nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-01046-9

Daily, weekly, seasonal and menstrual cycles in women's mood, behaviour and vital signs

Emma Pierson[1,2], Tim Althoff[3], Daniel Thomas[4], Paula Hillard[5] and Jure Leskovec[1,6]✉

Dimensions of human mood, behaviour and vital signs cycle over multiple timescales. However, it remains unclear which dimensions are most cyclical, and how daily, weekly, seasonal and menstrual cycles compare in magnitude. The menstrual cycle remains particularly understudied because, not being synchronized across the population, it will be averaged out unless menstrual cycles can be aligned before analysis. Here, we analyse 241 million observations from 3.3 million women across 109 countries, tracking 15 dimensions of mood, behaviour and vital signs using a women's health mobile app. Out of the daily, weekly, seasonal and menstrual cycles, the menstrual cycle had the greatest magnitude for most of the measured dimensions of mood, behaviour and vital signs. Mood, vital signs and sexual behaviour vary most substantially over the course of the menstrual cycle, while sleep and exercise behaviour remain more constant. Menstrual cycle effects are directionally consistent across countries.

Daily, weekly, seasonal and menstrual cycles in human behaviour, health and vital signs affect health and happiness. Daily cycles are implicated in sleep[1] and obesity[2]; seasonal cycles in mood disorders[3]; and the menstrual cycle in fertility[4], schizophrenia[5] and cancer[6].

However, three fundamental questions about cycles remain unanswered. First, no study has compared the magnitudes of the daily, weekly, seasonal and menstrual cycles, making it unclear which is associated most substantially with cyclic variation. Second, it is not clear which dimensions of behaviour, mood and vital signs actually cycle because previous studies of daily[7], seasonal[8] and menstrual cycles[9,10] have yielded conflicting results. For example, previous studies of peaks in negative mood have disagreed about the time of day they occur[7], the time of year they occur[8], where in the menstrual cycle they occur[9], and even whether they occur at all[9,11,12]. Third, because many previous datasets have been small scale and country specific, they have been unable to study how factors such as cultural background and age affect cycle dynamics, and it is unclear the extent to which their conclusions generalize. Uncertainty over how cyclic patterns generalize across cultures has clinical implications: for example, this was a primary argument against including premenstrual dysphoric disorder in the Diagnostic and Statistical Manual of Mental Disorders (DSM)[13].

The answers to these three questions are particularly unclear for the menstrual cycle, which has been understudied, to the detriment of women's health[14–17]. (Throughout this paper, we use the term 'women' to refer to women and people who menstruate, and 'women's health' analogously; however, not all menstruators are women, and not all women menstruate.) The menstrual cycle is more difficult to study because it is individual-specific: in contrast to many cycles which are synchronized across the entire population (such as a cyclic weekend increase in sleep duration), individual-specific cycles such as the menstrual cycle begin on different days for each person. While population-synchronized cycles can be studied by computing a population average at each timepoint[7,18], studying individual-specific cycles requires an additional piece of data: where each person is in their cycle at each timepoint (cycle phase). Failing to account for this renders individual-specific cycles invisible because cycles do not begin on the same day for each person, and so are averaged out.

Here we decompose women's behaviour, mood and vital signs into simultaneous daily, weekly, seasonal and menstrual cycles. We use cellphones as a data collection instrument, relying on an international dataset of 241 million observations from 3.3 million reproductive-age women who use the women's health mobile tracking application Clue by BioWink GmbH, which allows women across more than 100 countries to prospectively track more than 100 features (Fig. 1 and Supplementary Tables 1–3). In our primary analysis, we focus on 88 million observations from 499,000 women who log in at least 12 unique months (Supplementary Table 4). Such data from women's health tracking apps have only recently become available on a large scale as such apps have grown in popularity; a critical advantage that these datasets offer is information on menstrual cycle starts, enabling the computation of where each woman is in her cycle at each timepoint. Previous analyses of Clue data have reproduced known biological findings[19–22]. Using the Clue dataset, we analyse menstrual, daily, weekly and seasonal cycles in three dimensions of behaviour (sleep, exercise and sexual activity) nine dimensions of mood and three vital signs (resting heart rate (RHR), basal body temperature (BBT) and weight). We focus on mood, behaviour and vital signs because they are fundamental to human health and well-being and previous research has suggested that they may exhibit cycles across multiple timescales[7,18,23–28]. Within the three broad areas of mood, behaviour and vital signs, we choose our specific dimensions to analyse because they are logged by large numbers of women in our dataset (Supplementary Tables 3 and 4).

Results

We first illustrate that failing to account for the menstrual cycle substantially understates individual cyclic variation (Fig. 2). In Fig. 2,

[1]Department of Computer Science, Stanford University, Stanford, CA, USA. [2]Microsoft Research New England, Cambridge, MA, USA. [3]Paul G. Allen School of Computer Science & Engineering, University of Washington, Washington, WA, USA. [4]Clue by BioWink GmbH, Berlin, Germany. [5]Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA, USA. [6]Chan Zuckerberg Biohub, San Francisco, CA, USA. ✉e-mail: jure@cs.stanford.edu

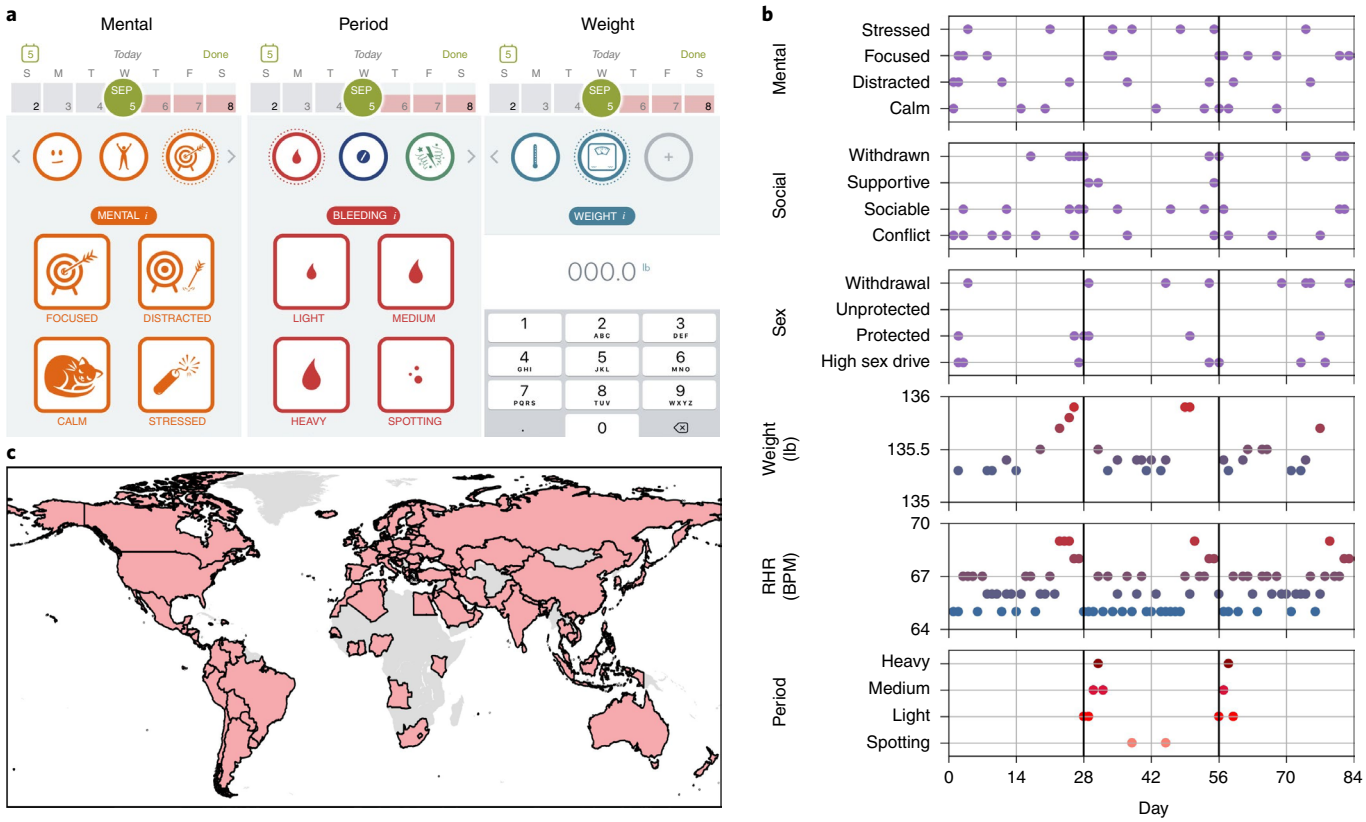

Fig. 1 | Overview of the dataset. a, Three screenshots from the women's health mobile app which was used to collect data, illustrating how women can enter logs about their mental state, period starts and weight. Mental and period logs are binary and can be entered by tapping a square on the screen; weight is continuous and entered as a numerical value. Screenshots reproduced with permission from the women's health mobile app Clue by BioWink GmbH. **b**, Illustrative simulated data for an individual woman; the horizontal axis is time, and the vertical black lines denote inferred menstrual cycle starts (Methods). Each dot represents the woman logging one feature on one day; only a subset of the features that can be logged are shown. The top three subplots illustrate the logging of binary features; the next two subplots illustrate weight and RHR, which are continuous; period logging data are shown in the bottom subplot. **c**, The 109 countries with at least 100 women and 1,000 observations in the dataset (across the more than 100 features which can be logged); Supplementary Table 2 reports the 25 countries with the most women in the dataset. Map shapefiles are from ArcGIS Hub (https://hub.arcgis.com/datasets/UIA::uia-world-countries-boundaries).

we show a daily signal (the happy–sad dimension of mood) averaged across the whole population. (To avoid falsely attenuating seasonal cycles, we filter for women in the Northern Hemisphere who log in at least 12 unique months; see Methods.) As expected, this reveals some population-synchronized cycles, such as weekly cycles, as well as outliers such as Christmas (when happiness increases) and the day after the 2016 US election (when happiness decreases). But averaging across the whole population conceals the menstrual cycle because menstrual cycles do not begin on the same day for every woman, and so menstrual variation is averaged out when all women are combined. To reveal this variation, we leverage the fact that our dataset contains the dates when menstrual cycles begin (that is, the dates when the period begins; see Methods) for each woman. This allows us to compute where each woman is in her menstrual cycle—the cycle phase—at each timepoint. We can then separate happy–sad mood into menstrual, daily, weekly and seasonal cycles by running a linear regression of happy–sad mood on categorical variables for day relative to period start, hour of day, day of week and month of year (Methods; Fig. 2b).

