Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Effects of menstrual cycle phase on interoception: A daily experience sampling study across two cycles

Rosie Donaghy

r.donaghy@surrey.ac.uk

University of Surrey

Matteo Lisi

Royal Holloway University of London

Jenny Hall

University College London

Rebecca Brewer

Royal Holloway University of London

Jennifer Murphy

University of Surrey

Article

Keywords: Interoceptive accuracy; Interoceptive Attention; Menstrual Cycle; Sex Differences; Intra-Individual Variability; Experience Sampling Method

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11612288/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Effects of menstrual cycle phase on interoception: A daily experience sampling study across two cycles

Rosie Donaghy[1]*, Matteo Lisi[2], Jenny Hall[3], Rebecca Brewer[2] & Jennifer Murphy[1]

[1.] School of Psychology, University of Surrey, Guildford, UK

[2.] Department of Psychology, Royal Holloway University of London, Egham, UK

[3.] Institute for Women's Health, University College London, UK

*Dr Rosie Donaghy

School of Psychology

University of Surrey

Stag Hill, Guildford, UK

r.donaghy@surrey.ac.uk

Abstract

Interoception – the processing of internal bodily signals – has been related to numerous aspects of mental and physical health. Sex differences have been reported in objective interoceptive accuracy and self-reported interoceptive attention, yet the factors underlying such differences remain poorly understood. One theory is that these differences arise from increased amount of physical and hormonal change females experience across the lifespan. The menstrual cycle provides a compelling test of this hypothesis and here we present the first investigation of interoception across the menstrual cycle. Twenty females completed daily interoceptive accuracy and attention assessments across two consecutive cycles, alongside matched exteroceptive control tasks (ISRCTN registration: 99520723). Cycle phases were determined using the counting method, with ovulation confirmed via cervical mucus ratings and basal body temperature. Interoceptive accuracy was higher during the periovulatory phase compared with the perimenstrual phase. Interoceptive attention was highest during the perimenstrual phase compared to all phases, though fully explained by physical symptom burden. Cycle-related changes were unique to interoception and emerged only within a complete menstrual cycle, highlighting the potential role of between-cycle hormonal variability. These data provide the first evidence that interoception varies across the menstrual cycle, with implications for sex differences in interoception and mental health.

Keywords: Interoceptive accuracy; Interoceptive Attention; Menstrual Cycle; Sex Differences; Intra-Individual Variability; Experience Sampling Method

Introduction

Interoception, the processing of internal bodily signals, has been linked to both physical and mental health conditions. Atypical interoception – for example, unusually high or low interoceptive accuracy and attention – has been reported across numerous disorders, where it is thought to explain both transdiagnostic and disorder-specific symptoms (for reviews see [1–3]). Sex appears to be one source of individual differences. Males tend to outperform females on laboratory-based tasks of interoceptive accuracy (e.g., cardiac, respiratory [4,5]) and tend to rely more on interoceptive vs exteroceptive cues when gauging their internal states outside of the laboratory [6]. In contrast, females consistently allocate greater attention to bodily states [7,8], report more somatic complaints [9,10] and are more negative when evaluating bodily signals [11,12]. In combination, it has been suggested that these patterns may partly explain sex differences in vulnerability to common mental health conditions linked to interoception that are often more prevalent in females (anxiety, depression [13–15]).

The potential mechanism(s) underlying sex differences in interoception remain unknown. Proposed explanations include functional and anatomical differences; stress; signal strength; and socialisation [5]. However, one common but previously unexamined theory is that sex differences arise from the greater degree of physical and hormonal change that females typically experience across the lifespan [13,14]. Indeed, although often conceptualised as a 'stable trait of an individual' [16], evidence does suggest state-variation in interoceptive processing [17]. Despite the theoretical and clinical importance of understanding such patterns of change [18], the factors underlying within-person variability remain underexplored.

The menstrual cycle provides a compelling first test of the hypothesis that some within-person variability in females may be attributable to hormonal fluctuations [14]. The healthy menstrual cycle produces circulating hormonal fluctuations that influence numerous physiological systems from which interoceptive signals arise (e.g., cardiovascular, gastric [19,20]). In the most studied interoceptive domain – the cardiovascular domain – the menstrual cycle has been associated with changes in resting heart rate (HR) [21,22] and heart rate variability (HRV) [23–27], factors that have also been linked to individual differences in cardiac interoceptive accuracy [28,29]. Yet, beyond direct physiological effects that may contribute to variability in interoceptive accuracy, cyclical hormonal shifts may also produce frequent mismatches between actual and expected bodily states due to variability in the sensory environment. This could result in the heightened attention to, and reduced propensity to utilise, internal signals observed in females [4].

Overall, there is good reason to expect that interoceptive accuracy and attention may vary across the menstrual cycle, yet no studies have investigated this. Although the menstrual cycle presents complex methodological challenges [30], it provides a valuable model for testing the hypothesis that hormonal change may contribute to within-person variability in interoception. In the current pre-registered study, we employed daily experience sampling methodology (ESM) and ecological momentary assessment (EMA) to intensively sample interoceptive accuracy and attention across two consecutive menstrual cycles in a sample of twenty healthy women using a combination of objective tasks and self-report measures, and wearable technology. To isolate interoception specific effects, all interoception measures were administered alongside daily non-interoceptive control measures. Based on the aforementioned literature, we hypothesised that interoceptive accuracy and attention

would vary across menstrual cycle phases, and that any cycle-related effects would be unique to interoception.

Methods

This pre-registered study was part of a longitudinal investigation into interoception and well-being across the menstrual cycle (ISCRTN 05/11/24: 99520723; OSF 14/05/25: https://osf.io/fsvyu/overview?view_only=785bf6d9bfe7440091e391fe492d98c2). Several methodological choices deviated from the initial design. We initially pre-registered that participants would complete tasks during estimated menstrual phase windows. However, following recent studies [31] and methodological recommendations for studying the menstrual cycle [30], we decided instead to adopt a daily experience sampling design. This involved substantially more repeated measures, and we adjusted our statistical approach and sample size accordingly (see Analysis Strategy). All other deviations from the planned protocol are detailed in text.

Participants

Sample size was determined a priori via power analysis. For a repeated-measures design with *N*=5 daily observations per menstrual phase (two menstrual cycles, 18% missing data), >80% power could be achieved to detect a medium effect size (f = 0.25; alpha = .05) with *N*=18 participants (estimated power 81.3%). Given anticipated dropout, we recruited 20 participants (estimated power 85.2%). Simulations were based on ICC values from a previous study [17] where performance on a similar heartbeat detection task was assessed at three time points in a pooled sample of females and males (ICC = .38).

Twenty female participants (M_{age} = 27.7 years; 20-36; N = 19 identifying as woman, N = 1 non-binary) with regular menstrual cycles ($M_{cycle\ length}$ = 27.82 days, 22-35, SD = 2.06) were recruited using flyers and social media within the Greater London/Surrey area. To participate, participants were required to own an iPhone that met software requirements for tasks (≥XS model, ≥iOS 16) and to have been actively tracking their menstrual cycles using a smartphone application (Clue or Apple Health) for at least the last 2 consecutive menstrual cycles. This prior tracking was a requirement that was pre-registered as optional but was considered essential for ensuring that 1) participants' cycle length fell within recommended limits for estimation (25-35 days [30]) and 2) variability was low enough to enable us to accurately allocate participants to specific starting phases (<3 days cycle variability; see Procedure for explanation). Only participants meeting these criteria were invited to participate. Eligibility was verified by assessing cycle characteristics based on menses start dates from participants' most recent consecutive cycles over a maximum 6 cycles (M=5.25; SD = 1.25; range = 2-6). A cut-off of 6 cycles was used to align with eligibility criteria requiring no hormonal contraception use in the preceding 6 months, which ensured cycle history reflected naturally occurring cycles. To ensure accuracy, creation dates were checked to verify that participants had entered the dates in the past (i.e., not at sign up) as prospective logging is more reliable. Further exclusion criteria included any known conditions affecting hormones (e.g., thyroid disorders), history of pelvic complaints (e.g., endometriosis, uterine fibroids), pregnancy/breastfeeding in the last 12 months, and a score ≥8 on the physical subscale of the Anxiety Sensitivity Index [32] (ASI-3). This ASI-3 cut off was used to reduce the likelihood that participants would experience distress when

attending to interoceptive signals in objective tasks (full exclusion criteria is listed in Supplementary Information; SI). Demographics can be found in SI (Table S1).

The study was conducted between September 2025 and March 2026. Participants were compensated up to £600, with the exact amount determined by engagement (£10 per daily task completion; ~1 hour per day). All procedures were approved by the University's Ethics Review Board (FHMS 23-24-203 EGA/ERM ID 630).

Materials

Questionnaire items and task descriptions can be found in Supplementary Files (Questionnaire+Task list.docx). All notification prompts were managed via Samply [33], behavioural data for smartphone-based tasks was collected in Firebase, and questionnaire responses collected in Qualtrics (Provo, UT).

Baseline survey

A baseline survey at enrolment included demographic questions and a series of standardised questionnaires, including pre-menstrual symptoms, alexithymic traits, paranoid thoughts, and more. Descriptive statistics on a selection of these data are included in SI Table S1 to characterise the sample but are not reported in any analyses as they are not the focus of this paper.

Daily tasks

Phase Adjustment Task (PAT): Participants completed the PAT [34,35] via smartphone application (Figure 1A). During the task, participants hold their finger over the camera and flash while heartbeats are recorded via photoplethysmography. A 180 second baseline recording of heart rate (HR) preceded 4 practice trials and 20 experimental trials. In the first two practice trials participants were asked to match

two tones (as per the Screener task; described below) and in the second two practice trials participants were asked to match tones to heartbeats as per the 20 experimental trials.

On each experimental trial, a series of 40 ms tones are presented that are triggered by heartbeats. These tones are presented at the rate of each participant's HR but are phase-shifted relative to the heartbeats. The initial offset is randomised across trials and the point on the dial where heartbeats and tones are synchronous is randomised across trials. Participants are instructed to place their finger on a virtual dial and rotate it to adjust the tones until they perceive the tones to be synchronous with their heartbeats. Once they are ready to submit their response, with their finger still on the dial, they use another finger of that hand to tap a lock button to confirm their response and slide their index finger from the dial to a submission button. Holding this submission button for 2 seconds submits their response and activates the next trial. To discourage 'cheating' (i.e., feeling their pulse with one hand during the task), a feature is implemented in the application which detects if the participant removes their finger from the dial before submitting their response. A trial is aborted if their finger is removed from the dial for longer than 1000 ms. In this case, the trial is discarded and repeated with a new starting phase. During practice trials, there are no limits on how often a participant triggers this "cheat-avoidance" feature. During experimental trials, participants can attempt a maximum of 40 trials (i.e., 20 valid trials and 20 "cheat-avoidance" trials), before the task terminates. A session is considered invalid if a participant exceeds this limit (see [35] for details). Notably in our sample the cheat avoidance feature was activated only in 3.3% of completions of the PAT (far fewer than the tone-flash control task (TCT) described below; 23.2% of completions).

