Biological Psychology 184 (2023) 108694

Contents lists available at ScienceDirect

Biological Psychology

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

Do Schulz et al.'s (2021) findings support the validity of the heartbeat counting task? Joint conclusion to commentaries

Olivier Desmedt [a,b,c,d,*], Olivier Corneille [c], Olivier Luminet [c,d], Pierre Maurage [c,d], Claus Vögele [e], André Schulz [e,f]

[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
[e] Clinical Psychophysiology Laboratory, Department of Behavioural and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[f] Institute for Cognitive and Affective Neuroscience, Trier University, Trier, Germany

The current synthesis aims to address issues of potentially contrasting theoretical considerations, findings, and interpretations presented in Schulz et al. (2021) original article, Desmedt et al.'s (Desmedt, Luminet, Maurage, & Corneille, 2023a) commentary, and Schulz et al. (2023) reply. The goal is also to guide future research on how to proceed with the HBCT and the development of novel methods to assess cardiac interoception.

1. Conceptual issue: measuring "interoceptive accuracy" in light of theoretical views

Interoceptive accuracy (IAcc) is one of the most central dimensions of conscious interoception and is assessed with performance-based tasks. While we first thought that our teams (Desmedt et al. and Schulz et al.) disagreed on the definition of this construct, the Commentary and the associated Response indicate that our views converge in defining IAcc as the objectifiable capacity to detect internal bodily signals. Thus, we concur that measures of IAcc should not be (substantially) influenced by guessing strategies and signal properties (e.g., frequency, intensity, duration). Nevertheless, measures of cardiac IAcc might be affected by signal properties and, in particular, signal intensity. Indeed, the stronger the heart contractions, the easier it is for participants to detect their heartbeats, independently of their cardiac IAcc (Schandry et al., 1993). Future studies should, therefore, find ways to address this issue (see Desmedt et al., 2023b for a discussion).

In the context of heartbeat detection/perception tasks, it is unreasonable to attribute variance in performance exclusively to either cardiac signal intensity or to guessing strategies. These contributors may both operate; yet, they are mutually exclusive in their implications. An association with cardiac signal intensity suggests participants' reliance on perceived heartbeats, while an association with guessing strategies indicates reliance on beliefs about heart rate. Hence, an association between task performance and signal intensity would represent evidence that participants at least partly relied on felt heartbeats to perform the task. Cardiac signal intensity and task performance may be inherently bound. However, this would provide an advantage to participants with stronger signal intensity, which may compromise task validity. Future studies should, therefore, control for this confound.

As it seems that cardiac contractility and indicators of peripheral sympathetic activation are amongst the most important correlates of cardiac IAcc (Eichler & Katkin, 1994; Schandry et al., 1993), the Heather Index (HI), pre-ejection period (PEP), or T-wave amplitude (TWA) might be promising candidates for indicators of cardiac signal intensity. If the impedance cardiography measures (HI and PEP) are unavailable, TWA (based on ECG) can be an alternative. Future research should clarify if cardiac IAcc scores increase in validity if they are inter or intra-individually standardized or residualized by HI, PEP, or TWA measures.

2. Coherence issue: convergence between tasks (or lack thereof)

As no consensus could be reached between Desmedt and colleagues, and Schulz and colleagues, on the question of whether the reported correlation between $IAcc_{HBCT}$ and $IAcc_{HBDT}$ scores of $r = .42$ should be interpreted as evidence for or against their validity, we attempted to exchange important arguments to develop recommendations for future research on cardiac interoception.

Regarding the potential inflation of the association between $IAcc_{HBCT}$

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

https://doi.org/10.1016/j.biopsycho.2023.108694
Received 8 August 2023; Received in revised form 22 September 2023; Accepted 24 September 2023
Available online 28 September 2023

