Estimating the stability of heartbeat counting in middle childhood: a twin study

Jennifer Murphy[1]*, Rosa Cheesman[1]* Alice M. Gregory[2], Jennifer Lau[3], Anke Ehlers[4], Caroline

Catmur[3], Geoffrey Bird[1,4] & Thalia C. Eley[1,a]

*equal contributions

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and

Neuroscience, King's College London

[2]Department of Psychology, Goldsmiths, University of London

[3]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's

College London

[4]Department of Experimental Psychology, University of Oxford

[a]Corresponding author: Thalia Eley

+44 (0)20 7848 0863

Thalia.eley@kcl.ac.uk

Social, Genetic and Developmental Psychiatry Centre

Institute of Psychiatry, Psychology and Neuroscience - PO80

De Crespigny Park,

Denmark Hill, London,

United Kingdom, SE5 8AF

Research Highlights

- We assess the etiology of stability in heartbeat counting performance in childhood

- Moderate stability of heartbeat counting is observed

- Non-shared environmental factors contribute most to stability

- Heartbeat counting is not associated with mental health or emotion recognition

Abstract

There is growing interest in interoception, the perception of the body's internal state, and its relevance for health across development. Most evidence linking interoception to health has used the heartbeat counting task. However, the temporal stability of the measure, particularly during childhood, and the etiological factors that underlie stability, remain largely unexamined. Using data from the ECHO twin sample we estimated the magnitude of genetic and environmental influences on the stability of heartbeat counting across two years (age 8-10), the longest time-frame examined. Heartbeat counting accuracy was modestly correlated across time, (r=.35), and accuracy improved with age. Non-shared environmental factors accounted for the most variance at both time points and were the main contributors to temporal stability of heartbeat counting. Future research should seek to identify these non-shared environmental factors and elucidate whether this relatively modest stability reflects variability of interoception across development or unreliability of the heartbeat counting task.

Keywords: Interoceptive accuracy; interoception; heartbeat counting; stability; heritability; Twins

Introduction

In recent years there has been a growing interest in the importance of interoception, the perception of the body's internal state (Craig, 2002; Khalsa et al., 2018), for health and aspects of higher order cognition (Khalsa & Lapidus, 2016; Khalsa et al., 2018). Indeed, atypical interoception (both unusually good or bad interoceptive ability) has been proposed to underlie a number of transdiagnostic and disorder-specific symptoms. For example, high anxiety has been linked to unusually high interoceptive ability, whereas depression is often associated with poor interoceptive ability (Murphy, Brewer, Catmur, & Bird, 2017; Khalsa & Lapidus, 2016; Khalsa et al., 2018). Within the typical population, poor interoceptive accuracy has also been linked to atypical cognition in domains as diverse as decision making (e.g., Dunn et al., 2010), theory of mind (e.g., Shah, Catmur, & Bird, 2017) and emotional processing (e.g., Terasawa, Moriguchi, Tochizawa, & Umeda, 2014).

Much of the research that has examined how individual differences in interoception are related to health (e.g., depression, anxiety or sleep problems) and aspects of higher order cognition (e.g., emotion recognition) has utilized the heartbeat counting task as a measure of interoception (Dale & Anderson, 1978; Schandry, 1981). In this task, participants are asked to count their heartbeat over a series of intervals whilst their objective heartbeat is recorded. The participant's count is then compared to the objective measure to determine its accuracy. Despite widespread use of this task for quantifying interoception, in recent years there has been increasing focus on the *validity* of heartbeat counting as a measure of interoception. Indeed, questions have been raised as to the validity of the task as a measure of interoceptive accuracy given evidence that individual differences in physiology, heart rate knowledge, differences in

task administration, and non-interoceptive factors may contribute towards task performance (e.g., Desmedt, Luminet, & Corneille, 2018; Khalsa, Rudrauf, Sandesara, Olshansky, & Tranel, 2009; Murphy, Brewer, Hobson, Catmur, & Bird, 2018; Ring, Brener, Knapp, & Mailloux, 2015; Zamariola, Maurage, Luminet, & Corneille, 2018).

Despite this research focus on the *validity* of the heartbeat counting task as a measure of interoception, surprisingly few studies have examined the *stability* of heartbeat counting across time. Whilst the short-term stability (e.g., test-re-test reliability) of heartbeat counting is likely a product of the reliability of the task, long-term stability presumably captures both the reliability of the task and the extent to which heartbeat counting is an enduring trait. Over the short-term (e.g., <6 months), estimates of the stability of heartbeat counting performance in adulthood range from approximately r=. 41 to .81 depending on the time period examined, intervention (e.g., meditative training) and participant group employed (e.g., Ehlers, Breuer, Dohn, & Fiegenbaum, 1995; Ferentzi, Drew, Tihanyi, & Köteles, 2018; Herbert, Herbert, & Pollatos, 2011; Mussgay, Klinkenberg, & Rüddel, 1999; Parkin et al., 2014; Wittkamp, Bertsch, Vögele, & Schulz, 2018; for an overview see Ferentzi et al., 2018). In contrast, few studies have examined the stability of heartbeat counting across long time periods (e.g., >6 months). Indeed, in adulthood it appears that 9 months is the longest time period that stability has been assessed, with estimated stability approximately *r*=.70 (Bornemann & Singer, 2017). Such evidence suggests that some trait-like factors are indexed by heartbeat counting scores in adulthood.

Compared to adulthood, surprisingly few studies have examined the temporal stability of heartbeat counting in childhood. Given increasing focus on interoception across development (e.g., Murphy et al., 2017; Khalsa et al., 2018; Murphy, Viding & Bird, 2019), understanding the stability of this commonly used measure across developmental periods is of crucial importance.

Indeed, increased understanding of the stability of scores on this measure across periods where interoception may change may shed light on whether heartbeat counting performance can be considered an enduring trait across the lifespan, or whether certain developmental periods are associated with changing performance (e.g., Murphy et al., 2017; Murphy et al., 2019). Given links between heartbeat counting and mental health, better understanding of the developmental trajectory of heartbeat counting may ultimately aid our understanding of the development of conditions thought to be associated with interoception (e.g., Murphy et al., 2017; Khalsa & Lapidus, 2016; Khalsa et al., 2018). To our knowledge, only one study has examined the stability of heartbeat counting in childhood. In a large sample (N = 1350) of children aged between 6-11 years, stability of only $r=.33$ was observed across a 1-year period (Koch & Pollatos, 2014). Such evidence suggests that the long-term stability of heartbeat counting performance may be lower in childhood than adulthood.