During the task, heartbeat detection was carried out using an automated peak-detection algorithm applied to the PPG waveform. To minimise noise, detected inter-beat intervals (IBIs) shorter than 400 ms or longer than 1.8 s were discarded. Additionally, any IBI that differed by more than 30% from the previous valid IBI was rejected to minimise the influence of transient artefacts or ectopic heartbeats. When an IBI was rejected, the algorithm accepted the next detected period that met the criteria described above. Offline, we employed several further quality checks [35]. To extract HR and HRV from baselines, we first screened IBIs ensuring that they corresponded to predefined lower and upper bounds (0.5–1.5 s; 40-120 bpm). For HRV, we identified the longest continuous segment of IBIs within this valid range to calculate RMSSD ensuring a minimum length of 30 seconds to ensure reliability [35] (89.3% of samples had valid PAT HRV data). For task trials, quality filters removed cheat avoidance trials, duplicates, practice trials, trials with zero dial movements or ≤3 recorded heartbeats, as well as trials where HR fell outside of physiologically plausible limits for recordings at rest (40–120 bpm). Completions with fewer than 17 valid trials were excluded from further analysis. As pre-registered, to balance trial counts across participants and completions, only the first 17 valid trials were retained for scoring. Engagement metrics were obtained for each session (i.e., time spent on trials, number of dial-turns per trial). Participants overall completed 645 PAT sessions across an average period of 55.9 days (*SD* = 6.14, range 46-70) before any post-hoc exclusions (see below). Note that confidence ratings are collected after the task but are not relevant to the present investigation.

As pre-registered, scoring followed the framework introduced by Palmer et al [35] which captures non-random responding based on the extent to which participants select a consistent delay or phase over trials. For both scoring methods, consistency

was defined using delay values as periodic functions of the heartbeat period where d is the delay and p is the heartbeat period.

$$consistency(d,p) = \frac{1}{n} mod\left(\sum_{j=1}^{n} e^{i2\pi\frac{d_j}{p}}\right)$$

Fixed-period (delay-based) scoring assesses how consistently a participant selects absolute delays across trials. For this approach, all delays are interpreted as an absolute value using a constant participant-specific period, calculated as the participant's overall median of the final sequence median IBIs. Because variability in IBIs affects the chance distribution under this model, an individualised chance threshold is derived from 10,000 simulations per participant using the participant's median IBI[1]. In this instance, a participant would score highly if they consistently selected a delay of 200 ms for example, as it suggests they perceive the tones and heartbeats as synchronous at this delay from the R-peak.

Phase-based consistency scoring assesses whether participants consistently select the same point in their cardiac cycle across trials, using a trial-by-trial heartbeat period and a group-level chance threshold derived from 100,000 simulations of random delay sequences and randomly sampled heartbeats (0.5s-1.5s; note that a group level cut off is used as IBI does not affect chance thresholds using this approach [35] where IBI is <.36[1]). For this approach, delays are expressed as phase angles relative to each trial's median inter-beat interval (IBI) at the final dial position (p). Under this approach, a participant would score highly if they consistently

[1] N = 11 participant sessions were identified as having IBI variability ≥.36 in the PAT and N = 13 for the TCT, and were removed following recommendations from Palmer et al[35]: Note increasing the number of simulations from 10,000 to 100,000 simulations for fixed delay scoring produces no differences in classification but increases processing speed.

selected delays that correspond to the same proportion of their cardiac cycle (e.g., 25%).

To combine these approaches, a binary score is used – interoceptive or non-interoceptive – reflecting above or below chance performance on either scoring methods. While it is recommended to use the binary classification scores [35] – and these are included in line with our pre-registered approach – as binary scores result in a substantial loss of power and we are interested in within person change (where the degree of change is relevant even in those scoring above chance), we also included unplanned analyses focusing on continuous consistency scores derived from the fixed-period (delay) scores. Importantly, we focused on delay-based consistency scores because phase-based and delay-based continuous scores are not directly comparable and therefore cannot be combined [35]. In addition, maximum phase-based consistency scores are affected by IBI variability, meaning that only the binary score is valid for phase-based scoring [35]. Therefore, in follow up analyses we retained only delay-based continuous scores, excluding participant sessions classified as interoceptive solely based on phase-based consistency (3.96% of sessions). Each participant (N = 20) contributed on average 24.9 valid PAT sessions (SD = 7.62; range = 10-37) and overall participants were classified as interoceptive on 23% of occasions after data cleaning and screening on TCT performance.

Tone-Flash Control Task (TCT): A matched non-interoceptive control task was developed for the purposes of this study (Donaghy et al. [in prep]; Figure 1B). This task is matched to the PAT, but in place of heartbeats, participants matched tones (40 ms) to a virtual circle which pulses (expands and shrinks) for 100 ms and remains for 40 ms. To match variable presentation in the TCT to variable

presentation on the PAT, the pulsing shape was triggered by the heartbeat and heartbeat was recorded throughout. Participants were not informed that the flashing shape was linked to their heartbeat. Participants read the same matched instructions (only “heartbeats” is replaced by “flashing circle”; see Donaghy et al. [in prep]; Figure 1C), they rotate the same virtual dial and the procedure is identical for trial submission (including the “cheat-avoidance” feature outlined above and cleaning; Figure 1D). Scoring followed the same framework as the PAT, as described above. As with the PAT, binary scores included all sessions, including phase-only classifications where IBI variability was <.36 (N = 13 ≥ .36). For continuous TCT scores, phase-only classification sessions were excluded as per the PAT, and fixed-period delay consistency scores were used. Before data cleaning and screening TCT performance, participants were classified as passing the TCT on 77.7% of occasions. Following screening, each participant (N = 20) contributed on average 25.25 valid (passed) TCT sessions (SD = 8.90; range = 7-39).

Screener Task: In the orientation, all participants completed a supervised screener task (see [34,35]). This was identical to the above tasks, but participants instead matched two tones (40 ms, 30 ms), with one tone (30 ms) unbeknownst to participants representing their heartbeat. This ensured participants understood instructions and could fulfil the basic requirements of the PAT and TCT before commencing the study. Those exhibiting poor performance on the screener task (N = 2) were replaced. Scoring was identical to the PAT and TCT as described above.

Following the above processing, the final number of observations consisted of 514 valid PAT observations and 520 TCT observations.

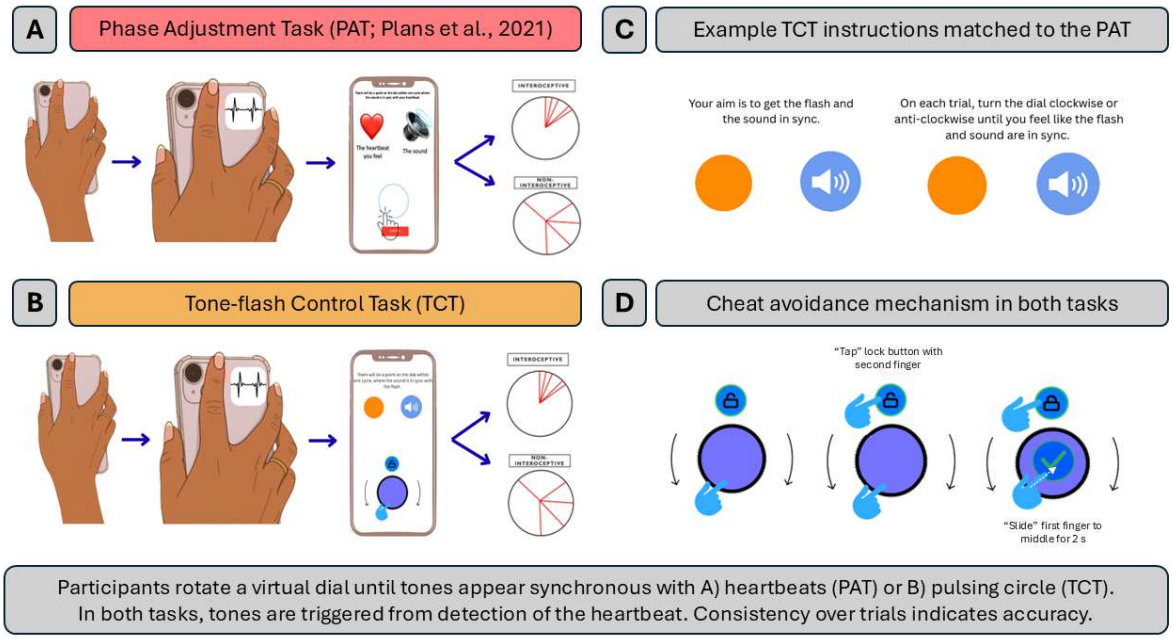


Figure 1. Illustration of the Phase Adjustment Task **(A)** and novel Tone-flash Control task (TCT) developed for the current study **(B)**, with example of matched instructions **(C)** and example of cheat avoidance mechanism implemented in both tasks to **(D)**.

Cervical mucus ratings, temperature and confounds: Each morning, participants rated their cervical mucus qualities using the following scale; *dry/no discharge (1)*, *sticky (2)*, *wet and watery (3)*, *creamy (4)*, *very wet and stretchy, resembles raw egg white (5)*, *menstruating i.e. on my period (6)*, *spotting (7)* and received guidance on this in the orientation (see Supplementary Files: Orientation-Slides.pdf). The ratings were developed from the Creighton Model and Billings Method for tracking the sensation, texture and appearance of cervical mucus qualities [36], and are in line with existing cervical mucus options available in menstrual cycle tracking apps (i.e., Apple Health: Dry, Sticky, Watery, Creamy, Egg White, Period, Spotting).

Participants recorded temperature from the mouth using a Numark Health thermometer and reported potential confounds affecting basal body temperature (BBT) in this morning questionnaire (alcohol use and illness [37,38] in the last 24

hours; yes/no). Temperature was also recorded by an Apple Watch worn over night to record BTT from the wrist (Series 10, version OS 11+). When estimating BTT, we focused on the Apple Watch recordings as temperature is sampled every 5 seconds during sleep and logged on wake up and such dense sampling is likely to provide more precise estimates compared to mouth temperature taken once per day. BBT estimates for days where confounds were reported in the past 24 hours (e.g., alcohol/illness) were excluded. In addition, every morning participants reported the presence of bothersome symptoms (yes/no) based on a combination of items from the Somatic Symptoms Scale 8-item (SS8 [39,40]) and the Premenstrual Symptoms Screening Tool (PSST [41]). For this study, we restricted to physical symptoms, and a total daily count was calculated after excluding affective items from the PSST (e.g. 'tearfulness/sensitivity to rejection').

Interoceptive attention ratings: Four items adapted from the Interoceptive Attention Scale (IATS [42]) assessed self-reported attention to heartbeats/breathing, hunger/thirst, urination/defecation, and stomach/gastric sensations in the past five minutes (yes/no). We selected these items based on relevance, as well as previous work suggesting that these are the most engaged interoceptive signals [43]. A yes/no response was used instead of a rating scale as previous work suggests that this is the most appropriate approach [43], and we collapsed items that may show redundancy [44] to reduce participant burden (e.g., heartbeats and breathing). A daily total score was calculated as the sum of all items.

