Respiratory and Cardiac Interoceptive Sensitivity in the First Two Years of Life

Tunte, Markus; Hohl, Stephanie; Wunderwald, Moritz; Bullinger, Johannes; Boyadziheva, Asena; Maister, Lara; Elsner, Birgit; Tsakiris, Manos; Kayhan, Ezgi

Elife

E-pub ahead of print: 18/12/2024

Peer reviewed version

eLife

Reviewed Preprint
v2 • September 10, 2024
Revised by authors

Reviewed Preprint
v1 • December 18, 2023

Neuroscience

Respiratory and Cardiac Interoceptive Sensitivity in the First Two Years of Life

Markus R Tünte, Stefanie Höhl, Moritz Wunderwald, Johannes Bullinger, Asena Boyadziheva, Lara Maister, Birgit Elsner, Manos Tsakiris, Ezgi Kayhan

University of Vienna, Faculty of Psychology, Department of Developmental and Educational Psychology, Austria • Vienna Doctoral School Cognition, Behavior and Neuroscience, University of Vienna, Vienna, Austria • Ludwig-Maximilians-Universität München, Munich, Germany • School of Human and Behavioural Sciences, College of Human Sciences, Prifysgol Bangor University, Gwynedd, Wales, LL57 2AS, United Kingdom • Department of Developmental Psychology, University of Potsdam, Germany • Department of Psychology, Royal Holloway University of London, the United Kingdom • Max Plank Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Abstract

Several recent theoretical accounts have posited that interoception, the perception of internal bodily signals, plays a vital role in early human development. Yet, empirical evidence of cardiac interoceptive sensitivity in infantsto date has been mixed. Furthermore, existing evidence does not go beyond the perception of cardiac signals and focuses only on the ageof 5–7months, limiting the generalizability of the results.Here, we used a modified version of the cardiac interoceptive sensitivity paradigm introduced by Maister et al. (2017) in 3-, 9-, and 18-month-old infantsusing cross-sectional and longitudinal approaches. Going beyond, we introduce a novel experimental paradigm, namely the iBREATH,to investigate respiratory interoceptive sensitivity in infants. Overall, for cardiac interoceptive sensitivity (*totaln*= 135) we find rather stableevidence across ages with infants on average preferring stimuli presented synchronously to their heartbeat. For respiratory interoceptive sensitivity (*totaln* = 120) our results show a similar pattern in the first year of life, but not at 18months.We did not observe a strong relationship between cardiac and respiratory interoceptive sensitivity at 3 and 9months butfoundsome evidence for a relationship at 18months. We validated our results using specification curve- and mega analytic approaches.Byexaminingearly cardiac and respiratory interoceptive processing,we provide evidence that infants are sensitive to their interoceptive signals.

eLife assessment

This study presents **important** findings on the early development of cardiac and respiratory interoceptive sensitivity based on an investigation of infants aged 3, 9 and 18 months and on extensive statistical analyses. The evidence supporting the conclusions are **convincing** although the research faced technical and recruitment challenges that limit the findings interpretation and generalizability. This study will be of significant interest to developmental psychologists and neuroscientists working on interoception and its influence on socio-cognitive development.

https://doi.org/10.7554/eLife.91579.2.sa3

Introduction

Bodily functionssuch as heartbeat and respiration are vital to the survival of living beings. The perception of signals arising from the body such as heartbeat, respiration, and hunger is called interoception (Craig, 2002). Individuals differ with regard to their interoceptive sensitivity, the degree to which they perceive their own bodily signals (Critchley & Harrison, 2013). Interoceptive sensitivity is related to human experience and behavior, such as the perception of emotions, mental health,and social cognition(Khalsa et al., 2018).Further, several recent theoretical accounts have highlighted that interoceptive sensitivity plays a vital role in early development in infancy,such asthe development ofthe self and early social abilities(Filippetti, 2021; Fotopoulou & Tsakiris, 2017; Musculus et al., 2021).As infants are born with limited ability to self-regulate bodily states giving rise to interoceptive sensations such as hunger,they rely on interactions with their primary caregiver for co-regulation. These interactions in turn play an important role in shaping early development in infancy. Despite these theoretical frameworks,we have little knowledge aboutinfants'sensitivity to theirinteroceptive signals.Recently, the first paradigm to assess cardiac interoceptive sensitivity in infancywas introduced (Maister et al. 2017). In the present study, we aim to replicate the experimental paradigm introduced by Maister et al. (2017) on cardiac interoceptive sensitivity in infants.Further, we aim at tracking the development of interoceptive sensitivity and related individual differences across the infancy period.Going beyond cardiac perception, we introduce a novel approach to measure respiratory interoceptive sensitivity in infants.

Most empirical investigations of interoceptive processing have focused on cardiac interoception(Khalsa et al., 2018). In adults, a large body of research has investigated cardiac interoceptive perception using paradigms in which participants are asked to count or detect their own heartbeat (Brener & Ring, 2016; Schandry, 1981). Using modified versions of the tasks for adult participants, studies with children have shown that stable and adult-like interoceptive skills can be measured already at 4–6 years of age (Schaan et al., 2019).

In contrast to the existing evidence on cardiac interoception in children, we know little about whether infants perceive their interoceptive signals. The first published empirical evidence on interoceptive sensitivity used an eye-tracking paradigm,namelythe iBEATs task,in which 5-month-old infants observed images on the screen such as clouds and stars that bounced either synchronouslyor asynchronouslyto the infant's heartbeat. Maister and colleagues (2017) found that infants on average looked longer at stimuli that moved asynchronously to their heartbeat as compared to stimuli moving synchronously. Furthermore, infants' cardiac interoceptive sensitivityscores were correlated with their heartbeat evoked potentials (HEPs), a neural marker of interoceptive processing (Coll et al., 2021).This study provided the first evidence that already

at 5months of age infants show sensitivity to their own cardiac signals. Further, this approach has also successfully been replicated with rhesus monkeys (Charbonneau et al., 2022) and a recent study using an adapted experimental paradigm in 6-month-old infants has found similar results (Imafuku et al., 2023)

Recently, however, no evidence of cardiac interoceptive sensitivity in 5- to 7-month-old infants was reported (Weijs et al., 2023). Despite some methodological differences,all studiesusedvery similar experimental paradigms in which infantswere presented with stimuli oscillating either synchronously or asynchronously to their heartbeat. It is unclear whether thenull findingsreported by Weijs and colleagues (2023) indicate that infants at this age do not show cardiac interoceptive sensitivity or whether methodological differences,such as the measurement method, outlier rejection criteria,and statistical power,might explain divergent results.In any case, the findings ofthe study by Weijsand colleagues(2023) highlight the importance of replicating the iBEATs paradigm (Maister et al. 2017) to advance the understanding of interoception in infants.

To gain a more comprehensive understanding of interoceptive processing in infancy, it iscrucial to investigate other interoceptive modalities. Especially, as different interoceptive modalities are not necessarily related and might have different functions or underlying neural signature (Allen et al., 2022; Garfinkel et al., 2016; Khalsa et al., 2018).One interoceptive signal that is closely related to cardiac processes is respiration. In fact, heartbeat and respiration are linkedfunctionally and anatomically(Draghici & Taylor, 2016; Garcia Ill et al., 2014).

In recent years, an increasing number of publications have focused on the perception of respiration in adults (Weng et al., 2021). Experimental paradigms that measure sensitivity to resistance in breathing, for instance, have highlighted the connection between respiratory interoception and emotional states such as anxiety (Harrison, Garfinkel, et al., 2021; Nikolova et al., 2021; Harrison, Köchli, et al., 2021). The neural network which links respiratory perception with emotional and cognitive processes has become a matter of scientific interest (Allen et al., 2022; Kluger et al., 2021). The emerging literature shows that breathing constitutes a fundamental process with functional significance for self-regulation (Boyadzhieva & Kayhan, 2021; Heck et al., 2017). In children, it has been shown that sensitivity to respiratory signals can be observed from at least 10 years of age onward (Nicholson et al., 2019). Given the relevance of respiration for self-regulation and social interaction, it is important to map out respiratory interoceptive sensitivity in infancy, a period that is especially relevant for development of self-related perception(Van Puyvelde et al., 2019; Weng et al., 2021).

Regarding the relationship between cardiac and respiratory interoception, empirical results have painted a mixed picture. In children it has been reported that respiratory interoception is not correlated to cardiac interoception (Nicholson et al., 2019).In adults it has been found that while accuracy in interoceptive domains is not related across cardiac and respiratory perception, meta-cognitive awareness for both modalities shows a significant relation(Garfinkel et al., 2016).In early infancy the relationship between different interoceptive modalities has not yet been investigated.Therefore, it is currently unclear whether sensitivity to different interoceptive modalities emerges at the same time, and in a similar manner.

In the present study, we aim to fill the knowledge gap on interoceptive sensitivity in early infancy byreporting results from two studies investigatingcardiac and respiratory interoceptive sensitivity in 3-,9- and 18- month-old infants. We investigated the age group of 3 months to provide evidence on the early emergence of interoceptive sensitivity, as 3 months is the earliest at which eye tracking paradigms can be reliably applied.We chose the groups of 9 and 18monthsas theseages mark important milestones in the development of abilities related to interoception, such asself-perception(Filippetti, 2021; Fotopoulou & Tsakiris, 2017; Musculus et al., 2021) and social

cognition(Carpenter et al., 1998; Repacholi & Gopnik, 1997). For instance, around 18 months of age explicit mirror self-recognition can be observed (Amsterdam, 1972).Moreover, around 9months of age, the ability to show joint attention drastically matures (Carpenter et al., 1998).

To measure cardiac interoceptive sensitivity, we used a modified version of the iBEATs paradigm originally developed by Maister and colleagues (2017), as it is the only task to measure cardiac interoceptive sensitivity in infants to date. Moreover, we developed a novel experimental paradigm that follows the logic of the iBEATs to investigate respiratory interoceptive sensitivity in infants: the iBREATH task. In contrast to previous tasks the present study used a fixed-experimental paradigm due to a technical error (i.e., the presentation sequence was consistent across participants; see Supplementary Materials A).In a first step, we conducted a study with a longitudinal design investigating 9- and 18-month-old-infants. We then replicated the experimental paradigms in a separate sample of 3-month-old infants.Our initial prediction concerning the 9- and 18-month-old sample were threefold: 1) For both tasks and in both age groups (9 and 18months), we expected to find the same preference in looking behavior as reported in Maister et al. (2017), that is, longer looking times to asynchronous trials as an indication of infants' detection of incongruency of the visual stimulus and their interoceptive signals. 2)We expected to find a positive correlation between performance in the cardiac and respiratory interoception tasksgiven the conceptual proximity between both tasks. Adult cardiac and respiratory interoception paradigms typically use two conceptually different paradigms. Thus, null results in the adult literature might be due to the unique characteristics of those paradigms. Last,3) we predicted an increase in individual interoceptive sensitivity from 9 to 18 months of age, as increased interoceptive sensitivity might be associated with the development of self-recognition and socio-cognitive skills.We tentatively predicted that 3-month-olds would already differentiate between visual displays that move in synchrony vs. asynchrony to their own cardiac and respiratory signals.

Results

Confirmatory Analyses: 9-month-old infants

First, we investigated whether 9-month-old infants displayed sensitivity to their cardiac and respiratory signals. Following our preregistered analysis plan (*https://aspredicted.org/QP9_6FP*) we computed paired t-tests to compare mean looking times between synchronous and asynchronous conditions for the iBEATs and the iBREATH. Wefound that for both tasks 9-month-old infants displayed a preference for stimuli presented synchronously with their own heartbeat (**Figure 1A**, N = 52, M_{synch} = 7020.62msSD_{synch} = 3790.00ms, M_{asynch} = 5496.7ms, SD_{asynch} = 2469.34ms,t = -2.96, p = .005, Cohens d = .48) and respiration (**Figure 1B**, N = 56, M_{synch} = 6336.21ms, SD_{synch} = 3017.37ms, M_{asynch} = 5483.77ms,SD_{asynch} = 2244.84ms,t = -2.80, p = .007, Cohens d = .32).These results on the one hand replicate the approach by Maister et al. (2017) in an older age group showing that infants are sensitive to their cardiac signals. Going beyond, using a novel paradigm, we further provide the first evidence that infants are also sensitive to their respiratory signals.Notably, mean preferences were switched compared to our expectations and the results of Maister et al. (2017) who reported a mean preference for stimuli presented asynchronously to the infants' heartbeats.

Interoceptive Sensitivity at 18months

Next, we followed up the same infants at 18months. Unfortunately, as the study was conducted during the Covid-19 pandemic, we had a large number of dropouts for the longitudinal follow-up. We conducted paired t-tests comparing looking times to synchronous and asynchronous stimuli at 18months following our approach of the 9-month-old sample for iBEATs (N = 28,M_{synch} = 6924.16ms SD_{synch} = 3833.01ms, M_{asynch} = 6427.68ms, SD_{asynch} = 3801.02ms,t(27) = -.75, p = .461, d =

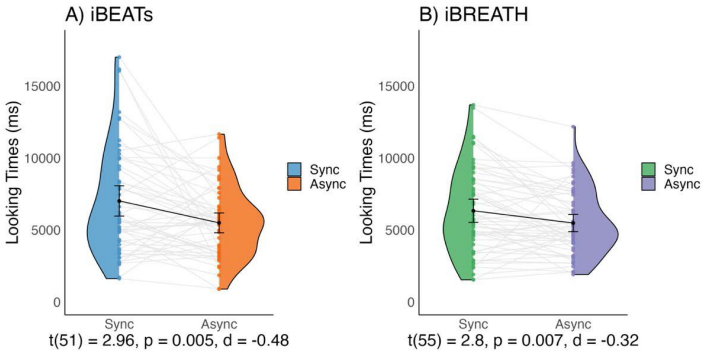


Figure 1.

Looking times for A) iBEATs and B) iBREATH.

Note. Looking times for the A) iBEATs and B) iBREATH tasks. In both tasks, 9-month-old infants looked significantly longer at stimuli presented synchronously to their own physiological signals. Black dots refer to the group mean. Black bars refer to the standard error of the mean. Grey lines and colorful dots refer to individual mean looking times per condition and infant.

.13) and iBREATH (N = 30,M_{synch} = 3612.95ms SD_{synch} = 1879.02ms, M_{asynch} = 4098.71ms, SD_{asynch} = 2074.11ms,t(29) = 1.09, p = .283, d = -.25) which did not indicate a significant mean preference. However, a non-significant result does not provide evidence for the absence of an effect (Lakens et al. 2017). Therefore, we conducted two equivalence tests using the effect size reported by Maister et al. (2017, d = .4) as equivalence bounds. Equivalence tests facilitate the interpretation of non-significant results by investigating whether a given confidence interval is too wide to discriminate between expected effect (= the equivalence bounds) and null effect, or whether one can rule out an effect at least as strong as we expected. The results of the equivalence tests indicate that we do not find conclusive evidence in favor of or against a mean preference effect in our 18-month-old sample for both the iBEATs(t(27) = .71, p = .242) or the iBREATH ((t(29) = 1.10, p = .141), potentially due to the small sample size.

When inspecting results from our analysis approach (e.g., **Figure 1**), as well as previous results (Maister et al. 2017, Weijs et al. 2023) it becomes evident that there are large individual differences in preferences (e.g., some infants prefer synchronous, some asynchronous trials). Thus, sample size might be an important factor in detecting a mean preference effect. To gain additional insights into the interplay of sample size and variability due to the large individual differences we conducted simulations which are reported in Supplementary Materials B. Overall, results from the simulation indicate that sample sizes of around 30 infants might be too small to reliably detect a mean preference effect in the version of the iBEATs task used here.

Interoceptive Sensitivity at 3months

Initially the present project was planned as a longitudinal approach spanning 3-, 9-, and 18months. However, difficulties in recruiting very young infants due to the Covid-19 pandemic precluded usfrom startingthe longitudinal assessment with 3-month-old infants. Still, we decided to test the iBEATs and iBREATH tasks in an additional 3-month-old sample once recruitment was possible again (pre-registration: *https://aspredicted.org/44L_QKH*). Data for this sample was collected after analysis of the 9- and 18-month-old data. Using our preregistered analysis approach, we found evidence for a group mean preference forsynchronous stimuli in the iBEATs (paired Bayesian t-test; BF = 2.02, mean difference: 793.95ms, 95% CI [108.63, 1388.69], N = 53, M_{synch} = 6131.13ms SD_{synch} = 5129.93ms, M_{asynch} = 5337.17ms, SD_{asynch} = 5044.83ms) but not in the iBREATH task (paired Bayesian t-test; BF = 0.23, mean difference: 502.21ms, 95% CI [-701.49, 1600.86], N = 40, M_{synch} = 7881.72ms SD_{synch} = 7641.50ms, M_{asynch} = 7379.50ms, SD_{asynch} = 7220.50ms) at 3months of age.Due to the absence of evidence for the iBREATHtask we conducted a test for practical equivalence similar to our approach for the 18-month-old'sdata(Harms & Lakens, 2018).We usedthe effect size of the iBREATH task at 9 months to investigate whether we can rule out an effect at least as strong as that. Results indicated that we cannot distinguish between absence or presence of an effect at least as strong as it was present in the 9-month-old's iBREATH sample(95%HDIs = [-711.41, 1606.80], region of practical equivalence: 77.08%). Reasons for the non-significant result might include the smaller sample size for the iBREATH at 3months (N = 40) compared to the iBEATs (N = 53), combined with a reduced signal to noise ratio for eye-tracking tasks in 3-month-olds compared to older infants, in general. In sum, wereplicate the results of our 9-month-old sample for the cardiac domain in 3-month-old infants, while finding inconclusive evidence regarding the respiratory domain.