The factors underlying the apparently reduced stability of heartbeat counting in childhood remain unknown. Furthermore, there is little research into the etiology of heartbeat counting at any developmental period. Twin studies enable the disentangling and estimation of genetic and environmental influences on traits, by comparing the similarity of monozygotic (identical) and dizygotic (non-identical) twins. Longitudinal twin studies can identify the extent to which genes and the environment influence stability and change of traits over time. For example, such studies indicate that the moderate stability of anxiety (and depression) from childhood through to adulthood is predominantly influenced by stability of genetic influences (Nivard et al., 2015; Waszczuk, Zavos, Gregory, & Eley, 2014). In contrast, environmental effects are primarily time specific, and are thus associated with change. To our knowledge only one twin study has examined the etiological factors underlying performance on the heartbeat counting task (Eley,

6

Gregory, Clark, & Ehlers, 2007). The authors observed a moderate genetic influence (~30%) on heartbeat counting at age 8. Non-shared environmental influences were substantial. However, no studies have used longitudinal data to assess whether these etiological influences remain stable over time. Longitudinal studies in childhood would aid our understanding of the etiological factors underlying both stability and change in heartbeat counting performance across a period where interoception may change (e.g., Murphy et al., 2017) and where performance is reportedly less stable (Koch & Pollatos, 2014).

In the present study we aimed to investigate the stability of the etiological influences on heartbeat counting across time. In addition to elucidating the factors underlying the long-term stability of heartbeat counting, we capitalised on the large sample to explore associations with other traits previously shown to covary with heartbeat counting. We aimed to estimate the magnitude of shared genetic and environmental influences between heartbeat counting and other traits associated with heartbeat counting in adulthood, such as anxiety and depression (see Khalsa & Lapidus, 2016), sleep problems (e.g., reduced sleep quality and insomnia; Ewing et al., 2017; Wei et al., 2016), and aspects of higher order cognition (e.g., emotion recognition; Terasawa et al., 2014) in this sample of children. Whilst few studies have examined these relationships in childhood, those that have typically observe similar associations with anxiety (e.g., Eley et al., 2007; Eley, Stirling, Ehlers, Gregory, & Clark, 2004). However, a recent study of pre-school children (aged 4-6 years) observed no relationship between cardiac interoception (using a modified version of the heartbeat counting task) and emotion recognition (Schaan et al., 2019). As such, it is unclear whether other relationships observed in adulthood (e.g., with sleep problems and emotion recognition) can be replicated in middle-childhood. Crucially, however, to our knowledge only one study has examined the etiology of these observed associations between

heartbeat counting and health (e.g., depression, anxiety, sleep problems) or higher order

cognition. In the only twin study of heartbeat counting described above, Eley et al. (2007)

observed that higher panic/somatic anxiety ratings were associated with lower error on the

heartbeat counting task (r= -.13). This relationship was partly explained by genetic factors

(genetic correlation = -.46 (95% CI: -1.00-1.00)) though this did not reach statistical

significance. Whether the etiology of this relationship remains stable over time, and can be

observed for other factors previously associated with heartbeat counting in adulthood (e.g., sleep

problems, emotion recognition, anxiety or depression), and childhood (e.g., anxiety) remains

unknown. However, better understanding of the shared etiology between heartbeat counting and

other traits across development would inform models of the etiology of mental health and may

ultimately provide insights into the potential efficacy of interventions aimed at improving

heartbeat counting and mental health.

This study first aimed to test the long-term stability of heartbeat counting in childhood

across a two-year period to examine whether the etiological factors change over time, and to

estimate to what extent genetic and environmental factors drive any observed stability. To this

end, we revisited data reported in Eley et al. (2007), and previously unexamined data collected in

the same twin sample two years later. This is the longest time period across which the long-term

stability of heartbeat counting has been assessed at any developmental stage. It is also the only

study to examine the etiology of the stability of heartbeat counting. Finally, we examined

previously-reported associations between heartbeat counting and aspects of health (anxiety,

depression, sleep problems) and higher order cognition (emotion recognition) to see whether 1)

associations observed in adulthood would be observed in childhood; and 2) these associations

were stable over time. Where significant relationships were observed, etiological factors

underlying these relationships could be assessed. In line with previous reports, we expected that the long-term stability of heartbeat counting would be low in childhood ($\sim r=.30$), with heartbeat counting expected to be associated with anxiety, depression, sleep problems and emotion recognition.

Methods

Participants

The ECHO study consists of 300 twin pairs from the Twins' Early Development Study, which recruited over 215,000 twin pairs born in England and Wales during 1994–96 (TEDS; Trouton, Spinath, & Plomin, 2002). Data were collected at the Institute of Psychiatry, London apart from a few families who were visited at home. Ethical approval was granted by the Maudsley Hospital Ethics Committee, London, UK. Informed consent from parents was obtained via postal methods in advance.

A selected extremes design was used when identifying the ECHO sample, to increase statistical power. This involved the selection of twin pairs where one or both twins scored high (top 15%) on parent-rated anxiety at age 7, plus pairs of controls where neither twin scored high on parent-rated anxiety at age 7. For the high anxiety group, 381 twin pairs were eligible and invited to participate, of these 247 twin pairs agreed to participate (65%). For control twin pairs, 92 pairs were eligible and invited to participate and of these 53 took part (58%). For further details regarding selection please see Gregory et al., (2006).

Following the selection of these pairs, 11 were removed due to mental or physical impairment. Zygosity was determined using parent-reported physical similarity plus DNA in

uncertain cases (see Price et al., 2000). One pair of unknown zygosity did not consent to provide

DNA and were excluded. In the final sample, 193 twin pairs completed both time points.

At Time point 1, twins were approximately 8 years of age ($M = 8.47$, $SD = 0.18$). Data

for Time point 2 was collected approximately 2 years later. Of the ECHO sample, at baseline

57% were female, and 33% were MZ twins.

Measures

The Heartbeat Counting Task

All participants completed the heartbeat counting task in a controlled laboratory setting.

Participants were asked to silently count the heartbeats they could feel during three intervals (of

35, 25 and 45 seconds) following a 10 second practice trial that was not analyzed (see Eley et al.,

2007 for further details). Objective heartbeat was recorded via a medical grade

electrocardiogram. During each trial, the electrocardiogram (ECG) was recorded and a custom

computer program (as used in Ehlers & Breuer, 1992) scored the number of R-waves (the largest

peak of the ECG QRS complex, with the number of peaks representing the actual number of

heartbeats). Participants were explicitly instructed not to take their pulse or to use any other

strategies such as holding their breath, which was visually checked by the researcher (trained

psychology graduates). At the start of each trial, participants heard a warning stimulus (800 Hz,

65 dB, 100 ms) to prepare them for the task (as in Ehlers & Breuer, 1992). This warning was

given 500 ms after an R wave was recorded on the participants ECG. The start signal (1000 Hz,

65 dB, 50 ms) was triggered immediately after the third R wave that followed the warning

stimulus. The tone signaling the end of the counting period (1000 Hz, 65 dB, 50 ms) was given

after the time interval for that trial was complete and 300 ms after the last R wave had elapsed.

