Biological Psychology 137 (2018) 12–17

Contents lists available at ScienceDirect

Biological Psychology

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

Interoceptive accuracy scores from the heartbeat counting task are problematic: Evidence from simple bivariate correlations

Giorgia Zamariola[a,*], Pierre Maurage[b,c], Olivier Luminet[a,c], Olivier Corneille[a]

[a] Psychological Science Research Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[b] Laboratory for Experimental Psychopathology, Psychological Science Research Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[c] Belgian Fund for Scientific Research (FNRS), Belgium

ARTICLE INFO

Keywords:
Heartbeat counting task
Interoceptive accuracy

ABSTRACT

Interoception, the capacity to perceive internal bodily states, is thought to influence cognitive, affective and interpersonal functioning. It is frequently assessed using the heartbeat counting task, introduced recently in interoceptive research. In this task participants are requested to count their heartbeats without relying on external cues. Interoceptive Accuracy (i.e., IAcc) scores are then computed based on absolute comparisons between actual and reported heartbeats. In a large sample of participants ($N = 572$), we observed that, whereas IAcc scores are meant to be theoretically agnostic to error type (i.e., over- or underestimation of heartbeats), these scores massively (i.e., > 95%) reflect under-reports. Of concern too, the correlation between actual and reported heartbeats is low overall ($r = .16$), varies non-linearly across IAcc score quantiles, and suggests undistinguishable interoceptive capacities within the top 60% IAcc scorers. We also found that IAcc scores, which are conceptually independent from actual heart rates, are structurally bound to them. Finally, we show that IAcc scores vary across the time intervals used in the task. We encourage researchers using this score for studying cognitive and emotional processes to reconsider its meaning.

1. Introduction

Interoception, defined as the ability to correctly perceive bodily sensations (Garfinkel, Seth, Barrett, Suzuki, & Critchley, 2015), is thought to regulate a wide range of psychosocial phenomena (Craig, 2002). Interoceptive impairments have been related to psychopathological conditions (Barrett & Simmons, 2015), underlining the theoretical and clinical importance of this process for psychological science. The key role played by interoception in emotional and cognitive processes was already stated in early theoretical accounts of emotion-cognition interactions. For instance, James (1884) suggested that a stimulus is, first, perceptually and cognitively processed in the brain. This initial step then gives place to the behavior and consequently to bodily sensations that are fed back to the brain, producing the feeling part of the emotion. A more recent illustration is offered by the 'somatic marker hypothesis' (Damasio, 1994) stating that feelings require the presence of bodily-anchored physiological activations in order to be recognized and processed. The role of interoception is also evident in embodied cognition theories (e.g., Barsalou, 2008; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005), which posit that cognition is built into the bodily structure of the organism. Additionally, it is integral to decision making theories interested in the role of affects in decisions, such as in research on intuitive judgments and the affect heuristic (e.g., Slovic, Peters, Finucane, & MacGregor, 2005). A central assumption in all these models is that the ability to perceive one's bodily sensations is of utmost importance for efficient psychological functioning. This underlines the need for researchers and clinical practitioners to dispose of reliable conceptualizations and measures of interoceptive abilities.

Although research on interoception has been accumulating for a long time, a systematic model of this construct appeared only recently. In an insightful conceptual effort, Garfinkel et al. (2015) divided interoception in three dimensions. First, *interoceptive accuracy* is the capacity to perceive internal signals coming from the body; it is evaluated using objective tests such as mental tracking (e.g., Dale & Anderson, 1978; Schandry, 1981) or heartbeat detection tasks (e.g., Katkin, Reed, & Deroo, 1983). Second, *interoceptive sensibility* is the subjective belief in one's ability to recognize bodily sensations; it is assessed via self-report using questionnaires, e.g., Porges Body Perception Questionnaire (Porges, 1993) or average confidence in the given response. Third, *interoceptive awareness* is the difference between how good an individual thinks he/she is (subjectively) and how good he/she actually is

* Corresponding author.
E-mail address: giorgia.zamariola@uclouvain.be (G. Zamariola).

https://doi.org/10.1016/j.biopsycho.2018.06.006
Received 7 May 2018; Received in revised form 13 June 2018; Accepted 18 June 2018
Available online 23 June 2018
0301-0511/

(objectively); it can be measured using the area under ROC curves which relates confidence to accuracy (Garfinkel & Critchley et al., 2013).

