or on the meta-analysis of Schulz et al.' studies, we can conclude that only 4–6 % of shared variance exist between $IAcc_{HBCT}$ and $IAcc_{HBDT}$ scores, which suggests that these tasks do not measure the same construct.

Moreover, a carry-over effect from the HBDT to the HBCT, due to a fixed-order design, may also have contributed to inflating the reported association between $IAcc_{HBDT}$ and $IAcc_{HBCT}$ scores. During the HBDT, participants listen to their heart rate. This feedback might increase their knowledge about their heart rate and subsequently increase their likelihood to correctly guess their heart rate when performing the subsequent HBCT. This could explain the high average HBCT performance in their sample ($M = 0.77$) despite the use of stricter instructions (i.e., asking participants to only report the felt heartbeats). Indeed, studies using modified instructions generally report lower average performance ($M < 0.50$; Desmedt et al., 2018; Ehlers et al., 1995; Ferentzi et al., 2021; Van Den Houte et al., 2021). However, again, one could argue that differences observed in mean $IAcc_{HBCT}$ scores are due to differences in the HBCT procedure or in sample characteristics. Finally, the HBDT with two intervals has important limitations and more optimal versions have been proposed (e.g., the method of constant stimuli with 6 intervals; Brener & Ring, 2016). This reduced validity further compromises the interpretation of the association observed between $IAcc_{HBCT}$ and $IAcc_{HBDT}$ scores.

In summary, we propose that the association between the two most frequently used measures of "interoception accuracy" is, at best, weak.

Biases issue: The contribution of estimation processes

Participants may achieve good performance on the HBCT by counting seconds, without relying on felt heartbeats, which is problematic for a task meant to measure the capacity to detect cardiac signals.[2] It is therefore important to control for the influence of time estimation. To do so, it is common practice to test the association between $IAcc_{HBCT}$ scores and time estimation accuracy (i.e., the absolute proportional difference between reported and actual seconds; Dunn et al., 2010), as it was done by Schulz et al. (2021). In their study, the authors did not find a significant association between these variables ($\beta = .19$, $t = 1.35$, $p = .184$). Therefore, they concluded that the discriminant validity of the task is supported. However, this raises several comments.

First, the authors used the modified HBCT instructions (vs. the original ones), asking participants to avoid using guessing strategies. We have previously shown that modified instructions greatly reduce the influence of guessing strategies on $IAcc_{HBCT}$ scores (Desmedt et al., 2020). Given that most HBCT studies did not use modified instructions, the issue might apply to those. Second, as explained by Desmedt et al. (2020), the simple correlation between the number of counted heartbeats per minute in the HBCT and the number of counted seconds per minute in the time estimation task is a more valid test of the contribution of time estimation strategies in the HBCT. We, therefore, encourage more compelling tests in future studies. Third, the sample size was very limited in Schulz et al. (2021), which weakens the strength of the demonstration. Finally, estimation processes are not limited to time estimation. Participants could indeed be able to estimate their heart rate because they (1) have already heard or seen their or others' heart rate (e.g., by putting one's ear on another's chest or by watching a heart rate monitor) and (2) have felt their heartbeats in the past (e.g., while doing a physical activity) or even during a part of the task. Therefore, the validity of a task cannot be ascertained by testing one source of biasing estimation only. For this reason, previous studies have considered other potentially biasing variables, and centrally the semantic knowledge about (one's) heart rate (Murphy et al., 2018), which has not been done by Schulz et al. (2021).

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors did not use generative AI technologies for the preparation of this work.

[2] In particular, participants could (1) simply count seconds, (2) count the seconds by adapting the pace based on their knowledge about their heart rate, and (3) count the seconds and subsequently estimate the number of heartbeats through mental arithmetic.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Pierre Maurage (Research Associate) and Olivier Luminet (Research Director) are funded by the Fund for Scientific Research – Belgium (FRSFNRS). Olivier Desmedt was funded by the Fund for Scientific Research – Belgium (FRS-FNRS) as a PhD student and is now funded by the Swiss National Science Foundation as a Post-doc researcher.

References

Brener, J., & Ring, C. (2016). Towards a psychophysics of interoceptive processes: The measurement of heartbeat detection. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 371*(1708), Article 20160015. https://doi.org/10.1098/rstb.2016.0015

Corneille, O., Desmedt, O., Zamariola, G., Luminet, O., & Maurage, P. (2020). A heartfelt response to Zimprich et al. (2020), and Ainley et al. (2020)'s commentaries: Acknowledging issues with the HCT would benefit interoception research. *Biological Psychology, 152*, Article 107869. https://doi.org/10.1016/j.biopsycho.2020.107869

Desmedt, O., Corneille, O., Luminet, O., Murphy, J., Bird, G., & Maurage, P. (2020). Contribution of time estimation and knowledge to heartbeat counting task performance under original and adapted instructions. *Biological Psychology, 154*, Article 107904. https://doi.org/10.1016/j.biopsycho.2020.107904