At the end of each trial, the child told the researcher how many heartbeats they had counted. To prevent distraction and remove the possibility of cheating, children were seated so that they could not see the computer screen or ECG during the task.

For each trial, percentage error scores were calculated by taking the absolute difference between the actual number of heartbeats recorded by the ECG (AB) and the number of heartbeats counted by the child (CB), as a percentage of the number of actual heartbeats (i.e., $(((AB–CB)/AB)*100$) as in previous work (Ehlers & Breuer, 1992)). At both time points, scores across the three intervals were highly correlated (all rs >.80). As is typical, an average score was then taken across the three trials completed. Accordingly, a score of zero reflects totally accurate performance, whereas a score of 100 reflects totally inaccurate performance (e.g., feeling no heartbeats at all)[1].

Questionnaire measures

At both time points, self-report and parent-report data were available for a number of measures of health and wellbeing that have previously been linked to cardiac interoception in either children or adults. For anxiety, data were available from the Screen for Childhood Anxiety Related Emotional Disorders (SCARED; for psychometric properties see Birmaher et al., 1999; Birmaher et al., 1997) and the Children's Anxiety Sensitivity Index (CASI; for psychometric properties see Silverman, Fleisig, Rabian, & Peterson, 1991; Silverman, Ginsburg, & Goedhart, 1999; Silverman, Goedhart, Barrett, & Turner, 2003). Total scores for these measures were computed, with high scores on both measures reflective of higher anxiety/ greater sensitivity to the physical symptoms of anxiety. For depression, data were available from the Children's

---

[1] Note that in this sample no children overestimated the amount of actual heartbeats and therefore scores fit within the range of 0-100.

Depression Inventory (CDI; for psychometric properties see Kovacs, 1985; Smucker, Craighead, Craighead, & Green, 1986). Total scores for this measure were computed with high scores reflecting higher depressive symptoms. For sleep problems, data were available from the Sleep Self-Report (SSR; for psychometric properties see Owens, Spirito, McGuinn, & Nobile, 2000) and the Children's Sleep Habits Questionnaire (CSHQ; for psychometric properties see Owens, Spirito, & McGuinn, 2000). Again, total scores for these measures were computed with high scores reflecting greater sleep problems. In addition to total scores, for the measures where data was available (SCARED and CSHQ), subscale scores were calculated.

Emotion recognition

At Time point 2, data for emotion recognition ability were also available. The emotion recognition task consisted of 160 trials (for further details see Lau et al., 2009). On each trial, participants were presented with a facial image that morphed from a neutral expression into one of 5 basic expressions (angry, fear, sad, disgust, happy). All facial expressions were taken from a standard set of pictures of facial affect. Facial expression morphs were displayed as animations changing from neutral to one of four levels of intensity (25%, 50%, 75%, 100%) with intensity adjusted for happy expressions (to 10%, 25%, 50%, 75%) given that they were easier to identify. Head orientation (facing towards or away from the camera) and gaze direction (towards and away from the camera) were also manipulated resulting in 4 different trial types.

Prior to the task, participants were read standardised instructions and were asked to provide a definition of each emotion to ensure they were familiar with the emotion labels. On each trial participants were instructed to name the expression using one of five labels corresponding to the different emotions, with five practice trials completed before commencement of the task.

Accuracy scores were summed across all trials for each of the 5 expressions. In addition to the individual expression scores, these scores were also averaged across the 5 expressions to create an overall score. This variable was taken as a measure of overall emotion recognition ability, with high scores for all expressions and the total score reflecting better performance.

Statistical analyses

Selection variable

All twin analyses were conducted jointly with the 7-year anxiety screening variable from TEDS, allowing us to control for selection bias. Specifically, we linked our heartbeat counting data to the original distribution of the selection variable in the entire sample (N>5000). We then ran trivariate (3-variable) twin models decomposing shared variation between the two heartbeat counting variables and the selection variable. This approach of including the selection variable in the model-fitting allows maximum likelihood to estimate the corrected distributions, variances and covariances of the heartbeat counting variables. This increases the statistical power and generalisability of the analyses.

Twin model-fitting

The classic twin study design capitalizes on the fact that 'identical' monozygotic (MZ) twins in principle share 100% of their genes, whereas 'non-identical' dizygotic (DZ) twins share on average 50% of their segregating genes. However, these types of twins are equally similar in terms of their environment. The degree of genetic contribution to variation in a particular phenotype in a population is estimated by comparing monozygotic to dizygotic resemblance. The

extent to which members of monozygotic pairs are more similar to one another than members of dizygotic pairs indicates the degree of genetic influence on the trait of interest, because degree of genetic sharing correlates with degree of phenotypic similarity. Specifically, the comparison between MZ and DZ twins is used to estimate the contribution of genetic (A), shared environmental (C), and non-shared environmental (E) influences to variation in the phenotype. The heritability of a trait (A) is the proportion of phenotypic variance that can be explained by genetic variation in the population under study. Shared environment refers to environmental influences that result in family members resembling one another. Non-shared environment refers to environmental influences that make family members different from one another. Here we are not talking about whether the environmental experiences are shared, but whether their *effects* are shared.

To investigate the influences on the covariance between heartbeat counting performance across time, we fitted a longitudinal twin model, the Cholesky decomposition (see Figure 1). The Cholesky decomposition allows the investigation of stability and innovation in the genetic and environmental influences on our measure of heartbeat counting across the two time points. The first genetic factor (A1) represents genetic influences on heartbeat counting at Time 1. The extent to which these same genes also influence heartbeat counting at Time 2 is also estimated and is represented by the diagonal pathway from A1 to Time 2. The second genetic factor (A2) represents genetic influences on heartbeat counting at Time 2 that are independent of those influencing Time 1. The Cholesky model allows the A, C and E factors underlying the first measured variable to influence the second variable, but not vice versa. The same decomposition is done for the shared environmental and non-shared environmental influences (C1–2 and E1–2, respectively). A series of maximum-likelihood nested models were applied to the data, allowing

14

point estimates and confidence intervals to be established for the variance component estimates, and the model fit to be tested. This test is achieved by comparing the fit statistics of the model (difference in log likelihood, p-value, AIC) to a fully saturated model in which all parameters are free to vary, and no structure is imposed on the data. If the fit of the constrained model is not significantly worse than that of the saturated model, it may be considered a good fit. See Supplementary Table 1 for more model fit information and results. All twin model fitting analyses used full-information maximum likelihood and were carried out with structural equation modelling software OpenMx (Neale et al., 2016).