The present research is interested in interoceptive accuracy and questions the validity of its most common index: IAcc scores derived from Schandry's heartbeat counting task. In this task, participants have to silently count their heartbeats without relying on external cues. Comparisons between actual and reported heartbeats provide a measure of interoceptive accuracy (IAcc). The heartbeat counting task has been extensively used in the last decades, and its related IAcc score still represents a core interoceptive measure in recent high-level publications (e.g., Shah, Hall, Catmur, & Bird, 2016; Yao et al., 2018).

A growing number of articles, however, have cast doubt on the reliability of this measure. The involvement of beliefs about heart rate in performance of the heartbeat counting task has been suggested repeatedly over the past 30 years. According to Brener and Ring (2016), the task does not allow disentangling real ability in detecting heartbeats from answering according to previous knowledge about the number of heartbeats occurring in a given time interval. Since the original task instruction requires participants to estimate the number of heartbeats, individuals may provide an answer based on beliefs without actually attempting to perceive their heartbeats. Consistent with this view, one study (Windmann, Schonecke, Fröhlig, & Maldener, 1999) showed that changing the heart rate in patients with cardiac pacemaker, setting them to low (50 beats per minute, bpm), medium (75 bpm), or high (110 bpm) heart rate, did not influence their reported number of heartbeats. This suggests that these patients performed the task by relying on previous knowledge instead of perception of their bodily states.

1.1. The present study

Consistent with concerns recently raised about the validity of the heartbeat counting task, we report data questioning the construct validity of the IAcc scores. Accuracy corresponds to the extent to which a performance is error-free. The following formula is generally used to quantify IAcc on data collected for three time durations: $1/3\ \Sigma\ (1-(|\text{actual heartbeats} - \text{reported heartbeats}|)/\text{actual heartbeats})$[1], with higher scores indexing higher interoceptive accuracy. We argue that these scores lack validity and explain why based on four main arguments.

First, several sources of errors can conceptually contribute to imperfect IAcc scores: normally distributed random errors (e.g., in reporting/recording heartbeats), under-detection (i.e., not reporting actual heartbeats) and over-detection (i.e., reporting inexistent heartbeats). The formula is designed such that under- and over-detection are conceptualized as equivalent. The latter choice is theoretically questionable. This is because abilities involved in not missing true heartbeats may differ from abilities involved in not over-interpreting heartbeats-unrelated signals. Importantly too, assuming this choice is theoretically pertinent, it would be questioned by evidence showing that IAcc scores largely depend on one error type only. We report here that this is the case and discuss why we think this is problematic.

Second, another critical issue concerns whether IAcc scores validly distinguish between respondents. If IAcc scores reflect people's ability to accurately perceive their inner states, a correlation between actual and reported heartbeats should be observed, and this correlation should linearly increase with higher IAcc scores (i.e., better IAcc scorers should better map actual and reported heartbeats). This is not the case. The correlation between actual and reported heartbeats is low overall. And, even more critically, it is higher at average than higher IAcc levels.

Third, a valid measure of interoception accuracy should not be structurally tied to heart condition. This is because heart condition (i.e. actual heartbeats) is not inherent to the definition of the interoceptive accuracy construct. In other words, it is essential for construct validity that people's accuracy at perceiving their inner life is not *structurally* bound to their cardiac condition. Instead, we find here that IAcc scores (which are mostly indexing under-reporting) are mechanically increased at slower heart rates.

Fourth, a wide range of time intervals have been previously used in heartbeat counting tasks (from 25 to 103 s, Murphy, Brewer, Hobson, Catmur, & Bird, 2018), but these time intervals are systematically merged to compute IAcc scores according to the above-mentioned formula, hampering to consider performance variation across different testing intervals. We find here that various time intervals (i.e. 25 s, 35 s, 45 s) lead to significantly different IAcc scores, due to reduced reported heartbeats per minute for longer time intervals.