As interoceptive attention may have different qualities [45,46], after each item participants were also asked to rate the quality of their interoceptive attention. This was done using a scale from Very Negative/Unhelpful (1) to Very Positive/Helpful (5) in response to the following item "How would you describe this attention/lack of

attention to… (e.g., heartbeats/breathing)". A variant of this question was asked regardless of whether participants reported attention to bodily signals or not (e.g., "lack of attention") to ensure a matched participant experience, but we only explored the affective quality of interoceptive attention where participants responded "Yes". For this we took all occasions where participants responded 'Yes' to any of the four interoceptive attention items (1 = Very Negative/Unhelpful to 5 = Very Positive/Helpful).

Exteroceptive attention

One item assessed daily self-reported exteroceptive attention [43]. As per the interoceptive attention questions, this question asked participants to report (yes/no) whether in the last 5 minutes they were paying attention to auditory signals (e.g., listening to music, the television or a podcast, as well as things in the background like birds singing, transport noises, the sound of people talking). This item was included to test the specificity of menstrual effects on interoceptive attention, given that cycle-related changes may be observed in general attentional processes rather than interoception specifically. As we did for interoceptive attention, we assessed the affective quality of exteroceptive attention by following Yes/No responses with a question to rate the quality of their exteroceptive attention. Scoring was identical to the procedure described above for interoceptive attention.

Weekly questions

To monitor ongoing participation, each week participants were sent one notification to report on changes in contraception use (emergency or long-term), pregnancy (suspected or confirmed) and vulva/mucus symptoms or medication (yes/no). New symptoms (e.g., odour, soreness, itching) and medication use (e.g., antibiotics) were

collected as additional information for post-hoc verification of the sampled menstrual cycles in case of unexpected characteristics (e.g., atypical mucus reporting). There were no reports of changes to contraception, pregnancy or medication. Four reports of vulva changes (vulva itching, irritation or redness *N*=2; unusual change in the consistency i.e., thick, white; *N*=2) were not considered having an influence on the sampled menstrual cycles because they otherwise followed the expected patterns and did not disrupt ovulatory indicators.

Other ratings

Several other daily and weekly questionnaires were included (see Supplementary Files: Questionnaire and Task list.docx) but are not reported here as they are not relevant to the questions explored in the present study.

Procedure

Participants who passed the eligibility screening survey completed the demographic questionnaire and were invited to attend a comprehensive individual orientation with the lead researcher. Participants completed the screener and TCT. As pre-registered, those who passed were invited to continue and were instructed to complete the PAT later that day. After the tasks, participants watched an instructional video on how to track and report cervical mucus and answered questions to consolidate understanding (see in Supplementary Files: Orientation-Slides.pdf). Participants were given an Apple Watch configured to record temperature each night and were enrolled in the notification app which began prompting questions and task reminders the next day. For the first 5 days, only morning prompts were sent to familiarise participants with the procedures for temperature recordings and reporting of cervical

mucus, and to allow the Apple Watch to establish BBT. From the sixth day, the remaining prompts began for all other daily questions and tasks (described below).

Using their previous menstrual cycle tracking data (described above), participants were enrolled ensuring that the sixth day of the study fell on a specific menstrual cycle phase (*N*=5 per phase: perimenstrual, midfollicular, periovulatory, midluteal). This counterbalancing was deemed essential to ensure that start date was not confounded with menstrual cycle phase as it was yet not known whether any of the interoceptive measures would exhibit practice/fatigue effects as no prior studies have so densely sampled interoceptive ability across such a protracted period. Allocation to starting phases was not randomised but was instead informed by the participant's next cycle phase and orientation availability. Note that all participants completed measures densely sampled in each phase on two occasions (i.e. from two cycles). However, due to counterbalancing, observations could span phases from three separate cycles, such that each participant contributed both a complete cycle sequence (perimenstrual to midluteal) and cross-cycle observations (e.g., perimenstrual from cycle 1 and remaining phases from cycle 3), hereafter referred to as a merged cycle.

For two menstrual cycles, participants received a series of prompts to complete measures throughout the day (outlined above): one was sent each morning at their specified average wake-up time, followed by a reminder 1 hour later (morning questions; e.g., temperature; cervical mucus tracking). A 10-hour window for completion was given before this expired. Two further notifications were sent randomly between 10am-8pm. One prompt asked participants to report on their interoceptive attention, while the other asked for reports of exteroceptive attention (as well as other measures; see attention sets in Questionnaire and Task list.docx).

The order of these prompts was randomised across the study, and participants were given a 2-hour window to respond before links expired. Each evening (randomly between 8.30pm-9pm), participants were also reminded to complete the TCT and PAT before the end of day. They were instructed to complete the tasks at least every 2 days to meet the study requirements. Engagement was monitored throughout and all participants were sent frequent reminders by email to alternate the order of the tasks (TCT/PAT first). To further ensure compliance, each week participants uploaded their Apple Health data to ensure participant compliance with the daily Apple Watch and cervical tracking.

Data availability statement

Scripts for analysis can be found at https://osf.io/fsvyu/overview?view_only=785bf6d9bfe7440091e391fe492d98c2 and data can be found at Figshare (*link inserted on acceptance of paper*).

Analysis strategy

All analyses were carried out in R version 4.4.3.

PAT scoring

Although we utilised all available data for analyses of the PAT and TCT, as pre-registered and recommended [35] our primary analyses of PAT scores are restricted to sessions with a corresponding TCT pass (though for transparency results from the full sample are provided in SI and referred to in the main text). This restriction ensures that differences in PAT performance are not due to an inability to match stimuli or disengagement in the tasks more generally (given that participants were completing both tasks daily). Thus, when examining PAT scores, we included all sessions where there was a corresponding TCT pass for that session (including

phase-only classifications) resulting in a total of 514 PAT sessions (21.6% TCT sessions were excluded).

Menstrual phases

Daily data were allocated to four menstrual phases following prior recommendations [30]. Phases were defined using the forward and backward counting method from menses onset and next observed menses collected from daily tracking data throughout the study: perimenstrual (-3 to +2 days; low estradiol and progesterone levels), midfollicular (days 4-7; low progesterone, estradioal increasing), periovulatory (-15 to -12 days; estradiol peaks and falls, progesterone increasing), and midluteal (-9 to -5 days; rising estradiol and progesterone) (see Figure 2A). Ovulatory cycles were verified in all data by peak BBT and sustained rise following nadir (see Figure 3A) and/or the presence of stretchy cervical mucus +/- 2 days periovulatory window (see Figure 3B).

As we employed daily experience sampling, participants completed measures on transitionary days between phases. Prior studies typically discard these data which does not fall within cycle windows, though this represents substantial data loss (in the present study 36.2% of observations for the PAT). Whilst there is currently no agreed method for making use of these data, to maximise power we used linear interpolation to assign proportional phase weights to data obtained during transitional days (weights summing to 1; see Figure 2B). Days falling within a pre-defined phase (as above) were assigned a weight of 1, whereas, for example, observations falling on day 3 (a transitionary day between the perimenstrual and midfollicular phases) would contribute equally to the perimenstrual and midfollicular phases (assigned weights 0.5) but not contribute to the periovulatory and midluteal phases (see

Supplementary Files: Exmaple-Phase-Weighting.csv). Given that ovulation is unlikely before cycle day 8 (even in the shortest cycles i.e <21 days) [24], periovulatory weights were constrained to zero prior to day 8, with interpolation beginning from day 8 onwards and the preceding midfollicular phase always assigned weight 1 due to it being defined as days 4-7. Across two menstrual cycles, this yielded a final sample of up to 520 observations (*N* = 514 PAT, 520 interoceptive attention). Note that due to this approach, rather than using cycle phase as a grouping factor in models, each phase is entered as a fixed effect containing continuous bounded weights between 0 and 1.

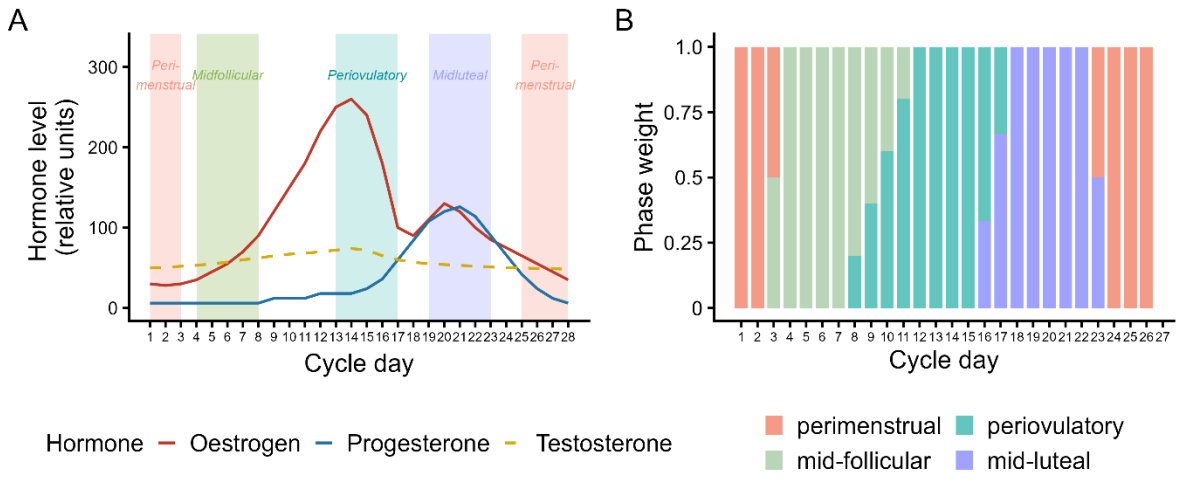


Figure 2 A) Illustration of key hormonal shifts in a typical 28-day menstrual cycle, corresponding to the four menstrual phases highlighted. Values are in relative units and based on reference values from Stricker et al [47]. **B)** Example of weighted phases from one participant across a single complete menstrual cycle, demonstrating phase weights for transitionary days between the phases. Full bars = days entirely within a phase window. Split bars = transitionary days.