Relationships with other measures

We assessed correlations between heartbeat counting, health variables (e.g., anxiety, depression, sleep) and emotion recognition ability for the data available at both time points. We also conducted prospective analyses, by examining the correlation between heartbeat counting at Time 1 and health (e.g., anxiety, depression, sleep) and emotion recognition ability at Time 2. Differences in accuracy across the two time points was also assessed using a paired samples t-test.

Sensitivity analysis

We tested whether results changed when body mass index (BMI) and sex were regressed out of the heartbeat counting scores given evidence that both body composition and sex are related to heartbeat perception ability (e.g., Murphy et al., 2018; Grabauskaitė, Baranauskas & Griškova-Bulanova, 2017; Rouse, Jones & Jones, 1988). As this had little influence on the pattern of results obtained these data are reported in the Supplementary Information (see

supplementary Figure [S1]). Analyses in the main text feature the original phenotypes, since the sample size and thus statistical power, was higher than for the residualised phenotypes.

Results

Phenotypic descriptive statistics

See Table 1 for descriptive statistics of the heartbeat counting measures (for one randomly selected twin from each twin pair). The full sample used for longitudinal twin modelling was 5579, including individuals with ECHO and/or TEDS data, and zygosity data. Note that the selection variable is not represented in the table but was included in analyses. The variable represents case/control anxiety status, based on maternal ratings when the children were aged 7. The sample size for the selection variable was 5345, and 16.8% met case status for anxiety.

At both Time 1 and Time 2 heartbeat counting error scores were high with very few children meeting the cut off to be considered good perceivers (defined as <20% error; see Eley et al., 2007). At Time 1, 31 children met cut off (5.6% of the total sample). At Time 2, 39 children met cut off (9.5% of the total sample).

Table 1. Descriptive statistics

	N	Mean	SD	Min	Max
Time 1. Heartbeat counting error	279	69.01	26.3	4.34	100
Time 2. Heartbeat counting error	204	57.93	27.29	5.22	100

Phenotypic correlations across time and across twins

Error scores for heartbeat counting were significantly lower at Time 2 ($M = 57.35$, $SD = 27.40$) in comparison to those at Time 1 ($M = 69.09$, $SD = 27.17$), $t(196) = -4.87$, $p<.001$. The overall phenotypic correlation between heartbeat counting scores across the two time points was $r = 0.35$, indicating moderate stability in task performance. Initial inspection of twin correlations for heartbeat counting performance suggested a heritability of ~42% at time-point 1 and ~4% at time-point 2.

Phenotypic correlations with other measures

Heartbeat counting performance was not significantly correlated with any of the examined variables (symptoms of anxiety, depression or sleep quality) at Time 1 nor with any of the examined variables (symptoms of anxiety, depression, sleep problems or emotion recognition) at Time 2 (Table 2). When considering prospective analyses between heartbeat counting at Time 1 and aspects of health (anxiety, depression and sleep) and higher order

cognition at Time 2, only emotion recognition performance at Time 2 was predicted by earlier heartbeat counting performance. Specifically, higher error of heartbeat counting at Time 1 negatively predicted subsequent emotion recognition ability (Table 2). However, this relationship was not significant after correction for multiple comparisons. Where subscales were available, we also examined the relationship between heartbeat counting and subscales for these measures. As reported in Eley et al., (2007) at Time 1 heartbeat counting was associated with panic/somatic anxiety symptoms. At Time 2, heartbeat counting was associated with social phobia and the recognition of happiness and sadness, specifically. However, none of these relationships were significant after correction for multiple comparisons (see supplementary information [S2-S4]). Given that no significant relationships were observed, it was not possible to examine the etiological factors underlying the predicted overlap between these factors and heartbeat counting.

Table 2. Correlations between heartbeat counting and other measures at Time 1 and Time 2

		Anxiety (SCARED)	Anxiety (CASI)	Depression (CDI)	Sleep problems (SSR)	Sleep Problems (CSHQ)	Emotion Recognition
Time 1	Heartbeat counting error	-.028 N = 279	-.104 N = 279	-.021 N = 277	-.011 N=277	-.015 N=269	N/A
Time 2	Heartbeat counting error	-.115 N = 204	.029 N = 203	.043 N = 204	-.007 N = 203	.044 N = 198	.079 N = 203
Time 1- > Time 2	Heartbeat counting error	.069 N = 240	.045 N = 239	.030 N = 240	-.041 N = 237	-.109 N = 231	-.152* N = 238

*Note. Time 1 -> Time 2 refers to phenotypic correlations between heartbeat counting error scores at Time 1 and questionnaire/cognition measures at Time 2. As shown, only emotion recognition scores at Time 2 were predicted by heartbeat perception at Time 1. *denotes significant at p<.05.*

Longitudinal twin model-fitting results

Figure 1 presents the results of the model-fitting analyses. First considering the total genetic, shared environmental and non-shared environmental influences at each time-point (represented on the vertical lines for heartbeat counting task at Time 1, and by the addition of the vertical and diagonal lines for Time 2), there was a moderate heritability of heartbeat counting at Time 1 (30%). At Time 2, heritability was much lower, at 6%. In contrast, shared environmental influences increased from 6% to 22% from Time 1 to Time 2. The nonshared environmental contributions were more similar across timepoints, being 64% and 73% at Time 1 and Time 2, respectively. It is important to note that the majority of these parameter estimates are non-significant (Figure 1; for unstandardised estimates see supplementary Figure [S5]). This is because, although the sample is well-powered for phenotypic correlation analyses, power for twin model-fitting, especially for distinguishing genetic from shared environmental influences, is low. In sensitivity analyses we tested whether all familial influences (i.e. genetic and shared environmental factors) on stability across time could be dropped from the model. We found that simultaneously removing both genetic and shared environmental factors significantly reduces model fit (see Supplementary Table [S6]), indicating the presence of familial influences on the longitudinal association.

Focusing on the level of innovation in genetic and environmental influences at Time 2 (represented by the vertical lines coming from A2, C2 and E2), there were no new genetic influences on heartbeat counting error scores at Time 2, indicating that genetic influences across this age-range were entirely stable (albeit low). In contrast, for both shared and non-shared environment there were new influences at Time 2. For example, at Time 2, 22% of the variance is accounted for by the shared environment, of which 82% (18 as a proportion of 22) is new