2. Method

2.1. Participants

Five hundred seventy-two healthy students (386 females, $M_{\text{age}} = 22.24$, $SD = 4.33$) tested individually were recruited using a Facebook page dedicated to paid studies or using advertisements at the Faculty. The studies received the approval from the Ethical Committee of the research Institute. Data from four experiments were combined for the present analyses.[2]

2.2. Material

2.2.1. Heartbeat counting task

Participants' heart rate was assessed using the Polar Watch RS800CX heart monitor (which derives heart rate from the placement of electrodes on wrists). Polar products have been used in previous studies, showing excellent validity and reliability in measuring heart rate and R-R interval data (e.g., Kingsley, Lewis, & Marson, 2005; Nunan et al., 2008; Quintana, Heathers, & Kemp, 2012; Weippert et al., 2010). Following the well-validated Mental Tracking Method by Schandry (1981), data were recorded during three randomly presented time intervals (25 s, 35 s, 45 s), each separated by a pause of 20 s. The software Polar ProTrainer5 was used to extract the actual number of heartbeats. One acoustic start cue was presented at the beginning of each time interval and another acoustic stop cue indicated the end of the interval. Throughout the experiment, participants were instructed to silently count (not estimate or guess) their heartbeats. At the end of each time interval, they were asked to verbally report how many heartbeats they counted. No feedback on the length of the counting phases or on performance was given. In order to quantify IAcc, the classic IAcc formula reported above was used.

3. Results

3.1. Data exclusion

1.38% of the data ($N = 8$) were removed of the analyses due to heartbeat recording issues.[3]

[1] Another formula has been proposed by Garfinkel et al. (2015), namely: $1-(|\text{actual heartbeats} - \text{reported heartbeats}|)/((\text{actual heartbeats} + \text{reported heartbeats})/2)$. It faces the same problems, as the correlation between IAcc scores obtained through the classical and new formula is $r = .985$ in our sample.

[2] Three of these studies were published (Zamariola, Vlemincx, Corneille, & Luminet, 2018a) that measured IAcc in relation to other factors (i.e., mainly alexithymia). A fourth study also served another research purpose (Zamariola, Luminet, Mierop, & Corneille, submitted).

[3] Data are available at the following address: https://doi.org/10.5281/zenodo.1288323.

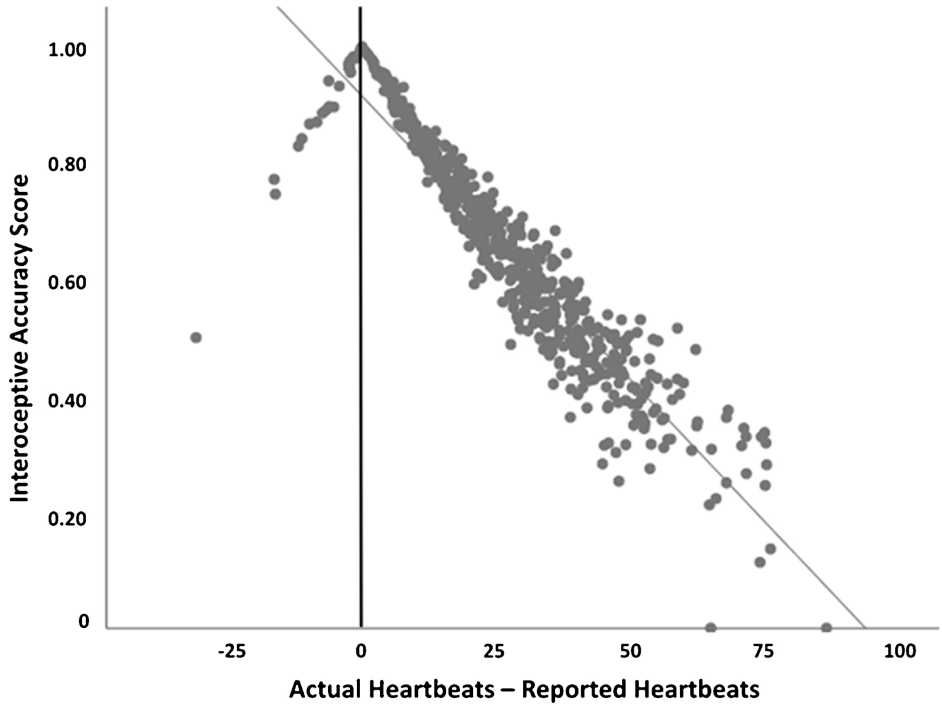

Fig. 1. Correlation between Interoceptive Accuracy scores and Actual/Reported Heartbeats differences.

3.2. Error types underlying IAcc scores

IAcc scores are mainly driven by differences in under-reporting heartbeats (r = -0.93, $p < .001$; Fig. 1). Less than 5% of the participants showed an overestimation (N = 25), 88% of which belong to the top 20% IAcc scorers. Hence, IAcc scores essentially inform us of how (un)willing participants are to report they perceived a heartbeat.