Interoceptive Sensitivity in the first Two Years of Life – A MEGA Analytic Approach

So far, we have presented results on cardiac and respiratory interoceptive sensitivity spanning three age groups in the first two years of life. We find some evidence that infants prefer stimuli presented synchronously with their respective physiological signal. However, we also find some inconclusive evidence, such that we cannot distinguish between a null-finding and a significant effect.This might be potentially due to a small number of observations in some of our samples, as indicated by equivalence tests and data simulation. So far, we have investigated all age groups

separately, building up on our preregistration, and the assumption that results might be different for age groups. An alternative approach that might help in adjusting for sample size issues is to pool together all age groups using an explorative MEGA-analytic approach (Koile & Cristia, 2021). Such an approach might give us the statistical power needed to make claims about absence or presence of acohesive effect in the first two years of age, i.e., whether the mean effect across age groups supports the conclusion of a shared effect.

We computed two mixed models with looking time as outcome, condition and age-group/experiment, as well as their interaction, as fixed effects, and participant as a random effect using the R-package "glmmTMB"(Brooks et al., 2017) utilizing a beta error distribution for the iBEATs (**Figure 2A** bottom, **Table 1** & 2) and the iBREATH (**Figure 2B** bottom, **Table 3** & **4**), respectively. This approach allowed us to include 135 observations for the iBEATs from 125 infants, and 120 observations for the iBREATH from 107 infants. The sample size differs slightly from our preregistered approach given that we used the same preprocessing approach for the MEGA-analysis for all samples.

First, we compared each model with a null model missing the condition term. For the iBEATs we find that the full model is statistically significant from the null model, suggesting a better fit (p = .012).For the iBREATH we do not find a statistically significant better fit for the full compared to the null model (p = .091). Still, the Bayesian information criterion (BIC), which can be interpreted similar to a Bayes Factor(Burnham & Anderson, 2004), related to this comparison is 15.1 smaller for the full (BIC_{full} = -1574.5), compared to the null model (BIC_{null} = -1559.4), giving some evidence for a better fit for the full model.

Next, we inspected the model output. For both models we did not find a significant interaction between age and conditionindicating that the effect of condition on age group does not significantly vary between age groups (iBEATs: **Table 1**, iBREATH: **Table 2**). Next, we computed posthoc comparisons using estimated marginal means from the MEGA-analysis across all age groups to investigate whether we find indications for a similar effect across ages. For the iBEATs, we found a significant main effect of condition on looking time in the combined sample indicating that infants show longer looking times for stimuli presented synchronously with their heartbeat over all ages (OR = 1.13, 95% CI [1.03, 1.25], t(1769) = 2.541, p = .011). In contrast, for the iBREATH, we did not find a significant effect of condition on looking time over all ages (OR = 1.07, 95% CI [0.96, 1.20], t(1284) = 1.192, p = .234). Interestingly, we find that all samples and tasks apart from the 18-month-olds iBREATH sample show a numerical preference for synchronous stimuli. In reporting these results we focus on whether we found evidence for interactions between age groups, and whether we found evidence for a general effect across age groups. In-depth results and tables can be found in Supplementary Materials C.

To sum up, regarding cardiac interoceptive sensitivity,results from the MEGA analysis support the notion that,across all age groups tested here,infants on average prefer stimuli presented synchronously with their own heartbeat. Regarding respiratory interoceptive sensitivity, we only found evidence in our 9-month-old sample, but not in the 3- and 18- month-olds, or the MEGA analysis.However, this latter result might be driven by the 18-month-olds iBREATH sample.

The Relationship Between Cardiac and Respiratory Interoceptive Sensitivity

Next, we investigated the relationship between cardiac and respiratory interoceptive sensitivity. First, we computed absolute proportional scores as individual difference scores for the iBEATs and the iBREATH following previous approaches and our preregistration. These scores range from 0 to 1, and a higher score indicates a stronger preference for either synchronous or asynchronous stimuli in the iBEATs or iBREATH, respectively. However, a difference score does not indicate the

Figure 2.

Results from the MEGA analysis for A) iBEATs and B) iBREATH.

Note. Plot of difference scores computed as mean synchronous looking times minus mean asynchronous looking times per individual for each age group, as well as the combined sample. Red dots refer to mean effects for the respective analysis as described above, red bars refer to 95% confidence/credible intervals. Dashed line indicates a difference of 0. For 3, 9, and 18months age groups our preregistered analysis is plotted. For the combined sample we computed alinear mixed model using lme4 for visualization purposes as results from a mixed model with a beta error distribution cannot easily be transformedback to the original scale.

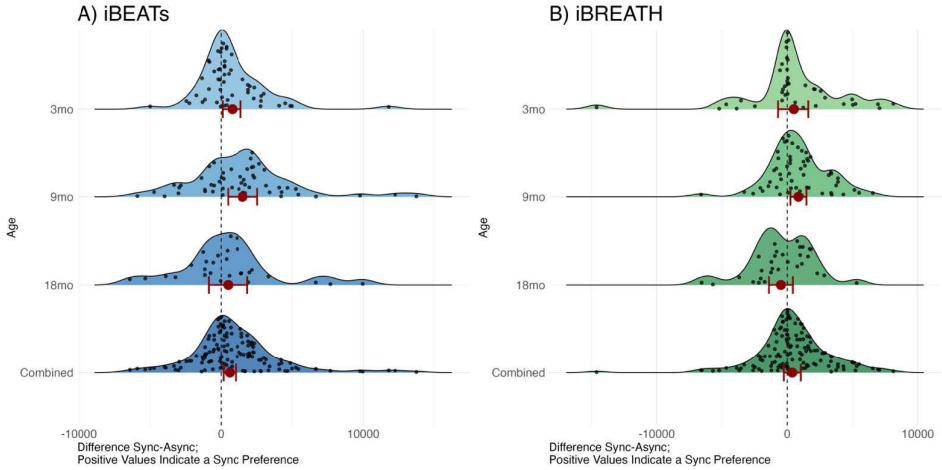


Table 1.

Interactions between condition and age for the iBEATs MEGA-analysis.

Term	Estimate	SE	z-value	p-value
Condition * age (3-vs 9 months)	0.00	0.11	0.02	.982
Condition * age(3- vs 18 months)	0.14	0.13	1.08	.283
Condition * age(9- vs 18 months)	0.14	0.12	1.12	.264

Term	Estimate	SE	z-value	p-value
Condition * age (3- vs 9 months)	0.02	0.12	0.17	.864
Condition * age (3- vs 18 months)	0.23	0.16	1.50	.134
Condition * age (9- vs 18 months)	0.21	0.15	1.43	.154

Table 2.

Interactions between condition and age for the iBREATH MEGA-analysis.

direction of the preference (synchronous or asynchronous). The reasoning behind the use of absolute proportional scores is that, in principle, both a preference for synchronous and for asynchronous stimuli indicates that the participant identified a (bodily) signal from noise. Importantly, all studies using iBEATs like paradigms in infants so far have used absolute proportional scores to investigate individual differences (Maister et al. 2017; Weijs et al. 2023). Further, visual inspection of the individual preferences in both paradigms (grey lines, **Figure 1**) reveals that, although the group mean difference displays a preference to the synchronous stimuli, in fact, looking preferences for both synchronous and asynchronous stimuli can be observed on an individual level.

We used a mega analytic approach, pooling together data from all age groups, to investigate the relationship between both tasks.We fitted a mixed modelusing a beta-error distribution with the iBREATH scores as outcome, the iBEATs, age, and their interaction as factors, and participant as a random factor (**Table 2**, for detailed results see Supplementary MaterialsC). We did not find a strong relationship between cardiac and respiratory interoceptive sensitivity across all ages (**Figure 4A**, N = 84), mirroring previous results in adults and children (Garfinkel et al., 2016; Nicholson et al., 2019). However, we found a significant interaction between theiBEATs scores and age, specifically comparing the 3- and 18-month-old groups (β = 3.13, SE = 1.41, p = .027). This interaction indicates that the relationship between iBEATs and iBREATH scores changes between 3 and 18months of age.

To follow up the interaction, we conducted a pairwise comparison which indicated that for the effect of iBEATs scores on the iBREATH scores there was a significant difference between 9- and 18-months of age (β = -0.60, SE = 0.24, p =.043), while there were no significant differences between 3- and 18-months (β = -0.60, SE = 0.25, p = .055) or the 3- and 9-month-olds (β = 0.00, SE = 0.23, p = .999). Still, coefficients indicate a similar strength and direction of the comparison between 9 and 18months as well as 3 and 18months.

Developmental Changes in Interoceptive Sensitivity

Next, we aimed at further investigating whether there are developmental changes in interoceptive sensitivity in the first two years of life. Initially, following our preregistration (*https://aspredicted.org/GMB_XCW*), weconducted a longitudinal analysis using the infants that participated both at 9 and 18months of age. Unfortunately, as described earlier, due to the study being conducted during the Covid-19 pandemic, only a subsample of infants could be re-invited to the lab in the targeted age range and contributed data suitable for longitudinal analyses.Comparing the absolute individual difference scores between both age groupsfound no evidence for a strong change in cardiac (paired Bayesian t-test; BF = 0.26, N = 20) or respiratory (paired Bayesian t-test; BF = 0.33, N = 19) interoceptive sensitivity, indicating that absolute individual difference scores in both domains do not change substantially from 9 to 18 months of age. Notably, a regions of practical equivalence follow-up analysis indicates that we cannot rule out an effect at least as strong as a change of .1 for the absolute proportional scores (iBEATs: ROPE [-0.10, 0.10], 97.53% inside ROPE, iBREATH: ROPE [-0.10, 0.10], 97.76% inside ROPE, 95% HDI [-0.11, 0.05]).Further, in an exploratory analysis we computed Spearman correlations between timepoints. We did not find evidence for the iBEATs (r(18) = .236, p = .315) and the iBREATH (r(17) = .195, p = .423) that individual difference scorescorrelatestrongly between timepoints.

To increase the number of observations, and statistical power, we conducted an exploratoryMEGA-analytic follow up in which we included all infants, not only those that contributed usable data to both timepoints. Results showed that individual difference scores increased significantly for the iBREATH (**Figure 4B**, **Table 7**) in the 18-month-olds compared to the 3-month-olds (OR = 0.544, SE = 0.12, p = .014), and the 9-month-olds (OR = 0.525, SE = 0.12, p = .004), but not for the iBEATs (**Figure 4A**, **Table 8**) indicating that respiratory, but not cardiac, interoceptive sensitivity increases at 18 months of age.

Figure 3.

Relationship between iBEATs and iBREATHusing a combined sample.

Note.Histogram with plotted line for individual performance on iBEATs and iBREATH using a beta regression. Following Maister et al. (2017), individual difference scores were computed as proportion of absolute difference between synchronous and asynchronous trials.

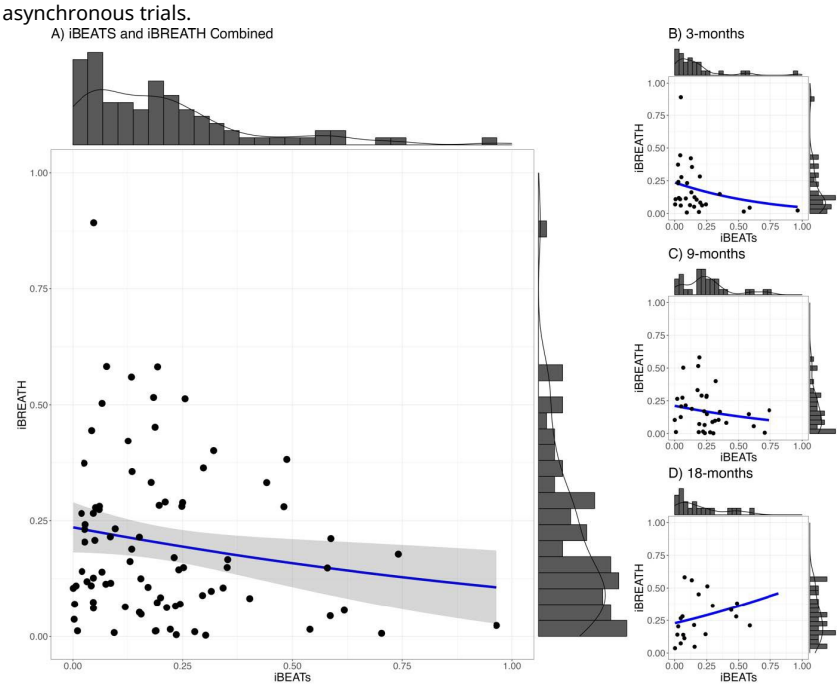


Table 2.

Effects of iBEATs on iBREATH for all age groups, as well as for interactions between iBEATs and age.

Term	Estimate	SE	z-value	p-value
iBEATs (3-months)	-1.83	0.97	-1.89	.059
iBEATs (9-months)	-1.16	0.90	-1.30	.192
iBEATs (18-months)	1.30	1.02	1.27	.204
iBEATs * age (3- vs 9 months)	-0.15	0.34	-0.42	.674
iBEATs * age (3- vs 18 months)	3.13	1.41	2.22	.027
iBEATs * age (9- vs 18 months)	2.45	1.36	1.81	.070

Figure 4.

Exploratory analysis for age effect.

*Note.*Absolute proportional scores for A) iBEATs and B) iBREATH plotted for each age group. Red dots refer to group means, and colorful dots to individual means.

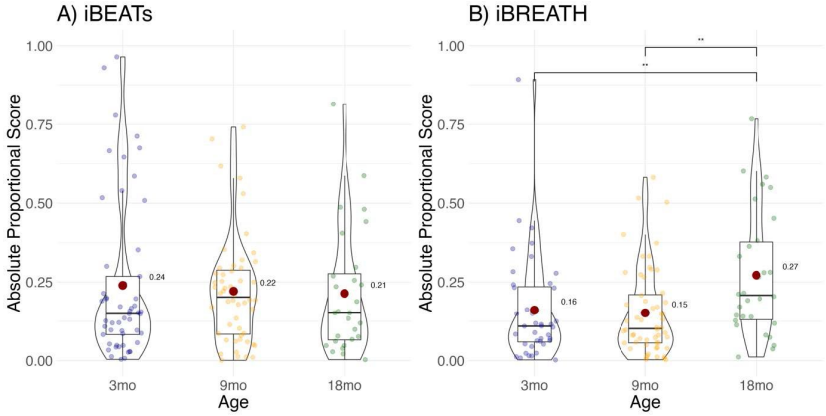


Table 7.

Change in absolute proportional scores across age groups for the iBEATs.

Term	Estimate	SE	z-value	p-value
Intercept	-1.11	0.14	-7.91	< .001
9-months	-0.04	0.19	-0.22	.826
18-months	-0.09	0.23	-0.41	.684

Term	Estimate	SE	z-value	p-value
Intercept	-1.61	0.15	10.55	< .001
9-months	-0.04	0.19	-0.18	.853
18-months	0.61	0.21	2.86	.004

Table 8.

Change in absolute proportional scores across age groups for the iBREATH.

Specification Curve Analysis

Notably, apart from the 18-month-old iBREATH sample,we found that (numerical)mean group preferences indicated a preference to stimuli presented synchronously with the respective bodily signal. Thus, mean group preferences were switched compared to our initial expectation and the original study by Maister et al. (2017) who found a mean group preference for stimuli presented asynchronously to the infant's heartbeat.In addition, other studies have failed to find evidence for cardiac interoceptive sensitivity in infants (Wejis et al. 2023). Further, a wide range of analytical choices have been reported in approaches on cardiac interoception in infants (Maister et al., 2017; Weijs et al., 2023) and nonhuman primates (Charbonneau et al., 2022) to date.

Therefore, it is important to further describe and validate our results. Using a specification curve analysis (Simonsohn et al., 2020), it is possible to map out the space of theoretically justified analysis strategies on a given dataset. Thus, it is possible to investigate whether analytical choices, such as differences in exclusion criteria and physical signal preprocessing, impact the results. Importantly, this method allows us to rule out that a group mean preference for synchronous stimuli is due to specific analytical choices of our preregistered analysis or whether a range of different analysis paths come to the same conclusion.

We ran a specification curve analysis following the approach outlined by Simonsohn et al. (2020). We used our 9-month-olds dataset as input dataset, as it shows the clearest evidence for infant interoceptive sensitivity (i.e., better data quality compared to the 3-months sample, and larger sample size compared the 18-months sample). First, we identified theoretically justified analysis paths applicable to the present dataset by comparing the approaches presented in Weijs et al. (2023), Maister et al. (2017) and Charbonneau et al. (2022). As the first step, we focused on the iBEATs, as already 3 different research groups have published experiments similar to the iBEATs and thus, a number of different specifications could be extracted from the literature (e.g., regarding physiological signal processing, in/exclusion criteria for infants and number of trials, or statistical analysis; for a full list see Supplementary Materials D). Combining all possible ways of analyzing the present dataset gave a number of 1024 possible analyses which we subsequently ran (**Figure 5A**). Next, we ran a specification curve for the iBREATH data of our 9-month-old sample by extracting and adapting analytical decisions we used for the iBEATs, which resulted in 1536 possible analyses (**Figure 5B**).

Our results indicated that for the iBEATs almost half (44.73%) of all analytical paths led to a significant result ($p < .05$), while for the iBREATH 17.51% of all analytical paths came to such a conclusion. Almost all specifications indicated a mean group preference for synchronous trials (43.16%, **Figure 5**, blue color). Interestingly, however, there were also a handful of specifications for the iBEATs ($n = 16$, 1.6%) that would have resulted in a mean group preference for asynchronous trials (**Figure 2A**, red color). In sum, these results can be seen as a validation of our preregistered analytical approach described above, as they highlighted that a mean group preference for synchronous trials is not dependent on the combination or interaction of specific analytical choices. Still, given that 1.6% of analysis paths would have come to a different conclusion might indicate that the influence of analytical choices is not completely negatable.