variance specific to this time-point. Non-shared environmental influences were primarily time-specific, and were the strongest contributor to change from Time 1 to Time 2.
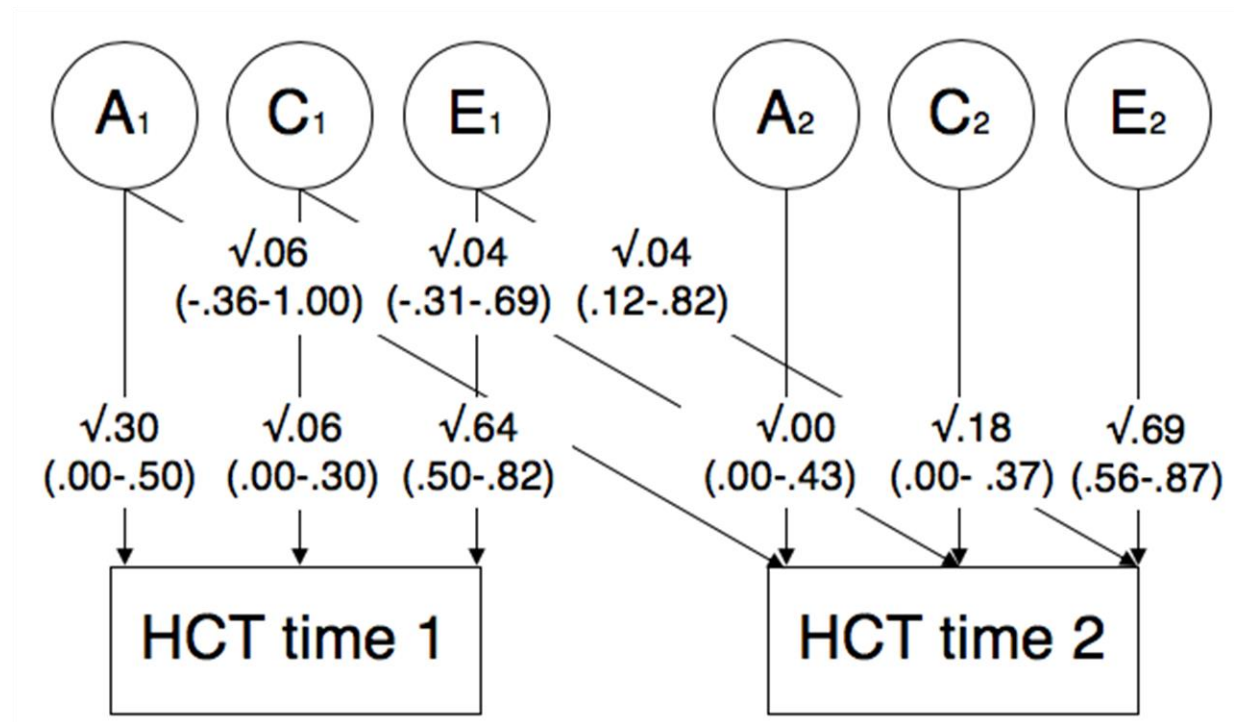

Figure 1: Standardised path estimates from the Cholesky decomposition, plus 95% confidence intervals. Note that A1, C1 and E1 represent the proportion of variance in heartbeat counting performance at Time 1 explained by genetic, shared environmental and non-shared environmental influences, respectively. A1, C1 and E1 sum to 100%. The diagonal paths show how much A1, C1 and E1 influence heartbeat counting at Time 2. These are added to A2, C2 and E2, respectively, to give the overall variance explained by genetic, shared environmental and non-shared environmental factors at Time 2 (the estimates in the figure add to 101% due to rounding). For example, the heritability is 6% at Time 2 (6% + 0%). The selection variable has not been represented, but was accounted for in analyses. See Supplementary Figure S10 for unstandardised estimates and Supplementary Table [S6] for model fit statistics.

Considering the diagonal lines running from A1, C1 and E1 to heartbeat counting task performance at Time 2, the data suggest that some of the shared environmental influence on heartbeat counting performance at Time 1 also influences the trait at Time 2 (4% out of the total 22% shared environmental influence). Non-shared environmental influence from Time 1 also influences trait variation at Time 2 (4% out of the 73%). In sum, genetic influences are largely

stable and environmental influences are largely time-specific. Indeed, of the phenotypic stability ($r = 0.35$), the percentages due to A, C and E were 37% [-36%-103%], 15% [-31%-69%] and 48% [12%-89%]. Here, E contributes most to the sharing of influences between the two time points because non-shared environments are the major sources of variation at both time-points, even if only 4% out of 73% of the non-shared environmental influences at Time 2 affected heartbeat counting at Time 1.

Our sensitivity analyses showed that results did not differ when controlling for BMI and sex (see Supplementary Figure [S1]).

Discussion

This study aimed to quantify the long-term stability of heartbeat counting across a 2-year period in childhood and examine the etiological factors underlying stability of performance. Additionally, we also examined the relationship between heartbeat counting and aspects of health (anxiety, depression and sleep) and higher-order cognition (emotion recognition) that have previously been associated with heartbeat counting in adulthood, but not previously examined in childhood.

First, considering the longitudinal stability of the heartbeat counting phenotype, results revealed a small but significant correlation between heartbeat counting at Time point 1 and Time point 2, with a reduction in error observed with age. In line with previous estimates across a 1-year period (Koch & Pollatos, 2014), these data suggest that the long-term stability of heartbeat counting in childhood is relatively low (~r=.35) in comparison to stability estimates in adulthood (r=~.41-.81; see Ferentzi et al., 2018). One explanation for this discrepancy is differences in the time period examined (to our knowledge 9 months is the longest time period across which heartbeat counting stability has been assessed in adulthood; Bornemann & Singer, 2017).

However, as estimates of stability in childhood over a 1-year period (~r=.33; Koch & Pollatos, 2014) are also lower than estimates in adulthood, this is an unlikely explanation for the pattern of results obtained. An alternative explanation is that heartbeat counting performance changes to a greater extent in childhood than in adulthood. Indeed, in the present study a reduction in error was observed from Time 1 to Time 2 suggesting that (in general), changes in heartbeat counting performance in this sample were driven by an improvement in performance with age. It is therefore possible that in childhood, heartbeat counting ability may not be fully developed or it may not have reached adult levels of stability. For example, anxiety is only moderately stable across childhood: correlations between measures at different time points are only ~r=0.30 (e.g., Cheesman et al., 2018). In contrast, stability in adulthood is greater, in the region of ~r=.50 (e.g., Nes, Røysamb, Reichborn-Kjennerud, Harris, & Tambs, 2007) with some evidence that stability increases with age (from around r=0.6 in adolescence to r=.80 in adulthood; e.g., Nivard et al., 2015). As such, whilst the stability of heartbeat counting is lower in childhood than adulthood, this is consistent with a body of literature that indicates that the stability of a number of traits is lower in childhood (a time of great developmental change) as compared to adulthood.