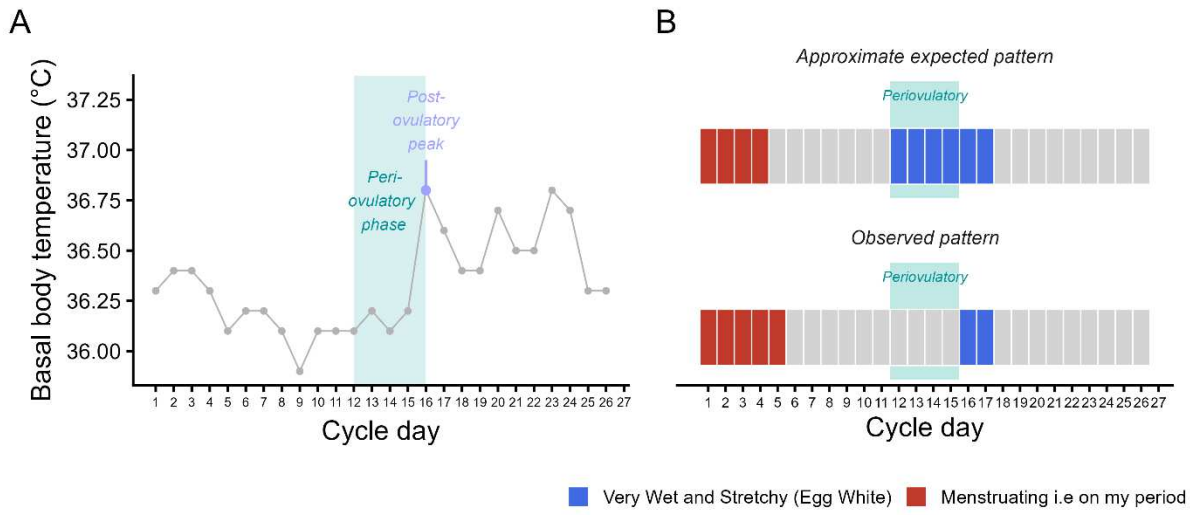


Figure 3. A) Example of BBT wrist temperature readings and **B)** cervical mucus ratings from the same participant in a single complete menstrual cycle. Turquoise shading highlights the periovulatory phase that precedes a temperature peak following ovulation, and ratings indicating more fertile cervical mucus; 'Very wet and stretchy, resembles raw egg white' (in blue) occurring on Days 16 and 17, falling in the expected approximate window.

Statistical models and data checks

As pre-registered, we used multi-level models and selected a series of generalised-linear multi-level models because the distributional properties of our outcome variables were not well suited to Gaussian assumptions. Specifically, posterior predictive checks of initial Gaussian models revealed insufficient fit, with predicted values falling outside the theoretical bounds of several outcomes (e.g., PAT consistency scores bounded between 0 and 1). We therefore selected model families appropriate to each outcome's distribution, including Beta, Bernoulli, and cumulative ordinal regression, implemented within a Bayesian framework, which provide a robust way to quantify uncertainty around the effects of interests. These models compute the posterior probability of different effects given the observed data, resulting in a distribution of possible values for the parameters [48]. From the

posterior distribution it is possible to derive useful scalar summaries, such as the probability of direction (*pd*), which reflects the probability that a parameter is strictly positive ($\Delta > 0$) or negative ($\Delta < 0$). A *pd* between 97.5% and 99.95% corresponds approximately to two-sided p-values between .05 and .001 in common frequentist approaches [49]. However, the primary criterion for a meaningful directional effect is whether the 95% credible interval (CrI), i.e., 95% of the posterior probability, does not include zero (or include one for binary models). The *pd* is used as an additional measure of evidential strength, with values ≥ 95%-99% or 0-5% considered indicative of moderate-to-strong evidence for a directional effect, depending on the direction i.e. positive (*pd*(> 0)) or negative (*pd*(< 0))[2].

To identify potential covariates and ensure the suitability of data for longitudinal analyses, we used Bayesian multilevel regression models to run preliminary checks prior to our main analyses. Specifically, we used gaussian models to examine whether cardiac variables (HR/HRV) and task engagement metrics (i.e., time spent on trials) varied across menstrual phases, given previous evidence of phase-related changes HR and HRV [21,22,27,50]; Beta regression models for exploring potential covariates in relation to consistency score outcomes (bounded 0-1) and; Poisson models for examining potential covariates in relation to count outcomes (interoceptive attention). These data checks examined 1) whether aforementioned variables varied across menstrual phases; 2) whether completion number, cardiac variables or task engagement predicted PAT and TCT performance; and 3) the

[2] Note that effect sizes were reduced in models including HRV, as HRV data were missing for 13.5% of PAT and 10.7% of TCT sessions. Importantly, when the PAT analyses were restricted to participants with valid HRV data, effect sizes were comparable before and after HRV was added as a covariate, suggesting that the attenuation reflected the reduced sample rather than adjustment for HRV itself.

relationship between the four interoception measures. Where a potential confound covaried with interoception measures or menstrual cycle phase, it was controlled for in sensitivity analyses.

Modelling across the menstrual cycle

We used Bayesian regression models for all main analyses. Beta models examined whether PAT and TCT performance varied across the menstrual cycle, given the bounded outcome [51] (0-1), with sensitivity analyses used to compare Beta regression models including and excluding identified covariates. Models were also applied to examine variation in self-reported interoceptive attention (Poisson; range 0-4) and exteroceptive attention (Bernoulli; 0,1) across the menstrual cycle (before and after including covariates). Exploratory analyses investigated the interoceptive attention models before and after including self-reported symptom burden, as well as whether participants' subjective appraisal of their interoceptive and exteroceptive attention varied across menstrual phases (i.e., the extent that they perceived their attention to bodily signals to be positive/helpful or negative/unhelpful), with quality ratings modelled as an ordered outcome (1 = Very Negative/Unhelpful to 5 = Very Positive/Helpful) using a Bayesian ordinal regression [51].

Models were fitted using the R package *brms* (v2.23.0 [52]) with rstan backend function and participant-level random intercepts. All continuous predictors and covariates were standardised (z-scored). Weakly informative priors were applied on all parameters: Normal(0, 1) for the intercept and fixed effects, Exponential(1) for the random effect standard deviation and the Beta precision parameter (φ). When modelling menstrual phase effects, cycle phase was dummy coded. To examine all contrast levels, each phase was modelled as a reference category with remaining

phases as fixed effects. Fit for all Bayesian models was confirmed via posterior predictive checks.

Given emerging evidence of intra-individual between-cycle variation in hormonal patterns [53], in exploratory analyses we repeated the above, stratifying the dataset to include observations from 1) a complete cycle (i.e., observations sequencing from perimenstrual to midluteal) and 2) across cycles (i.e., observations where at least one phase is drawn from a different menstrual cycle to the rest of the phases), referred to as a merged cycle. This allowed us to explore whether cycle variation may have influenced any observed phase effects.

Results

Data checks

Prior to examining variation in cardiac interoceptive accuracy and interoceptive attention across the menstrual cycle, we conducted a series of preliminary checks to identify physiological and task-related factors that might influence performance. Full results are reported in SI.

First, we examined whether relevant physiological and behavioural variables varied across menstrual-cycle phases. Resting heart rate (HR) was higher in the midluteal than midfollicular phase (SI Table S2), whereas there was no strong evidence for phase-related variation in heart rate variability (HRV) (SI Table S3). Task engagement also varied across the cycle: participants tended to spend less time on PAT and TCT trials during the perimenstrual phase than during midfollicular and periovulatory phases (SI Table S4).

We next examined whether these factors were related to PAT and TCT performance (SI Table S5). HR was not associated with PAT performance, but higher HR was

associated with poorer TCT performance when continuous delay scores were examined. Time spent on trials was positively associated with performance on both the PAT and TCT. HR was therefore included in sensitivity analyses of TCT performance, and trial duration was included in relevant sensitivity analyses for both tasks. Although evidence for associations between HRV and task performance was limited, HRV was also included in sensitivity analyses given previous reports linking HRV to cardiac interoceptive accuracy [28-29]. We also tested for practice or boredom effects. There was no reliable evidence that PAT performance, TCT performance, or self-reported interoceptive attention changed as a function of task completion date. Self-reported exteroceptive attention, however, decreased across repeated completions, and task completion was therefore included as a covariate in models involving exteroceptive attention.

Finally, we examined relationships among the interoceptive measures (SI Table S6). Evidence for an association between PAT and TCT performance was uncertain. There was little evidence that either PAT or TCT performance was associated with self-reported interoceptive attention. In contrast, there was probable evidence for a positive association between self-reported interoceptive and exteroceptive attention. There was also probable evidence for a positive association between exteroceptive attention and PAT performance, particularly for continuous PAT scores, whereas evidence for an association between exteroceptive attention and TCT performance was weak.

Modelling interoceptive accuracy by menstrual phase

In the full dataset (prior to TCT exclusions; *N*=656), we found suggestive but inconclusive evidence that PAT binary scores varied across menstrual cycle phases,

with the pattern showing higher accuracy in the periovulatory vs perimenstrual phase (OR = 1.65, 95% CrI [0.882, 2.838]; *pd*(OR > 94%; see Table S7). When restricting to PAT sessions with a corresponding TCT pass as pre-registered (*N*=516), evidence for a periovulatory advantage in the PAT remained probable but below threshold (OR = 1.69, 95% CrI [0.843, 3.104]; *pd*(OR > 1) = 92%; Table 1).

When examining continuous consistency scores, the full unscreened dataset (prior to TCT exclusions; *N*=633) again showed inconclusive evidence of a periovulatory advantage (Δ = 0.024, 95% CrI [−0.01, 0.057]; *pd*(Δ > 0) = 92%; see Table S8). When restricting to TCT-screened sessions as pre-registered (*N*=498; see Table 2), evidence strengthened for the periovulatory vs midluteal (Δ = 0.033, 95% CrI [−0.005, 0.072]; *pd*(Δ > 0) = 96%) and periovulatory vs perimenstrual phase contrasts (Δ = 0.031, 95% CrI [−0.006, 0.067]; *pd*(Δ > 0) = 95%), though credible intervals crossed zero in both cases.

Exploratory analysis in a complete cycle vs merged cycle

Exploratory analysis revealed notable differences between the complete-cycle and merged-cycle subsamples. PAT results from the complete-cycle sample were largely consistent with the results from two cycles. In contrast, the merged-cycle subset showed little evidence of phase-related differences in cardiac interoceptive accuracy. We detail these below, focusing on screened sessions (with a corresponding TCT pass) for brevity.

In the complete cycle, there was moderate-strong evidence that the likelihood of being classified as interoceptive was higher during the periovulatory vs perimenstrual phase (*N*=250, OR = 2.63, 95% CrI [0.971, 5.923]; *pd*(OR > 1) = 97%; see Table 1), though the credible interval marginally crossed 1. In contrast, there was no evidence

of a phase-effect in the merged-cycle (*N*=266; *pd*s(OR > 1) ≤64%, see Table 1). This pattern was also observed in the full dataset prior to TCT exclusions (complete cycle; *N*=325, merged cycle; *N*=331; see Table S7).

Table 1. Model estimates of differences in the odds ratio for PAT binary classification (1=pass, 0=fail) between menstrual phases, in two menstrual cycles, a complete cycle, and merged cycles. Data screened for a corresponding pass in the TCT control task. Odds ratios for PAT classification during each menstrual phase are presented.