Discussion

In the present study we investigated cardiac and respiratory interoceptive sensitivity in 3-, 9-,and 18-month-old infants utilizing a preregistered approach, validated by a specification curve analysis, and MEGA analytic approaches. Regarding cardiac interoceptive sensitivity we found evidence fora preference for stimuli presented synchronously in all three age groups. Regarding

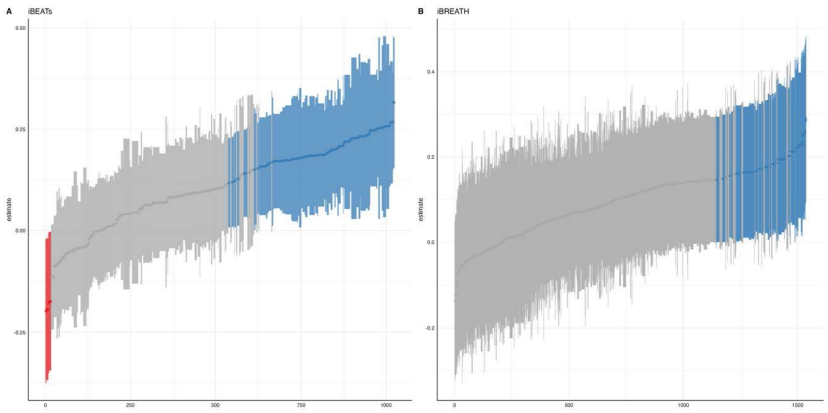


Figure 5.

Specification Curve Analysis for the A) iBEATs and B) iBREATH task.

Note. Specification curve analysis plotting standardized beta regression coefficients (y-axis) and number of analysis (x-axis) for A) iBEATs and B) iBREATH. Number of analysis (x-axis) are ordered increasing from lowest to highest standardized beta regression coefficient. Blue color indicates a significant effect (p<. 05) for a mean synchronous preference, red color indicates a significant effect (p< .05) for a mean asynchronous preference, and grey indicates a non-significant outcome

respiratory interoceptive sensitivity we find a more nuanced picture with infants showing a significant preference for stimuli presented synchronously at 9-months of age, but not at 3- and 18-months.We did not find strong evidence for a relationship between cardiac and respiratory interoceptive sensitivity in infants in the first year of life. However, we find some evidence for a positive relationship at 18months. Further, in an exploratory analysis we find indications that respiratory perception increases between 9- and 18-months. However, due to the small sample size at 18months the results regarding changes and stability of interoceptive sensitivity in the second year of life must be considered speculative and need to be validated in further research.

In recent years, new theoretical frameworks have deepened our understanding of interoceptive processing (Murphy et al., 2020; Suksasilp & Garfinkel, 2022). For instance, in their 2x2 model, Murphy et al. (2020) distinguished between two main factors of interoception – interoceptive accuracy (i.e., how exact one perceives internal bodily signals) and interoceptive attention (i.e., how often one thinks of internal bodily signals in everyday life). When applying the Murphy model to the iBEATs and the iBREATH, both aspects might be needed to show a preference. First, it is necessary to access one's own internal bodily signals to notice a difference between synchronous and asynchronous signals (i.e., interoceptive accuracy). Second, one also needs to pay attention to one's own bodily signals and compare them to what is happening on the screen (i.e., interoceptive attention). Thus, it is possible that the present task does not distinguish between both dimensions. Instead, the present task might measure a propensity to engage with own interoceptive signals (Murphy, 2023). In fact, when considering the potential impact of interoceptive sensitivity in real-world settings it is unlikely that "pure" interoceptive accuracy or attention can be differentiated, but that the interplay of both shapes the outcomes.

We do not find evidence for a strong relationship between cardiac and respiratory interoceptive sensitivity in the firstyear of life. This finding is in line with empirical results not finding a relationship between cardiac and respiratory interoception in adults (Garfinkel et al., 2016) and children (Nicholson et al., 2019). Further, these findings might be explained by accounts proposing different brain networks for processing of cardiac and respiratory information (Suksasilp & Garfinkel, 2022). Still, we find a relationship between cardiac and respiratory signals that in the oldest sample tested here, the 18-month-olds, which is closest to adults. Although this effect needs to be interpreted with caution due to the small sample size, this might indicate that using conceptually similar experimental paradigms might be a promising avenue to investigate relationships between different interoceptive modalities in adults.

To investigate individual differences, we used absolute proportional scores, following previous approaches (see **Figure 5**, Maister et al., 2017; Weijs et al., 2023). As a preference in any direction in the iBEATs or the iBREATH task can, in principle, be considered as evidence for the participant's ability to distinguish their own bodily signals from noise. However, it remains an open question if individual looking preferences for synchronous or asynchronous stimuli have a functional importance. In other studies,investigating infants' processing of information about body ownership, preferential looking paradigms similar to the iBEATs and iBREATH have been used. For instance, newborns prefer to look at synchronous visuo-tactile cues compared to asynchronous ones (Filippetti et al., 2013), similar to 7- and 10-month-olds (Zmyj et al., 2011). In other cases, older infants showed a looking preference for sensorimotor-incongruencies (Rochat, 1998). Furthermore, at 5 months of age infants recognize delays in visualization of their own leg-movements and prefer to look at stimuli that are asynchronous to their own movements (Bahrick & Watson, 1985). Thus, previous results are inconclusive as to whether infants generally prefer to look at stimuli that are synchronous or asynchronous to their own bodily movements and experiences. Yet, there is convincing evidence that they detect such (in-)congruencies. For instance, 14-month-old infants are more likely to help a person that had previously bounced in synchrony with them, compared to an asynchronously bouncing person (Cirelli et al., 2014).

Longer looking times for synchronous stimuli might indicate a familiarity preference (or more generally a preference for a stimulus that is easier to process).In this context familiarity might refer to the infant's perception of congruence between internal signal and external stimuli which might drive the infant's attention. Specifically, the synchronous condition should be easier to process due to the intersensory redundancy and predictability between interoceptive and exteroceptive signals. Longer looking times for asynchronous stimuli might indicate a novelty preference, that is,a preference for a stimulus that offers a learning opportunity (Hunter & Ames, 1988). According to the framework presented by Hunter and Ames (1988), the preference for novelty or familiarity depends on three factors that interact with each other: familiarization, age, and task difficulty. In short, the model proposes that less familiarization, a lower age, as well as an increase in task difficulty, facilitates a familiarity preference. However, it is important to consider that other cognitive and attentional mechanisms could also influence these responses.

Thus, when applying the framework to the present results it might be that certain details contributed to a familiarity preference displayed by most infants (as indicated by the synchronous preference). For instance, the data presented here at 9- and 18-months was collected as part of a larger study with several other paradigms, whereas the study of Maister et al. (2017) used the iBEATs as the first task of the session. Thus, this increased complexity for the infant in our setting might have impacted the task difficulty and potentially reduced the familiarization. Further, at 3-months of age the experimental setting might be more challenging thus leading to an increased complexity. However, the interpretation of looking time preferences in infancy research in general and the Hunter and Ames (1988) framework specifically, remains a topic of debate and further research(Bergmann et al., 2019; ManyBabies 5 Team, 2023).

Nevertheless, the switch in mean preference reported here regarding cardiac interoceptive sensitivity, compared to Maister et al. (2017), might also indicate a development around 5 months of age.Infants at 5 months of age might be more drawn to asynchrony between their cardiac signals and visual stimuli, while 3- and 9-month-old infants on average prefer synchrony. Such a developmental trajectory might also explain the null findings reported by Weijs et al. (2023), as infants tested in their study were in between 5 and 7 months of age. It is also possible that there are developmental windows in which the perception of bodily signals plays an important role. For instance, age groups in the present study were chosen to be in a similar range as the emergence of theoretical relevant constructs, such as mirror self-recognition. However, to disentangle such effectsadequately powered longitudinal studies are needed.

To validate the impact of analytical choices on mean group preferences we used a specification curve analysis. In the following, we will discuss impactful decisions and make recommendations for future approaches (see **Table 9**).Regarding analytical choices that had an impact on the results, we found that applying the same physiological data rejection criteria to synchronous and asynchronous trials led to more significant results (**Table 9**, 1st entry). The logic behind not removing asynchronous trials with physiological artifacts in the tasks described here is that in these trials the signal is not generated by real-time feedback of the physiological signal. Thus, it is not directly relevant for stimulus presentation. However, our results indicate that applying differing criteria for both trial categories might obscure effects.

Moreover, we found that for both tasks, in terms of physiological artifact rejection, including more data points led to more significant results (**Table 9**, 3rd entry). This might be explained by the inclusion of more data, and thus, greater statistcal power. For instance, in the iBEATstask,strict artifact rejection means that a trial is removed once a single R-peak is not (or falsely) detected. However, in such trials it might still be possible to recognize that the stimulus presentation is synchronous or asynchronous to one's heartbeat and it thus still holds information relevant for the task. For future studies, we would recommend a more fine-tunedapproach for removing trials based on physiological artifacts. Furthermore, we would advise employing the same criteria to all conditions used.

Regarding specifications that did not have a strong impact,we found that outlier criteria using standard deviationshad a negliable impact on the results (**Table 9**, 2nd entry). Such criterions are usually applied to remove extreme values in the data.In the paradigms described here, looking times were bound by trial length (e.g., in the iBEATs max. 20 s).Thus, rejecting trials based on standard deviation might not be useful in analyses of preferential looking paradigms that use maximum trial length. One reason might be that extremely large outliers in looking times are impeded already by the experimental design. We also did not find that an inclusion criterion regarding a minimal number of valid trials an infant had to contribute to be included in the analysis changed the number of significant results much.Such criterions are typically used to increase the reliability of results, as individual trial outliers' weight stronger when an infant only completes few trials. For instance, in our preregistered analysis, infants had to complete a minimum of 8 trials for the iBEATs or 4 trials for the iBREATH to be included in the analysis (**Table 9**, 6thentry).However, there were few infants who completed lessthan these minimum number of trials. For future approaches, we would advise against using exclusion criteria based on standard deviations or number of trials. Moreover, the statistical test used (paired t-test vs linear mixed model, **Table 9**, 7th entry) had a rather small impact on the results. However, given the large number of analyses conducted, this might be related not being able to precisely formulate the model to fit the complexity of the data for each specification.

Overall, the recommendations outlined above can be discussed within the scope of a fundamental challenge in experimental research– how to balance noise in a given dataset with losing statistical power by exclusion of participants and trials. This is especially relevant in infancy research that oftentimes deals with high drop-out rates and noisy datasets. For the present dataset, we find that leaning on the side of including more data points (e.g., regarding rejection of physiological artifacts, or exclusion criteria) might be more beneficial as long as the same criteria are applied to all data. Thus, exclusion of data points should be driven by trying to minimize the impact of erroneous or random datapoints, while still keeping those that have interesting characteristics (Leys et al., 2019). We want to stress that outlier criteria should ideally be formulated within a preregistration (Bakker & Wicherts, 2014).

Overall, we found more significant results for a group mean preference for the iBEATs (44.73%) compared to the iBREATH (17.51%) at 9 months of age. Given that our exploratory analysis indicated an increase of iBREATH difference scores from 9 to 18 months, respiratory interoceptive sensitivity might develop in this age range.However, it is also possible that the coupling of physiological signals with visual stimuli in infancy might produce stronger mean preferences for cardiac-, compared to respiratory signals. In sum, the results of the specification curve analysis validated our preregistered analysis, as almost all analysis paths resulted in anumerical mean preference for synchronous stimuli.

Ideas and Speculation – Development of Respiratory Interoceptive Sensitivity

While we had found consistentevidence for cardiac interoceptive sensitivity,whereby infants on average prefer stimuli presented synchronously with their heartbeat, the evidence regarding respiratory interoceptive sensitivity was more nuanced. In particular, the 18-month-olds sample for the iBREATH displayed three interesting characteristics: it was the only sample showing a (numerical) preference for the asynchronous condition, absolute proportional scores increased compared to 3, and 9months, and there was a positive relationship with cardiac interoceptive sensitivity scores at 18months (but not at 3 or 9months). To interpret these results, one might speculate that a maturation of respiratory interoceptive sensitivity towards 18monthsof age takes place.A hypothesis to be tested in future research is that developmental improvement in respiratory perception might be related to increases in other domains that show links to interoception. For instance self-perceptionmatures towards the second year of life and has been

Specification	iBEATs	iBREATH
1. Outlier rejection	128, only async	111 only async
	330, both	158 both
2. SD outlier rejection	107, no rejection	59 no rejection
	107, 2SD	70 2SD
	120, 2.5SD	104 2.5SD
	124, 3SD	36 3SD
3. Artifact trial rejection	224, 85% criterion	207, large artifacts
	162, small artifacts	included
	included	62, small artifacts included
	72, strict rejection	0, strict rejection
4. Data transformation	240, log transformed	162 log transformed
	218, not transformed	102 not transformed
5. Trial removal	308, 0LTs included	183, 0LTs included
	150, 0LTs excluded	86, 0LTs excluded
6. Min. number of trials	115, min. 2 trials	86, min. 2 trials
per id to be included	115, min. 4 trials	74, min. 4 trials
	113, min. 8 trials	24, min. 8 trials
	115, no criterion	85, no criterion
7. Statistical analysis	220, linear mixed model	89, linear mixed model
	238, paired t-test	180, paired t-test

Table 9.

Number of significant results for specifications for iBEATs and iBREATH

conceptually related to interoception (Fotopoulou & Tsakiris, 2017; Musculus et al., 2021). Further, gross motor development may be considered in future research, which drastically matures in the first two years of life(WHO Multicentre Growth Reference Study Group, 2006) has been shown to be related to respiratory function in children with cerebral palsy(Kwon & Lee, 2014). However, the result and interpretation warrant further follow-up given the small sample size of the 18-month-olds and exploratory nature of the respective analysis.

Limitations

The data presented in this study holds several limitations.First, due to an error in our experimental scripts we unintentionally used a fixed-order design, in which almost all infants saw the same fixed order of condition (always starting with a synchronous trial), image assigned to condition, and location of the image (left/right) instead of a semi-randomized/randomized design. Such a fixed-order design holds several important limitations as visual preferences might be influenced by the experimental design, i.e., the first trial always being synchronous might have influenced a mean group preference. Further, we cannot rule out that mean group preferences were influenced by the stimuli used (as in most cases the same stimuli were used for synchronous/asynchronous trials) or by the location of the image in a given trial (left/right). Still, there is no strong theoretical argument as to why image used or location should have an impact on infants' preferences. The stimuli were selected to be similar to each other, in order not to evoke a pioripreferences. To further illustrate the impact of the fixed order design we have conducted several additional analyses, which can be found in Supplementary Materials A, which do not indicate that there was an impact of the fixed-order design. Specifically, we find no evidence for systematic differences between infants tested with the fixed design and infants tested with a randomized design.

Despite these limitations fixed-order designs also holdadvantages, as they are more suitable to investigate individual differences (Dang et al., 2020; Hedge et al., 2018). When each participant is exposed to the same procedure, individual differences are less likely to be attributed to effects of randomization but are more likely to reflect real differences between participants. Also, when considering the impact of the randomization, one must consider our results in relation to earlier studies (Maister et al. 2017, Weijs et al. 2022, Imafuku et al. 2023), some of which used the exact same stimuli as we did (Maister et al., 2017),withsemi randomized designs. Results of these studies indicate no looking times differences depending on the stimulus assigned to each condition or systematic preferences for one of the stimuli.

Further, drop-out numbers must be discussed. For the 9-month-old sample ninety mother-infant dyads were invited to take part in the present study, but only 74/75 provided data for iBEATs and iBREATH, respectively. Further, only 52 (iBEATs) and 56 (iBREATH) could be included in the confirmatory analysis based on the predefined exclusion criteria, and only 34 contributed usable data for both paradigms. This might also be attributable to the paradigms being embedded in a data collection for a larger project.Similar, for the 3-month-old sample 80 infants were invited to the lab, however only 53 (iBEATs) and 40 (iBREATH) could be included in the analysis. Also, the 9- and 18-month-old samples were collected during the Covid-19 pandemic, which led to high dropout rates for the longitudinal follow-up at 18months, as lockdowns and Covid-19 cases made data collection challenging.Thus, we might not have hadsufficient statistical power to detect possible effects using our longitudinal sample. The reduced sample size might have impacted the statistical power to detect mean preferences for some age groups. Still, it must be noted that even the smaller sample sizes included were of similar size as used in previous studies on infant interoceptive sensitivity (Imafuku et al., 2023; Maister et al., 2017; Weijs et al., 2023).

To overcome some ofthese limitations, we have computed exploratory analysis using all data available, not just those infants that contributed data at both timepoints. However, such an approach can only provide correlational evidence.Regarding the specification curve analysis, it is

possible that there are specifications that might be relevant, which were not considered here. Furthermore, in the specification curve analysis, we did not inspect assumptions underlying the statistical tests in-depth.

Conclusion

To sum up, we present compelling evidence that infants are sensitive to their own cardiac signals in the first two years of lifeusing an adapted version of the paradigm introduced by Maister et al. (2017). Moreover, we present the first evidence that infants are sensitive to their respiratory signals using the iBREATH paradigm. By using a preregistered approach,a comparably large sample size and age rangespanning the first two years of life, and by extending the interoceptive modality assessed to respiration,we provided important empirical evidence for theoretical accounts highlighting the relevance of interoceptive sensitivity in infancy.Regarding longitudinal development, we found no evidence for a change of interoceptive sensitivity in our confirmatory longitudinal analysis. However, exploratory analysis using a between groups approach revealed evidence for an age-related increase in respiratory, but not cardiac, interoceptive sensitivity scores towards 18-months-of age.We did not find that cardiac and respiratory interoceptive sensitivity are strongly related, mirroring results in adults and children. However, we find exploratoryevidence for a relationship at 18-months.