Of course, we must also acknowledge that, in the context of questions over the validity of the heartbeat counting task as measure of interoception (e.g., Desmedt et al., 2018; Khalsa et al., 2009; Murphy et al., 2018; Ring et al., 2015; Zamariola et al., 2018), we cannot confidently infer the long-term stability of interoception across middle childhood from these data. Even if the heartbeat counting task can be considered a valid measure of interoception, it is also possible that that low long-term stability of heartbeat counting reflects unreliability of the measure, particularly across middle childhood. Indeed, as the long-term stability of heartbeat counting performance is likely a product of both the test re-test reliability of the measure and the extent to

which heartbeat counting is an enduring trait, further research into the short term test-re-test reliability of the measure across middle childhood is required to determine the extent to which task reliability and/or developmental change underlies the long-term stability of heartbeat counting in middle childhood. Alternatively, however, it may be that performance on the heartbeat counting task is largely state-dependent. Given some evidence that time-specific person-situational factors are related to performance in adulthood (e.g., Wittkamp et al., 2018), it may be that state-specific effects have a greater influence on performance in middle-childhood. Future research assessing both short- and long-term stability, using multiple measures of interoception, may help to disentangle these possibilities.

Turning to our twin model-fitting results, our main finding is that the primary factor influencing variation in heartbeat counting at each time-point is the non-shared environment. Such individual-specific environments accounted for the greatest proportion of variance at both time points, 64% and 73% at Time 1 and Time 2, respectively. In terms of the etiological factors underlying performance, some caution is required given that the sample size employed here is small for a twin study. As a result, we have limited statistical power to test the significance of genetic and shared environmental parameters. We therefore discuss the results for the genetic and shared environment components of variance with caution, as most estimates have confidence intervals that cross zero. The heritability of heartbeat counting dropped between Time 1 and Time 2 from 30% to 6%. No new genetic influences were observed at Time 2, suggesting that genetic influences across this age-range are entirely stable. In contrast, shared environmental influences increased from 6% to 22%, with most of the influence at Time 2 due to new environmental factors.

In terms of the etiological factors underlying long-term stability, results from genetic analyses suggest that the small amount of stability of heartbeat counting observed (r=.35) is driven by genetic, shared environmental and non-shared environmental factors (37%, 15% and 48%, respectively), with only non-shared environmental factors making a significant contribution to long-term stability. Although measurement error is captured in the non-shared environment component in twin studies, the large non-shared environment estimates are unlikely to be a product of measurement error alone, as this is unlikely to be stable across years. As such, it appears that time-specific non-shared environmental factors contribute significantly towards performance at both time points, but a large proportion of the observed stability is explained by the stable non-shared environmental factors. Such a pattern indicates that heartbeat counting performance, both performance at each time point and stability, is driven largely by child-specific factors – factors that make individuals in the same family different. Whilst these data cannot elucidate what factors these may be, various non-shared factors are likely to contribute to both time-specific performance and long-term stability. This may include factors related to the administration of the heartbeat counting task (e.g., Wittkamp et al., 2018), experience-based factors that may influence performance (e.g., heart rate knowledge; Ring et al., 2015), as well as factors that shape individual differences in the ability to perceive one's heartbeat (e.g. chance or environmentally-driven changes in blood pressure or resting heart rate; see Murphy et al., 2018). Future research should seek to replicate this finding of the primary importance of non-shared environmental influences, and to identify the specific physiological and psychological factors that contribute to this component of variance. In particular, further research using a control task would be useful to elucidate whether the stability of non-shared environmental factors is a

product of task administration (e.g., similarity of testing sessions) or reflects enduring child-specific factors that shape individual differences in heartbeat counting ability.

As noted above, several other findings emerge from the data pertaining to genetic and shared environmental influences, but statistical power is too low to have high confidence in these results. Whilst genetic and shared-environmental factors did not significantly contribute towards time-specific performance or stability, it is still notable that genetic factors remained stable over time – all of the genetic variance at Time 2 was explained by variance at Time 1 – though in comparison to Time 1 (reported in Eley et al., 2007) the genetic influence at Time 2 was much lower. In contrast, shared environmental factors were largely time-specific and showed an increase from Time 1 to Time 2. This observation must be treated with some caution: when power is low, it is difficult to distinguish genetic from shared environmental influences. As such, we cannot be certain which makes a greater contribution towards long-term stability. Nevertheless, these data suggest some role of genetic and shared-environmental factors to heartbeat counting performance and highlight a need for further research into the etiology of heartbeat counting performance, and the etiology of individual differences in interoception more broadly, in larger samples.

In addition to answering questions regarding the factors underlying stability, a secondary question concerned the relationship between heartbeat counting and health/cognition during middle childhood. Contrary to predictions based on previously-reported associations in adulthood (e.g., Ewing et al., 2017; Khalsa & Lapidus, 2016; Terasawa et al., 2014), no significant relationship between heartbeat counting and anxiety, depression, sleep problems or emotion recognition was observed in this sample of children at either Time 1 or Time 2 after correction for multiple comparisons. There are several possible explanations for this lack of

significant correlations. First, this sample was comprised mostly of highly-anxious individuals

and it is possible that the relationship between heartbeat counting and the factors examined here

may differ as a function of anxiety levels. However, as low-anxiety control participants were also

studied, and no relationship was observed between heartbeat counting and anxiety, it is unlikely

that this provides a full explanation of these findings. Second, it is of course possible that the

relationship between interoception and mental health/cognition emerges over the course of

development. If, as has been suggested by Murphy et al., (2017; 2019), adolescence is a sensitive

period for interoceptive development (and the heartbeat counting task can be considered a

measure of interoception), it is possible that the relationship between interoception and mental

health/cognition emerges at a later stage of development. Third, it must also be noted that in this

sample very few children met criteria to be considered a good heartbeat perceiver. As such, it is

possible that this also contributes towards the absence of previously reported relationships

between heartbeat counting and aspects of health and higher order cognition. Of course, a final

possibility is that unreliability of the heartbeat counting task may also contribute towards the lack

of significant associations between scores on this task and other measures. Indeed, as highlighted

throughout, the validity of this measure remains under question with various non-interoceptive

factors thought to contribute towards task performance (e.g., Desmedt et al., 2018; Khalsa et al.,

2009; Murphy et al., 2018; Zamariola et al., 2018). As a number of these possible confounds

(e.g., heart rate knowledge, systolic blood pressure) were not controlled for here (given that the

data were collected prior to the outlining of appropriate controls), it is not possible to conclude

that associations between heartbeat counting and aspects of health and cognition would not be

found in childhood if the full range of controls were employed. However, given that associations

in adulthood have been reported without using these controls, this again is unlikely to fully

explain the pattern of results unless it is assumed that these confounds are likely to have a larger impact in childhood than adulthood. These data highlight an urgent need to examine the relevance of interoception to health and cognition across development using valid measures of interoception.