PAT performance (binary classification scores) Note: Odds Ratios reported for phase contrasts

	Two cycles (N = 516)			Complete cycle (N = 250)			Merged cycles (N = 266)		
	Prob	**95% CrI**	**pd(OR > 1)**	**Prob**	**95% CrI**	**pd(OR > 1)**	**Prob**	**95% CrI**	**pd(OR > 1)**
Contrast (Odds Ratios)									
Periovulatory – Perimenstrual	1.687	[0.843, 3.104]	93%	2.630	[0.971, 5.923]	97%	0.930	[0.367, 1.918]	36%
Periovulatory – Midluteal	1.555	[0.708, 2.978]	84%	2.311	[0.727, 5.569]	91%	1.128	[0.394, 2.59]	51%
Periovulatory – Midfollicular	1.551	[0.68, 3.061]	83%	1.986	[0.583, 4.974]	84%	1.161	[0.363, 2.82]	51%
Midfollicular – Perimenstrual	1.180	[0.554, 2.218]	61%	1.560	[0.538, 3.597]	75%	0.929	[0.352, 2.018]	36%
Midfollicular – Midluteal	1.076	[0.498, 2.054]	50%	1.369	[0.41, 3.385]	62%	1.110	[0.39, 2.572]	48%
Perimenstrual – Midluteal	0.967	[0.458, 1.801]	39%	0.964	[0.329, 2.282]	37%	1.296	[0.502, 2.825]	64%
Menstrual Phase									
Perimenstrual	0.181	[0.094, 0.299]		0.133	[0.055, 0.25]		0.262	[0.135, 0.429]	
Midfollicular	0.197	[0.102, 0.327]		0.174	[0.071, 0.325]		0.233	[0.107, 0.407]	
Periovulatory	0.259	[0.147, 0.403]		0.260	[0.123, 0.441]		0.234	[0.113, 0.398]	
Midluteal	0.196	[0.103, 0.319]		0.153	[0.061, 0.293]		0.234	[0.111, 0.4]	

For continuous consistency scores, we found strong evidence that cardiac interoceptive accuracy was higher in the periovulatory phase than the perimenstrual phase in the complete cycle data (*N* = 239; Δ = 0.054, 95% CrI [0, 0.111], *pd*(Δ > 0) = 97%; see Table 2 and Figure 4A). This was also consistent in the complete cycle

data prior to TCT exclusions (N = 309, Δ = 0.052, 95% CrI [0.004, 0.101]; pd(Δ > 0) = 98%; Table S8). In contrast, there was no evidence of a phase-effect in the merged-cycle subsample (N=259; pds(Δ > 0) ≤73%, see Table 2).

The above effects were robust to the inclusion of covariates (HR, HRV, and time spent on trials; see SI Tables S9–14)[3].

Control task (TCT) performance across the menstrual cycle

In the full sample, there was no strong evidence of a phase effect on TCT scores across binary (N=665; pds ≤ 88%, see Table S15) or continuous scores (N=652; pds ≤ 77%, see Table S19). However, comparable analyses of TCT scores in the complete and merged cycles revealed evidence of phase-effects in the complete cycle only. Binary classification showed suggestive evidence of lower TCT performance in the midfollicular vs midluteal phase (N=327, OR = 0.517, 95% CrI [0.146, 1.108]; pd(OR > 1) = 96%; Table S15) which was strengthened after including HRV and time spent on trials as covariates (e.g., HRV Table S17; N=284, OR = 0.384, 95% CrI [0.123, 0.9]; pd(OR > 1) = 1%). Continuous delay scores showed the same pattern (Δ = −0.060, 95% CrI [−0.115, −0.005]; pd(Δ < 0) = 98%; see Tables S19-22). Neither effect was present in the merged cycle.

[3] Note that effect sizes were reduced in models including HRV, as HRV data were missing for 13.5% of PAT and 10.7% of TCT sessions. Importantly, when the PAT analyses were restricted to participants with valid HRV data, effect sizes were comparable before and after HRV was added as a covariate, suggesting that the attenuation reflected the reduced sample rather than adjustment for HRV itself.

Table 2. Model estimates of mean differences in PAT consistency scores following TCT screening, between menstrual phases. The table presents results from two menstrual cycles, a complete cycle, and merged cycles. Mean PAT consistency scores during each menstrual phase are also presented. Contrasts are ordered by descending pd(Δ > 0). Values > 50% indicate the first phase in the contrast is higher on average while values < 50% indicate the reverse. Clustering of pd values near 50% reflects the absence of a consistent directional pattern.

PAT consistency scores, following TCT screening

	Two cycles (N = 498)			Complete cycle (N = 239)			Merged cycles (N = 259)		
	Mean	**95% CrI**	**pd(Δ > 0)**	**Mean**	**95% CrI**	**pd(Δ > 0)**	**Mean**	**95% CrI**	**pd(Δ > 0)**
Contrast									
Periovulatory – Midluteal	0.033	[-0.005, 0.072]	96%	0.048	[-0.01, 0.104]	95%	0.016	[-0.042, 0.073]	70%
Periovulatory – Perimenstrual	0.031	[-0.006, 0.067]	95%	0.054	[0, 0.109]	**97%**	0.001	[-0.052, 0.054]	52%
Periovulatory – Midfollicular	0.019	[-0.023, 0.062]	82%	0.033	[-0.029, 0.097]	85%	-0.002	[-0.061, 0.056]	48%
Midfollicular – Midluteal	0.014	[-0.024, 0.053]	76%	0.014	[-0.043, 0.071]	69%	0.018	[-0.039, 0.075]	73%
Midfollicular – Perimenstrual	0.011	[-0.029, 0.051]	71%	0.021	[-0.035, 0.078]	77%	0.003	[-0.058, 0.062]	54%
Perimenstrual – Midluteal	0.003	[-0.035, 0.042]	55%	-0.006	[-0.061, 0.049]	40%	0.015	[-0.042, 0.071]	69%
Menstrual Phase									
Perimenstrual	0.271	[0.228, 0.316]		0.253	[0.203, 0.305]		0.293	[0.243, 0.345]	
Midfollicular	0.282	[0.238, 0.329]		0.274	[0.22, 0.332]		0.296	[0.246, 0.35]	
Periovulatory	0.302	[0.259, 0.348]		0.307	[0.25, 0.367]		0.294	[0.247, 0.343]	
Midluteal	0.268	[0.226, 0.312]		0.260	[0.21, 0.315]		0.278	[0.23, 0.329]	

Additional checks

As participants were instructed to counterbalance the order of the PAT and TCT, we also examined whether order interacted with the above by including order as a categorical predictor. Notably, all patterns were unchanged with the inclusion of order.

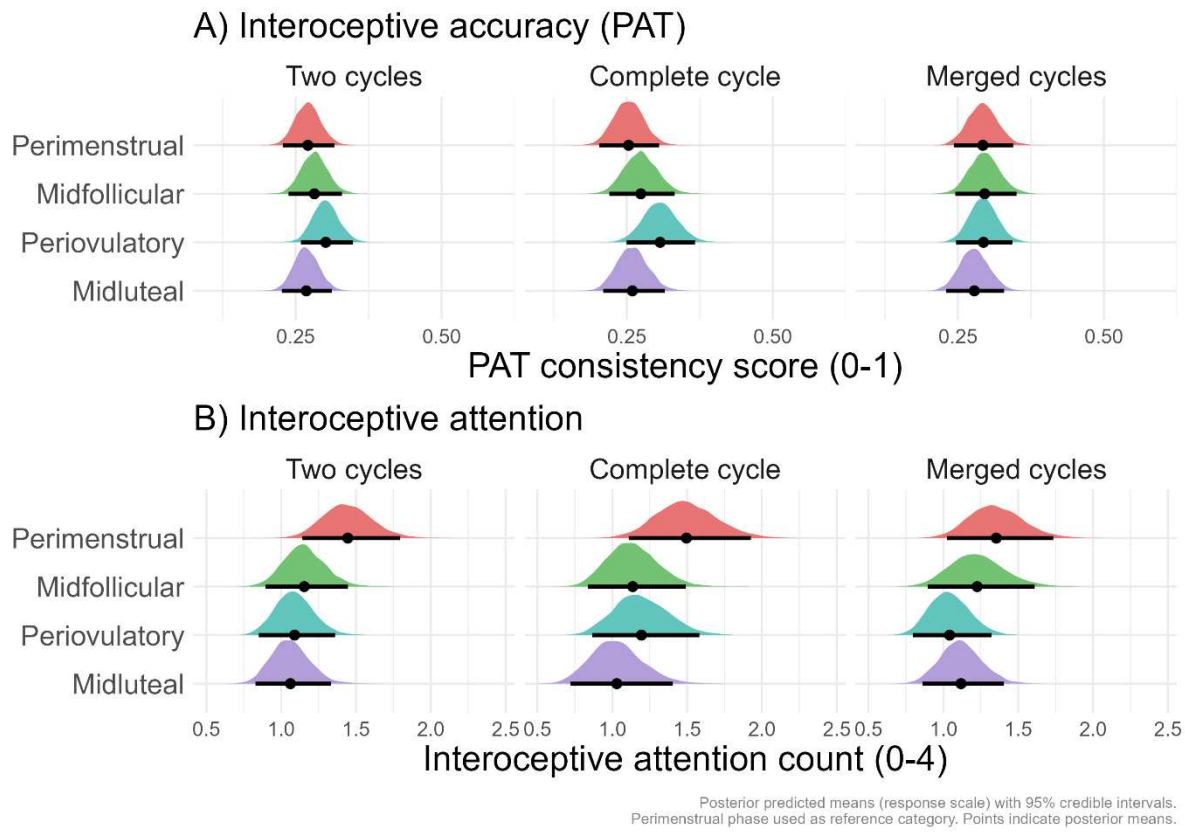


Figure 4. Plots showing the posterior estimates for interoceptive accuracy i.e. PAT score (**B**) and interoceptive attention (**A**) (x-axis) across menstrual phases (y-axis). Bars at each phase show the 95% credible intervals for the predicted (A) mean PAT consistency score and (B) interoceptive attention count at each phase in relation to each other. Note: Estimates were computed from models using perimenstrual phase as the baseline reference phase. Each panel represents the results from phases across two menstrual cycles, a complete cycle, and merged cycles.

Modelling interoceptive attention by menstrual phase

Table 3 shows the predicted average attention to interoceptive signals in each menstrual phase and posterior contrasts between each phase based on 520 observations. Estimated mean scores were highest during the perimenstrual phase (*M* = 1.45, 95% *CrI* [1.14, 1.81]) and there was strong evidence (*pd*s(Δ > 0) = 99%) that attention to interoceptive signals was higher in perimenstrual relative to all other phases (Figure 4B); midfollicular (Δ = 0.351, 95% CrI [0.090, 0.616]), periovulatory (Δ

= 0.337, 95% CrI [0.042, 0.648]) and midluteal (Δ = 0.396, 95% CrI [0.091, 0.715]). In contrast, there was no evidence that exteroceptive attention varied across menstrual cycle phases (Table S23), with this effect unchanged after controlling for completion number.