We used a specification curve analysis to validate our results and showed that a specification curve analysis is a suitable tool to investigate the impact of analysis choices in infancy research. Finally, we provided guidelines for the analysis of the two paradigms presented here, as well as for preferential looking time paradigms in general.Overall, our results demonstrate that infants' interoceptive sensitivity, measured through coupling a visual presentation to a physiological signal, is a replicable phenomenon, that can be generalized to different age groups,as well as to different interoceptive modalities. By providing empirical results that go beyond previously published studies on infant interoception, our results give an important empirical basis for theoretical approaches targeting interoception during development, as well as related constructs such as self-perception, in early infancy.

Materials and Methods

Sample

The data reported here was collected as part of a larger project involving a range of other measures. To stay coherent, we refer to each age group throughout the manuscript with regard to the lower end of the age range in which we included infants (e.g., we tested infants between 9 and 10 months, but refer to them as the 9-month-old group). For the 9-month-old sample in total, 90 infant-mother dyads were tested in the laboratory. Initially, we intended to invite mother-infant dyads when the infant was 9 to 10months of age. However, as this study was conducted during the Covid-19 pandemic, we extended the age range to 10-months and 15 days to be able to include a sufficient number of infants (M_{age}= 301.63 days,SD_{age} = 10.57). We followed up the same sample again when the infantswere 18–20 months of age (N = 54, M_{age} = 576.65 days,SD_{age} = 14.49).Data collection took place during the Covid-19 pandemic, from September 2020 to September 2021. The total sample size was based on a power analysis for an unrelated analysis. However, building up on the results reported by Maister et al. (2017; paired t-test; t = 3.267, n = 29, Cohen's d = .4), the study would have been adequately powered to detect an effect approx. 30% (Cohen's d = .3) smaller than reported by Maister et al. (2017).The 3-month-old sample was tested after completion of the 9- and 18-month-old samples. Initially, we had planned to start data collection

with the 3-month-old sample. However, due to the Covid-19 pandemic this was not possible. We invited 80 infant-mother dyads to the lab when the infant was 3–4 months old (M_{age}= 113.53 days, SD_{age} = 7.82).

Participants were recruited from an existing database of volunteer families and parents. We strived to include an equal number of boys and girls. All infants were born full term with normal birth weight and had no known developmental delays or neurological impairments. Experiments were approved by the ethics committee of the University of Vienna (reference no. 00504).

Experimental Procedures

Upon arrival in the laboratory, primary caregivers were asked to fill out an informed consent form. After a warm-up period, the infants performed several tasks in randomized order. In the current manuscript, we only report results from the iBEATs and the iBREATH tasks, as results from the other tasks will be presented in separate reports. The order of the tasks was counterbalanced across participants. As both the iBEATs and the iBREATH followed a similar structure and required similar equipment, the tasks were performed back-to-back in an alternating order.Between the iBEATs and the iBREATH, we additionally acquired 3-minutes of resting state data to analyze cardio-respiratory coupling while infants watched a neutral video. The procedure was the same for infants from all age groups. Notably, the infants participating at 3-months only did the iBEATs and iBREATH in alternating order.

iBEATs

To measure cardiac interoceptive sensitivity, we used the iBEATs paradigm (Maister et al., 2017). Three electrodes were attached to the infant's chest in a three-lead setup. We used an ADInstrumentsPowerlab and BioAmp equipment to monitor and to record cardiac activity (*www.adinstruments.com*). To identify R-peaks, we used the built-in hardware-based function, namely "fast response output", which sends a pulse to a presentation-computer via a custom-made Arduino set-up, once a predefined threshold is reached. The threshold was set individually for each infant.

Upon the placement of ECG electrodes and the adjustment of the fast response output, infants were placed in an infant chair roughly 60 cm away from an eye tracker sampling at 500 Hz (Eyelink 1000 plus). The caregiver was asked to sit right behind the infant. In case the infant got fussyor did not tolerate the infant chair, we offered the option to place the infant on the caregiver's lap. Following a 3-point calibration, infants were presented with trials in which visual stimuli (i.e., either ayellow cloud or a pink star) moved rhythmically up or down on the screen, either synchronous or asynchronous to the infant's heartbeat. Movements of the stimuli were accompanied with a jumping sound to attract infants' attention. In synchronous trials, the movement of the stimulus on the screen was coupled to each infant's R-peak. For asynchronous trials, first, mean inter-beat-interval of the preceding synchronous trial were computed for each infant. Movement of the stimuli then followed a predetermined rhythm that was either 10% faster or slower than the average inter-beat-interval of the last synchronous trial for that infant.

There was a maximum of 80 trials in the task. The first trial was always synchronous.Before each trial, an attention getter was displayed. Once the infant looked at the screen, a trial started lasting for a minimum of 5 seconds and a maximum of 20 seconds. After the initial 5 seconds, the duration of the trials was infant controlled. The ongoing trial automatically terminated, and the next trial started, if the infant looked away from the screen longer than 2 consecutive seconds or the maximum trial duration of 20 seconds was reached. The task was terminated, if the infant looked away from the screen longer than four consecutive trials (i.e., a total duration of 20 consecutive seconds) or the infant became fussy or tired.

Initially, we had intended to counterbalance the stimuli across experimental conditions and infants. However, when conducting additional analysis during the review process we noticed an error in our randomization scripts (for a detailed description and additional analyses regarding its impact see Supplementary MaterialsA) which led to the iBEATs following a fixed-order experimental design for most participants across all age groups. Thus, for all infants apart from a small subsample of 9-month-olds randomization was fixed, so that the same stimulus was synchronous or asynchronous, with a fixed order of locations (stimuli appeared left and right) and a fixed trial order. However, the randomization was not completely alternating, so that in some cases two synchronous or asynchronous trials could follow each other. The stimulus presentation was performed using a custom-made script in MATLAB (Matlab 2018b).

iBREATH

To measure respiratory interoceptive sensitivity, we developed and used the iBREATH paradigm, which followed a similar logic to the iBEATs task. A respiratory belt connected to an ADInstrumentsPowerlab was attached to the infant's torso (*www.adinstruments.com*). Once a stable signal was obtained, infants were seated in an infant chair roughly 60 cm away from an eye tracker sampling at 500 Hz (Eyelink 1000 plus). The caregiver was asked to sit right behind the infant. The signal of the respiration belt was sent to a presentation computer using a custom-made Arduino set-up.

Similar to the iBEATs procedure, during a 3-point calibration, infants observed moving circles accompanied by a sound. Following calibration, infants were presented with an infant-friendly neutralstimulus (i.e., a red strawberry or agreen apple), which increased and decreased in size, either synchronous or asynchronous to that infant's respiratory rhythm. Stimuli presentation was accompanied by an infant-friendly sound. The volume of the sound was adjusted in relation to the size of the stimuli, thus, increasing and decreasing as the image got bigger and smaller, respectively. In synchronous trials, the stimuli on the screen expanded and shrank in synchrony with each infant's respiration rhythm. In asynchronous trials, movement of the stimulus was either 10% faster or slower than the average breathing frequency of the last trial for that individual infant.

To generate asynchronous trials, two components of the immediately preceding synchronous trial were used to compute a sinusoidal signal that was either 10% faster or slower than the signal in the previous trial. First, the average breathing frequency of the last trial was extracted, which was either speeded up or slowed down based on the asynchronous trial type. Then, the average respiratory amplitude in the last synchronous trial was extracted, which was used to set the amplitude of the asynchronous trial. By combining frequency and amplitude, the sinusoidal signal was created.

The iBREATH paradigm consisted of a maximum of 80 trials. The first trial always was synchronous. Before each trial, an attention getter was displayed. A trial was displayed for a minimum of 5 seconds and a maximum of 30 seconds. Following the initial 5 seconds, the duration of the trials was infant controlled. An ongoing trial was terminated automatically, and the next trial started, if the infant looked away from the screen longer than 2 consecutive seconds or when the maximum trial duration of 30 seconds was reached. The task was terminated, if the infant looked away from the screen longer than four consecutive trials (i.e., a total duration of 20 consecutive seconds) or when the infant became fussy or tired.

Initially, we had intended to counterbalance the visual across experimental conditions and infants. However, when conducting additional analysis during the review process we noticed an error in our randomization scripts (for a detailed description and additional analyses regarding its impact see Supplementary Materials A) which led to the iBREATH following a fixed-order experimental design for most participants across all age groups. Thus, for all infants apart from a small

subsample of 9-month-olds randomization was fixed, so that the same stimulus was synchronous or asynchronous, with a fixed order of locations (stimuli appeared left and right) and a fixed trial order.

Confirmatory Analysis

This study was preregistered on aspredicted.org. The preregistration for the 9-month-old sample can be accessed here: *https://aspredicted.org/QP9_6FP*. The preregistration for the longitudinal analysis can be assessed here: *https://aspredicted.org/GMB_XCW*. The preregistration for the 3-month-old sample can be accessed here: *https://aspredicted.org/44L_QKH*. Data, analysis-, and experimental scripts are available here:*https://osf.io/jy5fe/?view_only=6199b7c7e7f34599a10ccaf25d5e33d8*.

Pre-processing

In a first step, we visually inspected each trial of the iBEATs and the iBREATH tasks to exclude trials in which stimulus presentation was impacted by technical problems or physiological artifacts. We excluded trials for technical problems if transmission of the physiological signal was interrupted during a trial (e.g., an electrode was removed, a cable got unplugged etc.) or stimulus presentation was interrupted (e.g., there was a problem in connecting to the stimulus presentation computer).

Next, we excluded trials with physiological artifacts. In the iBEATs, we excluded a trial if not all R-peaks were picked up by the fast-response-output. In the iBREATH, we excluded a trial if movement or other technical artifacts were visible in the respiratory signal during a trial. Furthermore, in the iBEATs, infants were included if they completed a minimum of 8 trials. In the iBREATH, we adapted this criterion as respiration is a slower signal than the heartbeat and maximum trial durations were longer. As this might result in fewer total number of trials in the iBREATH task as compared to the iBEATs task, we adjusted the cut-off number for the iBREATH task and included data of infants who completed a minimum of 4 trials in the analysis. For the longitudinal analysis, we used a less strictcriterion to increase our potential sample size as outlined in our preregistration. Thus, infants were included when they completed at least 4 trials in either task.

Pre-processing of looking times-data

We defined an area of interests (AOIs) based on the maximum coordinates of the animated character on the screen. We took the maximum movement range of the animated character and computed looking times in each trial as the summed duration of all eye-tracking samples falling in that AOI. Because we aimed to replicate the study by Maister and colleagues (2017), we followed the same analysis approach as they did in the original paper. Accordingly, we excluded trials with looking times two standard deviations away from the condition's (i.e., synchronous or asynchronous trials) group mean. To compare cardiac and respiratory interoceptive sensitivity, we computed individual discrimination scores defined as the absolute proportion of looking time difference between synchronous and asynchronous conditions, again following the procedure by Maister and colleagues (2017). For both tasks we excluded trials with looking times of 0, as it is not clear whether infants did not look at the screen in these trials, or whether there were technical issues in these trials.

Statistical analysis

All statistical analysis reported here were computed in R (R Core Team, 2022) using the packages "pwr"(Champely, 2020), "TOSTER" (Lakens et al. 2018), "ggstatsplot"(Patil, 2021), "BayesFactor"(Morey & Rouder, 2022), "specr"(Masur & Scharkow, 2019), "lme4"(Bates et al., 2015), "afex"(Singmann et al., 2022), "psych"(Revelle, 2022), "broom.mixed"(Bolker & Robinson, 2022), "bayestestR" (Makowski et al., 2019), "DHARMa"(Hartig, 2022),

"glmmTMB"(Brooks et al., 2017)and "faux"(DeBruine, 2023). To compute the Stouffer's z indices for the specification curve analysis we used the function provided in Simonsohn et al. (2021).

Out of the 90 mother-infant dyads invited to participate in the study, for the 9-month-old sample,74 infants contributed any data for the iBEATs task and 75 to the iBREATH task. For the iBEATs task,3additional infants were excluded due to technical errors. Furthermore, following our preregistered analysis, 2 infants were excluded for the iBEATs task due to not reaching the minimum of 8 trials, 9 due to noisy ECG data, and 8 due to the +/- 2SD outlier rejection criterion, leaving a final sample of 52 infants. In comparison, for the iBREATH task, 10 infants were excluded due to technical errors, 3 due to not reaching at least 4 trials, 3 due to noisy respiratory belt data, and 3 due to the +/- 2 SD outlier rejection criterium, leaving a final sample of 56 infants.

As outlined in our preregistration, we lowered the threshold for outlier rejection in the longitudinal analysis to increase the sample size. Thus, for all analysis infants who completed at least 4 trials per task were included. For the 9-month-old data this would have slightly changed the iBEATs analysis plan. However, this criterion did not lead to the inclusion of additional infants in the final sample. For the 18-month-olds' iBEATs data, no infants were excluded due to not reaching at least 4 trials, 4 infants were excluded due to quality of the ECG signal, and 2 infants were excluded due to the +/- 2SD outlier rejection criterium, resulting in a final sample of 28. For the 18-month-olds' iBREATH data,1 infant was excluded due to not reaching at least 4 trials, 3 infants were excluded due to noisy physiological data, and 3 infants were excluded due to the +/- 2SD outlier rejection criterium, leaving a final sample of 30 infants. Means and SDs for number of trials completed for infants included in the analysis can be found in **Table 7**.

Out of the 80 mother-infant dyads invited to participate in the 3-month-old study, for, 77 infants contributed any data for the iBEATs task and 71 to the iBREATH task. Furthermore, following our preregistered analysis, 1 infant was excluded for the iBEATs task due to noisy ECG data, and 23 due to problems with the eye-tracking giving a sample of 53 infants. In comparison, for the iBREATH task, 2 infants were excluded due to not reaching at least 4 trials, 10 due to noisy respiratory belt data, and 19 due to problems with the eye-tracking resulting in a final included sample of 40 infants.

All statistics in our confirmatory analysis using null hypothesis testing were evaluated against a two-tailed significance level of $p < .05$. In case of non-significant results, if possible we aimed at following up the respective analysis with an equivalence or region of practical equivalence test (Lakens et al., 2018).To compare synchronous and asynchronous trials at 9 and 18months, both for the iBEATs and the iBREATH tasks, we computed two separate paired t-tests (Maister et al., 2017; see Supplementary Materials E for more information on asynchronous trials). At 3months we used a Bayesian paired t-test as the data collection was done after having collected the 9- and 18-month-old samples. Our intention in the analysis of the 3-month-old sample was to focus on strength of evidence in favor of/against an effect instead of a binary classification. We preregistered to correlate iBEATs and iBREATH scores at 9months. However, in the manuscript we only report the details of the MEGA-analysis (see next paragraph). To investigate the longitudinal development of cardiac and respiratory interoceptive sensitivity, we computed a Bayesian paired t-test comparing absolute proportional scores between 9 and 18months.

MEGA-analysis

We computed three MEGA-analyses pooling together data from all three age groups – to investigate a mean preference effect, the relation between the iBEATs and the iBREATH, as well as the development over age groups. First, to investigate whether there is a mean preference in the iBEATs and the iBREATH tasks, we computed mixed models using the R-package "glmmTMB" utilizing a beta-error distribution and logit-link function. We used looking time as outcome, condition, age, and their interaction as fixed effect, and participant as a random effect. We

Paradigm, Age Group	$M_{completed}$	$SD_{completed}$	$M_{included}$	$SD_{included}$
iBEATs, 3-months	13.97	7.08	9.82	7.44
iBEATs, 9-months	18.16	6.35	9.52	6.63
iBEATs, 18-months	15.62	6.31	10.90	7.54
iBREATH, 3-months	13.00	5.84	9.16	6.52
iBREATH, 9-months	13.25	4.85	10.10	5.21
iBREATH, 18-months	12.52	7.63	6.88	5.56

Table 7.

Descriptive information for number of trials completed and included

transformed age into a factor with 3-levels (3, 9, 18months), whereby 3months was set as reference level. After fitting the model, we visually inspected assumptions using the check_model function of the R-package “performance”. In addition, we checked for overdispersion using the “DHARMa” package (iBEATs: dispersion = 1.07, p = .168; iBREATH: dispersion = 1.12, p = .120). Further, we checked a reduced model lacking the interaction term for issues of collinearity (iBEATs, VIF = 1.00; iBREATH: VIF = 1.00). We then conducted full-null model comparisons by fitting a null-model that excluded the condition factor.

To investigate whether there is a relationship between the iBEATs and the iBREATH absolute proportional scores we computed a mixed model using the R-package “glmmTMB” using a beta error distribution. We used the iBREATH scores as outcome variable, the iBEATs scores, age group and the interaction as factors, and participant as a random intercept. Age was included as a factor with 3months as reference level. After fitting the model, we visually inspected assumptions using the check_model function of the R-package “performance”. We also did not find evidence for overdispersion (dispersion = 1.03, p = .800). Further, we checked a reduced model lacking the interaction term for issues of collinearity (VIF = 1.04).

Last, to investigate whether there is a difference between absolute proportional scores in the iBEATs and the iBREATH we computed two mixed models using the R-package “glmmTMB” with a beta error distribution. We used the iBEATs or the iBREATH absolute proportional scores as outcome, age as factor, and participant as a random effect. Age was included as a factor with 3 levels. After fitting the model, we visually inspected assumptions using the check_model function of the R-package “performance”. Further, we checked for absence of overdispersion (iBEATs: dispersion = 1.07, p = .560; iBREATH: dispersion = 1.09, p= .600). We found that for the iBEATs, the full model did not significantly improve fit over the null model ($\chi^2(3) = 0.170$, p< .919), but for the iBREATH, the full model did provide a significantly better fit than the null model ($\chi^2(3) = 10.60$, p = .005).