Our central finding is that the menstrual cycle is larger in magnitude than the other three cycles for most dimensions of mood, behaviour and vital signs. Specifically, it is the cycle with the largest amplitude (as measured by the cycle maximum minus its minimum) for seven out of nine dimensions of mood, sexual behaviour and all three vital signs (Fig. 2c). For happy–sad mood, the amplitude of the menstrual cycle (5.5%, 95% CI 5.4–5.7%) is 1.4× the amplitude of the daily cycle (95% CI 1.2–1.5×, $P<0.001$; amplitude of daily cycle, 4.0%, 95% CI 3.7–4.6%), 3.3× the amplitude of the weekly cycle (95% CI 3.2–3.6×, $P<0.001$; amplitude of weekly cycle 1.7%, 95% CI 1.6–1.8%) and 2.3× the amplitude of the seasonal cycle (95% CI 2.2–2.5×, $P<0.001$; amplitude of seasonal cycle, 2.4%, 95% CI 2.2–2.5%). Importantly, the amplitude of the menstrual cycle is also substantial relative to the overall mean: on average, 25% of happy–sad logs are sad, so the 5.5% menstrual cycle amplitude represents a 22% relative change in the probability of logging sadness. The amplitude of the menstrual cycle is also substantial relative to the effects of outlier events: it is about 1.7× the Christmas effect (95% CI 1.4–2.2×) and 0.6× the 2016 US election effect (95% CI 0.6–0.7×). (Supplementary Table 5 performs this analysis separately for each of the the five best-represented countries in our dataset, showing similar results for each country: the country-specific menstrual cycle amplitude is substantial relative to events which have large effects on happiness in each country.) Consequently, if menstrual variation could be observed in the population signal (Fig. 2a, red line) it would substantially increase cyclic variability. Studies of cyclic variation in mood, behaviour and vital signs will be more accurate if menstrual cycles are accounted for.

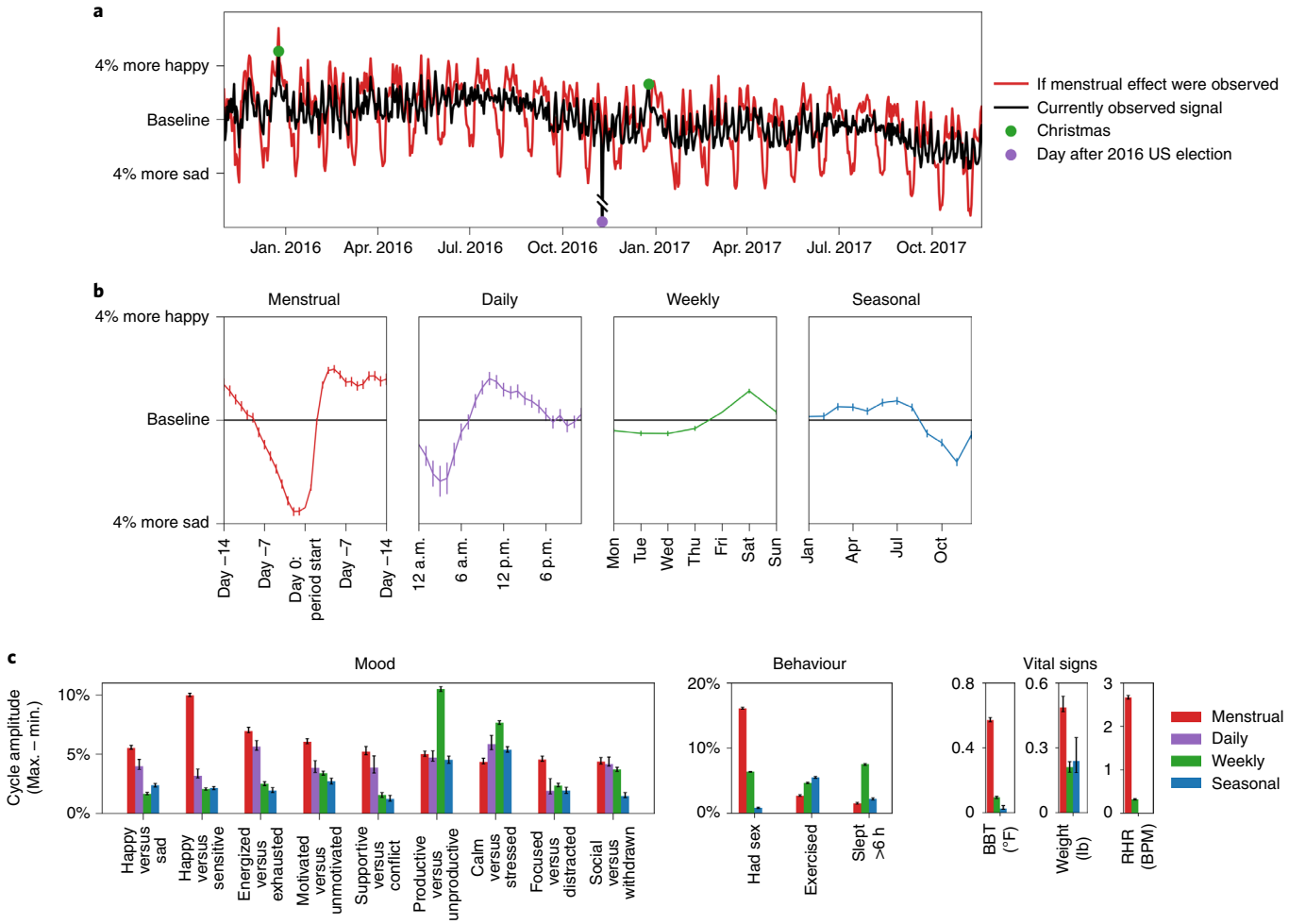

Fig. 2 | Decomposition of women's mood, behaviour and vital signs into daily, weekly, seasonal and menstrual cycles. a, Averaging across the entire population conceals menstrual cycles in happy–sad mood. The black line plots the fraction of happy–sad logs which are happy by day, after subtracting the baseline (individual mean) for each woman to reveal within-individual variation. Weekly cycles are apparent, as well as outlier events such as Christmas and the day after the 2016 US election, but this view is also limited: menstrual cycles cannot be seen because cycles do not begin on the same day for each woman, and so are averaged out. The red line plots what the black line would look like if the menstrual cycle could be observed (Methods), providing a more accurate picture of individual variability. The amplitude of the menstrual cycle is 1.7× the Christmas effect (95% CI 1.4–2.2×) and 0.6× the 2016 US election effect (95% CI 0.6–0.7×). (To focus on the predominant range of variation, the vertical axis truncates the US election outlier, which actually appears 10.3% below baseline.) **b**, Happy–sad mood can be broken down into menstrual, daily, weekly and seasonal cycles (Methods). **c**, Amplitude of cycles for all mood, behaviour and vital sign dimensions. Menstrual cycles (red bars) have the largest amplitude for seven out of nine dimensions of mood, sexual behaviour and all three vital signs: BBT, weight and RHR. The weekly cycle (green) is the largest cycle for productivity, stress and sleep, the daily cycle (purple) is prominent for most mood dimensions, and the seasonal cycle (blue) is relatively small in magnitude. For behaviour and vital signs, no hourly information is available, so the daily cycle is not shown; for RHR, less than a year of data is available, so the seasonal cycle is not shown.

We emphasize that we analyse variation within people, as opposed to between people[7,10,18]. For example, if a given person weighs 20 lb less than the population mean, most of that variation is likely due not to cyclic variation but to between-person variation, since the menstrual weight cycle amplitude is roughly half a pound. Our analysis implies that the menstrual cycle is a primary contributor to within-individual cyclic variation. As our analysis makes no comparison across genders or sexes, our results should not be taken to imply that women are more volatile than men, as discussed further below.

We next analyse the four-cycle decompositions of all 15 dimensions in our analysis (Supplementary Figs. 1–4), revealing intriguing variation. The menstrual cycle is greater in amplitude than the other three cycles for seven out of nine dimensions of mood (happy–sad, happy–sensitive, energized–exhausted, focused–distracted, motivated–unmotivated, sociable–withdrawn, supportive social–conflict social; for the seven dimensions where the menstrual cycle is the largest cycle, all P-values for amplitude differences are statistically significant except for the daily–menstrual comparison for sociable–withdrawn). The daily cycle in mood is also prominent, with mood becoming more negative between midnight and 6 a.m. (Supplementary Fig. 1). While seasonal and weekly cycles in mood are generally smaller than the daily and menstrual cycles, they are large for the calm–stressed and productive–unproductive dimensions: calm peaks during the summer and end-of-year (likely due to school holidays, since a large fraction of the population in the dataset is school-age) and on weekends, and productivity declines. These results speak to the importance of modelling multiple dimensions of mood, which may show different cyclic patterns. For behaviour dimensions, the menstrual cycle is the largest cycle for sexual activity ($P<0.001$), with a large decrease in logged sexual activity immediately after the period begins; sexual activity

also increases on the weekends (Supplementary Fig. 2). For sleep, the weekly cycle is most prominent, with an increase in sleep on the weekends; for exercise, both weekly and seasonal cycles are prominent, with decreases in exercise on the weekends and during the winter (Supplementary Fig. 2). The menstrual cycle is the largest cycle for all three vital signs (all $P<0.001$): BBT, RHR and weight (Supplementary Fig. 3).