Exploratory analysis in a complete-cycle vs merged-cycle

The results from complete-cycle data were largely consistent with the pattern from two cycles (see Table 2). Attention to bodily sensations was highest during the perimenstrual phase (*M* = 1.54, 95% *CrI* [1.162, 1.98]), again with the largest positive differences observed against the midluteal (Δ = 0.515, 95% CrI [0.067, 0.983]) and midfollicular phases (Δ = 0.435, 95% CrI [0.012, 0.88]). All contrasts met the threshold for strong evidence of a directional effect where the 95% CrI did not cross zero (*pd*(Δ > 0) = 98%). In contrast, when examining the merged-cycle observations, there was little evidence of phase-related differences in interoceptive attention. Although the pattern and contrasts were broadly similar to the complete-cycle subsample, the results did not reach thresholds for probable evidence of directional effects (*pd*s <90%). In contrast, there was no evidence that exteroceptive attention varied across menstrual cycle phases in the complete or merged cycle data (Table S23) and this was robust after controlling for completion number.

Table 2. Model estimates of mean differences in daily interoceptive attention between menstrual phases in two menstrual cycles, a complete cycle, and merged cycles. Mean interoceptive attention during each menstrual phase is presented.

Interoceptive attention by menstrual phase

	Two cycles (N = 520)			Complete cycle (N = 242)			Merged cycles (N = 278)		
	Mean	**95% CrI**	**pd(Δ > 0)**	**Mean**	**95% CrI**	**pd(Δ > 0)**	**Mean**	**95% CrI**	**pd(Δ > 0)**
Contrast									
Perimenstrual – Midluteal	0.399	[0.099, 0.728]	**100%**	0.515	[0.067, 0.983]	**99%**	0.196	[-0.218, 0.625]	82%
Perimenstrual – Midfollicular	0.319	[0.002, 0.647]	**98%**	0.435	[0.012, 0.88]	**98%**	0.089	[-0.386, 0.577]	64%
Perimenstrual – Periovulatory	0.338	[0.028, 0.656]	**98%**	0.335	[-0.124, 0.787]	93%	0.235	[-0.16, 0.662]	87%
Midfollicular – Midluteal	0.080	[-0.191, 0.355]	71%	0.079	[-0.323, 0.477]	66%	0.107	[-0.284, 0.515]	70%
Periovulatory – Midluteal	0.061	[-0.213, 0.343]	66%	0.180	[-0.266, 0.63]	78%	-0.039	[-0.393, 0.325]	41%
Midfollicular – Periovulatory	0.019	[-0.293, 0.33]	55%	-0.101	[-0.55, 0.343]	33%	0.146	[-0.285, 0.609]	74%
Menstrual Phase									
Perimenstrual	1.450	[1.146, 1.811]		1.541	[1.162, 1.98]		1.303	[0.97, 1.695]	
Midfollicular	1.131	[0.88, 1.423]		1.105	[0.815, 1.441]		1.214	[0.888, 1.59]	
Periovulatory	1.112	[0.866, 1.396]		1.206	[0.866, 1.606]		1.068	[0.816, 1.355]	
Midluteal	1.051	[0.807, 1.313]		1.026	[0.72, 1.393]		1.107	[0.849, 1.39]	

Exploratory analysis of the role of physical symptoms in interoceptive attention

To explore whether phase-related differences in interoceptive attention were accounted for by the number of physical symptoms reported each day, we repeated the aforementioned models controlling for daily symptom count. Once physical symptoms were included in the model, effect sizes reduced and there was no longer strong evidence for directional effects (see Table S24). The perimenstrual-midluteal and the perimenstrual-midfollicular contrasts remained the largest differences with

probable evidence (*Pd*s = 91-97%), however the credible boundaries indicated greater uncertainty.

Exploratory analysis of the affective evaluation of interoceptive attention

Across the full dataset[4], affective evaluations of one's own interoceptive attention were rated more positively during the periovulatory phase than the remaining phases (see Table 3); perimenstrual (Δ = 0.233, 95% CrI [-0.002, 0.468]), midfollicular (Δ = 0.352, 95% CrI [0.027, 0.592]), and midluteal (Δ = 0.221, 95% CrI [-0.039, 0.476). The periovulatory-midfollicular contrast provided strong evidence for a directional effect (*pd* = 98%) with certain evidence based on the credible boundary. The other contrasts provided moderate evidence (95-97%; see Table 3), although the credible boundaries indicated greater uncertainty. The complete and merged cycle data, however, provided little evidence of these patterns likely due to the small number of observations (as participants only reported paying attention to bodily signals on 20% of occasions). Evidence that the quality of exteroceptive attention varied across menstrual cycle phases was limited. There was probable evidence for more favourable attention in midfollicular vs periovulatory, but the credible interval was wide and crossed zero (see Table S25). This was robust after controlling for completion number and consistent across the complete and merged cycles (Table S26).

[4] Each participant could contribute up to 4 response observations per day, as inclusion in the analysis was based on whether a participant reported paying attention to an interoceptive item (heartbeats/breathing, hunger/thirst, stomach/gastric or urination/defecation). If yes, their appraisal to the interoceptive attention was examined.

Table 3. Model estimates of mean differences in the appraisal of interoceptive attention between menstrual phases in two menstrual cycles, a complete cycle, and merged cycles. Mean affective appraisal to interoceptive attention during each menstrual phase is presented (1 = very negative/unhelpful, 5 = very positive/helpful).

Affective appraisal of interoceptive attention by menstrual phase (1 = very negative, 5 = very positive)

	Two cycles (N = 634)			Complete cycle (N = 307)			Merged cycles (N = 327)		
	Mean	**95% CrI**	**pd(Δ > 0)**	**Mean**	**95% CrI**	**pd(Δ > 0)**	**Mean**	**95% CrI**	**pd(Δ > 0)**
Contrast									
Periovulatory – Midfollicular	0.312	[0.027, 0.592]	**98%**	0.296	[-0.123, 0.713]	91%	0.161	[-0.215, 0.541]	80%
Periovulatory – Perimenstrual	0.233	[-0.002, 0.468]	97%	0.134	[-0.212, 0.472]	78%	0.216	[-0.111, 0.532]	91%
Periovulatory – Midluteal	0.221	[-0.039, 0.476]	95%	0.127	[-0.305, 0.555]	71%	0.283	[-0.037, 0.593]	96%
Midluteal – Midfollicular	0.091	[-0.167, 0.356]	75%	0.169	[-0.22, 0.558]	80%	NA	NA	
Perimenstrual – Midfollicular	0.080	[-0.179, 0.329]	74%	0.162	[-0.172, 0.498]	83%	NA	NA	
Midluteal – Perimenstrual	0.012	[-0.225, 0.254]	54%	0.007	[-0.361, 0.379]	52%	NA	NA	
Menstrual Phase									
Perimenstrual	3.145	[2.857, 3.407]		3.100	[2.77, 3.419]		3.243	[2.914, 3.548]	
Midfollicular	3.065	[2.763, 3.362]		2.938	[2.597, 3.295]		3.298	[2.939, 3.614]	
Periovulatory	3.377	[3.092, 3.637]		3.234	[2.868, 3.593]		3.459	[3.149, 3.729]	
Midluteal	3.156	[2.873, 3.439]		3.107	[2.714, 3.492]		3.176	[2.865, 3.473]	

Discussion

This study aimed to investigate whether objective interoceptive accuracy and self-reported interoceptive attention vary across the menstrual cycle. Using an experience sampling approach and matched non-interoceptive control measures, we collected daily measures from twenty females across two consecutive menstrual cycles. As hypothesised, both interoceptive accuracy and attention varied across the

menstrual cycle. Importantly, these patterns were unique to interoception, and for both interoceptive accuracy and attention the strongest evidence was observed when sampling from a complete menstrual cycle. Below we consider the implications of these novel findings.

Overall, we found evidence that interoceptive accuracy was higher during the periovulatory phase than during the perimenstrual phase, but only when data were sampled from a complete menstrual cycle. This was broadly consistent across continuous and binary scores, despite the loss of power associated with binary scores [54], and remained after controlling for HR, HRV, and engagement metrics, despite both HR and engagement metrics showing cycle-related variation. While we did not predict differences across complete and merged cycles, and thus results require replication, the consistent pattern for both interoceptive accuracy and attention (as well as the TCT) suggests that the effect of cycle type is unlikely to be an artefact of the PAT. Instead, it is more likely that this pattern reflects known variability across menstrual cycles, whereby physiological markers show limited predictive value from one cycle to the next [53]. Notably, the TCT phase effect differed from that observed for PAT, with lower performance during the midfollicular relative to the midluteal phase, underscoring the specificity of the periovulatory advantage for interoceptive accuracy. This highlights the value of including control tasks which are often absent from the literature, not only for studies of between-person differences, but also, critically, in studies examining within-person change (Donaghy et al., in prep; [18]).

Although we did not measure hormones due to cost constraints, the difference between the periovulatory and perimenstrual phase provides some insights into the hormones that may be driving observed effects. Both oestradiol and progesterone

are low during the perimenstrual phase and rise substantially across the follicular and luteal phases, though with different temporal profiles [55] (see Figure 2A). The periovulatory peak in cardiac interoceptive accuracy aligns with the surge in oestradiol that precedes ovulation, though it should be noted that ovulation is associated with a rise in several other hormones (e.g., progesterone and testosterone [31,47]) that may instead be driving observed effects. Given known sex differences in interoceptive accuracy (with males outperforming females on lab-based tasks [4,5]), it is conceivable that changes in accuracy may be driven by changes testosterone, rather than oestradiol, though future research is needed to disentangle these possibilities. Nevertheless, given observed patterns, these data tentatively suggest that changes in accuracy are not likely due to changes in progesterone as if progesterone were driving effects, we would have expected to see elevation in the mid-luteal phase.

Notably, these data also cannot determine whether cycle-related changes in interoceptive accuracy are driven by cardiovascular parameters or perceptual changes. This is because cardiac interoceptive accuracy measures do not enable separation of signal strength from perception [56]. Although we selected the PAT as it is the only valid measure that can be administered remotely, we cannot rule out that changes in cardiac parameters may be driving observed effects. Indeed, periovulatory increases in oestradiol are associated with changes in cardiac parameters such as blood pressure [58,59], which has also been linked with improved cardiac interoceptive accuracy [60]. Whilst lab-based assessments incorporating concurrent measurements (e.g., of blood pressure, cardiac output etc.) would be required to address this possibility, it is reassuring that our effects remained consistent after controlling for HR, which systematically varied across the menstrual

cycle (as previously observed [21,22]), as well as HRV. As such, future research would benefit from replicating these findings using a lab-based design that enables the collection of multiple physiological parameters to determine whether perceptual or physiological changes underlie observed. Although our dense sampling approach would be unfeasible for a lab-based study, our findings highlighting the value of sampling from a complete menstrual cycle, and the absence of practice effects on the PAT, provides reassurance that this could be practically achieved using the same phasing approach, but with all participants beginning in the perimenstrual phase (i.e. at menstruation). Such studies would also benefit from assessing interoceptive accuracy across multiple domains given evidence of domain-specificity [62–64], including tasks that enable separation of stimulus strength from perception (e.g. [65]), and taking lab-based assessments of physiological measures (e.g., HRV) that did not covary strongly with the menstrual cycle in our sample (in contrast to previous literature [24]), potentially due to the use of phone-based photoplethysmography. Overall, whilst more work is required, these data provide important practical and theoretical insights to inform future work.