Acknowledgements

This research was funded in whole or in part by the Austrian Science Fund (FWF) [Project Number: P33486]. Ezgi Kayhan was funded by the DFG (Project number: 402789467). We want to thank all infants and mother who participated in this project. We also want to thank Monica Vanoncini and LiesbethForsthuber, as well as all research assistants, interns, and master students for their help in data collection and preparation of the experiment: Sandra Gaisbacher, Laura Neumann, Julia Otter, Lisa Triebenbacher, Jakob Weickmann, Felicia Wittmann, Gesine Jordan, Nina Maier, Rebecca Lutz, Celine Dorczok, Ann-CathrineGärtner, Maria Baumann, Nadine Pointner.

Contributions:

Markus R. Tünte: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft, writing – review & editing, visualization, supervision, project administration, funding acquisition. **Stefanie Höhl**: conceptualization, methodology, validation, resources, writing - original draft, writing – review & editing, supervision, project administration, funding acquisition. **Moritz Wunderwald**: conceptualization, methodology, software, validation, visualization. **Johannes Bullinger**: validation, investigation, writing – original draft, writing – review & editing. **AsenaBoyadziheva**: validation, investigation, writing – original draft, writing – review & editing. **Lara Maister**: methodology, software, validation. **Birgit Elsner**: conceptualization, supervision, project administration, funding acquisition. **Manos Tsakiris**: conceptualization, validation, writing –

original draft, writing – review & editing, supervision, funding acquisition. **Ezgi Kayhan**: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft, writing – review & editing, visualization, supervision, project administration, funding acquisition.

Conflict of Interest Statement:

We have no conflicts of interest to disclose.

Supplementary Material

A) Randomization Error in iBEATs and iBREATH Scripts

When conducting additional analysis regarding trial order effects during the review process, we noticed an error in the randomization procedure of our experimental scripts, which were custom Matlab scripts (all scripts are publicly available in the corresponding osf project). The error stems from the use of Matlab functions designed to return random structures, such as the 'randperm' function, without the specification of a random seed. If a random seed is not specified before running these functions, they will always return predefined sequences that are coupled to the respective computer (i.e., the first "random" sequence generated once starting Matlab will always be the same, and the second "random" sequence generated will always be the same). Thus, when repeatedly running this command in Matlab without restarting Matlab, seemingly random sequences will be returned. However, the first time a script is ran when the program is started will always result in the same sequence.

Apart from a few testing sessions with 9-month-olds we always restarted Matlab before every paradigm was run in order to make sure that the connection to the Arduino, which we used to synchronize physiological signals with the Eye tracking, was reset. The reason for this was that the Matlab would sometimes crash in case the connection to the Arduino was not established before every paradigm. This means, that in contrast to our intended experimental approach of randomizing stimuli location and image used, almost all infants watched the same combination of image used for synchronous and asynchronous condition, as well as order of location. Also, in contrast to our intended experimental approach for the 3-month-olds sample, in which we wanted to counterbalance synchronous and asynchronous conditions for the first trial, the first trial was always synchronous.

Alternative Randomizations

One potential issue with such a fixed-order design is that our experimental effects might have been impacted by the lack of randomization. As most infants were presented with the same image in synchronous or asynchronous conditions it is possible that mean group differences between conditions are influenced by the stimulus assigned to the respective condition (stimuli for iBEATs and iBREATH are displayed in **Figure 1**). However, for some infants in the 9-month-olds group we did not restart Matlab after every experimental paradigm. Thus, they were presented with a different randomization. To illustrate the impact of the randomization on our results we decided to repeat our main analysis reported in the manuscript for the 9-month-olds but split the samples apart according to their randomization (labeled as randomization 1 and 2 in the following). In addition, we did not reject trials using standard deviations as we wanted to retain as much data as possible. Results are displayed in **Table 1**. For the smaller samples we used a Wilcoxon ranked sign test instead of the paired t-test.

Figure 1.

Stimuli used for A) iBEATs and B) iBREATH.

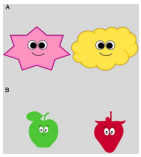


Table 1.

Main analysis for the 9-month-olds for iBEATs and iBREATH split by randomization.

Experiment	Randomization	*N*	*Mean* LT Synch (*SD*)	*Mean* LT Async (*SD*)	Comparison
iBEATs	1	46	6792 (3473)	5727 (2508)	t(45) = 2.37, *p* = .022
iBEATs	2	6	8775 (5810)	3734 (1175)	V = 21, *p* = .031
iBREATH	1	53	6307 (3090)	5563 (2282)	t(52) = 2.37, *p* = .021
iBREATH	2	3	6847 (1321)	4085 (331)	V = 6, *p* = .25

For the iBEATs we find that 6 infants, and for the iBREATH 3 infants, completed an alternative randomization (marked as randomization 2) and had enough data to be included in the analysis. Further, when considering mean preferences, we find that the exclusion of the infants with alternative randomization (so only considering randomization 1) does not change the results of our main analysis, as for both paradigms a significant mean preference for synchronous trials is still present. Next, when just considering the alternative randomization, we find that the numerical preference is the same for both iBEATs and iBREATH, with the iBEATs showing a significant mean preference for synchronous trials. Notably, these results need to be interpreted with caution due to the small sample sizes for randomization 2. Last, we aimed at investigating whether there is a significant difference between infants in randomization 1 or randomization 2 for iBEATs and iBREATH, respectively. To do so we computed a difference score for each infant by subtracting mean looking times in synchronous trials minus mean looking time in asynchronous trials. Then we used a Wilcoxon rank sign test to investigate whether the two groups differed from each other. We do not find evidence for a significant difference between randomization groups for iBEATs (W = 91, p = .188, **Figure 1A**), or iBREATH (W = 34, p =.102, **Figure 1B**). Following, these results indicate that the stimuli assigned to synchronous or asynchronous trials did not exhibit a impact on infant preference.

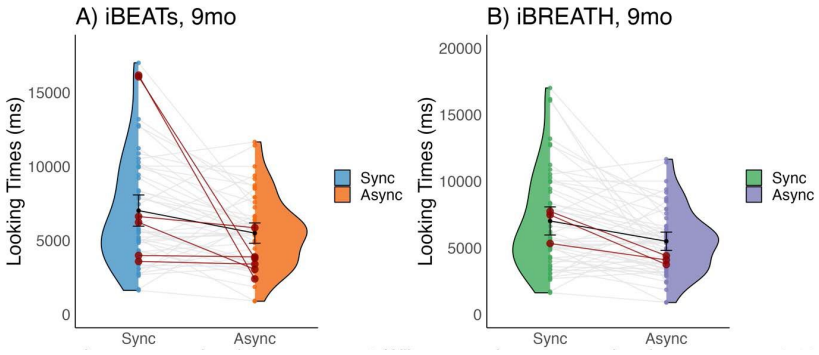


Figure 2.

Mean differences in A) iBEATs and B) iBREATH for the 9-month-olds.

Note. Infants with alternative randomization are highlighted in red.

Preferences Over Time

Next, we aimed at investigating whether the infant's mean preference for synchronous or asynchronous trials changed over time in both tasks. For this analysis we used data from the confirmatory mean group preference analysis from all age groups. We computed two different analyses. First, in a cumulative approach (**Figures 3** & **4**, left column), we computed paired t-tests (and corresponding 95% confidence intervals) with a cumulative increasing number of trials included. This means that the first point displayed corresponds to the mean difference found in a paired t-test using only the first two trials, while the last point displayed corresponds to a paired t-test using all trials (our confirmatory analysis). This analysis gives us an indication of how the mean effect varies with the inclusion of an increasing number of trials.

In the second analysis we used a sliding window approach of 3 trials regarding trial inclusions (**Figures 3** & **4**, right column). This means that each point shows the result of a paired t-test (and corresponding 95% confidence intervals) using the three trials before and after the corresponding trial (e.g., for trial 10, trials 7–13 are included). This analysis gives us an indication of how the infant's preference changes throughout the paradigm.

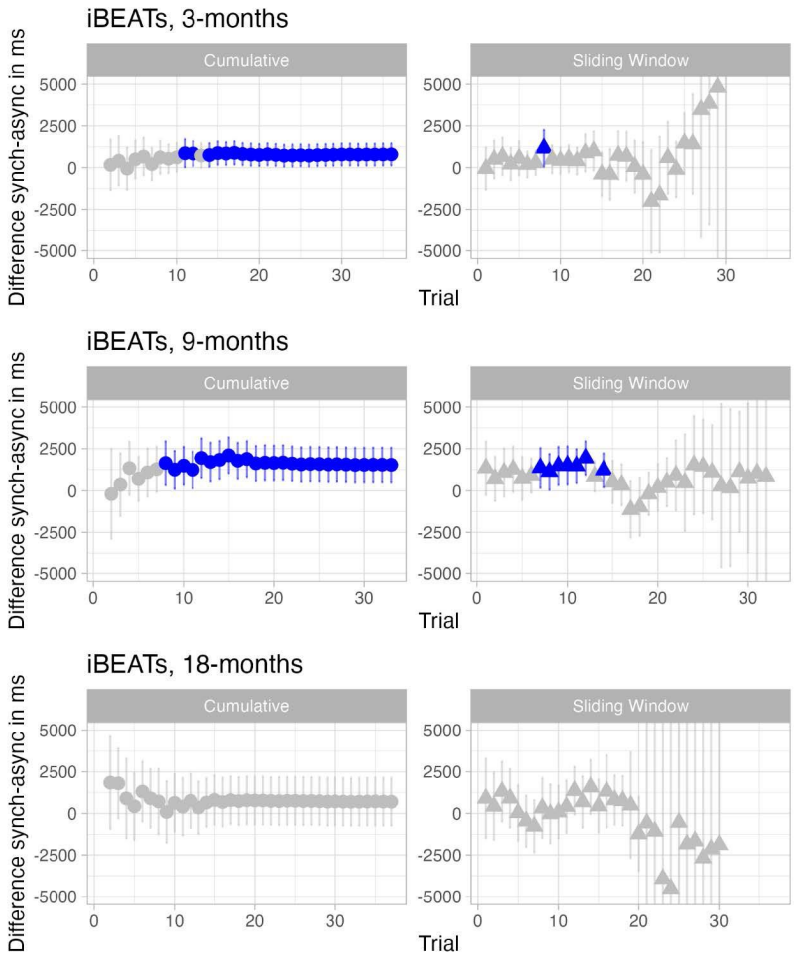


Figure 3.

Infants' Preference Over Time in the iBEATs Paradigm for All Age Groups.

Note. Here results of paired t-tests for the iBEATs with different inclusion criteria for trials are displayed for all age groups. The left column shows a cumulative analysis in which all previous trials are included. The right column shows a 3-trial sliding window analysis in which the three preceding and 3 proceeding trials are included. Blue indicates a significant result of the paired t-test.

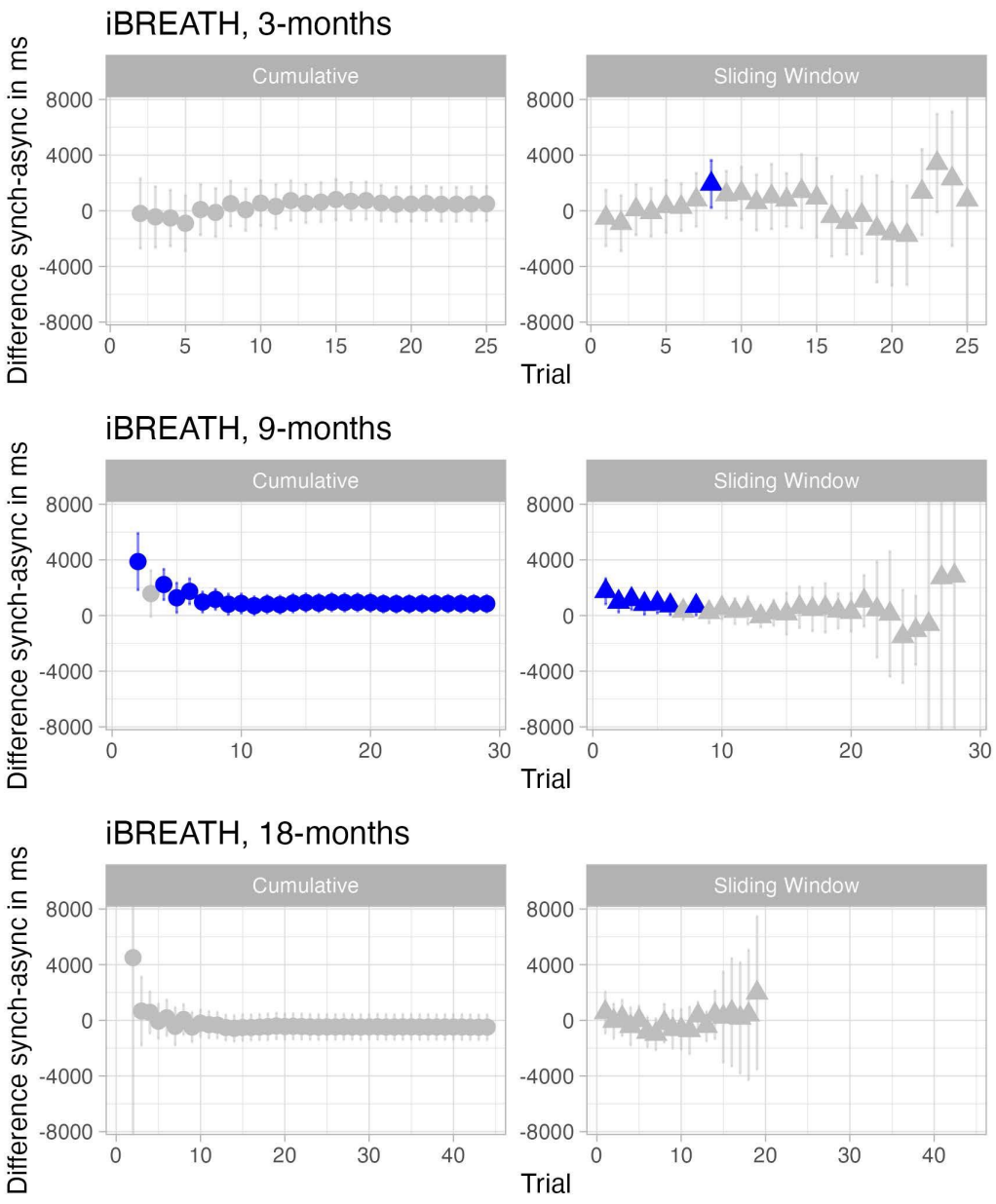


Figure 4.

Infants' Preference Over Time in the iBREATH Paradigm for All Age Groups.

Note. Here results of paired t-tests for the iBREATH with different inclusion criteria for trials are displayed for all age groups. The left column shows a cumulative analysis in which all previous trials are included. The right column shows a 3-trial sliding window analysis in which the three preceding and 3 proceeding trials are included. Blue indicates a significant result of the paired t-test.

For the cumulative analysis (**Figure 3** & **4**, left columns) we find that for the iBEATs (**Figure 3**) in the 3- and 9-month-olds mean preferences seem to stabilize around 7–8 trials, with the early trials not showing a clear preference. For the 18-month-olds we find a stronger preference in the first trials, which then shifts towards no preference. For the cumulative analysis in the iBREATH (**Figure 4**) we find that in the 9- and 18-month-olds the first two trials have an impact, while the mean preference stabilizes relatively fast and stays rather constant. In turn, for the 3-month-olds we do not see an impact of the first trial in the iBREATH.

For the sliding window analysis (**Figure 3** & **4** right column) we find that for the iBEATs (**Figure 3**) in the 3- and 9-month-olds the strongest preferences are around the trials 7 to 13. Interestingly, this coincides with the trial numbers in which the cumulative analysis finds a stabilization of preferences. While for the 18-month-olds we find that preferences are not stable and drift around. For the iBREATH (**Figure 4**) we find that rather strong preferences can be observed for the first trials, with later trials not showing strong effects. In general, for the sliding window analyses it must be noted that later trials show very unreliable effects with large confidence intervals. This is probably due to fewer number of infants included as mean numbers of trials completed ranged from 12.52 (iBREATH 18-months) to 18.26 (iBEATs 9-months).

Taken together, we find that the effect of the trials on mean group preferences varies both across tasks and age groups. If present, for the iBEATs (3- and 9-month-olds) mean group effects seem to emerge in later trials (around trial 7), while for the iBREATH (9 and 18-month-olds) the first trials were more impactful. At the same time, we do not see a coherent effect across all age groups, in the iBEATs the 18-month-olds, and in the iBREATH the 3-month-olds, show more diffuse patterns.

If the randomization of the trials, and first trial being fixed, had had a big impact on infants' looking patters, similar patterns should have emergedacross tasks and age groups. Given that we do not find such a coherent pattern, our results probably capture sufficient variation with regard to the interoceptive manipulation present in the tasks.

Conclusion

In sum, we conducted several analyses to investigate the impact of the experimental design on infant looking behavior. First, we considered those infants from the 9-month-old subgroup with an alternative randomization. We do not find that the image used for synchronous or asynchronous trials had an impact on mean preferences in our analysis. Next, we considered the trial order and investigated whether we find patterns regarding the emergence of preferences. We find that infants' preferences vary over tasks and age groups. Thus, it is unlikely that the randomization itself was driving looking behavior.