We confirm that the amplitudes and cyclic patterns we observe remain stable under a number of robustness checks (Methods). First, we repeat our analysis under alternate statistical models, including two other cycle decomposition methods (a mixed model and a model-free estimation method), alternate parameterizations of the seasonal cycle (using a linear time trend rather than an indicator for year, and using indicator variables for week of year rather than month of year) and alternate parameterizations of the menstrual cycle (using day relative to last cycle start, day relative to next cycle start and fraction of the way through the cycle). As a second robustness check, we repeat our main analyses across subsets of the dataset broken down by app usage variables and demographics. While, as expected, we observe some variation across subsets of the population, we find that the ordering of cycle amplitudes remains generally stable (Supplementary Figs. 5 and 6) and, importantly, the prominence of the menstrual cycle is not driven by a particular subgroup. This shows that, while the population in our dataset is likely non-representative, the prominence of the menstrual cycle persists across subpopulations, and our conclusions are more likely to generalize.

Having established the importance of the menstrual cycle, we next examine how it varies across countries. Because most previous studies of the menstrual cycle have been small scale and country specific, it has been unclear the extent to which menstrual effects persist across countries. This uncertainty has clinical implications. For example, whether premenstrual mood disorder should be included in the DSM-V has been disputed on the grounds that it might be culturally specific[13]. To examine how menstrual effects vary across countries, we define the 'premenstrual effect' for a dimension as the change in the mean value of the dimension during its premenstrual peak or trough. (We allow the timing of this peak or trough to vary across dimensions, as described in Methods and Supplementary Table 6.) Contrary to previous concerns that premenstrual effects are culturally specific, we find that they are directionally consistent across countries. For example, the premenstrual decrease in happiness occurs across all 87 countries we examine (Fig. 3). The other large premenstrual effects in mood, sexual behaviour and vital signs also remain directionally consistent across countries (Supplementary Fig. 7), although for RHR and BBT, lack of data reduces the number of countries we are able to study and our results should be interpreted with more caution. We confirmed that our estimates of country-specific premenstrual effects remained consistent (Methods and Supplementary Fig. 8) when we controlled for demographic covariates, behaviour covariates and app usage covariates.

We next examine how premenstrual effects vary by age (Fig. 4). Menstrual cycle dynamics are known to change with age[29–33], and understanding normal ageing-related changes is important for characterizing healthy menstrual patterns[31]; however, it has not been possible to study ageing trends in all the dimensions we consider on the scale of our dataset. The premenstrual negative mood effect increases with age: from 3.6% (95% CI 3.5–3.7%) in 15–20-year-olds to 5.4% in 30–35-year-olds (95% CI 5.1–5.7%), a relative increase of 51% (95% CI 41–61%, $P<0.001$). This is consistent with prior reports that premenstrual dysphoria can increase during the late reproductive years[31]. We also observe age trends in all three vital signs: With increasing age, the premenstrual effects for RHR and weight decrease in magnitude, while the premenstrual effect for BBT increases (all $P<0.001$). The increase in the premenstrual BBT

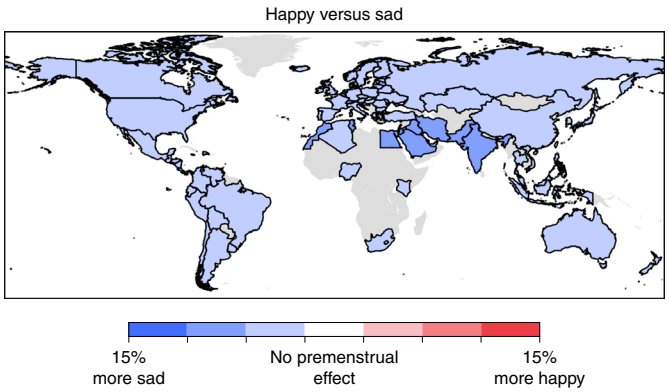

Fig. 3 | Premenstrual mood effects are directionally consistent across countries. The premenstrual effect for a dimension is defined as the change in the mean value of the dimension during its premenstrual peak or trough (Methods). All countries show a negative premenstrual mood effect. Colour bins are equally sized increments: blues indicate a negative premenstrual change in mood, the central white bin is centred around no premenstrual change and reds indicate a positive premenstrual change. Countries with fewer than 1,000 observations and 100 women in the dataset are shown in grey. Premenstrual effects in the other dimensions are directionally consistent as well (Supplementary Fig. 7). Map shapefiles are from ArcGIS Hub (https://hub.arcgis.com/datasets/UIA::uia-world-countries-boundaries).

effect with age is consistent with the fact that the fraction of cycles in which ovulation occurs increases with age[32], and BBT rises at ovulation[23]. While all these trends are robust to the inclusion of controls for demographics, behaviour and app usage (Supplementary Fig. 9), it is possible that unobserved heterogeneity also contributes to the age trends: for example, the age trend in BBT may also be driven in part by the fact that BBT is difficult to measure properly, and young women may be less skilled at it. (We mitigate this by filtering for regular BBT loggers; see Methods.) Because our analysis is cross-sectional (since our median follow-up time, of only 1.47 years for each person, is too short to allow longitudinal analysis), longer follow-up on large longitudinal datasets should further investigate the trends we observe.

Discussion

We use a large international dataset collected via a health tracking app to decompose women's behaviour, mood and vital signs into simultaneous daily, weekly, seasonal and menstrual cycles. We find that the menstrual cycle is the largest cycle for most of the 15 dimensions of mood, behaviour and vital signs that we study. Reassuringly, our study finds menstrual patterns that are consistent with the previous literature. For example, the cycle phase most commonly associated with negative mood in the previous literature is the premenstrual phase[9], which is consistent with our results (Fig. 2b and Supplementary Fig. 1). Similarly, our results concord with past findings on menstrual cycles in sexual activity, BBT, RHR and weight (Methods). However, importantly, the large size of our dataset also helps resolve ambiguities in the previous literature by more precisely pinning down how dimensions of mood, behaviour and vital signs fluctuate over the course of the menstrual cycle. For example, some studies have failed to find that negative mood fluctuates over the course of the menstrual cycle[9,11,12]. This is likely in part because of small sample size in previous studies—often only a few dozen people[9]—which limits statistical power to precisely resolve fluctuations.

Our dataset has limitations. First, the population in our dataset—smartphone users using a women's health app—is not representative

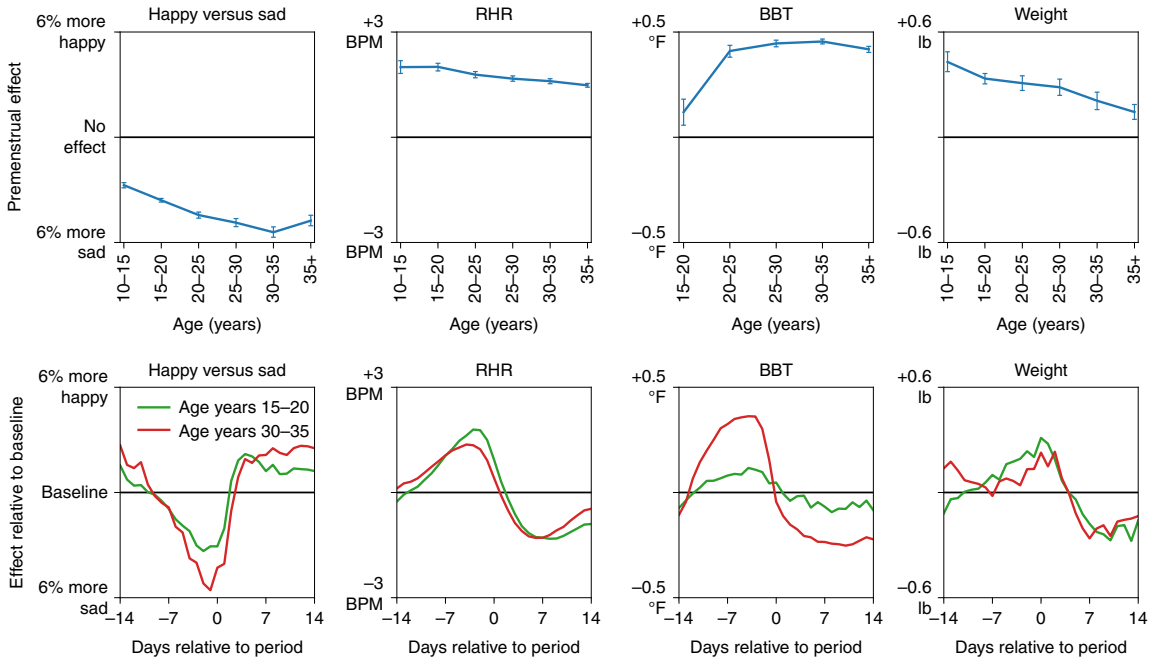

Fig. 4 | Premenstrual effects change with age for happy–sad mood, RHR, BBT and weight. The top row shows premenstrual effects (vertical axis) as a function of age (horizontal axis). The premenstrual effect for happy–sad mood grows larger with age: from 3.6% (95% CI 3.5–3.7%) in 15–20-year-olds to 5.4% in 30–35-year-olds (95% CI 5.1–5.7%), a relative increase of 51% (95% CI 41–61%, $P < 0.001$). The premenstrual effect for RHR grows smaller with age: from 2.0 BPM in 15–20-year-olds (95% CI 1.9–2.1 BPM) to 1.6 BPM in 30–35-year-olds (95% CI 1.5–1.7 BPM), a relative decrease of 20% (95% CI 15–25%, $P < 0.001$). The premenstrual effect for BBT increases with age: from 0.12 °F in 15–20-year-olds (95% CI 0.06–0.18 °F) to 0.45 °F in 30–35-year-olds (95% CI 0.44–0.47 °F), a 3.8× increase (95% CI 2.5–7.9×, $P < 0.001$). The premenstrual effect for weight decreases with age: from 0.33 lb in 15–20-year-olds (95% CI 0.30–0.36 lb) to 0.21 lb in 30–35-year-olds (95% CI 0.16–0.26 lb), a relative decrease of 38% (95% CI 21–54%, $P < 0.001$). Estimates of age effects are robust to inclusion of controls for demographics, behaviour and app usage (Supplementary Fig. 9). The bottom row shows the change in each dimension over the course of the menstrual cycle for 15–20-year-olds (green line) and 30–35-year-olds (red line). (All age groups have at least 1,000 observations and 100 women in the dataset; for BBT, the age group 10–15 is not shown because women in this age group log BBT too infrequently).