3.3. Differences between IAcc scorers

A weak positive correlation ($r = .16, p < .001$) was found between actual and reported heartbeats overall, suggesting a small overlap (i.e., 2.56% shared variance) between what people report experiencing and what is actually happening inside their body. When quintiles were created based on IAcc scores (N = 114–115 per quintile), correlations between actual and reported heartbeats varied non-linearly between lowest (r = .50), low (r = .67), average (r = .95), high (r = .83) and highest (r = .88) IAcc scorers (Table 1 and Fig. 2). Correlation coefficients were compared across groups using Fisher's r to z transformation. All subgroup correlations significantly differ ($z > 1.94$, $p < .03$), except for the high-to-highest comparison (z = 1.40, p = .08). This suggests that IAcc scores fail to validly differentiate individuals in their ability to accurately perceive their inner states within the top 60% IAcc scorers.

Table 1
Correlation coefficients between Actual and Reported heartbeats and corresponding slopes/intercepts of the regression for the whole sample and for each of the five IAcc score percentile groups.

	Whole Sample	Lowest Scorers	Low Scorers	Average Scorers	High Scorers	Highest Scorers
R	0.160	0.503	0.668	0.948	0.827	0.876
R^2	0.025	0.253	0.446	0.898	0.684	0.768
Slope	0.138	0.746	0.997	1.388	0.971	0.823
Intercept	70.72	60.11	37.67	8.54	19.66	17.02

3.4. Relationship between IAcc score and heart rate

IAcc scores were negatively correlated with heart rate (r = -0.36, $p < .001$; Fig. 3).

3.5. Variation of IAcc score across time intervals

As reported in Table 2, IAcc scores were significantly lower for 45 s interval than for 25 s [$t(571) = 6.61$, $p < .001$] and 35s [$t(571) = 7.68$, $p < .001$] intervals, which did not significantly differ from each other [$t(571) = 0.71$, $p = .48$]. This result may be explained based on lower reported bpm for the 45 s interval than for 25 s [$t(571) = 10.38$, $p < .001$] and 35s [$t(571) = 12.21$, $p < .001$] intervals, the 35s interval also being associated with higher reported bpm than the 25s one [$t(571) = 2.34$, $p = .02$]. In contrast, Fisher's r to z transformation showed that intervals did not significantly differ regarding correlations between actual and reported heartbeats (25s–35s: $z = 0.05$, $p = .96$; 25s–45s: $z = 0.92$, $p = .36$; 35s–45s: $z = 0.97$, $p = .33$).

4. Discussion

Interoception is a theoretically and practically critical construct for understanding people's psychological functioning in both healthy and clinical populations. It is largely assumed to be involved in the way people apprehend their world and make decisions about it. Interoceptive accuracy is arguably the most important dimension of interoception, not only because it tells us about people's ability to access their inner life, but also because we may know about the third dimension of interoception (i.e., interoceptive awareness) only by knowing about the first. In the absence of valid measure of interoceptive accuracy, one is left with only the subjective dimension of interoception (i.e., interoceptive sensibility), whose measurement is likely to suffer from the usual problems associated with self-reports (e.g., self-esteem, social desirability, accessibility effects).

The most frequently used measure of interoceptive accuracy is the heartbeat counting task. This task is very convenient because it requires no costly, complex, or ethically problematic assessment procedure. This