Taken together, we do not find evidence that our results have been impacted by the use ofa fixed experimental approach. Given that the images used in the experimental paradigms presented here were chosen with the intention of not inducing an a priori preference in the infants, this result is not necessarily surprising. Further, as our initial intention was to use a semi-random paradigm, with the first trial always being synchronous, and never more than two trials of the same type following each other, it is also unlikely that the trial order used here is substantially different from one used in more randomized approaches. However, the fixed experimental design is a major limitation of the present study, and we cannot fully rule out that it influenced infants' looking behavior to some extent.

B) Data Simulation for Different Sample Sizes

In our main analysis we have found that some of the tests investigating mean difference between conditions were not statistically significant. Follow-up analysis using equivalence tests and region of practical equivalence approaches showed that sample size might have played a role, as statistical power could have been too low to detect an effect. Here, we aimed at further investigating the absence of a significant effect for a mean group preference due to reduced statistical power in smaller samples. To investigate this hypothesis, we decided to run simulations building up on our data. Our aim was to characterize how statistical power to detect an effect is impacted by different levels of sample sizes. This is relevant as sample sizes in infancy research tend to be low, and all non-significant results reported so far for iBEATs-like paradigms had a sample size of roughly 30 infants (Weijs et al. 2023, the 18-month-old samples reported here). Further, the results of such simulations might be very informative for researchers planning to use experimental paradigms like iBEATs or iBREATH in infant samples in the future.

In a first step, we used the R-package "faux" to simulate data sets using the 9-month-old data from the iBEATs as input. The package "faux" simulates data that has the same properties as an existing data set. We simulated data sets ranging from 5 to 125 participants and generated 50 data sets for each number of participants, giving a total of 6000 datasets. We used the data sets processed according to our preregistration as input data set (n = 52, Cohen's d = .48).

Next, we ran analyses with the generated data which are visualized in **Figure 1**. We computed a t-test following our preregistered analysis strategy comparing mean looking times for synchronous and asynchronous conditions for different sample sizes. In 1A) mean differences for the paired t-test (y-axis) are plotted against the sample size for the simulated data obtained building up on our 9-month-old sample. Red color refers to a significant result from the paired t-test, while blue color refers to a non-significant effect. In B) the percent of significant results for the paired t-tests (y-axis) are plotted against the sample size (x-axis).

The results from the simulation give us an idea about the chance to find a significant result given an effect of d = .48 in data sets with varying sample sizes. We observed that for the data building up on our 9-month-old sample, upon approaching a sample size of 50–60, 80–90% of results are significant. Interestingly, the proportion of significant results for a sample size of 30 are a little bit above 50%, indicating that with a sample size of 30 infants the chance to find a significant mean difference is roughly that of a coin flip. Thus, the absence of a significant result in samples of 30 infants might not necessarily indicate the absence of a mean group preference in general, but the sample size might not have been sufficient to detect a significant result.

C) Detailed results for the MEGA-analyses

Here, we provide detailed results for the MEGA-analyses for the iBEATs and iBREATH.

iBEATs

Table 1 reports the full-null model comparison for the iBEATs MEGA-analysis. **Table 2** reports the coefficients for the iBEATs MEGA-analysis with 3-month as reference group, while **Table 3** reports the coefficients for the iBEATs MEGA-analysis with 9-months as reference group and **Table 4** with 18-months as reference group.

Figure 1.

Simulating data frames for sample sizes from 15 to 125 building up on the iBEATs data from the 9-month-olds.

Note. Results from the simulations. In A) mean effects and 95% confidence intervals are plotted for the different sample sizes. Red color indicates a significant effect, while blue indicates a non-significant result. In B) the percent of significant results are plotted with a fitted line.

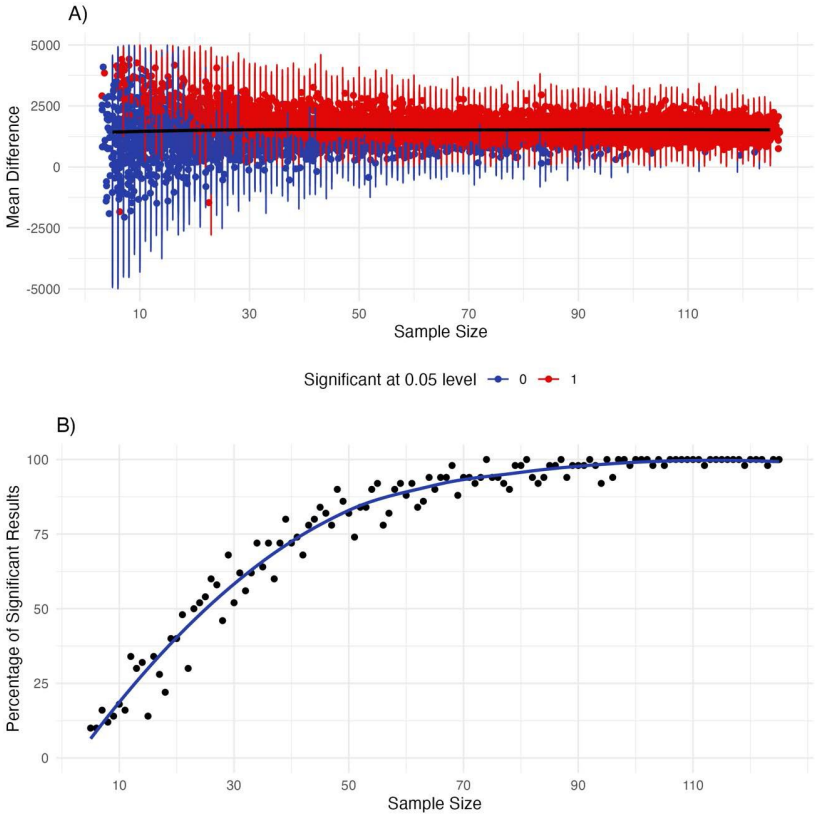


Table 1.

Full-null model comparison for the iBEATs model.

Model	Df	AIC	BIC	logLik	Deviance	Chisq	Chi Df	p-value
Null	5	-1222.9	-1195.5	616.46	-1232.9			
Full	8	-1227.8	-1184.0	621.92	-1243.8	10.91	3	.012

Table 2.

Results for the MEGA analysis of the iBEATs data with 3-months as reference group.

Term	Estimate	SE	z-value	p-value
Intercept	-1.08	0.12	-8.69	< .001
Condition asynchronous	-0.17	0.08	-2.15	.031
9 months	0.57	0.16	3.48	< .001
18 months	0.44	0.18	2.49	.013
Condition * 9 months	0.00	0.11	0.02	.982
Condition * 18 months	0.14	0.13	1.08	.283

Table 3.

Results for the MEGA analysis of the iBEATs data with 9-months as reference group.

Term	Estimate	SE	z-value	p-value
Intercept	-0.50	0.11	-4.66	< .001
Condition asynchronous	-0.17	0.07	-2.49	.013
3 months	-0.57	0.16	-3.48	< .001
18 months	-0.13	0.10	-1.22	.221
Condition * 3 months	-0.00	0.11	-0.02	.982
Condition * 18 months	0.14	0.12	1.12	.264

Term	Estimate	SE	z-value	p-value
Intercept	-0.63	0.13	-4.91	< .001
Condition asynchronous	-0.03	0.10	-0.31	.756
3 months	0.44	0.18	2.49	.013
9 months	-0.13	0.10	-1.22	.221
Condition * 3 months	-0.14	0.13	-1.08	.283
Condition * 9 months	-0.14	0.12	-1.12	.264

Table 4.

Results for the MEGA analysis of the iBEATs data with 18-months as reference group.

iBREATH

Next, we report the same results for the iBREATH MEGA-analysis. **Table 5** reports the full-null model comparison, while **Table 6** reports coefficients with 3-months as reference, **Table 7** with 9-months as reference, and **Table 8** with 18-months as reference.

Relationship between iBEATs and iBREATH

Regarding the relationship between iBEATs and iBREATH we conducted a beta regression with iBEATH as outcome, and iBEATs, age, as well as the interaction between iBEATs and age as predictors. Here, we report detailed results for each age group (**Table 9**: 3-months, **Table 10**: 9-months, **Table 11**: 18-months).

D) Specification Curve Analysis

For the specification curve analysis, we followed the approach outlined in Simonsohn et al. (2020).First, we identified the subset of suitable analytical choices byreviewing all available papers that used a task similar to the iBEATs published so far (Charbonneau et al., 2022; Maister et al., 2017; Weijs et al., 2023). Building up, we extracted potential analytical decisions applicable to our dataset (**Table 1**, **Table 2**). Second, we ran all suitable analysis and plotted the results (**Figure 1**, **Figure 2**). Third, we used a permutation approach to investigate how inconsistent the obtained results were with the null hypothesis of no effect (**Table 3**). Next, we will discuss results of the specification curve analysis for the iBEATs and the iBREATH, respectively, and make recommendations for future data analysis for projects using iBEATs/iBREATH as well as preferential looking paradigms in general.

Specification Curve Analysis iBEATs

An overview over analytical choices can be found in **Table 1**.For the iBEATs we identified 7 different categories, yielding a total of 1024 potential analyses (**Figure 2**). We found that 458 (44.73%) analyses led to a significant effect. Most of these (442, 43.16%) yieldeda significant effect for a mean synchronicity preference. However, there are also a few analysis paths (16, 1.56%) that we could have chosen that would have resultedin a mean preference for asynchronous stimuli.

Specification Curve Analysis iBREATH

As this is the first paper on respiratory interoceptive sensitivity in infants, for the iBREATH we adapted the choices made for the iBEATs paradigm. An overview can be found in **table 2**. The only difference between the iBEATs and the iBREATH specification curve analysis concerns artifact removal for the physiological data. For the iBEATs, there were 4 different categories, while for the iBREATH we identified 6 different categories. There were 1536 potential analyses for the iBREATH (**Figure 3**). We found that 269 (17.51%) analytical choices led to a significant effect. Further, all analyses rendering a significant effect revealed a mean synchronous preference.

E) Slow and Fast Asynchronous Trials

Asynchronous trials in the iBEATs and the iBREATH could be either faster or slower than the infant's respective physiological signal. In an exploratory analysis, we computed t-tests to investigate whether looking times differed between slow and fast asynchronous trials for all age groups and paradigms (**Table 1**). Overall, we did not find evidence for a difference between looking times to fast and slow asynchronous trials.

Table 5.

Full-null model comparison for the iBREATH model.

Model	Df	AIC	BIC	logLik	Deviance	Chisq	Chi Df	p-value
Null	5	-1600.3	-1574.5	805.14	-1610.3			
Full	8	-1600.7	-1559.4	808.37	-1616.7	6.45	3	.091

Table 6.

Results for the MEGA analysis of the iBREATH data.

Term	Estimate	SE	z-value	p-value
Intercept	-1.35	0.13	-10.35	< .001
Condition asynchronous	-0.15	0.09	-1.74	.082
9 months	0.25	0.17	1.47	.141
18 months	-0.15	0.19	-0.77	.440
Condition * 9 months	0.02	0.12	0.17	.864
Condition * 18 months	0.23	0.16	1.50	.134

Table 7.

Results for the MEGA analysis of the iBREATH data.

Term	Estimate	SE	z-value	p-value
Intercept	-1.10	0.11	-9.76	< .001
Condition asynchronous	-0.13	0.08	-1.75	.080
3 months	-0.25	0.17	-1.47	.141
18 months	-0.40	0.13	-3.12	.001
Condition * 3 months	-0.02	0.12	-0.17	.864
Condition * 18 months	0.21	0.15	1.43	.154

Table 8.

Results for the MEGA analysis of the iBREATH data.

Term	Estimate	SE	z-value	p-value
Intercept	-1.35	0.13	-10.35	< .001
Condition asynchronous	-0.15	0.09	-1.74	.082
3 months	0.25	0.17	1.47	.141
9 months	-0.15	0.19	-0.77	.440
Condition * 3 months	-0.23	0.16	-1.50	.134
Condition * 9 months	-0.21	0.15	-1.43	.154

Table 9.

MEGA analysis for the relationship between iBEATs and iBREATH with 3-months as reference group.

Term	Estimate	SE	z-value	p-value
Intercept	-1.17	0.22	-5.25	< .001
iBEATs score	-1.83	0.97	-1.89	.059
9-months	-0.15	0.34	-0.42	.674
18-months	-0.05	0.35	-0.15	.880
iBEATs * 9-months	0.67	1.31	0.51	.610
iBEATs * 18-months	3.13	1.41	2.22	.027

Table 10.

MEGA analysis for the relationship between iBEATs and iBREATH with 9-months as reference group.

Term	Estimate	SE	z-value	p-value
Intercept	-1.13	0.27	-4.89	< .001
iBEATs score	-1.16	0.90	-1.30	.192
3-months	0.15	0.34	0.42	.674
18-months	0.09	0.38	0.24	.810
iBEATs * 3-months	-0.67	1.31	-0.51	.610
iBEATs * 18-months	2.45	1.36	1.81	.070

Term	Estimate	SE	z-value	p-value
Intercept	-1.22	0.28	-4.39	< .001
iBEATs score	1.30	1.02	1.27	.204
3-months	0.05	0.35	0.15	.880
9-months	-0.09	0.38	-0.24	.810
iBEATs * 3-months	-3.13	1.41	-2.22	.027
iBEATs * 9-months	-2.45	1.36	-1.81	.070

Table 11.

MEGA analysis for the relationship between iBEATs and iBREATH with 18-months as reference group.

Category	Implementationspecification curve analysis	Number of analytical choices
1. Outlier rejection	Only for sync trials, apply to all trials	2
2. SD outlier rejection	2SD, 2.5SD, 3SD, no criterion	4
3. ECG artifact trial rejection	All R-peaks identified, missed single R-peaks, missed R-peaks in last two seconds, identified at least 85% of R-peaks	4
4. Data transformation	Log-transformation, non-transformed data	2
5. Trial removal	Remove trials with 0 looking times, keep trials with 0 looking times	2
6. Min. number of trials per id to be included	8, 4, 2, 0	4
7. Statistical analysis	Paired t-test, linear mixed model (trials clustered in id)	2

Table 1.

Analytical decisions for the iBEATs.

Category	Implementation specification curve analysis	Number of analytical choices
1. Outlier rejection	Only for sync trials, apply to all trials	2
2. SD outlier rejection	2SD, 2.5SD, 3SD, no criterion	4
3. Respiration artifact trial rejection	Only good signals, include deep breaths, artifacts in ECG but not respiration, deep breaths & artifacts in ECG, short flat lines in respiratory signal, flat lines & deep breaths and/or ECG artifacts	6
4. Data transformation	Log-transformation, non-transformed data	2
5. Trial removal	Remove trials with 0 looking times, keep trials with 0 looking times	2
6. Min. number of trials per id to be included	8, 4, 2, 0	4
7. Statistical analysis	Paired t-test, linear mixed model (trials clustered in id)	2

Table 2.

Analytical decisions for the iBREATH.

Test Statistic Used	Observed Result	p-value (% of shuffled results as or more extreme than observed results)
A. iBEATs		
1. Median effect size	.111	.008
2. Share of significant results	470	.042
3. Aggregate all p-values	Stouffer Z = -39.91	.068
B)iBREATH		
1. Median effect size	.109	.032
2. Share of significant results	269	.088
3. Aggregate all p-values	Stouffer Z = -24.32	.120

Table 3.

Inference of the Specification Curve Analysis

Figure 1.

Specification Curve Analysis for the iBEATs.

Note. Descriptive results from the specification curve analysis for the iBEATs task. Blue coloring in A) and B) refers to a significant result for a mean synchronous preference, while red color indicates to a significant result for a mean asynchronous preference ($p < .05$) for the specification and test. In A) standardized beta regression estimates are plotted. In B) an overview for a range of analytical choices is given. In C) analytical choices are further decomposed.

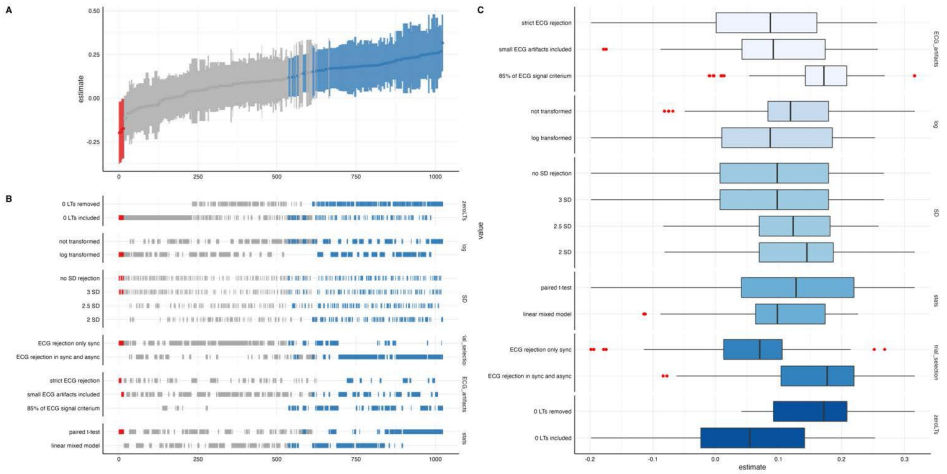
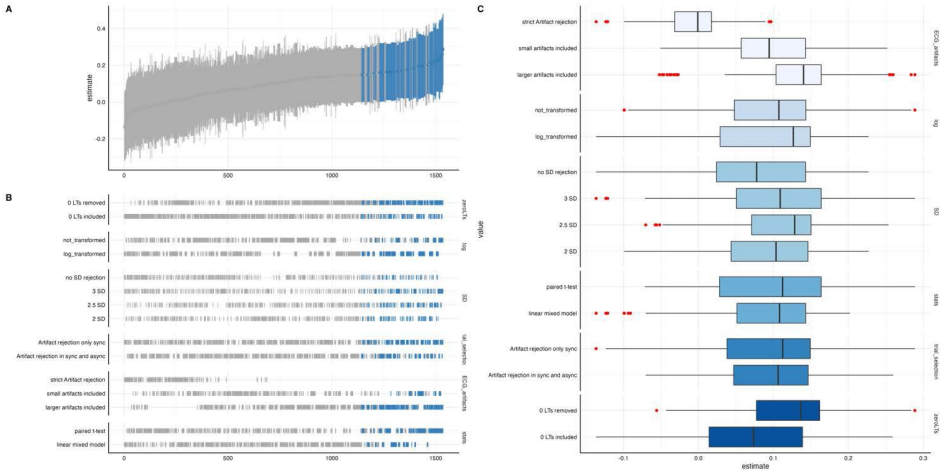


Figure 2.