Beyond accuracy, we found strong evidence for heightened interoceptive attention during the perimenstrual phase relative to all other phases, a pattern unique to interoceptive attention and strongest in the complete cycle data. As these changes were accounted for by the presence of physical symptoms (e.g. headaches, back pain), this suggests that phase-related shifts in interoceptive attention are likely driven by the presence of bodily signals capturing attention, rather than an increased tendency to focus on the body in the absence of physical changes. We also found preliminary evidence that attention to bodily signals may be evaluated more favourably during the periovulatory phase, which was again unique to interoception,

though this did not vary across the complete and merged cycles. Although such findings must be treated as preliminary given the low incidence of reports of interoceptive vs exteroceptive attention (~20% vs ~80%, respectively), this pattern is intriguing given previous links between interoceptive attention and negative mood [43], as well as evidence of elevated mood during the periovulatory phase [66] (but see [67]). The extent to which such patterns are related, and the directionality of such effects, remains a future question.

Taken together, these data have important theoretical implications for interoceptive measurement and theories of sex differences. For interoceptive measurement, consistent with prior work [17], we found large intra-individual variation in interoceptive accuracy and attention suggesting that single time point assessments may not adequately capture a person's interoceptive ability [18,34]. Systematic covariation with the menstrual cycle suggests these patterns are not due to measurement noise or poor measure reliability, which in turn has critical implications for how interoception is measured and conceptualised. More broadly, it is conceivable that treating interoception as a stable trait [16] and prioritising measures with high test-reliability (e.g., the Heartbeat Counting Task [68]), has contributed towards known replication issues [18,55,69]. Given debate regarding whether test-retest reliability should be used to assess the suitability of interoceptive accuracy tasks [55,70], our data suggest that measures exhibiting very high test-retest reliability should be interpreted with caution. For example, although the HCT exhibits high test-retest reliability (r = .81 [71] depending on time period; one week r=.60 [72]; one day ICC=.88 [73]), there remains debate regarding whether the HCT captures perceptual accuracy or one's ability to estimate heartbeats or count seconds at a rate consistent with one's expected heart rate [69]. Our findings suggest that if a task

primarily captures interoceptive processes (whether perceptual or physiologically driven), rather than stable factors such as beliefs about heart rate, greater within-person variation would be expected. As argued previously [34], these data highlight a need for caution when interpreting between-person differences from single time-point assessments (or changes in interoception where single time-point assessments are used at baseline in pre-post designs). Indeed, they suggest that when relating individual differences in interoception to theoretically relevant outcomes, it may be more informative to consider the stability of interoceptive ability over time, situations, and domains. Whilst this argument relates to the value of adequately capturing between person differences using repeated assessments and measures that are sensitive to change, our data also highlight the potential value of exploring within-person change, which may prove to better relate to theoretically relevant outcomes than between-person differences.

Beyond measurement, these data also have important implications for theories of sex differences in interoception. Although sex differences were not examined here, by providing the first evidence of menstrual cycle-related change on interoception in females, these data provide support for the proposal that sex differences in interoception may arise from the increased amount of physical and hormonal change that females experience across the lifespan [14]. Whilst we focused on individuals with regular cycles to enable counterbalancing of starting phases, future work would benefit from extending this approach to individuals with irregular menstrual cycles, those taking hormonal contraceptives (who have suppressed endogenous sex steroids), and males. This would allow direct testing of the hypothesis that greater variability in the physiological environment contributes to variation in interoception and, in turn, to sex differences.

Beyond phase effects, these data also provide insights into the structure of interoceptive abilities and the suitability measures for longitudinal testing. Consistent with previous work [42,74], interoceptive accuracy and attention were not associated, suggesting that these aspects of interoception dissociate when examining within-person differences, as well as between-person differences, as previously observed [42,74]. Similarly consistent with prior literature, interoceptive and exteroceptive attention were associated, but differing patterns across the menstrual cycle suggest they also reflect distinct processes. As partially reported in Donaghy et al., [in prep], interoception measures did not exhibit practice effects, and the PAT was not associated with HR or HRV. Notably, however, both PAT and TCT performance were associated with engagement metrics. This contrasts with previous between-subjects work [34] and highlights the value of examining associations within individuals. Whilst both exteroceptive measures exhibited some influence of nuisance variables that may require further consideration if replicated, overall, these data offer support for previous theoretical proposals arguing for the separation of interoceptive accuracy and attention and underscore the suitability of the interoception measures for repeated sampling.

Despite the theoretical and methodological importance of these data, limitations must be acknowledged. First, whilst the recommended four-phase approach [30] offers preliminary insights into the hormonal shifts that may underlie patterns, without hormonal assays, it is not possible to determine which hormonal shifts underlie observed effects. However, given the cost and participant burden associated with repeated hormonal assays, establishing whether interoception varies systematically across the menstrual cycle was an important first step before attempting to isolate specific hormones. Second, cycle phase allocation relied on counting methods and

Apple Watch BBT recordings. Whilst this approach is broadly consistent with recommendations for estimating cycle phases and ovulation (given the reasonably consistent length of the luteal phase, particularly in those with regular cycles [30]), future research would benefit from more precise assessment of ovulation using urine luteinizing hormone kits. Third, whilst we decided to counterbalance participants to begin in specific cycle phases as practice effects for our interoception measures were not yet known, this counterbalancing approach required participants with highly regular cycles, limiting generalisability to the broader population. Fourth, although interoceptive accuracy effects were not explained by HR and HRV, as described above, due to our dense sampling approach we were unable to capture all physiological parameters that may underlie effects and lab-based studies are required to address this possibility. Finally, whilst we decided to match the TCT and PAT by triggering the flashing shape from the heartbeat, ensuring that the participant experience – e.g., interruptions due to finger movement; HR and HRV – was matched across tasks, we cannot rule out that some participants may have relied on interoceptive information during the TCT. Although unlikely given the difference in pass rates (TCT 77.7%; PAT 23%), future work should examine the best way to deploy matched control tasks (see Donaghy et al., in prep). Nevertheless, given that few studies of interoception employ control tasks [18], the inclusion of this measure was a strength.

In summary, despite limitations, this study presents the densest sampling of interoception to date and the first to study interoception across the menstrual cycle. We found that interoceptive accuracy was enhanced during the periovulatory phase relative to the perimenstrual phase, while interoceptive attention was heightened during the perimenstrual phase relative to all other phases. These findings

demonstrate substantial intra-individual variability in interoceptive processing and highlight the menstrual cycle as a meaningful source of state-level variation.

References

1. Brewer, R., Murphy, J. & Bird, G. Atypical interoception as a common risk factor for psychopathology: A review. *Neurosci. Biobehav. Rev.* **130**, 470–508 (2021).

2. Khalsa, S. S. & Lapidus, R. C. Can interoception improve the pragmatic search for biomarkers in psychiatry? *Front. Psychiatry* **7**, 187423 (2016).

3. Paulus, M. P. & Stein, M. B. Interoception in anxiety and depression. *Brain Structure and Function 2010 214:5* **214**, 451–463 (2010).

4. Prentice, F. & Murphy, J. Sex differences in interoceptive accuracy: A meta-analysis. *Neurosci. Biobehav. Rev.* **132**, 497–518 (2022).

5. Ma-Kellams, C., Prentice, F., Spooner, R. & Murphy, J. Demographic Differences in Interoception. in *Interoception: A Comprehensive Guide* 357–403 (Springer International Publishing, 2024). doi:10.1007/978-3-031-68521-7_11/SAVE-RESEARCH.

6. Pennebaker, J. W. & Roberts, T.-A. Toward a His and Hers Theory of Emotion: Gender Differences in Visceral Perception. *J. Soc. Clin. Psychol.* **11**, 199–212 (1992).

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

8. Longarzo, M., Mele, G., Alfano, V., Salvatore, M. & Cavaliere, C. Gender Brain Structural Differences and Interoception. *Front. Neurosci.* **14**, 586860 (2021).

9. Gijsbers Van Wijk, C. M. T. & Kolk, A. M. Sex differences in physical symptoms: The contribution of symptom perception theory. *Soc. Sci. Med.* **45**, 231–246 (1997).

10. Barsky, A. J., Peekna, H. M. & Borus, J. F. Somatic symptom reporting in women and men. *J. Gen. Intern. Med.* **16**, 266–275 (2001).

11. Franzoi, S. L., Kessenich, J. J. & Sugrue, P. A. Gender differences in the experience of body awareness: An experiential sampling study. *Sex Roles* **21**, 499–515 (1989).

12. Stewart, S. H., Taylor, S. & Baker, J. M. Gender differences in dimensions of anxiety sensitivity. *J. Anxiety Disord.* **11**, 179–200 (1997).

13. Harshaw, C. Interoceptive dysfunction: Toward an integrated framework for understanding somatic and affective disturbance in depression. *Psychol. Bull.* **141**, 311–363 (2015).

14. Murphy, J., Viding, E. & Bird, G. Does Atypical Interoception Following Physical Change Contribute to Sex Differences in Mental Illness? *Psychol. Rev.* **126**, 787–789 (2019).

15. Li, S. *et al.* Sex difference in incidence of major depressive disorder: an analysis from the Global Burden of Disease Study 2019. *Ann. Gen. Psychiatry* **22**, 1–9 (2023).

16. Garfinkel, S. N., Eccles, J. A. & Critchley, H. D. The Heart, the Brain, and the Regulation of Emotion. *JAMA Psychiatry* **72**, 1071–1072 (2015).

17. Wittkamp, M. F., Bertsch, K., Vögele, C. & Schulz, A. A latent state-trait analysis of interoceptive accuracy. *Psychophysiology* **55**, e13055 (2018).

18. Murphy, J. Interoception: Where do we go from here? *Quarterly Journal of Experimental Psychology* **77**, 223–229 (2024).

19. Lim, A. H. et al. Effect of menstrual cycle and menopause on human gastric electrophysiology. *Am J Physiol Gastrointest Liver Physiol* **1**:327, G47-G56 (2024).

20. Rosen Vollmar, A. K., Mahalingaiah, S. & Jukic, A. M. The menstrual cycle as a vital sign: a comprehensive review. *F. S. Rev.* **6**, (2025).

21. Shilaih, M., De Clerck, V., Falco, L., Kübler, F. & Leeners, B. Pulse Rate Measurement During Sleep Using Wearable Sensors, and its Correlation with the Menstrual Cycle Phases, A Prospective Observational Study OPEN. https://doi.org/10.1038/s41598-017-01433-9 doi:10.1038/s41598-017-01433-9.

22. Moran, V. H., Leathard, H. L. & Coley, J. Cardiovascular functioning during the menstrual cycle. *Clinical Physiology* **20**, 496–504 (2000).

23. McKinley, P. S. *et al.* The impact of menstrual cycle phase on cardiac autonomic regulation. *Psychophysiology* **46**, 904–911 (2009).

24. Schmalenberger, K. M. *et al.* A Systematic Review and Meta-Analysis of Within-Person Changes in Cardiac Vagal Activity across the Menstrual Cycle: Implications for Female Health and Future Studies. *J. Clin. Med.* **8**, 1946 (2019).