Desmedt, O., Luminet, O., & Corneille, O. (2018). The heartbeat counting task largely involves non-interoceptive processes: Evidence from both the original and an adapted counting task. *Biological Psychology, 138*, 185–188. https://doi.org/10.1016/j.biopsycho.2018.09.004

Dunn, B. D., Stefanovitch, I., Evans, D., Oliver, C., Hawkins, A., & Dalgleish, T. (2010). Can you feel the beat? Interoceptive awareness is an interactive function of anxiety- and depression-specific symptom dimensions. *Behaviour Research and Therapy, 48*(11), 1133–1138. https://doi.org/10.1016/j.brat.2010.07.006

Ehlers, A., Breuer, P., Dohn, D., & Fiegenbaum, W. (1995). Heartbeat perception and panic disorder: Possible explanations for discrepant findings. *Behaviour Research and Therapy, 33*(1), 69–76.

Ferentzi, E., Vig, L., Lindkjølen, M. J., Lien, M. E., & Köteles, F. (2021). Mental heartbeat tracking and rating of emotional pictures are not related. *Psychological Research*, 1–8. https://doi.org/10.1007/s00426-021-01593-4

Forkmann, T., Scherer, A., Meessen, J., Michal, M., Schächinger, H., Vögele, C., & Schulz, A. (2016). Making sense of what you sense: Disentangling interoceptive awareness, sensibility and accuracy. *International Journal of Psychophysiology, 109*, 71–80. https://doi.org/10.1016/j.ijpsycho.2016.09.019

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

Hickman, L., Seyedsalehi, A., Cook, J. L., Bird, G., & Murphy, J. (2020). The relationship between heartbeat counting and heartbeat discrimination: A meta-analysis. *Biological Psychology, 156*, Article 107949. https://doi.org/10.1016/j.biopsycho.2020.107949

Legrand, N., Nikolova, N., Correa, C., Brændholt, M., Stuckert, A., Kildahl, N., … Allen, M. (2022). The heart rate discrimination task: a psychophysical method to estimate the accuracy and precision of interoceptive beliefs. *Biological Psychology, 168*, 108239.

Michal, M., Reuchlein, B., Adler, J., Reiner, I., Beutel, M. E., Vögele, C., Schächinger, H., & Schulz, A. (2014). Striking discrepancy of anomalous body experiences with normal interoceptive accuracy in depersonalization-derealization disorder. *PLOS ONE, 9*(2), Article e89823. https://doi.org/10.1371/journal.pone.0089823

Murphy, J., Brewer, R., Hobson, H., Catmur, C., & Bird, G. (2018). Is alexithymia characterised by impaired interoception ? Further evidence, the importance of control variables, and the problems with the Heartbeat Counting Task. *Biological Psychology, 136*, 189–197. https://doi.org/10.1016/j.biopsycho.2018.05.010

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology, 18*(4), 483–488. https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schulz, A., Back, S. N., Schaan, V. K., Bertsch, K., & Vögele, C. (2021). On the construct validity of interoceptive accuracy based on heartbeat counting: Cardiovascular determinants of absolute and tilt-induced change scores (Scopus) *Biological Psychology, 164*. https://doi.org/10.1016/j.biopsycho.2021.108168.

Schulz, A., Lass-Hennemann, J., Sütterlin, S., Schächinger, H., & Vögele, C. (2013). Cold pressor stress induces opposite effects on cardioceptive accuracy dependent on assessment paradigm. *Biological Psychology, 93*(1), 167–174.

Schulz, A., Rost, S., Flasinski, T., Dierolf, A. M., Lutz, A. P. C., Münch, E. E., Mertens, V.-C., Witthöft, M., & Vögele, C. (2020). Distinctive body perception mechanisms in high versus low symptom reporters: A neurophysiological model for medically-unexplained symptoms. *Journal of Psychosomatic Research, 137*, Article 110223. https://doi.org/10.1016/j.jpsychores.2020.110223

Van Den Houte, M., Vlemincx, E., Franssen, M., Van Diest, I., Van Oudenhove, L., & Luminet, O. (2021). The respiratory occlusion discrimination task: A new paradigm to measure respiratory interoceptive accuracy. *Psychophysiology, 58*(4), Article e13760. https://doi.org/10.1111/psyp.13760

Whitehead, W. E., Drescher, V. M., Heiman, P., & Blackwell, B. (1977). Relation of heart rate control to heartbeat perception. *Biofeedback and Self-Regulation, 2*(4), 371–392. https://doi.org/10.1007/BF00998623

Wittkamp, M. F., Bertsch, K., Vögele, C., & Schulz, A. (2018). A latent state-trait analysis of interoceptive accuracy. *Psychophysiology, 55*(6), Article e13055. https://doi.org/10.1111/psyp.13055