Specification Curve Analysis for the iBREATH.

Note. Descriptive results from the specification curve analysis for the iBREATH task. Blue coloring in A) and B) refers to a significant result for a mean synchronous preference, while red color indicates to a significant result for a mean asynchronous preference ($p < .05$) for the specification and test. In A) standardized beta regression estimates are plotted. In B) an overview for a range of analytical choices is given. In C) analytical choices are further decomposed.



	Mean Slow (SD)in ms	Mean Fast (SD)in ms	Comparison
iBEATs, 3-months	5960 (5885)	5651 (5731)	t(299.15) = 0.46, *p* = .64
iBEATs, 9-months	6121 (4537)	6368 (5044)	t(277.21) = -0.44, *p* = .66
iBEATs, 18-months	6520 (5620)	7072 (5593)	t(151.22) = 0.61, *p* = .54
iBREATH, 3-months	6632 (7359)	7719 (8882)	t(102.45) = -0.72, *p* = .47
iBREATH, 9-months	5810 (4956)	6576 (5416)	t(292.78) = -1.29, *p* = .20
iBREATH, 18-months	4969 (4610)	6078 (6053)	t(99.74) = -1.11, *p* = .27

Table 1.

Looking times for slow and fast asynchronous trials

References

Allen M., Varga S., Heck D. H. (2022) **Respiratory rhythms of the predictive mind** *Psychological Review* https://doi.org/10.1037/rev0000391

Amsterdam B. (1972) **Mirror self-image reactions before age two** *Developmental Psychobiology* **5**:297–305 https://doi.org/10.1002/dev.420050403

Bahrick L. E., Watson J. S. (1985) **Detection of intermodal proprioceptive–visual contingency as a potential basis of self-perception in infancy** *Developmental Psychology* **21**:963–973 https://doi.org/10.1037/0012-1649.21.6.963

Bakker M., Wicherts J. M. (2014) **Outlier removal and the relation with reporting errors and quality of psychological research** *PLoS ONE* **9**:1–9 https://doi.org/10.1371/journal.pone.0103360

Bates D., Mächler M., Bolker B., Walker S. (2015) **Fitting Linear Mixed-Effects Models Using lme4** *Journal of Statistical Software* **67**:1–48 https://doi.org/10.18637/jss.v067.i01

Bergmann C., Rabagliati H., Tsuji S. (2019) **What's in a looking time?** *PsyArXiv Preprints* **1** https://doi.org/10.31234/osf.io/6u453

Bolker B., Robinson D. (2022) **broom.mixed: Tidying Methods for Mixed Models**

Boyadzhieva A., Kayhan E. (2021) **Keeping the Breath in Mind: Respiration, Neural Oscillations, and the Free Energy Principle** *Frontiers in Neuroscience* **15**:1–13 https://doi.org/10.3389/fnins.2021.647579

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

Brooks M. E., Kristensen K., van Benthem K. J., Magnusson A., Berg C. W., Nielsen A., Skaug H. J., Mächler M., Bolker B. M. (2017) **glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling** *R Journal* **9**:378–400 https://doi.org/10.32614/rj-2017-066

Burnham K. P., Anderson D. R. (2004) **Multimodel inference: Understanding AIC and BIC in model selection** *Sociological Methods and Research* **33**:261–304 https://doi.org/10.1177/0049124104268644

Carpenter M., Nagell K., Tomasello M., Butterworth G., Moore C. (1998) **Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age** *Monographs of the Society for Research in Child Development* **63**:i–174 https://doi.org/10.2307/1166214

Champely (2020) **pwr: Basic Functions for Power Analysis**

Charbonneau J. A., Maister L., Tsakiris M., Bliss-Moreau E. (2022) **Rhesus monkeys have an interoceptive sense of their beating hearts** *Proceedings of the National Academy of Sciences of the United States of America* **119**:1–8 https://doi.org/10.1073/pnas.2119868119

Cirelli L. K., Einarson K. M., Trainor L. J. (2014) **Interpersonal synchrony increases prosocial behavior in infants** *Developmental Science* **17**:1003–1011 https://doi.org/10.1111/desc.12193

Coll M. P., Hobson H., Bird G., Murphy J. (2021) **Systematic review and meta-analysis of the relationship between the heartbeat-evoked potential and interoception** *Neuroscience and Biobehavioral Reviews* **122**:190–200 https://doi.org/10.1016/j.neubiorev.2020.12.012

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

Critchley H. D., Harrison N. A. (2013) **Visceral Influences on Brain and Behavior** *Neuron* **77**:624–638 https://doi.org/10.1016/j.neuron.2013.02.008

Dang J., King K. M., Inzlicht M. (2020) **Why Are Self-Report and Behavioral Measures Weakly Correlated?** *Trends in Cognitive Sciences* **24**:267–269 https://doi.org/10.1016/j.tics.2020.01.007

DeBruine L. (2023) **faux: Simulation for Factorial Designs** https://doi.org/10.5281/zenodo.2669586

Draghici A. E., Taylor J. A. (2016) **The physiological basis and measurement of heart rate variability in humans** *Journal of Physiological Anthropology* **35** https://doi.org/10.1186/s40101-016-0113-7

Filippetti M. L. (2021) **Being in Tune With Your Body: The Emergence of Interoceptive Processing Through Caregiver–Infant Feeding Interactions** *Child Development Perspectives* **15**:182–188 https://doi.org/10.1111/cdep.12420

Filippetti M. L., Johnson M. H., Lloyd-Fox S., Dragovic D., Farroni T. (2013) **Body perception in newborns** *Current Biology* **23**:2413–2416 https://doi.org/10.1016/j.cub.2013.10.017

Fotopoulou A., Tsakiris M. (2017) **Mentalizing homeostasis: The social origins of interoceptive inference-replies to Commentaries** *Neuropsychoanalysis* **19**:71–76 https://doi.org/10.1080/15294145.2017.1307667

Garcia A., Koschnitzky J. E., Dashevskiy T., Ramirez J.-M. (2014) **Cardiorespiratory Coupling in Health and Disease** *Autonomic Neuroscience* **46**:759–785 https://doi.org/10.1146/annurev-cellbio-092910-154240.Sensory

Garfinkel S. N., Manassei M. F., Hamilton-Fletcher G., den Bosch Y. I., Critchley H. D., Engels M. (2016) **Interoceptive dimensions across cardiac and respiratory axes** *Philosophical Transactions of the Royal Society B: Biological Sciences* **371** https://doi.org/10.1098/rstb.2016.0014

Harms C., Lakens D. (2018) **Making "null effects" informative: statistical techniques and inferential frameworks** *Journal of Clinical and Translational Research* **3**:382–393 https://doi.org/10.18053/jctres.03.2017s2.007

Harrison O. K. *et al.* (2021) **The Filter Detection Task for measurement of breathing-related interoception and metacognition** *Biological Psychology* **165** https://doi.org/10.1016/j.biopsycho.2021.108185

Harrison O. K. *et al.* (2021) **Interoception of breathing and its relationship with anxiety** *Neuron* **109**:4080–4093 https://doi.org/10.1016/j.neuron.2021.09.045

Hartig F. (2022) **DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models**

Heck D. H. *et al.* (2017) **Breathing as a fundamental rhythm of brain function** *Frontiers in Neural Circuits* **10**:1–8 https://doi.org/10.3389/fncir.2016.00115

Hedge C., Powell G., Sumner P. (2018) **The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences** *Behavior Research Methods* **50**:1166–1186 https://doi.org/10.3758/s13428-017-0935-1

Hunter M. A., Ames E. W. (1988) **A multifactor model of infant preferences for novel and familiar stimuli** *Advances in Infancy Research* **5**:69–95

Imafuku M., Yoshimoto H., Hiraki K. (2023) **Infants ' interoception is associated with eye contact in dyadic social interactions** *Scientific Reports* **123456789**:1–8 https://doi.org/10.1038/s41598-023-35851-9

Khalsa S. S. *et al.* (2018) **Interoception and Mental Health: A Roadmap** *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* https://doi.org/10.1016/j.bpsc.2017.12.004

Kluger D. S., Balestrieri E., Busch N. A., Gross J. (2021) **Respiration aligns perception with neural excitability** *eLife* **10**:1–19 https://doi.org/10.7554/eLife.70907

Koile E., Cristia A. (2021) **Toward Cumulative Cognitive Science: A Comparison of Meta-Analysis, Mega-Analysis, and Hybrid Approaches** *Open Mind* **5**:154–173 https://doi.org/10.1162/opmi_a_00048

Kwon Y. H., Lee H. Y. (2014) **Differences of respiratory function according to level of the gross motor function classification system in children with cerebral palsy** *Journal of Physical Therapy Science* **26**:389–391 https://doi.org/10.1589/jpts.26.389

Lakens D., Scheel A. M., Isager P. M. (2018) **Equivalence Testing for Psychological Research: A Tutorial** *Advances in Methods and Practices in Psychological Science* **1**:259–269 https://doi.org/10.1177/2515245918770963

Leys C., Delacre M., Mora Y. L., Lakens D., Ley C. (2019) **How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration** *International Review of Social Psychology* **32**:1–10 https://doi.org/10.5334/irsp.289

Maister L., Tang T., Tsakiris M. (2017) **Neurobehavioral evidence of interoceptive sensitivity in early infancy** *eLife* **6**:1–12 https://doi.org/10.7554/eLife.25318

Makowski D., Ben-Shachar M., Lüdecke D. (2019) **bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework** *Journal of Open Source Software* **4** https://doi.org/10.21105/joss.01541

ManyBabies 5 Team (2023) **ManyBabies 5: A large-scale investigation of the proposed shift from familiarity preference to novelty preference in infant looking time Pre-data collection manuscript for peer-review The ManyBabies 5 Team** *PsyArXiv* https://doi.org/10.31234/osf.io/ck3vd

Masur P. K., Scharkow M. (2019) **specr: Statistical functions for conducting specification curve analyses**

Morey R. D., Rouder J. N. (2022) **BayesFactor: Computation of Bayes Factors for Common DesignsNo Title**

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

Murphy J., Brewer R., Plans D., Khalsa S., Catmur C., Bird G. (2020) **Testing the independence of self-reported interoceptive accuracy and attention** *Quarterly Journal of Experimental Psychology* **44** https://doi.org/10.1177/1747021819879826

Musculus L., Tünte M. R., Raab M., Kayhan E. (2021) **An Embodied Cognition Perspective on the Role of Interoception in the Development of the Minimal Self** *Frontiers in Psychology* **12** https://doi.org/10.3389/fpsyg.2021.716950

Nicholson T., Williams D., Carpenter K., Kallitsounaki A. (2019) **Interoception is Impaired in Children, But Not Adults, with Autism Spectrum Disorder** *Journal of Autism and Developmental Disorders* **49**:3625–3637 https://doi.org/10.1007/s10803-019-04079-w

Nikolova N., Harrison O., Toohey S., Brændholt M., Correa C., Vejlø M., Jensen M. S., Fardo F., Allen M. (2021) **The Respiratory Resistance Sensitivity Task : An Automated Method for Quantifying Respiratory Interoception and Metacognition** *bioRxiv* :0–41 https://doi.org/10.1101/2021.10.14.464418

Patil I. (2021) **Visualizations with statistical details: The "ggstatsplot" approach** *Journal of Open Source Software* **6** https://doi.org/10.21105/joss.03167

R Core Team (2022) **R: A language and environment for statistical computing**

Repacholi B. M., Gopnik A. (1997) **Early reasoning about desires: Evidence from 14- and 18-month-olds** *Developmental Psychology* **33**:12–21 https://doi.org/10.1037/0012-1649.33.1.12

Revelle W. (2022) **psych: Procedures for Psychological, Psychometric, and Personality Research**

Rochat P. (1998) **Self-perception and action in infancy** *Experimental Brain Research* **123**:102–109 https://doi.org/10.1007/s002210050550

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 https://doi.org/10.1016/j.ijpsycho.2019.02.001

Schandry R. (1981) **Heart Beat Perception and Emotional Experience** *Psychophysiology* **18**:483–488 https://doi.org/10.1111/j.1469-8986.1981.tb02486.x

Simonsohn U., Simmons J. P., Nelson L. D. (2020) **Specification curve analysis** *Nature Human Behaviour* **4**:1208–1214 https://doi.org/10.1038/s41562-020-0912-z

Singmann H., Bolker B., Westfall J., Aust F., Ben-Shachar M. S. (2022) **afex: Analysis of Factorial Experiments**

Suksasilp C., Garfinkel S. N. (2022) **Towards a comprehensive assessment of interoception in a multi-dimensional framework** *Biological Psychology* **168** https://doi.org/10.1016/j.biopsycho.2022.108262

Van Puyvelde M., Gorissen A. S., Pattyn N., McGlone F. (2019) **Does touch matter? The impact of stroking versus non-stroking maternal touch on cardio-respiratory processes in mothers and infants** *Physiology and Behavior* **207**:55–63 https://doi.org/10.1016/j.physbeh.2019.04.024

Weijs M. L., Daum M. M., Lenggenhager B. (2023) **Cardiac interoception in infants: Behavioral and neurophysiological measures in various emotional and self-related contexts** *Psychophysiology* :1–13 https://doi.org/10.1111/psyp.14386

Weng H. Y., Feldman J. L., Leggio L., Napadow V., Park J., Price C. J. (2021) **Interventions and Manipulations of Interoception** *Trends in Neurosciences* **44**:52–62 https://doi.org/10.1016/j.tins.2020.09.010

WHO Multicentre Growth Reference Study Group (2006) **WHO Motor Development Study: Windows of achievement for six gross motor development milestones** *Acta Paediatrica, International Journal of Paediatrics* **95**:86–95 https://doi.org/10.1080/08035320500495563

Zmyj N., Jank J., Schütz-Bosbach S., Daum M. M. (2011) **Detection of visual-tactile contingency in the first year after birth** *Cognition* **120**:82–89 https://doi.org/10.1016/j.cognition.2011.03.001

Editors

Reviewing Editor
Jessica Dubois
Inserm Unité NeuroDiderot, Université Paris Cité, Paris, France

Senior Editor
Floris de Lange
Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

Reviewer #1 (Public review):

Summary:

The authors of this study investigated the development of interoceptive sensitivity in the context of cardiac and respiratory interoception in 3-, 9-, and 18-month-old infants using a combination of both cross-sectional and longitudinal designs. They utilised the cardiac interoception paradigm developed by Maister et al (2017) and also developed a new paradigm to investigate respiratory interoception in infants. The main findings of this research are that 9-month-old infants displayed a preference for stimuli presented synchronously with their own heartbeat and respiration. The authors found less reliable effects in the 18-month-old group, and this was especially true for the respiratory interoceptive data. The authors replicated a visual preference for synchrony over asynchrony for the cardiac domain in 3-month-old infants, while they found inconclusive evidence regarding the respiratory domain.

Considering the developmental nature of the study, the authors also investigated the presence of developmental trajectories and associations between the two interoceptive domains. They found evidence for a relationship between cardiac and respiratory interoceptive sensitivity at 18 months only and preliminary evidence for an increase in respiratory interoception between 9 and 18 months.

Strengths:

The conclusions of this paper are mostly well supported by data, and the data analysis procedures are rigorous and well justified. The main strengths of the paper are:

- A first attempt to explore the association between two different interoceptive domains. How different organ-specific axes of interoception relate to each other is still open and exploring this from a developmental lens can help shed light into possible relationships. The authors have to be commended for developing a novel interoceptive tasks aimed at assessing respiratory interoceptive sensitivity in infants and toddlers, and for trying to assess the relationship between cardiac and respiratory interoception across developmental time.
- A thorough justification of the developmental ages selected for the study. The authors provide a rationale behind their choice to examine interoceptive sensitivity at 3, 9, and 18-months of age. These are well justified based on the literature pertaining to self- and social development. Sometimes, I wondered whether explaining the link between these self and social processes and interoception would have been beneficial as a reader not familiar with the topics may miss the point.
- An explanation of direction of looking behaviour using latent curve analysis. I found this additional analysis extremely helpful in providing a better understanding of the data based on previous research and analytical choices. As the authors explain in the manuscript, it is often difficult to interpret the direction of infant looking behaviour as novelty and familiarity preferences can also be driven by hidden confounders (e.g. task difficulty). The authors provide compelling evidence that analytical choices can explain some of these effects. Beyond the field of interoception, these findings will be relevant to development psychologists and will inform future studies using looking time as a measure of infants' ability to discriminate among stimuli.
- The use of simulation analysis to account for small sample size. The authors acknowledge that some of the effects reported in their study could be explained by a small sample size (i.e. the 3-month-olds and 18-month-olds data). Using a simulation approach, the authors try to overcome some of these limitations and provide convincing evidence of interoceptive abilities in infancy and toddlerhood (but see also my next point).