of the global population. Supplementary Table 2 illustrates that users from the USA are overrepresented in the dataset; further, the dataset is potentially biased towards women of higher socioeconomic status or who are particularly interested in women's health. This raises the question of whether our findings generalize. As we discuss in more detail in the Methods section, several lines of evidence mitigate this concern: (a) menstrual cycle apps are increasingly widely used[34], offering a data source which is arguably more representative than previous studies using small and non-representative populations[14], (b) previous analyses of this dataset[19–22] have reproduced numerous results consistent with previous findings, and we further confirm a number of previous results in our present analysis (Methods and Supplementary Fig. 10 and 11), and (c) the menstrual cycle remains prominent across countries and other subsets of the dataset, suggesting that this finding is not driven by a single, non-representative population. The size of the dataset enables us to confirm that our findings generalize across subsets of the population. For example, even though the USA is overrepresented in the data, we still have enough data to confirm that the premenstrual decrease in happiness occurs across all 87 countries we examine (Fig. 3). Perceptions of and stigma towards the menstrual cycle vary across cultures and can influence experiences and recording of the cycle (and indeed, in our data, the probability of using the app fluctuates over the course of the cycle; Supplementary Fig. 12)[16,35–38]. Cultural expectations for women to become sad during the premenstrual period may shape how mood is experienced and recorded. One promising direction for future work on datasets of this size is to study how menstrual cycle amplitudes vary across cultures, and how that correlates with other cultural characteristics.

A second limitation of our dataset is that it relies on self-reported data, which may not always be reliable, particularly for dimensions such as BBT which require skill to measure accurately. To mitigate this concern, we apply numerous quality control filters to increase the accuracy of logged data. The medical community is already making use of menstrual app tracking data[14,19–22,39], suggesting that neither of these limitations precludes its usefulness.

We emphasize that our results should not be taken to imply that women are more volatile than men. Our analysis, of course, makes no comparison across genders or sexes. Indeed, previous research in mice, rats and humans does not show sex differences in volatility, in part because males also experience hormone cycles[40]. Rather, our results illustrate the easily concealed importance of individual-specific cycles, a fact with implications beyond male and female hormone cycles. For example, genetics causes individual-specific variation in both daily and seasonal cycles[41]; genetic data allow study of such individual-specific variation.

Our results also do not imply that the menstrual cycle explains all the variation in women's mood, behaviour and vital signs. As discussed above, we analyse variation within people, as opposed to between people; further, we analyse cyclic variation specifically, but there are many other sources of within-person variation. As such, our results should not be taken to imply that attempts to target people based on their menstrual cycle phase—for example, for advertising—are likely to be effective[42,43] (separate, of course, from the substantial privacy concerns these practices raise). Our main finding is that the menstrual cycle is a primary contributor to within-individual cyclic variation.

This finding has implications for data collection in studies of women's health and behaviour, for clinical practice and for the broader cultural perception of the menstrual cycle. In terms of data collection, menstrual cycle information is often lacking from medical records[14] and global health data[44]; health-tracking apps have similarly been slow to incorporate menstrual tracking[45]. Without collecting menstrual cycle data, it is impossible to account for and study this fundamental source of variation in health and behaviour. Our findings also have implications with respect to clinical practice: clinicians do not always consider menstrual health[15], in spite of the fact that the menstrual cycle is considered to be a vital sign which can be used to monitor numerous health conditions[17,33]. Our findings emphasize the role of the menstrual cycle not just in disease, but in many dimensions of variation relevant to patients' well-being and behaviour. Finally, the menstrual cycle has been poorly understood and even stigmatized in popular culture, harming our understanding of women's health and behaviour[14,17]. Our analysis shows that the menstrual cycle, as a primary contributor to cyclic variation in women's mood, behaviour and vital signs, must be normalized and understood just as daily, weekly and seasonal cycles are.

Methods

User consent for research. All data used in this analysis are fully de-identified. All users in the dataset consent to use of data for research purposes: Clue obtains users' specific General Data Protection Regulation-compliant consent to the use of their de-identified data in scientific research collaborations whenever users create an account with Clue via Clue's privacy policy (https://helloclue.com/privacy). Analysis of this dataset was determined to be exempt from review by the Stanford Institutional Review Board (IRB-35159).

Clue app interface. Figure 1 shows screenshots from the Clue app interface to illustrate how women using the app log symptoms. Related categories of logs appear on the same screen. For example, all period log types (spotting, light, medium and heavy bleeding) appear on the same screen. All log types except for weight, BBT and RHR are binary, and can be logged simply by tapping the screen. Each binary log can be entered only once per day. Women can backfill information (that is, enter logs for previous days) by tapping the date at the top of the screen.

Data processing. *Definition of menstrual cycle start and date relative to period.* Consistent with the Clue app, we define a woman to be on her period when she logs light, medium or heavy bleeding. We define a period start as a start of bleeding when the woman has recorded no bleeding for at least 7 days. So if the woman records no bleeding in April, and bleeding on May 1, 2, 3 and 29, the period start dates would be May 1 and May 29. We remove the first logged period start for every woman because some women 'backfill' their first period start, potentially creating unreliable dates. We confirm that the distribution of period lengths is consistent with the previous literature (Supplementary Fig. 10). Given the period starts for each woman, we define the woman's 'day relative to period start' on each day d as $d-p$ (where p is the nearest period start as measured by absolute difference in days). So if the closest period start occurred on May 1, the day relative to period on April 30 would be −1, on May 1 would be 0 and on May 2 would be 1. We confirm that we observe similar cycle trajectories when we compute day relative to period start using only previous period starts or only subsequent period starts (so day relative to period start is always non-negative or always non-positive, respectively).

Location information and privacy. Our dataset contains women from the 25,000 cities with the most Clue users, covering 97% of the Clue population. Each city was mapped to its latitude (rounded to the nearest 5°), country and time zone; Clue removed latitude–time zone pairs with fewer than 1,000 women to protect user privacy, and provided us with the rounded latitude, country and time zone for each woman. We use each woman's time zone to compute the local time for each log, and use the woman's local time in all analyses.

Data filtering. For all dimensions, we filter for logs on dates between 1 November 2015 (since on earlier dates, not all features were available to be logged on the Clue app[46], rendering the data incomparable), and 20 November 2017 (when our dataset ends). We also filter out logs that are more than 20 days from a period start, since we will be unable to match these reliably to a period. In our primary analysis, we analyse women who log in at least 12 unique months, whom we term 'long-term loggers'. We do this to ensure that we have long enough time spans to estimate seasonal cycles and that women log reliably, mitigating missing data concerns. We verify that our estimates of cycle amplitudes do not change substantially when we examine the remainder of the population (Supplementary Fig. 5).

Because seasonal cycles could potentially be reversed or otherwise altered in the Southern and Northern hemispheres, in our analyses which compare with seasonal cycles (for example, Fig. 2) we analyse only women in the Northern Hemisphere to avoid falsely attenuating the seasonal cycle. We verify that cycle amplitudes remain stable when we substratify by latitude subgroups (for example, women between 25 °S and the Equator). In all analyses which focus on the menstrual cycle and do not consider seasonal cycles (for example, Figs. 3 and 4), we analyse all women and do not filter for the Northern Hemisphere.

Binary dimensions: mood and behaviour. To ensure that we have hourly information for logs and that logs are prospective (as opposed to logged weeks after the fact, potentially rendering memory unreliable), we filter for logs for which we have time information on when the logging session started and ended. We remove a very small fraction (<0.1%) of logs where the session lasts more than an hour, since this renders time information unreliable. To ensure logs are prospective, we remove logs where the woman enters a date for the log which is not the same day as the session start.