Despite the absence of significant relationships between heartbeat counting and aspects of health (anxiety, depression or sleep) and higher order cognition after correction for multiple comparisons, it is notable that when considering the relationship between heartbeat counting at Time 1 and emotion recognition ability at Time 2 a significant relationship emerged. Better heartbeat counting performance at Time 1 was marginally associated with better emotion recognition performance at Time 2. Likewise, certain anxiety subscales (e.g., social phobia) were associated with heartbeat counting and an association between heartbeat counting and the recognition of sadness and happiness was observed. Although these relationships did not survive correction for multiple comparisons, and therefore some caution is required, they are consistent with the proposed relationship between interoception and social abilities (emotion recognition and social phobia; e.g., Terasawa et al., 2014; Clark & Wells, 1995). Given that some of these relationships (e.g., emotion recognition and interoception) have not been observed at earlier stages of development (4-6 years; Schaan et al., 2019) these data may be taken to suggest that these relationships may begin to emerge in middle-childhood. As such, further research into the relationship between social cognition and interoception in childhood is warranted and would benefit from the use of multiple measures of interoceptive ability.

Notwithstanding the importance of these data we must acknowledge certain limitations. First, as this study involved re-examination of historical data, a number of factors were not controlled for (e.g., differences in heart rate physiology, knowledge of resting heart rate or

physical activity), and a control task was not utilized. However, given that few studies employ the full range of control variables advocated by Murphy et al., (2018), these data provide a crucial understanding of the etiological factors underlying the long-term stability of the heartbeat counting task, and the relationship between the measure and health and cognition, as the task is routinely administered. Second, it is important to acknowledge that given low statistical power the results of the genetic analyses must be treated with some caution. Nevertheless, this one of the largest samples available in studies of heartbeat counting performance in childhood, and is a rare genetically-sensitive resource. We were able to draw several conclusions with confidence, particularly that non-shared environmental factors are the driver of stability and time-specific differences in performance. These data provide the first description of the etiological factors underlying the long-term stability of performance on the heartbeat counting task, highlighting that individual-specific factors play a fundamental role over time.

In conclusion, this study sought to examine the long-term stability of heartbeat counting over a 2-year period in childhood and the etiological factors underlying stability. Results revealed low stability in childhood, with non-shared environmental factors substantially contributing to both time-specific performance and stability. Contrary to predictions, heartbeat counting was not associated with health or higher order cognition in this sample. These data contribute towards the growing debate surrounding the heartbeat counting task, suggesting that low stability may reflect either the unreliability of the measure, or that heartbeat counting is not a stable trait. Identification of the individual-specific factors contributing to stability of performance may shed light on the validity of the measure for quantifying individual differences in interoception.

Acknowledgements

Disclosures:

Alice Gregory is an advisor for a project sponsored by Johnson's Baby. She has written a book Nodding Off (Bloomsbury Sigma, 2018) and has a contract for a second book Sleepy Pebble (Nobrow). She is a regular contributor to BBC Focus magazine and has contributed to other outlets (such as The Conversation, The Guardian and Balance Magazine). She occasionally receives sample products related to sleep (e.g. blue light blocking glasses) and has given a paid talk to a business.

Data Availability Statement:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*(10), 1230–1236. doi:10.1097/00004583-199910000-00011

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., & Neer, S. M. (1997). The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(4), 545-553.

Bornemann, B., & Singer, T. (2017). Taking time to feel our body: Steady increases in heartbeat perception accuracy and decreases in alexithymia over 9 months of contemplative mental training. *Psychophysiology*, *54*(3), 469–482. doi:10.1111/psyp.12790

Cheesman, R., Purves, K. L., Pingault, J. B., Breen, G., Plomin, R., & Eley, T. C. (2018). Extracting stability increases the SNP heritability of emotional problems in young people. *Translational psychiatry, 8*(1), 223.

Clark, D.M., & Wells, A. A cognitive model of social phobia. R.G Heimberg, M.R Liebowitz, D.A Hope, F. R Schneier (Eds.), Social phobia: diagnosis, assessment and treatment, Guilford Press, New York (1995), pp. 69-93

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews. Neuroscience*, *3*(8), 655–666. doi:10.1038/nrn894

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*.

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. doi:10.1016/j.biopsycho.2018.09.004

Dunn, B. D., Galton, H. C., Morgan, R., Evans, D., Oliver, C., Meyer, M., … Dalgleish, T. (2010). Listening to your heart. How interoception shapes emotion experience and intuitive decision making. *Psychological Science*, *21*(12), 1835–1844. doi:10.1177/0956797610389191

Ehlers, A., & Breuer, P. (1992). Increased cardiac awareness in panic disorder. *Journal of abnormal psychology*, *101*(3), 371-382.

Ehlers, A., & Breuer, P. (1992). Increased cardiac awareness in panic disorder. *Journal of Abnormal Psychology*, *101*(3), 371–382. doi:10.1037/0021-843X.101.3.371

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

Eley, T. C., Gregory, A. M., Clark, D. M., & Ehlers, A. (2007). Feeling anxious: a twin study of panic/somatic ratings, anxiety sensitivity and heartbeat perception in children. *Journal of*

Child Psychology and Psychiatry, and Allied Disciplines, *48*(12), 1184–1191. doi:10.1111/j.1469-7610.2007.01838.x

Eley, T. C., Stirling, L., Ehlers, A., Gregory, A. M., & Clark, D. M. (2004). Heart-beat perception, panic/somatic symptoms and anxiety sensitivity in children. *Behaviour Research and Therapy*, *42*(4), 439–448. doi:10.1016/S0005-7967(03)00152-9

Ewing, D. L., Manassei, M., Gould van Praag, C., Philippides, A. O., Critchley, H. D., & Garfinkel, S. N. (2017). Sleep and the heart: Interoceptive differences linked to poor experiential sleep quality in anxiety and depression. *Biological Psychology*, *127*, 163–172. doi:10.1016/j.biopsycho.2017.05.011

Ferentzi, E., Drew, R., Tihanyi, B. T., & Köteles, F. (2018). Interoceptive accuracy and body awareness - Temporal and longitudinal associations in a non-clinical sample. *Physiology & Behavior*, *184*, 100–107. doi:10.1016/j.physbeh.2017.11.015

Grabauskaitė, A., Baranauskas, M., & Griškova-Bulanova, I. (2017). Interoception and gender: What aspects should we pay attention to?. *Consciousness and cognition, 48*, 129-137.

Gregory, A. M., Rijsdijk, F. V., Dahl, R. E., McGuffin, P., & Eley, T. C. (2006). Associations between sleep problems, anxiety, and depression in twins at 8 years of age. *Pediatrics*, *118*(3), 1124-1132.