Weaknesses:

- While the research question is timely and the methodology is detailed, there is a critical flaw in the experimental design: the lack of randomization of stimuli due to an error in the programming script. The authors very honestly report this error and have performed additional analyses to investigate its potential impact on the study's results. Unfortunately, I am not fully convinced these analyses provide enough reassurance and I believe the technical error still undermines the validity of the findings, making it difficult to draw meaningful conclusions.

https://doi.org/10.7554/eLife.91579.2.sa2

Reviewer #2 (Public review):

Summary:

This study by Tünte et al. investigated the development of interoceptive sensitivity during the first year of life, focusing specifically on cardiac and respiratory sensitivity in infants aged 3, 9, and 18 months. The research employed a previously developed experimental paradigm for the cardiac domain and adapted it for a novel paradigm in the respiratory domain. This approach assessed infants' cardiac and respiratory sensitivity based on their preferential looking behavior toward visuo-auditory stimuli displayed on a monitor, which moved either in sync or out of sync with the infants' own heartbeats or breathing. The results in the cardiac domain showed that infants across all age groups preferred stimuli moving synchronously rather than asynchronously with their heartbeat, suggesting the presence of cardiac sensitivity as early as 3 months of age. However, it is noteworthy that this preference direction contradicts a previous study, which found that 5-month-old infants looked longer at stimuli moving asynchronously with their heartbeat (Maister et al., 2017). In the respiratory domain, only the group of 9-month-old infants showed a preference for stimuli presented synchronously with their breathing. The authors conducted various statistical analyses to thoroughly examine the obtained data, providing deeper insights valuable for future research in this field.

Strengths:

Few studies have explored the early development of interoception, making the replication of the original study by Maister et al. (2017) particularly valuable. Beyond replication, this study expands the investigation into the respiratory domain, significantly enhancing our understanding of interoceptive development. The provision of longitudinal and cross-sectional data from infants at 3, 9, and 18 months of age is instrumental in understanding their developmental trajectory.

Weaknesses:

Due to a technical error, this study failed to counterbalance the conditions of the first trial in both the iBEAT and iBREATH tests. Although the authors addressed this issue as much as possible by employing alternative analyses, it should be noted that this error may have critically influenced the results and, thus, the conclusions.

https://doi.org/10.7554/eLife.91579.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Reviewer #1 Public:***
>
> *- The authors should carefully address the potential confounding of not counterbalancing the conditions of the first trial in both interoceptive tasks for the 9-month and 18-month age groups. The results of these groups could indeed be driven by having seen the synchronous trial first.*

Upon addressing this comment, we noticed an error in our presentation scripts that resulted in a fixed-experimental design for most of the infants. Therefore, it is crucial to investigate the impact of the fixed-experimental design on our results. We have conducted extensive additional analyses comparing data from infants with the inadvertent fixed design to data from infants for whom the randomization was achieved as intended, which can be found in Supplementary Materials A. In summary, we do not find that the fixed order design had a strong impact on the findings, as we do not find that looking behavior differed systematically between different randomization orders, while also looking patterns across ages and tasks

indicate that we were able to adequately capture variance associated with these features. Further, we have adapted the interpretation of the results across the manuscript to acknowledge the experimental error and its implications on the interpretation of the results.

For instance, on pages 30 and 31 we have added the following paragraphs:

"The data presented in this study holds several limitations. First, due to an error in our experimental scripts we unintentionally used a fixed-order design, in which almost all infants saw the same fixed order of condition (always starting with a synchronous trial), image assigned to condition, and location of the image (left/right) instead of a semi-randomized design. Such a fixed-order design holds several important limitations as visual preferences might be influenced by the experimental design, i.e., the first trial always being synchronous might have influenced a mean group preference. Further, we cannot rule out that mean group preferences were influenced by the stimuli used (as in most cases the same stimuli were used for synchronous/asynchronous trials) or by the location of the image in a given trial (left/right). Still, there is no strong theoretical argument as to why image used or location should have an impact on infants' preferences. The stimuli were selected to be similar to each other, in order not to evoke a piori preferences. To further illustrate the impact of the fixed order design we have conducted several additional analyses, which can be found in Supplementary Materials A, which do not indicate that there was a strong impact of the fixed-order design. Specifically, we find no evidence for systematic differences between infants tested with the fixed design and infants tested with a randomized design.

Despite these limitations fixed-order designs also hold advantages, as they are more suitable to investigate individual differences (Dang et al., 2020; Hedge et al., 2018). When each participant is exposed to the same procedure, individual differences are less likely to be attributed to effects of randomization but are more likely to reflect real differences between participants. Also, when considering the impact of the randomization, one must consider our results in relation to earlier studies (Maister et al. 2017, Weijs et al. 2022, Imafuku et al. 2023), some of which used the exact same stimuli as we did (Maister et al., 2017), with fully randomized designs. Results of these studies indicate no looking times differences depending on the stimulus assigned to each condition or systematic preferences for one of the stimuli."

> *- The conclusion that cardiac interoception remains stable across infancy is not fully warranted by the data. Given the small sample size of 18-month-old toddlers included in the final analyses, it might be misleading to state this without including the caveat that the study may be underpowered. In other words, the small sample size could explain the direction of the results for this age group.*

We agree with the reviewer and explicitly acknowledge this issue now in the discission, p. 23:

"However, due to the small sample size at 18 months the results regarding changes and stability of interoceptive sensitivity in the second year of life must be considered speculative and need to be validated in further research."

> ***Reviewer #1 (Recommendations For The Authors):***
>
> *Below are some comments that the authors may wish to take into account:*
>
> *- Why did the authors choose to apply different statistical analyses across the dataset (i.e. Bayesian t-test is used with the 3-month-old sample, whereas a paired t-test is used for the 9 and 18-month-olds)?*

The use of different statistical analyses was driven by the timeline of the project, as we had to update our initial plans. Due to challenges related to the Covid-19 pandemic, it was not possible to recruit 3-month-old babies for out study at the time we started the data collection.

Thus, we first collected the 9- and 18-month-olds, and the 3-month-olds later. For the 9- and 18-month-old samples we aimed at directly replicating the approach by Maister et al. (2017). However, for the 3-month-olds we wanted to focus more on classification of the strength of evidence in favor/against an effect, taking the results of the equivalence tests for the 9- and 18-month-olds into account.

The following parts have been added to the manuscript to clarify our approach:

Sample (p 33): “The 3-month-old sample was tested after completion of the 9- and 18-monthold samples. Initially, we had planned to start data collection with the 3-month-old sample.

However, due to the Covid-19 pandemic this was not possible.”

Statistical analysis (p. 41): “At 3 months we used a Bayesian paired t-test as the data collection was done after having collected the 9- and 18-month-old samples. Our intention in the analysis of the 3-month-old sample was to focus more strongly on strength of evidence in favor of/against an effect instead of a binary classification for/against an effect.”

> *- I found the way in which sample sizes are reported a little unclear. This may be due to having the Results section before the Methods section (in line with journal requirements), but it would be helpful if the authors could clarify their sample size from the outset. For example, sample size for the 3-month-olds first says N = 80 (page 9), but then it becomes apparent that N = 53 completed the iBEAT and N = 40 completed the iBREATH. I think for the purpose of explaining the results, it might be more helpful to the reader to only know the final sample size and then specify recruited participants and dropout in the Methods.*

We have adapted the description of sample sizes in the Results section. We now only refer to the number of infants included in a given analysis when reporting the results of the analysis. In addition, we have added the following clarification for the MEGA analysis (p. 11): “This approach allowed us to include 135 observations for the iBEATs from 125 infants, and 120 observations for the iBREATH from 107 infants. The sample size differs slightly from our preregistered approach given that we used the same preprocessing approach for the MEGAanalysis for all samples. “

In addition, we now refer to the sample of the MEGA-analysis in the abstract, to make the understanding of our approach more intuitive.

> *- I think the sentence "Interestingly, we find evidence for a positive relationship between cardiac and respiratory perception in our 18-month-old sample" at page 25 could be deleted given that the small sample size of 18-month-olds suggests this result should be interpreted with caution. The authors already explained this in the earlier paragraph (page 24) and simply re-stating this (weak) effect without further elaborating may not be necessary.*

We have removed the sentence.

> *- In multiple places in the manuscript, the authors hint at the association between interoception and certain social and self-related abilities (e.g. joint attention, mirror self-recognition), however, these are not fully elaborated on. Could the authors elaborate on the relation between mirror self-recognition and respiratory interoception (page 30)? Why would the ability to recognise the self-face be associated with the individual's ability to perceive their breathing pattern? How these two processes may be linked is not immediately obvious.*

We have rephrased the sentence on page 30 to highlight that the increase in respiratory perception found in our results happens at a similar age as increases in other domains that might be related to interoception. “A hypothesis to be tested in future research is that developmental improvement in respiratory perception might be related to increases in other domains that show links to interoception. For instance, self-perception matures towards the end of the second year of life and has been conceptually related to interoception (Fotopoulou & Tsakiris, 2017; Musculus et al., 2021). Further, gross motor development may be considered in future research, which drastically matures in the first two years of life (WHO Multicentre Growth Reference Study Group, 2006) and has been shown to be related to respiratory function in children with cerebral palsy (Kwon & Lee, 2014).”

> *- Aren't the 18-month-old infants effectively 19-month-olds? The mean age is 576.65 days, and the age window of recruitment was between 18 and 20 months.*

We have added a sentence clarifying how we refer to the infants age ranges. “To stay coherent, we refer to each age group throughout the manuscript with regard to the lower end of the age range in which we included infants (e.g., we tested infants between 9 and 10 months, but refer to them as the 9-month-old group).”

> ***Reviewer #2 Public:***
>
> *Weaknesses:*
>
> *(1) My primary concern is that this study did not counterbalance the conditions of the first trial in both iBEAT and iBREATH tests for the 9-month and 18-month age groups. In these tests, the first trial invariably involved a synchronous stimulus. I believe that the order of trials can significantly influence an infant's looking duration, and this oversight could potentially impact the results, especially where a marked preference for synchronous stimuli was observed among infants.*

Upon conducting further analyses to address this comment, we noticed an error in our presentation scripts that resulted in the inadvertent use of a fixed-experimental design for most infants. Therefore, we have conducted extensive additional analysis which can be found in Supplementary Materials A. Specifically, we compared data from infants who were tested with the inadvertent fixed design to data from infants for whom the randomization was achieved as intended. Further, we have adapted the interpretation of the results across the manuscript to acknowledge the experimental error and its potential implications for the interpretation of the results.

> *(2) The analysis indicated that the study's sample size was too small to effectively assess the effects within each age group. This limitation fundamentally undermines the reliability of the findings.*

We have added a statement addressing this issue to the limitation section: “The reduced sample size might have impacted the statistical power to detect mean preferences for some age groups. Still, it must be noted that even the smaller sample sizes included were of similar size as used in previous studies on infant interoceptive sensitivity (Imafuku et al., 2023; Maister et al., 2017; Weijs et al., 2023).”

> *(3) The authors attribute the infants' preferential-looking behavior solely to the effects of familiarity and novelty. However, the meaning of "familiarity" in relation to external stimuli moving in sync with an infant's heartbeat or breathing is not clearly defined. A deeper exploration of the underlying mechanisms driving this behavior, such as from the perspectives of attention and perception, is necessary.*

We have adapted the respective paragraph in the discussion to clarify the term familiarity, and to also address that other aspects of attention and perception, might be relevant (p. 25):

“In this context familiarity might refer to the infant’s perception of congruence between internal signal and external stimuli which might drive the infant’s attention. Specifically, the synchronous condition should be easier to process due to the intersensory redundancy and predictability between interoceptive and external signals. “

“However, it is important to consider that other cognitive and attentional mechanisms could also influence these responses.”

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *Introduction:*
>
> *(1) The relevance of respiration to self-regulation and social interaction was not clearly described.*

We have rephrased the relevant section to highlight that the increase in respiratory perception found in our results happens at a similar age as increases in other domains that might be related to interoception. “A hypothesis to be tested in future research is that developmental improvement in respiratory perception might be related to increases in other domains that show links to interoception. For instance, self-perception matures towards the end of the second year of life and has been conceptually related to interoception (Fotopoulou & Tsakiris, 2017; Musculus et al., 2021). Further, gross motor development may be considered in future research, which drastically matures in the first two years of life (WHO Multicentre Growth Reference Study Group, 2006) and has been shown to be related to respiratory function in children with cerebral palsy (Kwon & Lee, 2014).”

> *(2) In the last line of page 5, it might be more appropriate to use the term "meta-cognitive awareness" instead of "meta-perception," as the latter can refer to a different concept.*

We have changed the word as recommended.

> *(3) The authors predicted a positive correlation in sensitivity between the cardiac and respiratory domains, despite studies in adults suggesting these are not related. How did the authors arrive at this prediction, and how do they interpret the results showing a correlation only in 18-montholds, the age group closest to adults in this study?*

We have elaborated on our reasoning for our prediction (p. 7): “Adult cardiac and respiratory interoception paradigms typically use two conceptually different paradigms. Thus, null results in the adult literature might be due to the unique characteristics of those paradigms.”

Further, we have expanded on this result in the discussion (p. 24): “Still, we find a relationship between cardiac and respiratory signals in the oldest sample tested here, the 18-month-olds, which is closest to adults. Although this effect needs to be interpreted with caution due to the small sample size, this might indicate that using conceptually similar experimental paradigms might be a promising avenue to investigate relationships between different interoceptive modalities in adults.”

> *Results:*
>
> *(4) Please provide the descriptive statistics (means and standard deviations of looking time) for each independent condition, especially for the 18-month and 3-month age*

> *groups where this information is missing and only differences in looking times between conditions were mentioned. Furthermore, since the asynchronous condition includes both fast and slow stimuli, descriptive statistics for each should be included to help readers determine whether effects are due to synchronicity or stimulus speed.*

We have added the information on mean and sd of looking times to synch and asynch trials to the results section. Mean looking times to both types of asynchronous trials can be found in supplementary materials C. We have added the information about standard deviations to this part.

> *(5) Regarding the MEGA analysis for iBEATs, where a main effect of condition was found (OR = 1.13, t(1769) = 2.541, p = .011), are these t-value and p-value based on the GLMM analysis, or did the authors conduct a separate t-test? This query arises because the p-value of the main effect differs from that in Table 2. Also, is it conventional to present GLMM results in the manner of Table 2, comparing specific level combinations (i.e., synchronous condition and 3month age group), instead of listing main effects and interactions?*

Thank you very much for pointing out that the results of the GLMM were not reported as precise as possible, which might lead to confusion over the presented p-values. The main effect of condition refers to a post-hoc comparison using estimated marginal means from the GLMM across all age groups, while Table 2 refers to the main effect of condition for age group 3 months.

To make the results more accessible we have restructured parts of the manuscript following your suggestions: In the main manuscript we now focus on the interaction effects for condition and age, as well as the post hoc comparison, while we now report null-full model comparison, and tables for all age groups in the supplements.

We have added the following clarifying sentences to the manuscript, p. 12:

“In reporting these results we focus on whether we found evidence for interactions between age groups, and whether we found evidence for a general effect across age groups. In-depth results and tables can be found in Supplementary Materials C.

[…]

Next, we computed post hoc comparisons using estimated marginal means from the MEGAanalysis across all age groups to investigate whether we find indications for a similar effect across ages.”

> *(6) I am confused about the results indicating a significant effect of condition for the iBREATH dataset excluding 18-month-olds (Table 5, OR = 1.15, t(1050) = 2.397, p = .017), as the description in Table 5 suggests no statistical significance (p = .070). The decision to exclude the 18-month group seems arbitrary, particularly since the age-by-condition interaction was not significant in the GLMM across all three age groups.*

Thank you very much for the comment, we have removed the analysis excluding the 18-month-old group

> *(7) Regarding the relationship between cardiac and respiratory interoceptive sensitivity, the statement "However, we found a significant interaction between iBEATs scores and age at the 18-month level" (p16) seems unclear. Clarification is needed, as mentioning age interaction at a specific age stage is unusual. A pairwise comparison between 3 and 9 months should also be included.*

Thank you for pointing out that the results could be presented more clearly! Similar to the other MEGA analyses we have put detailed tables of the results of the beta regression in the supplements and have kept a single table with the most important results in the main manuscript. Further, we have clarified the text passage as follows: “However, we found a significant interaction between the iBEATs scores and age, specifically comparing the 3- and 18-month-old groups ($\beta = 3.13$, $SE = 1.41$, $p = .027$). This interaction indicates that the relationship between iBEATs and iBREATH scores changes between 3 and 18 months of age.” Also, we have now included a pairwise comparison between 3- and 9-month-olds.

> *Discussion:*
>
> *(8) In pages 27-28, the authors discuss the results of the specification curve analysis, but there is no explanation for the 7th entry (statistical analysis) in Table 9. This entry seems particularly important.*

We did not include an explanation for the 7th entry, as the impact of the statistical test used was comparatively less pronounced. However, to acknowledge this result we have added the following sentence to the discussion: “Moreover, the statistical test used (paired t-test vs linear mixed model, Table 9, 7th entry) had a rather small impact on the results. However, given the large number of analyses conducted, this might be related to not being able to precisely formulate the model to fit the complexity of the data for each specification.”

> *Methods:*
>
> *(9) What were the colors of the stimuli?*

We have added the colors of the stimuli to the methods section. Further, the stimuli can be found in the osf project associated with the manuscript.

> *(10) The percentage of trials excluded during preprocessing should be stated. Additionally, the number of trials included in the statistical analyses for each condition (including synchronous, fast, and slow) should be detailed separately.*

We have added information on numbers of trials completed and included in Table 7.

https://doi.org/10.7554/eLife.91579.2.sa0