For each dimension of mood and behaviour, we must define a way of converting the categorical data that users log (for example, tapping 'calm' or 'happy' on the app screen) to a numerical value on which we can perform quantitative analysis. To define this value for each mood dimension, we pair each mood with an opposing mood: for example, happy with sad, or calm with stressed. (For all nine mood dimensions, both the mood and its opposing mood appear on the same screen on the app; in almost all cases, each mood feature has an unambiguous opposite—for example, calm–stressed, happy–sad, or motivated–unmotivated. One exception to this is sensitive mood, for which happy mood is only an approximate opposite; we include sensitive mood in the analysis in spite of this caveat because it is one of the most frequently logged symptoms, and is thus important to include; see Supplementary Table 4). The value for a given log is 1 if the woman logs the first mood and 0 if she logs the opposite mood, for example, 1 if the woman logs happy, and 0 if she logs sad. (We examine all mood dimensions except for premenstrual syndrome because it is obviously menstrual cycle specific, and high/low energy, because it is redundant with the energized–exhausted dimension we analyse). Thus, the average value of the happy–sad mood dimension can be interpreted as the 'fraction of happy–sad logs which are happy'. We define mood dimensions in this way to control for the fact that a woman's probability of logging fluctuates considerably over time of day and point in the menstrual cycle (Supplementary Fig. 12). For example, simply examining whether a woman logs a mood, without normalizing in some way, would show large spikes in all emotions near the menstrual cycle start, due simply to the fact that many women use the app primarily to track menstrual cycle starts and thus are more likely to log other features then as well.

Pairing emotions with their opposites is consistent with previous conceptualizations of emotion[47,48] which have regarded them as opposites. However, because previous investigations[7] have noted that opposite emotions may fluctuate somewhat independently of each other, we verify that our primary conclusions remain unchanged when we instead normalize using all loggable symptoms on the screen on which a mood appears—in essence, computing the probability that a woman logged a given mood given that she looked at the screen on which it was an option to log. Our primary conclusions remain unchanged under this parameterization: the observed cycle trajectories remain consistent (with positive moods declining, and negative moods increasing, before period start), and the menstrual and daily cycles remain most prominent.

We now describe how we map from categorical log data to numerical values for each of the three behaviour dimensions: sleep, sex and exercise. For sleep, the app offers users four logging categories: 0–3 h, 3–6 h, 6–9 h and more than 9 h. (We note that this is an imperfect measure of sleep for multiple reasons—sleep duration is self-reported, the categories are not that fine-grained and this measure also does not measure awakenings or sleep quality.) We define the value of the sleep dimension as 1 if the woman logged sleeping 6–9 h or more than 9 h, and 0 otherwise. We choose the cutoff of 6 h because it is closest to the threshold for a normal amount of sleep for an adult.[49] For the sex and exercise behaviour dimensions, we use logs of any type as a normalizer: that is, the sex and exercise dimensions are 1 on a day if a woman logged that behaviour on that day, and 0 if she logged only something other than that behaviour. (For the sex dimension, we include only protected, unprotected and withdrawal sex, but do not include the 'high sex drive' feature which the app also allows women to log.) To ensure that women are tracking the relevant data, we analyse only women who log the behaviour at least once. We do not analyse the time at which behaviours are logged, since a woman may log at 9 p.m. that she slept, had sex or exercised that day, but this does not necessarily indicate that she slept, had sex or exercised at 9 p.m.

Continuous dimensions: vital signs. For the three vital sign dimensions—BBT, RHR and weight—our dataset does not contain the time at which the log was entered, so we apply no time filtering and do not analyse daily cycles. For RHR, we have less than a year of reliable data, and so we do not analyse seasonal cycles. BBT and weight are self-reported; RHR is measured using a wearable device. We apply basic quality control filters to each vital sign to ensure data are reliable (since, in contrast to the binary mood and behaviour logs, for vital signs women can enter implausible values).

For weight, we filter out weights less than 50 or greater than 500 lb, women whose weight fluctuates by more than 50 lb (since this indicates women who may be trying to lose or gain weight or who may have different cyclic patterns, and may also indicate women who log unreliably) and women with fewer than five logs. For RHR, which is automatically measured by a heart rate monitor, we filter out women with logs on fewer than 50 days and women whose RHR fluctuates by more than 50 BPM (all RHR observations are in a biologically plausible range, so we do not filter for a maximum or minimum value). For BBT data, we filter out values below 90 °F or greater than 110 °F, women with fewer than 50 logs and women with fewer than 5 unique values of BBT (since BBT is somewhat difficult to measure, and we observe that some people log only implausibly constant values of BBT). A small fraction of women have multiple readings on a single day; we average these together so that there is only a single observation for each woman on each day.

Decomposition into four cycles. We separate the overall signal into daily, weekly, seasonal and menstrual cycles as follows: Because we are interested in within-individual variation, we first remove individual means, following previous literature[7,18]: that is, for each woman and each dimension, we subtract the woman's mean for that dimension, so each each woman has zero mean. We then run a linear regression of the observations x, where i indexes observations:

$$x_i \sim \mathcal{N}(\alpha + \beta_{y[i]} + \gamma_{p[i]} + \delta_{h[i]} + \eta_{w[i]} + \kappa_{m[i]}, \sigma^2)$$

where $y[i]$ is the year of the ith observation, $p[i]$ is its day relative to period, $h[i]$ is its hour, $w[i]$ is its weekday, $m[i]$ is its month and all are encoded as categorical variables (with a distinct coefficient for each value). y ranges from 2015 to 2017; p ranges from −20 to 20; h ranges from 0 to 23; w is the seven days of the week; and m ranges from 1 to 12. In all four-cycle decomposition plots (for example, Fig. 2b), we extract the relevant coefficients and zero-mean them; the 'Baseline' label on the plot indicates a coefficient of zero. (Similarly, when we plot the daily signal in Fig. 2a, we subtract off the mean signal across days, so the average signal value is zero; we label this as 'Baseline' on the plot. To plot the red line in Fig. 2a—the counterfactual world where menstrual cycle effects are observed rather than averaged out—we add the inferred menstrual cycle effect, as plotted in Fig. 2b, to the observed signal (black line) in Fig. 2a.) Error bars are 95% confidence intervals; throughout the paper, all confidence intervals are computed using cluster-robust procedures, clustering at the person level. This accounts for the fact that each person in the dataset may provide multiple observations over time. We plot only estimates for day relative to period from −14 to 14 because the average menstrual cycle is roughly 28–29 days long. Day −14 is not equivalent to day 14, since many women have cycles longer or shorter than the typical length of 28–29 days, so we would not necessarily expect the regression coefficients for day −14 and day 14 to be equivalent; some discontinuities at the graph boundaries are expected. One advantage of our large dataset is that it allows us to precisely estimate day-specific coefficients for the menstrual cycle, rather than making potentially false parametric assumptions about how mood, behaviour and vital signs will change over the course of the cycle[10].

We define the amplitude of a cycle as the difference between the maximum and minimum coefficient values for the cycle. We compute confidence intervals on this amplitude (Fig. 2c) by bootstrapping replicate datasets, recomputing regressions and amplitudes for each replicate, and computing the 95% confidence interval of the bootstrapped amplitudes.

Computation of premenstrual effects. We define the premenstrual effect for a dimension as follows:

1. We first determine the premenstrual interval, that is, the week-long interval (beginning up to 2 weeks before period start) in which the dimension's mean value (as measured by the regression coefficients shown in Supplementary Figs. 1–3) differs most dramatically from its overall mean (Supplementary Table 6). Essentially, this captures the interval in which the cycle curve reaches its most pronounced peak or trough. As Supplementary Figs. 1–3 illustrate, dimensions display pronounced peaks or troughs beginning during the 2-week premenstrual interval, justifying our examination of a premenstrual effect; however, the exact timing of this peak or trough varies by dimension, as previous authors have also observed[21], justifying our use of a specific week-long period for each dimension. For example, for sexual activity, this interval begins at day −1 and ends at day 6 (where 0 denotes the period start date), and for BBT the interval begins at day −8 and ends at day −1.
2. After defining the premenstrual interval for each dimension, we define the premenstrual effect as the dimension's average value during the premenstrual interval minus the dimension's average value not during the interval. (As with the four-cycle analyses, we subtract each woman's mean before taking the averages). This is equivalent to performing a linear regression

$$x_i \sim \mathcal{N}(\alpha + \beta_{p[i]}, \sigma^2)$$

where $p[i]$ indicates that the ith observation occurs during the premenstrual interval. (We note that simply computing the amplitude of the premenstrual peak or trough, while intuitive for the population as a whole, would not allow us to perform a multiple regression, and thus to determine the effects of multiple covariates such as age or country.)

To compute country-specific premenstrual effects, we perform the regression

$$x_i \sim \mathcal{N}(\alpha + \beta_{p[i]} + \gamma_{c[i]} + \delta_{p[i]c[i]}, \sigma^2)$$

where $c[i]$ denotes the country of the ith observation, and the interaction term $\delta_{p[i]c[i]}$ allows premenstrual effects to differ by country. To confirm that our country-specific effects are robust to inclusion of other covariates (for example, age) we also fit models

$$x_i \sim \mathcal{N}(\alpha + \beta_{p[i]} + \gamma_{c[i]} + \delta_{p[i]c[i]} + \eta_{a[i]} + \kappa_{p[i]a[i]}, \sigma^2)$$

where $a[i]$ denotes the age group of the ith observation, and κ_{pa} allows for age-specific premenstrual effects. Besides age, the additional covariates we include are behaviour controls (if the woman has ever logged consuming alcohol, consuming cigarettes, exercise, taking hormonal birth control, taking a birth control pill or using an intrauterine device) and app usage controls (number of symptom categories used, start year and total symptoms logged). The goal of this analysis is to ensure that our country-specific estimates are not driven by other country-specific differences.

For each fitted model, we compute the country-specific premenstrual effect for each country c by setting the country for all women to c (keeping their other covariates the same) and computing the difference between their model-predicted value during the premenstrual interval and not during the premenstrual interval. This can be interpreted as the model's predicted premenstrual effect for each country if the distribution of all non-country covariates were that of the population as a whole. (For computational tractability, given the large number of countries in the dataset, we compute this quantity on a random sample of 100,000 women; our confidence intervals, which are very small, are thus conservative. We compute 95% confidence intervals by resampling the entire dataset using bootstrapping, repeating the premenstrual effects estimation procedure on each bootstrapped replicate and computing the 95% confidence interval of the bootstrapped premenstrual effects.)