Herbert, B. M., Herbert, C., & Pollatos, O. (2011). On the relationship between interoceptive awareness and alexithymia: is interoceptive awareness related to emotional awareness? *Journal of Personality*, *79*(5), 1149–1175. doi:10.1111/j.1467-6494.2011.00717.x

Khalsa, S S, Rudrauf, D., Sandesara, C., Olshansky, B., & Tranel, D. (2009). Bolus isoproterenol infusions provide a reliable method for assessing interoceptive awareness. *International Journal of Psychophysiology*, *72*(1), 34–45. doi:10.1016/j.ijpsycho.2008.08.010

Khalsa, Sahib S, & Lapidus, R. C. (2016). Can Interoception Improve the Pragmatic Search for Biomarkers in Psychiatry? *Frontiers in Psychiatry*, *7*, 121. doi:10.3389/fpsyt.2016.00121

Khalsa, Sahib S, Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., … Interoception Summit 2016 participants. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry : Cognitive Neuroscience and Neuroimaging*, *3*(6), 501–513. doi:10.1016/j.bpsc.2017.12.004

Koch, A., & Pollatos, O. (2014). Interoceptive sensitivity, body weight and eating behavior in children: a prospective study. *Frontiers in Psychology*, *5*, 1003. doi:10.3389/fpsyg.2014.01003

Kovacs, M. (1985). The children's depression, inventory (CDI). *Psychopharmacology Bulletin*, *21*(4), 995–998.

Lau, J. Y. F., Burt, M., Leibenluft, E., Pine, D. S., Rijsdijk, F., Shiffrin, N., & Eley, T. C. (2009). Individual differences in children's facial expression recognition ability: the role of nature and nurture. *Developmental Neuropsychology*, *34*(1), 37–51. doi:10.1080/87565640802564424

Murphy, J, Brewer, R., Catmur, C., & Bird, G. (2017). Interoception and psychopathology: A developmental neuroscience perspective. *Developmental Cognitive Neuroscience*, *23*, 45–56. doi:10.1016/j.dcn.2016.12.006

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.

Murphy, J., Viding, E & Bird, G., (2019). Does atypical interoception following physical change contribute to sex differences in mental illness? *Psychological Review.*

Mussgay, L., Klinkenberg, N., & Rüddel, H. (1999). Heart Beat Perception in Patients with Depressive, Somatoform, and Personality Disorders. *Journal of Psychophysiology*, *13*(1), 27–36. doi:10.1027//0269-8803.13.1.27

Neale, M.C., Hunter, M.D., Pritikin, J.N., Zahery, M., Brick, T.R., Kirkpatrick, R.M., Estabrook, R., Bates, T.C., Maes, H.H. and Boker, S.M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. Psychometrika, 81(2), pp.535-549.

Nes, R. B., Røysamb, E., Reichborn-Kjennerud, T., Harris, J. R., & Tambs, K. (2007). Symptoms of anxiety and depression in young adults: genetic and environmental influences on stability and change. *Twin Research and Human Genetics*, *10*(3), 450-461.

Nivard, M. G., Dolan, C. V., Kendler, K. S., Kan, K. J., Willemsen, G., Van Beijsterveldt, C. E. M., ... & Middeldorp, C. M. (2015). Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychological medicine*, *45*(5), 1039-1049.

Nivard, M. G., Dolan, C. V., Kendler, K. S., Kan, K. J., Willemsen, G., van Beijsterveldt, C. E. M., … Boomsma, D. I. (2015). Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychological Medicine*, *45*(5), 1039–1049. doi:10.1017/S003329171400213X

Owens, J. A., Spirito, A., & McGuinn, M. (2000). The Children's Sleep Habits Questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. *Sleep*, *23*(8), 1043–1051.

Owens, J. A., Spirito, A., McGuinn, M., & Nobile, C. (2000). Sleep habits and sleep disturbance in elementary school-aged children. *Journal of Developmental and Behavioral Pediatrics*, *21*(1), 27-36.

Parkin, L., Morgan, R., Rosselli, A., Howard, M., Sheppard, A., Evans, D., … Dunn, B. (2014). Exploring the relationship between mindfulness and cardiac perception. *Mindfulness*, *5*(3), 298–313. doi:10.1007/s12671-012-0181-7

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research and Human Genetics*, *3*(3), 129-133.

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. doi:10.1016/j.biopsycho.2014.12.010

Rouse, C. H., Jones, G. E., & Jones, K. R. (1988). The effect of body composition and gender on cardiac awareness. *Psychophysiology, 25*(4), 400-407.

Schaan, L., Schulz, A., Nuraydin, S., Bergert, C., Hilger, A., Rach, H., & Hechler, T. (2019). Interoceptive accuracy, emotion recognition, and emotion regulation in preschool children. *International Journal of Psychophysiology, 138*, 47-56.

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

Shah, P., Catmur, C., & Bird, G. (2017). From heart to mind: linking interoception, emotion, and theory of mind. *Cortex, 93*, 220-223.

Silverman, W. K., Fleisig, W., Rabian, B., & Peterson, R. A. (1991). Childhood Anxiety Sensitivity Index. *Journal of Clinical Child Psychology*, *20*(2), 162–168. doi:10.1207/s15374424jccp2002_7

Silverman, W. K., Ginsburg, G. S., & Goedhart, A. W. (1999). Factor structure of the childhood anxiety sensitivity index. *Behaviour research and therapy*, *37*(9), 903-917.

Silverman, W. K., Goedhart, A. W., Barrett, P., & Turner, C. (2003). The facets of anxiety sensitivity represented in the Childhood Anxiety Sensitivity Index: Confirmatory analyses of factor models from past studies. *Journal of Abnormal Psychology*, *112*(3), 364-374.

Smucker, M. R., Craighead, W. E., Craighead, L. W., & Green, B. J. (1986). Normative and reliability data for the Children's Depression Inventory. *Journal of abnormal child psychology*, *14*(1), 25-39.

Terasawa, Y., Moriguchi, Y., Tochizawa, S., & Umeda, S. (2014). Interoceptive sensitivity predicts sensitivity to the emotions of others. *Cognition & Emotion*, *28*(8), 1435–1448. doi:10.1080/02699931.2014.888988

Waszczuk, M. A., Zavos, H. M. S., Gregory, A. M., & Eley, T. C. (2014). The phenotypic and genetic structure of depression and anxiety disorder symptoms in childhood, adolescence, and young adulthood. *JAMA Psychiatry*, *71*(8), 905–916. doi:10.1001/jamapsychiatry.2014.655

Wei, Y., Ramautar, J. R., Colombo, M. A., Stoffers, D., Gómez-Herrero, G., van der Meijden, W. P., … Van Someren, E. J. W. (2016). I keep a close watch on this heart of mine: increased interoception in insomnia. *Sleep*, *39*(12), 2113–2124. doi:10.5665/sleep.6308

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

Zamariola, G., Maurage, P., Luminet, O., & Corneille, O. (2018). Interoceptive accuracy scores from the heartbeat counting task are problematic: Evidence from simple bivariate correlations. *Biological Psychology*, *137*, 12–17. doi:10.1016/j.biopsycho.2018.06.006