25. Schmalenberger, K. M. *et al.* Menstrual Cycle Changes in Vagally-Mediated Heart Rate Variability Are Associated with Progesterone: Evidence from Two Within-Person Studies. *Journal of Clinical Medicine 2020, Vol. 9, Page 617* **9**, 617 (2020).

26. Serrano, H.M., Gueugneau, N. & Ravier, G. Influence of menstrual cycle on autonomic nervous system, muscular strength and mood states. *Sci Rep* **16**, 28294 (2026). https://doi.org/10.1038/s41598-026-56802-0

27. Schmalenberger, K.M., Eisenlohr-Moul, T.A., Jarczok, M.N. *et al.* Associations of luteal phase changes in vagally mediated heart rate variability with premenstrual emotional changes. *BMC Women's Health* **24**, 448 (2024). https://doi.org/10.1186/s12905-024-03273-y

28. Knapp-Kline, K. & Kline, J. P. Heart rate, heart rate variability, and heartbeat detection with the method of constant stimuli: Slow and steady wins the race. *Biol. Psychol.* **69**, 387–396 (2005).

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

30. Schmalenberger, K. M. *et al.* How to study the menstrual cycle: Practical tools and recommendations. *Psychoneuroendocrinology* vol. 123 Preprint at https://doi.org/10.1016/j.psyneuen.2020.104895 (2021).

31. Doornweerd, A. M. & Gerritsen, L. 28 days later: A prospective daily study on psychological well-being across the menstrual cycle and the effects of hormones and oral contraceptives. *Psychol. Med.* **55**, e19 (2025).

32. Taylor, S. *et al.* Robust Dimensions of Anxiety Sensitivity: Development and Initial Validation of the Anxiety Sensitivity Index-3. *Psychol. Assess.* **19**, 176–188 (2007).

33. Shevchenko, Y., Kuhlmann, T. & Reips, U.-D. Samply: A user-friendly smartphone app and web-based means of scheduling and sending mobile notifications for experience-sampling research. https://doi.org/10.3758/s13428-020-01527-9/Published doi:10.3758/s13428-020-01527-9/Published.

34. Plans, D. *et al.* Measuring interoception: The phase adjustment task. *Biol. Psychol.* **165**, (2021).

35. Palmer, R. *et al.* Refinements of the phase adjustment task (PAT 2.0). *Front. Psychol.* **17**, 1677186 (2026).

36. Billings, E. L., Brown, J. B., Billings, J. J. & Burger, H. G. Symptoms and hormonal changes accompanying ovulation. *The Lancet* **299**, 282–284 (1972).

37. Bull, J. R. *et al.* Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. *npj Digital Medicine 2019 2:1* **2**, 83- (2019).

38. Evans, S. S., Repasky, E. A. & Fisher, D. T. Fever and the thermal regulation of immunity: The immune system feels the heat. *Nat. Rev. Immunol.* **15**, 335–349 (2015).

39. Gierk, B., Kohlmann, S., Hagemann-Goebel, M., Löwe, B. & Nestoriuc, Y. Monitoring somatic symptoms in patients with mental disorders: Sensitivity to change and minimal clinically important difference of the Somatic Symptom Scale-8 (SSS-8) ☆. https://doi.org/10.1016/j.genhosppsych.2017.07.002 (2017) doi:10.1016/j.genhosppsych.2017.07.002.

40. Gierk, B. *et al.* The Somatic Symptom Scale-8 (SSS-8) A Brief Measure of Somatic Symptom Burden. *JAMA Intern Med* **174**, 399–407 (2014).

41. Steiner, M., Macdougall, M. & Brown, E. The premenstrual symptoms screening tool (PSST) for clinicians. *Arch. Womens Ment. Health* **6**, 203–209 (2003).

42. Gabriele, E., Spooner, R., Brewer, R. & Murphy, J. Dissociations between self-reported interoceptive accuracy and attention: Evidence from the Interoceptive Attention Scale. *Biol. Psychol.* **168**, 108243 (2022).

43. Poerio, G. L., Klabunde, M., Bird, G. & Murphy, J. Interoceptive attention and mood in daily life: an experience-sampling study. https://doi.org/10.1098/rstb.2023.0256 (2024).

44. Lin, X., Shen, H., Lin, J., Murphy, J., Wang, Y, Sun, Y., Wang, N., Wang, J., Wei, G., & Luo, F. Psychometric Validation and Refinement of the Chinese Interoceptive Accuracy Scale (IAS) in General and Clinical Populations. (2023). *Journal of Psychosomatic Research, 175,* 111541.

45. Trevisan, D. A., Mehling, W. E. & McPartland, J. C. Adaptive and Maladaptive Bodily Awareness: Distinguishing Interoceptive Sensibility and Interoceptive Attention from

Anxiety-Induced Somatization in Autism and Alexithymia. *Autism Research* **14**, 240–247 (2021).

46. Clemente, R., Murphy, A. & Murphy, J. The relationship between self-reported interoception and anxiety: A systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **167**, 105923 (2024).

47. Stricker, R. *et al.* Establishment of detailed reference values for luteinizing hormone, follicle stimulating hormone, estradiol, and progesterone during different phases of the menstrual cycle on the Abbott ARCHITECT® analyzer. *Clin. Chem. Lab. Med.* **44**, 883–887 (2006).

48. Makowski, D., Ben-Shachar, M. & Lüdecke, D. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *J. Open Source Softw.* **4**, 1541 (2019).

49. Geller, J., Kubinec, R., Parlett, C. & Vuorre, M. A Beta Way: A Tutorial on Bayesian Beta Regression for Psychological Research. https://doi.org/10.31234/OSF.IO/D6V5T_V7 (2026) doi:10.31234/OSF.IO/D6V5T_V7.

50. Tenan, M. S., Brothers, R. M., Tweedell, A. J., Hackney, A. C. & Griffin, L. Changes in resting heart rate variability across the menstrual cycle. *Psychophysiology* **51**, 996–1004 (2014).

51. Ferrari, S. L. P. & Cribari-Neto, F. Beta Regression for Modelling Rates and Proportions. *J. Appl. Stat.* **31**, 799–815 (2004).

52. Bürkner, P. C. & Vuorre, M. Ordinal Regression Models in Psychology: A Tutorial. *Adv. Methods Pract. Psychol. Sci.* **2**, 77–101 (2019).

53. Liu, K. R. *et al.* Individual variation of follicular phase changes in endothelial function across two menstrual cycles. *Exp. Physiol.* **106**, 1389–1400 (2021).

54. Cohen, J. The Cost of Dichotomization. *Appl. Psychol. Meas.* **7**, 249–253 (1983).

55. Hawkins, S. M. & Matzuk, M. M. Menstrual Cycle: Basic Biology Overview of the Menstrual Cycle in Humans NIH Public Access. *Ann N Y Acad Sci* **1135**, 10–18 (2008).

56. Desmedt, O., Luminet, O., Walentynowicz, M. & Corneille, O. The new measures of interoceptive accuracy: A systematic review and assessment. *Neuroscience and Biobehavioral Reviews* vol. 153 Preprint at https://doi.org/10.1016/j.neubiorev.2023.105388 (2023).

57. Maric-Bilkan, C., Gilbert, E. L. & Ryan, M. J. Impact of ovarian function on cardiovascular health in women: focus on hypertension. *Int. J. Womens Health* **6**, 131 (2014).

58. Edwards, N., Wilcox, I., Polo, O. J. & Sullivan, C. E. Hypercapnic blood pressure response is greater during the luteal phase of the menstrual cycle. *https://doi.org/10.1152/jappl.1996.81.5.2142* **81**, 2142–2146 (1996).

59. Minson, C. T., Halliwill, J. R., Young, T. M. & Joyner, M. J. Influence of the menstrual cycle on sympathetic activity, baroreflex sensitivity, and vascular transduction in young women. *Circulation* **101**, 862–868 (2000).

60. O'Brien, W. H., Reid, G. J., & Jones, K. R. (1998). Differences in heartbeat awareness among males with higher and lower levels of systolic blood pressure. *International Journal of Psychophysiology,* **29**, 53-63.

61. Engelen, T., Solcà, M. & Tallon-Baudry, C. Interoceptive rhythms in the brain. *Nature Neuroscience 2023 26:10* **26**, 1670–1684 (2023).

62. Banellis, L. *et al.* Interoceptive ability is uncorrelated across respiratory and cardiac axes in a large scale psychophysical study. *Communications Psychology 2026 4:1* **4**, 43- (2026).

63. Crucianelli, L., Enmalm, A. & Ehrsson, H. H. Interoception as independent cardiac, thermosensory, nociceptive, and affective touch perceptual submodalities. *Biol. Psychol.* **172**, 108355 (2022).

64. Ferentzi, E. *et al.* Multichannel investigation of interoception: Sensitivity is not a generalizable feature. *Front. Hum. Neurosci.* **12**, 362681 (2018).

65. Gaigg, S. B., Cornell, A. S. F. & Bird, G. The psychophysiological mechanisms of alexithymia in autism spectrum disorder. *Autism* **22**, 227–231 (2018).

66. Hromatko, I. & Mikac, U. A Mid-Cycle Rise in Positive and Drop in Negative Moods among Healthy Young Women: A Pilot Study. *Brain Sciences 2023, Vol. 13, Page 105* **13**, 105 (2023).

67. Laessle, R. G., Tuschi, R. J., Schweiger, U. & Pirke, K. M. Mood changes and physical complaints during the normal menstrual cycle in healthy young women. *Psychoneuroendocrinology* **15**, 131–138 (1990).

68. Schandry, R. Heart Beat Perception and Emotional Experience. *Psychophysiology* **18**, 483–488 (1981).

69. Ring, C. & Brener, J. Influence of beliefs about heart rate and actual heart rate on heartbeat counting. *Psychophysiology* **33**, 541–546 (1996).

70. Murphy, J. & Bird, G. Auditing measures of interoceptive accuracy: Important clarifications. *Neurosci. Biobehav. Rev.* **158**, 105557 (2024).

71. Pollatos, O., Traut-Mattausch, E., Schroeder, H. & Schandry, R. Interoceptive awareness mediates the relationship between anxiety and the intensity of unpleasant feelings. *J. Anxiety Disord.* **21**, 931–943 (2007).

72. Ferentzi, E., Drew, R., Tihanyi, B. T. & Köteles, F. Interoceptive accuracy and body awareness – Temporal and longitudinal associations in a non-clinical sample. *Physiol. Behav.* **184**, 100–107 (2018).

73. Santos, L. E. R. *et al.* Reliability of the Heartbeat Tracking Task to Assess Interoception. *Applied Psychophysiology and Biofeedback 2022 48:2* **48**, 171–178 (2022).

74. Murphy, J. *et al.* Testing the independence of self-reported interoceptive accuracy and attention. *Quarterly Journal of Experimental Psychology* **73**, 115–133 (2020).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- InteroceptionMenstrualCycle031126SupplementaryInformation.docx
- ExamplePhaseWeighting.csv
- OrientationSlides.pdf
- QuestionnaireTasklist.docx