Our computations for age-specific premenstrual effects are analogous. In both cases, we find that inclusion of other covariates does not substantially change our age- or country-specific estimates. Our country-specific premenstrual effects are robust to inclusion of other covariates (Supplementary Fig. 8); similarly, we infer the same age trends regardless of which other covariates we include (Supplementary Fig. 9).

Identification of country-specific outlier events. In Supplementary Table 5, we identify events in each country which have a large effect on happy–sad mood and compare this effect with the amplitude of the menstrual cycle. To identify the most important event for each country, we compute the fraction of happy–sad logs which are happy by day, after subtracting the baseline (individual mean) for each woman to reveal within-individual variation (for example, the time series plotted in Fig. 2a, black line). To avoid noisy single-day outliers, we analyse only the five largest countries in Supplementary Table 2 and filter for days with at least 250 logs; to ensure that this filter does not create holes in our time period, we analyse only the last 18 months of data, which has more logs. We then identify the date, for each country, with the largest deviation from the overall time series mean. In the USA, this is the day after the 2016 presidential election; in three countries (France, Mexico and Britain) it is the day after a terrorist attack or natural disaster; and in Brazil, it is New Year's Day. We compare the country-specific amplitude of the menstrual cycle with the country-specific effect of this event (as measured by the difference in happy–sad mood on that date and the average happy–sad mood in the month in which the event occurred). For all five countries, the amplitude of the menstrual cycle is substantial relative to the event.

Robustness checks. *Alternate four-cycle decomposition methods.* We compare our linear regression four-cycle decomposition with two other cycle decomposition methods: taking the means for each cycle separately, and fitting a mixed model.

- **Means by group**. Rather than fitting a linear regression, we fit a simpler model where, after subtracting the means for each individual, we simply compute the means for each day relative to period, hour, weekday and month. This does not allow us to control for year or account for correlations between cycles (for example, if women tend to log when they are on their period only if it is a Sunday) but yields similar results to linear regression.
- **Mixed model**. We fit a mixed model which includes the same fixed effects as in the linear regression, but rather than removing the individual mean, we fit a random-effects intercept term for each individual. The motivation is that, while simply removing individual means is interpretable, scalable to large datasets and used in previous studies[7], it can also potentially lead to misleading estimates of cyclic effects by incorrectly attributing cyclic variation to between-individual variation. For example, if each individual logs for only a week, removing individual means may attenuate seasonal cycle estimates. We use a linear mixed model even though some of the dimensions are binary because we want the fixed-effects coefficients to be interpretable and

comparable to our estimates from the other two methods. We downsample the data for each dimension to a maximum of 100,000 individuals (randomly selected) for computational tractability.

Both methods yield very similar results to linear regression. The mixed model yields slightly larger estimates of cycle estimates (the estimated menstrual cycle amplitudes are 27% larger on average across all dimensions), so our estimates of cycle effects should be regarded as conservative. However, our main conclusions remain similar under the two alternative specifications: in particular, menstrual cycle amplitudes remain generally larger than those of the other three cycles. We therefore favour the linear regression model for its simplicity and because it scales to the entire dataset. (The only case in which the models yield somewhat different results is for seasonal cycles in weight, for which the mixed model and linear regression model estimate qualitatively similar trends but the mixed model estimates a considerably larger amplitude—0.65 versus 0.24 lb for the linear regression model. This discrepancy occurs because there is substantial change in the average weight of the population over time, likely caused by the expansion in the population using the app, and the mixed model attributes this to seasonal variation. Our estimates of seasonal changes in weight thus ought to be regarded as conservative, and it is possible that the true amplitude of the seasonal cycle effect is somewhat larger.)

Alternate parameterization of seasonal cycles. Because there are multiple ways of parameterizing seasonal cycles, we assess how our results vary under two alternate regression parameterizations: (1) replacing the month-of-year indicator variable κ_m with a week-of-year indicator variable and (2) replacing the year indicator variable β_y with a linear time trend. Our results remain similar under these alternate parameterizations. Replacing month-of-year with week-of-year slightly increases the amplitude of the seasonal cycle, as expected, due both to noise and to greater sensitivity to transient events such as Christmas, but estimates for the seasonal cycle are similar and estimates for other cycles are nearly identical; we favour the month-of-year parameterization because it allows us to more robustly estimate the seasonal cycle for small subgroups of the population in our substratification robustness analysis (Supplementary Fig. 5). Replacing the year indicator variable β_y with a linear time trend also produces generally similar results.

Alternate parameterization of menstrual cycles. There are multiple ways of parameterizing the menstrual cycle; we explore three alternative parameterizations in addition to our primary parameterization (ranging from −20, where the nearest cycle start is 20 days in the future, to 20, where the nearest cycle start is 20 days in the past):

- **Day relative to last cycle start**: for every date on which a woman logged, we compute the date of the last cycle start on or before the log date and take the difference in days. To ensure we have reliable information for cycle day, in this analysis we discard logs with more than 40 days since last cycle start.
- **Day relative to next cycle start**: for every date on which a woman logged, we compute the date of the next cycle start on or after the log date and take the difference in days. We discard logs with more than 40 days until the next cycle start.
- **Fraction of the way through cycle**: because average cycle length varies across women, for each woman and each log, we compute the log's cycle day (as in our primary parameterization) and divide by the woman's average cycle length. For example, if the log occurred on cycle day 9 and the woman's average cycle was 30 days long, the fraction of the way through the cycle f would be $\frac{9}{30} = 0.3$. We analyse logs with $-0.5 \leq f \leq 0.5$.

Our estimated menstrual cycle amplitudes remain stable under these three alternate parameterizations of the menstrual cycle, confirming that our conclusions are robust to our parameterization of the menstrual cycle. Specifically, in all cases our estimated menstrual cycle amplitude is within 20% of our original estimate, and the average difference between the original and alternate parameterizations is 4.8%.

Substratification robustness checks. The population using the app is likely non-representative; we therefore verify that the menstrual cycle remains prominent when we substratify the dataset by demographics (age group, country and latitude), app usage variables (number of categories logged and number of symptoms logged) and whether women logged in at least 12 unique months and were included in the main analysis, to ensure that this filtering does not change conclusions (Supplementary Fig. 5). For each subgroup, we compute the amplitude of the menstrual, daily, weekly and seasonal cycles for that subgroup alone. Unsurprisingly, we observe some variation in the cycle amplitudes across subgroups (variation across subgroups is, after all, the phenomenon that our age subgroup analysis highlights). However, importantly, the overall ordering of amplitudes remains largely stable across subgroups for each dimension, and in particular, the prominence of the menstrual cycle does not appear to be driven by any particular subgroup. This indicates that our central finding of the prominence of the menstrual cycle is unlikely to be driven by a non-representative subgroup.

While some previous analyses of the menstrual cycle[19,29,50] have studied only women who are not taking hormonal birth control (and are thus experiencing their natural hormone cycle), we do not apply this filter in our main analyses for two reasons. First, the goal of our analysis is to describe cycles in the general population of women, not to describe the effect of the natural hormone cycle specifically. Second, women in our dataset do not reliably log hormonal birth control information, so a hormonal birth control filter would be highly imperfect. Specifically, our dataset only contains information on whether women logged using birth control on each day, but many women who use daily birth control may not bother logging it. Thus, we cannot reliably differentiate between women who were truly not taking birth control and those who merely did not log it. This caveat notwithstanding, birth control status is an important covariate in analyses of the menstrual cycle, and we therefore plot how cycle amplitudes vary as a function of birth control status in Supplementary Fig. 6. We examine two subgroups: those who have ever logged taking any kind of hormonal birth control (the data include categories for 'injection', 'patch', 'pill' and 'ring') and those who have logged taking a hormonal birth control pill specifically, since this is the most commonly logged hormonal birth control category. While, as expected[51], the amplitude of the menstrual cycle changes with birth control status for some dimensions (most notably, sexual behaviour, BBT and RHR), the overall ordering of amplitudes remains stable across subgroups.

Reproducing previously known results. Here we describe our procedures for replicating previously known findings. The purpose of this analysis is to verify that our dataset is sufficiently reliable to reproduce previous results.

Menstrual cycle lengths. The means and standard deviations of menstrual cycle lengths in our dataset match previous findings by Chiazze et al.[52] (Supplementary Fig. 10). Following those authors, we filter for cycles between 15 and 45 days in length and stratify by age group. Both the mean and standard deviation for each age group closely match the previous estimates; our data also recapitulate the slight decrease in mean and standard deviation by age.

Country-specific patterns. We compare our dataset's country-specific measures of happiness and weight with previous country-specific measures of happiness and weight. Our source for previous country-specific happiness data is the Gallup World Poll, which has been used in a previous study of happiness[53]. Following those authors, we use Gallup's measures of real-time positive/negative experiences (as measured by experiences on the day before the survey) and overall life evaluation. We examine the correlations between Gallup's measures and our dataset's fraction of happy–sad logs which are happy across the 79 countries with Gallup data and at least 1,000 happy–sad logs from 100 unique women in our dataset. All correlations are statistically significant and have the expected sign: the correlation with Gallup's real-time positive experience index is $r = 0.51$ ($P < 0.001$), with Gallup's real-time negative experience index is $r = -0.38$ ($P < 0.001$) and with Gallup's overall life evaluation is $r = 0.47$ ($P < 0.001$).

For weight, we compare the average weight of women in our dataset who provide weight data with previous country-specific measures of weight and obesity. We study the 38 countries for which we have weight/obesity data from an external dataset and at least 1,000 weight logs from 100 women in our dataset. The average weight of women in each country (in our dataset) is significantly correlated with 2016 World Health Organization estimates of the fraction of women over 18 who are overweight[54] (body mass index, BMI > 25 kg m^{-2}) ($r = 0.53, P < 0.001$), the fraction of women over 18 who are obese[55] (BMI > 30 kg m^{-2}) ($r = 0.59, P < 0.001$) and the average weight of adults in that country[56] ($r = 0.51, P = 0.001$).