and $IAcc_{HBDT}$ scores, two arguments were presented. First, the two teams agreed on the fact that this association ($r = .42$) was unusual, as demonstrated by the small association ($r = .21$) found by Hickman et al. (2020) in their meta-analysis. This can either be explained by the small sample size or the specific conditions of this study (e.g., strict instructions). Among these conditions, Desmedt and colleagues noted that the fixed order design (the HBCT was always preceded by the HBDT) could have inflated the association between the two tasks. As noted by Schulz and colleagues, however, this does not qualify the convergent validity of the task. Instead, one could argue that administering the HBDT before the HBCT increased the validity of the latter. This is because, in the HBDT, participants can hear their heart rate, which could have helped them to feel their heartbeats (via biofeedback) and subsequently increased their ability to detect and count their heartbeats in the HBCT. In other words, the HBDT may have increased participants' ability to detect their heartbeats, and this may have facilitated heartbeat perception during the HBCT (vs. the use of guessing strategies). This, however, would suggest that the HBCT may be too demanding in usual testing conditions. That is, when its completion is not preceded by the completion of the HBDT. Validity issues of the HBCT may thus be partly explained by the too-high difficulty of the task. This means that future research should develop IAcc tasks that can be reasonably achieved by most participants (i.e., whose difficulty is compatible with participants' abilities). Moreover, it is necessary to have tasks with different levels of difficulty so that we can better discriminate between participants.

The disagreement on the minimum threshold for convergent validity is easily explained by the lack of consensual guidelines in the literature regarding this question. Schulz and colleagues consider that $r = .42$ supports the convergent validity of the two tasks based on Cohen's conventions following which this correlation is a medium to large effect size (Cohen, 1992). Desmedt and colleagues consider that this correlation does not reach the minimum threshold for convergent validity, as guideline papers suggest, notably based on statistical simulations, that an $r \geq .50$ would be the minimum effect size to support convergent validity and even more ($r \geq .70$) when two measures are meant to assess the exact same construct (e.g., Carlson & Herdman, 2012; Chmielewski et al., 2016). One should acknowledge that it may be more difficult to reach high convergence between performance tasks than between self-reports, as the former show relatively larger structural variations. However, it would still be problematic if a too-low convergence is reached (see also below).

Furthermore, Schulz and colleagues argue that conceptual, methodological, and empirical papers also suggest that the thresholds for validity measures to be interpreted as the same or a different construct are domain-specific and should be based on empirical findings in the respective area (Biesanz & West, 2004; Campbell & Fiske, 1959; Cicchetti, 1994; Cronbach & Meehl, 1955; DeVellis, 2016; Gliner et al., 2017). On the one hand, some previous studies showing significant correlations between indicators of different interoceptive tasks (within or across organ domains), which was typically interpreted as supporting their validity, reported (after recoding[1]) coefficients of $.30 \leq r \leq .59$ (e.g., Herbert et al., 2012; Knoll & Hodapp, 1992; van Dyck et al., 2016; Whitehead & Drescher, 1980), and, therefore, in a comparable range as the previously reported $r = .42$ (Schulz et al., 2021) – but never in a range of $r \geq .70$. On the other hand, these correlations can neither ensure that the respective tasks are valid nor can they determine a definite validity threshold. Schulz and colleagues thus recommend systematic multi-trait (e.g., different organ domains or different interoceptive facets) and multi-method (e.g., different tasks or observational level: behavior vs. self-report) studies (Campbell & Fiske, 1959) for the future to develop interoception-specific validity thresholds for interoceptive indicators. Desmedt and colleagues, nevertheless, invite caution while endorsing this domain-specific approach because, no matter the domain, suboptimal convergence ($< r = .50$–$.70$) between measures dramatically increases the risk of replication failure (as demonstrated by Carlson & Herdman, 2012).

Finally, even if the two teams agreed on the fact that the convergent validity of the HBCT is not optimal, Schulz and colleagues argue that no behavioral task is perfect (i.e., they are all more or less affected by other factors) and that we should, therefore, not have such expectations for the HBCT. Desmedt and colleagues agree that no behavioral measure is process-pure, but argue that the HBCT is far from reaching the minimal validity criteria. This being said, they acknowledge that this is a pervasive problem in psychology.

3. Biases issue: the contribution of estimation processes

Regarding the involvement of time estimation strategies while performing the HBCT, two key messages should be remembered. First, consistent with Desmedt et al. (2020), the correlation between average counted heartbeats in the HBCT and average counted seconds in the time estimation task seems to be a more valid test of the contribution of time estimation in the task than the correlation between $IAcc_{HBCT}$ and TEAcc scores. While the correlation between $IAcc_{HBCT}$ and TEAcc scores was not significant, the correlation between counted heartbeats and seconds was. This indicates that, in this study, despite the use of modified instructions, the involvement of time estimation in HBCT performance cannot be excluded – although this correlation is lower than what is generally observed with original instructions. Nevertheless, following Schulz and colleagues, time estimation might play a less prominent role than the shared variance of the HBCT and the HBDT.

Second, and related to the above issue, caution is needed when interpreting the correlation between $IAcc_{HBCT}$ and TEAcc scores, and also between counted heartbeats and seconds. While these correlations could indicate the use of estimation strategies in the performance of the HBCT, it could also be explained by the contribution of interoception in time estimation or by the fact that both abilities (i.e., counting heartbeats and counting seconds) share a common contributor (e.g., the capacity to track the rhythm of a stimulus). For this reason, this correlation should not be interpreted in isolation as an univocal indicator of task validity. Moreover, other estimation strategies (e.g., knowledge about heart rate) should be considered (Murphy et al., 2018).

Funding

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.

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.

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.

[1] Interoceptive sensitivity as estimated by the amount of ingested water (Herbert et al., 2012; van Dyck et al., 2016) is inversely coded (i.e., less water ingested = higher sensitivity) and has to be recoded to ensure that correlation coefficients are comparable across studies.

References

Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the big five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality, 72*(4), 845–876. https://doi.org/10.1111/j.0022-3506.2004.00282.x

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105.

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

Chmielewski, M., Sala, M., Tang, R., & Baldwin, A. (2016). Examining the construct validity of affective judgments of physical activity measures. *Psychological Assessment, 28*(9), 1128–1141. https://doi.org/10.1037/pas0000322

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302. https://doi.org/10.1037/h0040957

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., Maurage, P., & Corneille, O. (2023a). The validity issues of the heartbeat counting task are not ruled out by Schulz et al. (2021): A commentary. *Biological Psychology*. https://doi.org/10.1016/j.biopsycho.2023.108693

Desmedt, O., Luminet, O., Maurage, P., & Corneille, O. (2023b). Discrepancies in the Definition and Measurement of Human Interoception: A Comprehensive Discussion and Suggested Ways Forward. *Perspectives on Psychological Science*, 17456916231191537.

DeVellis, R. F. (2016). *Scale development: Theory and applications*. SAGE Publications.

Eichler, S., & Katkin, E. S. (1994). The relationship between cardiovascular reactivity and heartbeat detection. *Psychophysiology, 31*(3), 229–234. https://doi.org/10.1111/j.1469-8986.1994.tb02211.x

Gliner, J. A., Morgan, G. A., & Harmon, R. J. (2017). Measurement validity and reliability. *Research methods in applied settings: An integrated approach to design and analysis* (3e éd..,, pp. 105–146). Routledge.

Herbert, B. M., Muth, E. R., Pollatos, O., & Herbert, C. (2012). Interoception across modalities: On the relationship between cardiac awareness and the sensitivity for gastric functions. *PloS One, 7*(5), Article e36646. https://doi.org/10.1371/journal.pone.0036646

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

Knoll, J. F., & Hodapp, V. (1992). A comparison between two methods for assessing heartbeat perception. *Psychophysiology, 29*(2), 218–222.

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. (Scopus) *Biological Psychology, 136*, 189–197. https://doi.org/10.1016/j.biopsycho.2018.05.010.

Schandry, R., Bestler, M., & Montoya, P. (1993). On the relation between cardiodynamics and heartbeat perception. *Psychophysiology, 30*(5), 467–474. https://doi.org/10.1111/j.1469-8986.1993.tb02070.x

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. *Biological Psychology, 164*. https://doi.org/10.1016/j.biopsycho.2021.108168

Schulz, A., & Vögele, C. (2023). Why Desmedt et al.'s commentary does not apply to the findings of Schulz et al. (2021) concerning the validity of the heartbeat counting task. *Biological Psychology*. https://doi.org/10.1016/j.biopsycho.2023.108689

van Dyck, Z., Vögele, C., Blechert, J., Lutz, A. P. C., Schulz, A., & Herbert, B. M. (2016). The water load test as a measure of gastric interoception: Development of a two-stage protocol and application to a healthy female population. *PloS One, 11*(9), Article e0163574. https://doi.org/10.1371/journal.pone.0163574

Whitehead, W. E., & Drescher, V. M. (1980). Perception of gastric contractions and self-control of gastric motility. *Psychophysiology, 17*(6), 552–558. https://doi.org/10.1111/j.1469-8986.1980.tb02296.x