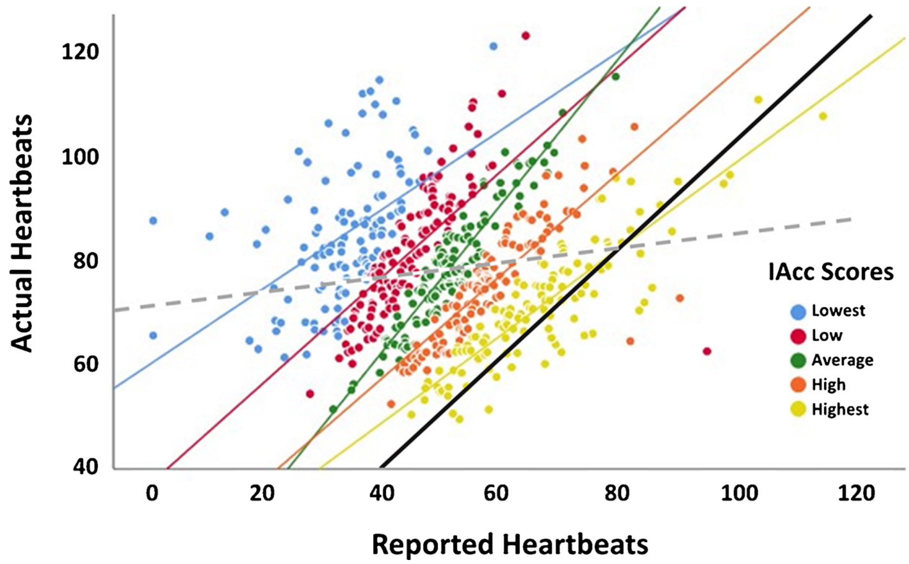

Fig. 2. Correlation between Actual and Reported heartbeats, with regression lines across all groups (grey dashed line) and within five IAcc score percentile groups (colored lines). The black line indicates perfect performance (i.e., perfect correspondence between Actual and Reported heartbeats).

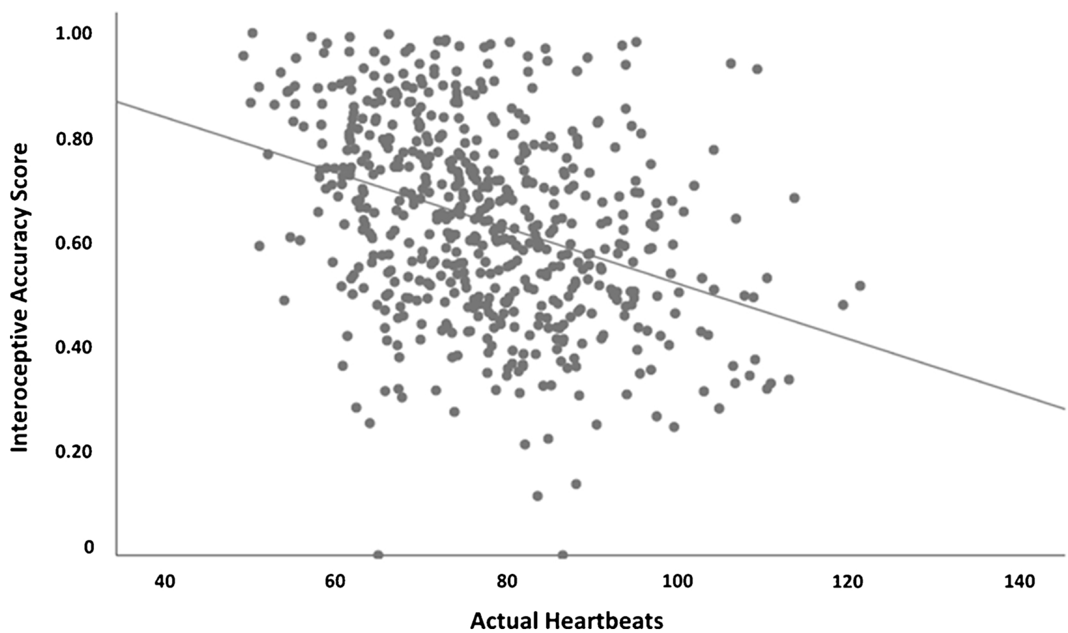

Fig. 3. Correlation between IAcc scores and participants' Actual heartbeats.

Table 2
Actual heartbeats, Reported heartbeats, IAcc score and Actual-Reported correlation for each testing time interval (25 s, 35 s, 45 s).

	25 s interval	35 s interval	45 s interval
Actual Heartbeats[a]	77.13	78.87	76.55
Reported Heartbeats[a]	50.49	51.50	45.96
IAcc score	0.646	0.650	0.605
Actual-Reported Correlation	0.158	0.161	0.105

[a] in Beats Per Minute.

convenience, along with the capacity of this task to tap into a precisely measurable physiological activity, has contributed to its success. This task also has face validity: most researchers would readily admit that heartbeats and the ability to perceive them tell us something about people's inner life and their capacity to access it. Extending the recent concerns raised about the validity of the task, we found that IAcc scores derived from this task suffer from four major issues:

(1) Whereas IAcc scores are meant to reflect people's level of interoceptive accuracy, irrespective of error type, these scores essentially index under-reporting, found in 95% of the participants. Under-reporting heartbeats may be due to a weaker sensitivity in detecting heartbeat signals, which is entirely relevant to interoceptive abilities. However, it may also be due to a large number of accuracy-irrelevant factors, such as decision threshold, beliefs about heart rate at rest, or time-perception based strategies. That *over*-reporting was massively represented amongst the best IAcc quintiles supports the first possibility (without excluding the others). The best IAcc scorers may simply be more willing to report perceiving a heartbeat. As a result, they would miss less of them but at the cost of over-reporting them. Clearly, relying on the latter strategy would be hardly informative of interoceptive accuracy. Even more important, and as just pointed out, it may also very well be the case that a host of perception-unrelated process are

driving reports of heartbeat perception in this task (for a recent discussion, see Murphy, Brewer et al., 2018). For instance, participants who report heartbeats by relying on personal beliefs about their heart rate at rest would naturally receive lower IAcc scores when their beliefs are such that they underestimate their personal heart rate.

If interoceptive accuracy is meant to reflect accuracy in detecting interoceptive signals, then what essentially is a measure of "misses" (in the present sample, IAcc scores were correlated by 93% with the magnitude of under-reporting) cannot reliably capture this ability. Instead, a signal detection paradigm is needed where decision threshold (i.e., the willingness to report perceiving a signal) and sensitivity (i.e., the ability to tease apart the presence and the absence of a signal) are independently estimated. This probably requires having experimental control over the presence or absence of the interoceptive signal. In the absence of such paradigm, the scores are hardly interpretable as they blend decision threshold and sensitivity.

(2) The correlation between actual and reported heartbeats is low overall. These two factors were poorly related (i.e., only 2.56% of shared variance) at the aggregate level. This suggests that this task does not clearly capture the ability to correctly perceive internal bodily signals. Would participants centrally base, as assumed, their estimation on actual heartbeats perception, a higher correlation should be found between this factor and its estimation. Admittedly, the correlation between actual and reported heartbeats may be modest overall, but higher as one goes from low and moderate to high IAcc scorers. This was not the case. To the contrary, the correlation between objective and reported heartbeats was higher for average than for higher and highest IAcc scorers.

Differences in correlation magnitude at the aggregate and sub-population levels are reminiscent of the Simpson's paradox wherein correlations may actually reverse at these levels. This type of pattern should be paid more attention to in psychological research (e.g., Kievit, Frankenhuis, Waldorp, & Borsboom, 2013). That the correlation between actual and reported heartbeats was lower at low than average/high levels of IAcc may be indicative of some validity in this task. However, the pattern found here still suggests that IAcc scores fail to validly discriminate among the top 60% respondents in their ability to map subjective and actual heartbeats. An important implication of this is that IAcc scores, if used for indexing interoceptive accuracy, should not be treated as a linear predictor in statistical models. At best, two broad groups of performers should be identified: low and high. A dichotomous treatment of this interoceptive measure would likely be associated with relatively low statistical power. Finally, one may also wonder why correlations were higher for average than for high and highest IAcc scorers. We can only speculate on this, but we surmise this may have to do with the presence of a mix of under- and over-reporting in the higher IAcc regions: while average scorers are all characterized by under-estimation, the merging of under- and over-estimators in high/higher scorers might mechanically reduce correlation rates.

Would correlations between actual and reported heartbeats then represent a better measure of interoceptive accuracy? In order to answer that question, one first needs to consider whether such measure would be used for making comparisons between groups (e.g., clinical versus non-clinical samples) or as an individual difference measure. If used for examining group differences, this correlation may be easily computed and reliably so given the potentially high number of data points involved. Whether it would have high (predictive, convergent, divergent) validity, however, it is an empirical question we cannot answer based on the data collected here. Turning to a correlation index as an individual measure, it would require a lot of observations within each participant. If data are collected at rest, actual heart rate is unlikely to vary, resulting in weak correlations. Alternatively, several observations may be collected within each participant under conditions that are conducive to heterogeneous heart rates (e.g., at rest, after a weak, mild, strong, intense effort). This may be an interesting way to go. Again, however, the validity of such index should be tested.

Finally, it is important to note here that the correlation found between actual and reported heartbeats within all quintiles does not necessarily indicate that IAcc scores reflect interoceptive processes. As discussed above, such correlation would also be obtained if participants' belief about their heart rate at rest is relatively accurate and if participants report their heartbeats counting based on these beliefs.

(3) IAcc scores decrease with actual heart rate. This finding is structurally tied to the finding that the IAcc scores massively reflect under-reporting. Hence, the lower the heart rate, the smaller the difference between actual and reported heartbeat and the larger the IAcc score. This finding confirms previous research on the link between lower heart rate and better performance at the heartbeat counting task (Ring, Liu, & Brener, 1994). It is also consistent with studies showing how body posture affects performance on the heartbeat counting task. Specifically, Bestler, Weitkunat, Keller, and Schandry, (1988) and Ring, Brener, Knapp, and Mailloux, (2015) obtained higher scores in supine than standing position. This finding, first attributed to an increase in stroke volume, which makes easier to detect the cardiac stimuli, was later found to be due to a lowered heart rate in supine position. That heart rate is structurally bound to IAcc scores threatens the construct validity of these scores. This is because the rate at which a heart beats is not conceptually part of people's ability at perceiving heartbeat signals

(4) IAcc scores are reduced for longer time intervals: significantly reduced IAcc scores were found here for longer intervals (45 s) than for shorter ones (25 s, 35 s). As actual heart rate is quite constant across intervals, this variation is related to a reduction of reported bpm at the longest interval. An increase of under-estimation with longer time intervals is classically found in time-estimation research (e.g. Lejeune & Wearden, 2009). This suggests that heartbeats counting reports may be partly based on time estimation. This would be the case, for instance, if people come up with a heartbeats estimate that is based on the use of heart rate beliefs (as necessarily adjusted by time estimate). At the empirical level, this underestimation issue may be further aggravated for wider range of time intervals (up to 103 s in Murphy, Brewer et al., 2018). Upcoming work should thus report distinct IAcc scores for distinct time instead of aggregating them into a global IAcc score. As we now discuss, however, one would still have to identify which of these various time intervals provides the best assessment of interoceptive accuracy.

Collectively, the findings reported and discussed here cast serious doubts on the construct validity of IAcc scores. This adds to growing concerns about the validity of the heartbeat counting task itself. Although the heartbeat counting task is convenient and has face validity, keeping using it is likely to slow down research efforts, at the cost of theoretical advances and effective clinical support to populations at risk. We strongly encourage researchers who relied on IAcc scores to examine whether the problems found here in a very large sample also applies to their (usually much smaller) dataset. If so, instead of indexing participants' interoceptive abilities, the IAcc scores they computed may have essentially reflected under-reporting heartbeats, whatever set of processes contributed to the latter outcome, and in particular belief-based inferences, decision threshold, or strategies based on time-perception. Future research may address what specific process underlies IAcc performance on this task. Alternatively, it may be that the problems found here differently apply to less common variants of this task or when covariates of interest are included in statistical model (see Murphy, Brewer et al., 2018). As we discuss now, however, we would rather like to encourage researchers to invest their efforts in the development of new tasks.

In our view, new measurement tools are needed for reliably assessing this important construct; one that is critical to physical and psychological health. Several alternative measures have been recently proposed, focusing on the detection of other bodily signals (e.g., Murphy, Catmur, & Bird, 2018). As these tools still strongly rely on the computation of scores comparing actual and estimated physiological variables, however, they may face issues similar to those evidenced here in the heartbeat counting task. We thus strongly encourage future studies to conduct an in-depth theoretical, experimental and psychometric evaluation of alternative interoceptive accuracy measures before using them in healthy or clinical populations. As interoceptive accuracy relies on the ability to detect bodily-related signals, a promising perspective could be to replace the classical "actual versus reported" score computation by tasks capitalizing on signal detection theory. By teasing apart decision criteria and accuracy, these tasks would allow to better assessing and interpreting abilities in detecting experimentally induced physiological modifications, making it possible to overcome interpretational pitfalls inherent to the currently available measures. To our knowledge, the study that comes closest to this is that by Windmann et al. (1999), where heart rate was changed in patients with pacemakers, leaving heartbeats counting unchanged for different heart rates.

Acknowledgments

This research was supported by an Excellence research grant funded by the Walloon region (Belgium), (FOOD4GUT: Innovative nutrition research on obesity based on colic nutriments. Biological, behavioral and societal aspects, project # 1318148). We are grateful to Olivier Desmedt, Mandy Hütter, Olivier Klein, Adrien Mierop, Christoph Stahl and Ali Gholamrezaei for their input.

References

Brener, J., & Ring, C. (2016). Towards a psychophysics of interoceptive processes: The measurement of heartbeat detection. *Philosophical Transactions of the Royal Society, 371*(1708), 20160015.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16*, 419.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Bestler, M., Weitkunat, R., Keller, W., & Schandry, R. (1988). Heartbeat perception and body position. *Journal of Psychophysiology, 2*, 123.

Craig, A. D. (2002). How do you feel? Interoception: The sense of The physiological condition of The body. *Nature Reviews Neuroscience, 3*(8), 655.

Dale, A., & Anderson, D. (1978). Information variables in voluntary control and classical conditioning of heart rate: Field dependence and heart-rate perception. *Perceptual and Motor Skills, 47*, 79–85.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason and the human brain.* New York, USA: Grosset/Putnam.

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.

Garfinkel, S. N., & Critchley, H. D. (2013). Interoception, emotion and brain: New insights link internal physiology to social behaviour. commentary on: "Anterior insular cortex mediates bodily sensibility and social anxiety" by terasawa et al. (2012). *Social Cognitive and Affective Neuroscience, 8*(3), 231–234.

James, W. (1884). What is an emotion? *Mind, 9*(34), 188–205.

Katkin, E. S., Reed, S. D., & Deroo, C. (1983). A methodological analysis of 3 techniques for the assessment of individual-differences in heartbeat detection. *Psychophysiology, 20*(4) 452-452.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology, 4*, 513.

Kingsley, M., Lewis, M. J., & Marson, R. E. (2005). Comparison of polar 810s and an ambulatory ECG system for RR interval measurement during progressive exercise. *International Journal of Sports Medicine, 26*(1), 39–44.

Lejeune, H., & Wearden, J. H. (2009). Vierordt's the experimental study of the time sense (1868) and its legacy. *European Journal of Cognitive Psychology, 21*(6), 941–960.

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*. http://dx.doi.org/10.1016/j.biopsycho.2018.05.010.

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.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*(3), 184–211.

Nunan, D., Jakovljevic, D. G., Donovan, G., Hodges, L. D., Sandercock, G. R., & Brodie, D. A. (2008). Levels of agreement for RR intervals and short-term heart rate variability obtained from the polar S810 and an alternative system. *European Journal of Applied Physiology, 103*(5), 529–537.

Porges, S. (1993). *Body perception questionnaire: Laboratory of development assessment.* University of Maryland.

Quintana, D. S., Heathers, J. A., & Kemp, A. H. (2012). On the validity of using the polar RS800 heart rate monitor for heart rate variability research. *European Journal of Applied Physiology, 112*(12), 4179–4180.

Ring, C., Brener, J., Knapp, K., & Mailloux, J. (2015). Effects of heartbeat feedback on beliefs about heart rate and heartbeat counting: A cautionary tale about interoceptive awareness. *Biological Psychology, 104*, 193–198.

Ring, C., Liu, X., & Brener, J. (1994). Cardiac stimulus intensity and heartbeat detection: Effects of tilt induced changes in stroke volume. *Psychophysiology, 31*, 553–564.

Schandry, R. (1981). Heart beat perception and emotional experience. *Psychophysiology, 18*(4), 483–488.

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex, 81*, 215–220.

Slovic, P., Peters, E., Finucane, M. L., & MacGregor, D. G. (2005). Affect, risk, and decision making. *Health Psychology, 24*(4S), S35.

Weippert, M., Kumar, M., Kreuzfeld, S., Arndt, D., Rieger, A., & Stoll, R. (2010). Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. *European Journal of Applied Physiology, 109*(4), 779–786.

Windmann, S., Schonecke, O. W., Fröhlig, G., & Maldener, G. (1999). Dissociating beliefs about heart rates and actual heart rates in patients with cardiac pacemakers. *Psychophysiology, 36*(3), 339–342.

Yao, S., Becker, B., Zhao, W., Zhao, Z., Kou, J., Ma, X., et al. (2018). Oxytocin modulates attention switching between interoceptive signals and external social cues. *Neuropsychopharmacology, 43*(2), 294.

Zamariola, G., Vlemincx, E., Corneille, O., & Luminet, O. (2018a). Relationship between interoceptive accuracy, interoceptive sensibility, and alexithymia. *Personality and Individual Differences, 125*, 14–20.

Zamariola G., Luminet O., Mierop A., & Corneille O. Do interoceptive accuracy and sensibility moderate the affective consequences of negative experiences?. (Manuscript submitted for publication).