Previously known cycles. In addition to the dimensions of mood, behaviour and vital signs considered in our primary analysis, we investigate whether our dataset displays expected seasonal, weekly and menstrual cycles for additional dimensions (Supplementary Fig. 11). (We do not investigate additional dimensions in daily cycles because the dimensions for which we have the most reliable time information—mood symptoms—are included in our primary analysis.) As with all our analyses which include seasonal cycles, we filter for women in the Northern Hemisphere.

For menstrual cycles, we examine the four physical pain symptoms that the Clue app allows women to log—cramps, headache, ovulation pain and tender breasts—since these are some of the best-studied and most commonly logged menstrual symptoms. Consistent with prior research, we observe in our dataset that cramps[57,58], breast pain[59] and headache[58] are more commonly reported by women near period start. In contrast, women in our dataset are most likely to report ovulation pain at about day −14 (that is, 14 days before the start of their next period), consistent with prior findings of 'Mittelschmerz' pain occurring near ovulation[60]. (Ovulation pain also shows a slightly smaller peak at day 14 in our dataset, equivalent to day −14 for women with cycles near the typical length of 28 days.)

For weekly cycles, we examine patterns in alcohol consumption. Previous findings indicate that alcohol consumption increases on the weekend[61,62]; consistent with this, we observe that women are more likely to report attending parties with alcohol on the weekends, with logs of hangovers peaking on Sundays.

For seasonal cycles, we investigate allergies, cold/flu symptoms and vacationing. Seasonal allergies have been previously found to peak in spring

and summer because of higher pollen counts[63–65]. Incidence of flu peaks in the winter[66,67]. Vacation is more commonly taken during the summer months and near the winter holidays[68]. Our dataset displays patterns consistent with all these prior findings.

Generalizability of findings. Since women who use menstrual cycle tracking apps likely differ from the general population, it is natural to ask how our findings generalize. While a population of millions of women is arguably large enough to be worth studying in and of itself even if it is somewhat non-representative, three lines of evidence support the generalizability of our findings:

1. Menstrual tracking apps are increasingly widely used[34], offering a data source which is plausibly more representative than those used in previous studies of the menstrual cycle, which have used small and non-representative populations[31].
2. Previous studies of the dataset have found that it replicates known biology—for example, menstrual cycle lengths[19] and premenstrual symptoms[20,21]. Our present analysis similarly produces numerous findings consistent with previous studies: a late-night shift towards negative mood[7,69]; estimates of menstrual BBT, RHR and weight cycle amplitudes[23,50,70,71]; decreases in exercise on the weekends and during the winter[24,25]; and decreases in sexual activity immediately after the period begins and during the weekdays[26]. In addition to our main analysis, we replicate a number of other previous results on our dataset. These replications suggest that our dataset is sufficiently reliable to reproduce previous results.
3. Our results remain stable and the menstrual cycle remains prominent when we substratify the dataset by demographics and app usage variables (Supplementary Fig. 5), indicating that our results are unlikely to be driven, for example, by very young women or by women who use the app more frequently because they have particularly severe menstrual symptoms.

Reliability of self-reported data. We apply several quality control filters to ensure the accuracy of logged data: for mood and behaviour dimensions, we filter for logs that are prospectively recorded (logged on the same day on which they occur), since prospective logging is more reliable[7]. For vital sign dimensions, we filter for women who log frequently, and remove biologically implausible values. For mood dimensions, one concern is that, because the app interface renders the menstrual cycle more salient, it primes women to report subjective period-related mood changes. This is an unavoidable concern in studies of menstrual mood changes; further, if menstrual cycle apps are priming large fractions of the population to experience negative mood near their periods, that is itself a phenomenon worth studying. The more objectively recorded behaviour and vital sign dimensions should be unaffected by this phenomenon.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data that support the findings of this study are available from the corresponding author (J.L.) with appropriate permission from Clue. The data are not publicly available to preserve the privacy of Clue users.

Code availability

Code to reproduce the findings of this study is available at https://github.com/epierson9/four-cycles.

Received: 14 April 2020; Accepted: 23 December 2020;
Published online: 01 February 2021

References

1. Van Dongen, H. P. & Dinges, D. F. Circadian rhythms in fatigue, alertness, and performance. *Princ. Pract. Sleep. Med.* **20**, 391–399 (2000).
2. Roenneberg, T., Allebrandt, K. V., Merrow, M. & Vetter, C. Social jetlag and obesity. *Curr. Biol.* **22**, 939–943 (2012).
3. Partonen, T. & Lönnqvist, J. Seasonal affective disorder. *CNS Drugs* **9**, 203–212 (1998).
4. Wilcox, A. J., Dunson, D. & Baird, D. D. The timing of the 'fertile window' in the menstrual cycle: day specific estimates from a prospective study. *BMJ* **321**, 1259–1262 (2000).
5. Bergemann, N., Parzer, P., Runnebaum, B., Resch, F. & Mundt, C. Estrogen, menstrual cycle phases, and psychopathology in women suffering from schizophrenia. *Psychol. Med.* **37**, 1427–1436 (2007).
6. Badwe, R. et al. Timing of surgery during menstrual cycle and survival of premenopausal women with operable breast cancer. *Lancet* **337**, 1261–1264 (1991).
7. Golder, S. A. & Macy, M. W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* **333**, 1878–1881 (2011).
8. Murray, G., Allen, N. B. & Trinder, J. A longitudinal investigation of seasonal variation in mood. *Chronobiol. Int.* **18**, 875–891 (2001).
9. Romans, S., Clarkson, R., Einstein, G., Petrovic, M. & Stewart, D. Mood and the menstrual cycle: a review of prospective data studies. *Gend. Med.* **9**, 361–384 (2012).
10. Lorenz, T. K., Gesselman, A. N. & Vitzthum, V. J. Variance in mood symptoms across menstrual cycles: implications for premenstrual dysphoric disorder. *Women's Reprod. Health* **4**, 77–88 (2017).
11. Bosman, R. C., Albers, C. J., de Jong, J., Batalas, N. & aan het Rot, M. No menstrual cyclicity in mood and interpersonal behaviour in nine women with self-reported premenstrual syndrome. *Psychopathology* **51**, 290–294 (2018).
12. Hengartner, M. P. et al. Negative affect is unrelated to fluctuations in hormone levels across the menstrual cycle: evidence from a multisite observational study across two successive cycles. *J. Psychosom. Res.* **99**, 21–27 (2017).
13. Hartlage, S. A., Breaux, C. A. & Yonkers, K. A. Addressing concerns about the inclusion of premenstrual dysphoric disorder in DSM-5. *J. Clin. Psychiatry* **75**, 70–76 (2014).
14. Chakradhar, S. Discovery cycle. *Nat. Med.* **24**, 1082–1085 (2018).
15. Elford, K. & Spence, J. The forgotten female: pediatric and adolescent gynecological concerns and their reproductive consequences. *J. Pediatr. Adolesc. Gynecol.* **15**, 65–77 (2002).
16. Johnston-Robledo, I. & Chrisler, J. C. The menstrual mark: menstruation as social stigma. *Sex. Roles* **68**, 9–18 (2013).
17. Hillard, P. J. A. Menstruation in adolescents: what do we know? and what do we do with the information? *J. Pediatr. Adolesc. Gynecol.* **27**, 309–319 (2014).
18. Park, M., Thom, J., Mennicken, S., Cramer, H. & Macy, M. Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nat. Hum. Behav.* **3**, 230–236 (2019).
19. Hillard, P. J. A. & Vlajic Wheeler, M. Data from a menstrual cycle tracking app informs our knowledge of the menstrual cycle in adolescents and young adults. *J. Pediatr. Adolesc. Gynecol.* **30**, 269–270 (2017).
20. Pierson, E., Althoff, T. & Leskovec, J. in *WWW: Proceedings of the 2018 World Wide Web Conference*, 107–116 (2018).
21. Alvergne, A., Vlajic Wheeler, M. & Högqvist Tabor, V. Do sexually transmitted infections exacerbate negative premenstrual symptoms? Insights from digital health. *Evol. Med. Public Health* **2018**, 138–150 (2018).
22. Liu, B. et al. in *WWW: Proceedings of the 2019 World Wide Web Conference* (2019).
23. Johansson, E., Larsson-Cohn, U. & Gemzell, C. Monophasic basal body temperature in ovulatory menstrual cycles. *Am. J. Obstet. Gynecol.* **113**, 933–937 (1972).
24. Tudor-Locke, C. et al. Descriptive epidemiology of pedometer-determined physical activity. *Med. Sci. Sports Exerc.* **36**, 1567–1573 (2004).
25. Shephard, R. J. & Aoyagi, Y. Seasonal variations in physical activity and implications for human health. *Eur. J. Appl. Physiol.* **107**, 251–271 (2009).
26. Caruso, S. et al. Do hormones influence women's sex? Sexual activity over the menstrual cycle. *J. Sex. Med.* **11**, 211–221 (2014).
27. Quer, G., Gouda, P., Galarnyk, M., Topol, E. J. & Steinhubl, S. R. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: retrospective, longitudinal cohort study of 92,457 adults. *PLoS ONE* **15**, e0227709 (2020).
28. Ma, Y. et al. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. *Eur. J. Clin. Nutr.* **60**, 519–528 (2006).
29. Dunson, D. B., Colombo, B. & Baird, D. D. Changes with age in the level and duration of fertility in the menstrual cycle. *Hum. Reprod.* **17**, 1399–1403 (2002).
30. Batista, M. C. et al. Effects of aging on menstrual cycle hormones and endometrial maturation. *Fertil. Steril.* **64**, 492–499 (1995).
31. Soules, M. R. et al. Executive summary: stages of reproductive aging workshop (STRAW). *Climacteric* **4**, 267–272 (2001).
32. Metcalf, M. G. & Mackenzie, J. A. Incidence of ovulation in young women. *J. Biosoc. Sci.* **12**, 345–352 (1980).
33. Committee on Adolescence, American College of Obstetricians and Gynecologists, Committee on Adolescent Health Care, AmericanAcademyofPediatrics Menstruation in girls and adolescents: using the menstrual cycle as a vital sign. *Pediatrics* **118**, 2245 (2006).
34. Epstein, D. A. et al. in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6876–6888 (ACM, 2017).
35. Bramwell, R. & Zeb, R. Attitudes towards and experience of the menstrual cycle across different cultural and religious groups. *J. Reprod. Infant Psychol.* **24**, 314–322 (2006).
36. Hoerster, K. D., Chrisler, J. C. & Rose, J. G. Attitudes toward and experience with menstruation in the US and India. *Women Health* **38**, 77–95 (2003).
37. Uskul, A. K. Women's menarche stories from a multicultural sample. *Soc. Sci. Med.* **59**, 667–679 (2004).
38. Chandra, P. S. & Chaturvedi, S. K. Cultural variations in attitudes toward menstruation. *Can. J. Psychiatry* **37**, 196–198 (1992).

39. Li, K. et al. Characterizing physiological and symptomatic variation in menstrual cycles using self-tracked mobile-health data. *NPJ Digit. Med.* **3**, 1–13 (2020).
40. Shansky, R. M. Are hormones a 'female problem' for animal research? *Science* **364**, 825–826 (2019).
41. Johansson, C. et al. Circadian clock-related polymorphisms in seasonal affective disorder and their relevance to diurnal preference. *Neuropsychopharmacology* **28**, 734 (2003).
42. Saad, G. & Stenstrom, E. Calories, beauty, and ovulation: the effects of the menstrual cycle on food and appearance-related consumption. *J. Consum. Psychol.* **22**, 102–113 (2012).
43. Rosato, D. What your period tracker app knows about you. *Consumer Reports* (28 January 2020).
44. Sommer, M. Menarche: a missing indicator in population health from low-income countries. *Public Health Rep.* **128**, 399–401 (2013).
45. Lerner, S. Over 5 years later, Fitbit finally adds menstrual cycle and text response features. *Tech Times* https://www.techtimes.com/articles/227079/20180507/over-5-years-later-fitbit-finally-adds-menstrual-cycle-and-text-response-features.htm (7 May 2018).
46. 23 new iOS tracking categories and a more accurate algorithm. *Clue by BioWink GmbH* http://helloclue.com/articles/about-clue/23-new-ios-tracking-categories-more-accurate-algorithm (2015).
47. Plutchik, R. The nature of emotions. *Am. Sci.* **89**, 344–350 (2001).
48. Plutchik, R. Outlines of a new theory of emotion. *Trans. N. Y. Acad. Sci.* **20**, 394–403 (1958).
49. How many hours of sleep are enough for good health? *The Mayo Clinic* https://www.mayoclinic.org/healthy-lifestyle/adult-health/expert-answers/how-many-hours-of-sleep-are-enough/faq-20057898 (2018).
50. Moran, V. H., Leathard, H. L. & Coley, J. Cardiovascular functioning during the menstrual cycle. *Clin. Physiol.* **20**, 496–504 (2000).
51. Yonkers, K. A., O'Brien, P. S. & Eriksson, E. Premenstrual syndrome. *Lancet* **371**, 1200–1210 (2008).
52. Chiazze, L., Brayer, F. T., Macisco, J. J., Parker, M. P. & Duffy, B. J. The length and variability of the human menstrual cycle. *JAMA* **203**, 377–380 (1968).
53. Jebb, A. T., Tay, L., Diener, E. & Oishi, S. Happiness, income satiation and turning points around the world. *Nat. Hum. Behav.* **2**, 33 (2018).
54. World Health Organization Global Health Observatory (GHO). Prevalence of overweight among adults, BMI ≥ 25 (age-standardized estimate), age group: 18+ years, Sex: Female (2016).
55. World Health Organization Global Health Observatory (GHO). Prevalence of obesity among adults, BMI ≥ 30 (age-standardized estimate), age group: 18+ years, Sex: Female (2016).
56. Walpole, S. C. et al. The weight of nations: an estimation of adult human biomass. *BMC Public Health* **12**, 439 (2012).
57. Durain, D. Primary dysmenorrhea: assessment and management update. *J. Midwifery Women's Health* **49**, 520–528 (2004).
58. Mannix, L. K. Menstrual-related pain conditions: dysmenorrhea and migraine. *J. Women's Health* **17**, 879–891 (2008).
59. Salzman, B., Fleegle, S. & Tully, A. S. Common breast problems. *Am. Fam. Phys.* **86**, 343–349 (2012).
60. Won, H. R. & Abbott, J. Optimal management of chronic cyclical pelvic pain: an evidence-based and pragmatic approach. *Int. J. Women's Health* **2**, 263 (2010).
61. Haines, P. S., Hama, M. Y., Guilkey, D. K. & Popkin, B. M. Weekend eating in the United States is linked with greater energy, fat, and alcohol intake. *Obes. Res.* **11**, 945–949 (2003).
62. Finlay, A. K., Ram, N., Maggs, J. L. & Caldwell, L. L. Leisure activities, the social weekend, and alcohol use: evidence from a daily study of first-year college students. *J. Stud. Alcohol Drugs* **73**, 250–259 (2012).
63. Seasonal allergies: nip them in the bud. *The Mayo Clinic* https://www.mayoclinic.org/diseases-conditions/hay-fever/in-depth/seasonal-allergies/art-20048343 (2019).
64. Lebowitz, M. D., Collins, L. & Holberg, C. J. Time series analyses of respiratory responses to indoor and outdoor environmental phenomena. *Environ. Res.* **43**, 332–341 (1987).
65. Andersen, T. B. A model to predict the beginning of the pollen season. *Grana* **30**, 269–275 (1991).
66. Finkelman, B. S. et al. Global patterns in seasonal activity of influenza A/H3N2, A/H1N1, and B from 1997 to 2005: viral coexistence and latitudinal gradients. *PloS ONE* **2**, e1296 (2007).
67. The flu season. *Centers for Disease Control and Protection* https://www.cdc.gov/flu/about/season/flu-season.htm (2018)
68. Baum T. & Lundtorp S. (eds) *Seasonality in Tourism* (Elsevier, 2001).
69. Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C. & Leskovec, J. in *CSCW: Proceedings of the Conference on Computer-Supported Cooperative Work*, vol. 2017, 1217 (NIH Public Access, 2017).
70. Symul, L., Wac, K., Hillard, P. & Salathe, M. Assessment of menstrual health status and evolution through mobile apps for fertility awareness. *NPJ Digit. Med.* **2**, 64 (2019).
71. Pliner, P. & Fleming, A. S. Food intake, body weight, and sweetness preferences over the menstrual cycle in humans. *Physiol. Behav.* **30**, 663–666 (1983).

Acknowledgements

The authors thank D. Jurafsky, P. Wei Koh, S. Li, A. Mastrioanni, M. Mateen, L. Pierson, R. Pierson, N. Roth, C. Ruiz, R. Sosic, P. Thaker, C. Yau, M. Zitnik and J. Zou for helpful comments. For financial support, they thank the Hertz and NDSEG Fellowships; the SAP Stanford Graduate Fellowship; NIH grant U54 EB020405; the Bill & Melinda Gates Foundation (INV-004841); the Stanford Data Science Initiative; NSF Grants IIS-1901386; OAC-1835598 (CINES), OAC-1934578 (HDR), CCF-1918940 (Expeditions) and IIS-2030477 (RAPID); Wu Tsai Neurosciences Institute; and the Chan Zuckerberg Biohub. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. J.L. is a Chan Zuckerberg Biohub investigator.

Author contributions

E.P. performed the statistical analysis. E.P., T.A., D.T., P.H. and J.L. jointly analysed the results and wrote the paper.

Competing interests

E.P. is employed by Microsoft Research. D.T. is employed by Clue by BioWink GmbH. P.H. is on the medical advisory board of Clue by BioWink GmbH. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-020-01046-9.

Correspondence and requests for materials should be addressed to J.L.

Peer review information Primary Handling Editor: Jamie Horder

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Jure Leskovec

Last updated by author(s): Sep 11, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	N/A; no custom code was used to collect the data in the study.
Data analysis	All analysis was performed in Python using standard libraries. Code will be made publicly available on GitHub upon publication of this manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data that support the findings of this study are available from the corresponding author (J.L.) with appropriate permission from Clue. The data are not publicly available to preserve the privacy of Clue users.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Longitudinal, quantitative cohort study.
Research sample	3.3 million people who used a female health-tracking app to monitor feature of their vital signs, mood, behavior, and menstrual cycle.
Sampling strategy	We did not subsample the dataset and used data from all participants who provided the necessary data, after filtering to ensure data quality. Further details about this filtering procedure are provided in the Methods section.
Data collection	We did not collect any data, but relied on previously collected and de-identified data; the app used to collect data is described in the Methods section.
Timing	The dataset spans November 1, 2015 to November 20, 2017.
Data exclusions	Data exclusions are described in the Methods, and are performed to ensure data quality.
Non-participation	N/A (analysis was limited to people who chose to used the app, so there was no way to "decline participation")
Randomization	This is not a randomized controlled trial and participants were not randomized into experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging