Hormones, ovulatory cycle phase and pathogen disgust: A longitudinal investigation of the Compensatory Prophylaxis Hypothesis

Julia Stern[1,2] & Victor Shiramizu[3]

1: Department of Psychology, University of Bremen, Grazer Strasse 2c, 28359 Bremen, Germany

2: Department of Psychology, University of Goettingen, Gosslerstrasse 14, 37073 Goettingen, Germany

3: School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK

Corresponding author: Julia Stern (jstern@uni-bremen.de)

This is an unedited manuscript accepted for publication in Hormones and Behavior. The manuscript will undergo copyediting, typesetting, and review of resulting proof before it is published in its final form. Please cite as:

Stern, J. & Shiramizu, V. (2022). Hormones, ovulatory cycle phase and pathogen disgust: A longitudinal investigation of the Compensatory Prophylaxis Hypothesis. *Hormones and Behavior*, *138*, 105103. https://doi.org/10.1016/j.yhbeh.2021.105103

Declaration of Conflicting Interests

The authors declare that they have no conflicts of interests.

Abstract

Multiple studies have argued that disgust, especially pathogen disgust and contamination sensitivity, change across women's ovulatory cycle, with higher levels in the luteal phase due to an increase in progesterone levels. According to the Compensatory Prophylaxis Hypothesis (CPH), women have a higher disgust sensitivity to pathogen cues when in the luteal phase (or when progesterone levels are higher), because progesterone is associated with suppressed immune responses. Evidence for this hypothesis is rather mixed and uncertain, as the largest study conducted so far reported no compelling evidence for an association between progesterone levels and pathogen disgust. Further, ovulatory cycle research has been criticized for methodological shortcomings, such as invalid cycle phase estimates, no direct hormone assessments, small sample sizes or between-subjects studies. To address these issues and to contribute to the literature, we employed a large, within-subjects design ($N = 257$ with four sessions each), assessments of salivary hormone levels and cycle phase estimates based on luteinizing hormone tests. A variety of multilevel models suggest no compelling evidence that self-reported pathogen disgust or contamination sensitivity is upregulated in the luteal phase or tracks changes in women's hormone levels. We further found no compelling evidence for between-subjects associations of pathogen disgust or contamination sensitivity and hormone levels. Results remain robust across different analytical decisions (e.g. in a subsample of women reporting feeling sick). We discuss explanations for our results, limitations of the current study and provide directions for future research.

Keywords: pathogen disgust, contamination sensitivity, progesterone, hormones, ovulatory cycle, compensatory prophylaxis hypothesis

Introduction

The Compensatory Prophylaxis Hypothesis (CPH) states that female disgust sensitivity is adjusted as a function of the current level of immunocompetence to avoid contact to pathogens and thus, potential infections (e.g. Fessler et al., 2005). More precisely, as immunosuppression is upregulated in the luteal phase of the ovulatory cycle, elevated by rising levels of progesterone, it is assumed that female disgust sensitivity should increase accordingly (Fessler & Navarette, 2003). A number of previous studies investigated whether human or non-human females show, indeed, higher levels of disgust sensitivity when in the luteal phase, or during the first trimester of pregnancy, or when progesterone levels were higher. However, these studies yield fairly mixed findings and were often criticized for their methods, such as small sample sizes, cross-sectional designs or not assessing hormone levels. In the current study, we aim to contribute to the previous literature by testing the CPH in a large within-subjects study including direct assessments of hormone levels.

Women's ovulatory cycle can roughly be separated in two different phases: the follicular phase, from menstrual onset to the day of ovulation (thus, ending with a fertile period), and the luteal phase, from the day following ovulation to the next menstrual onset. These distinct phases are characterized by massive fluctuations in hormone levels, mainly estradiol and progesterone. Whereas rising levels of estradiol and lower levels of progesterone characterize the follicular phase, levels of estradiol are lower and levels of progesterone much higher during the luteal phase, except for a second smaller estradiol peak mid-luteal (Roney & Simmons, 2013). Importantly, in the luteal phase, the female body prepares for pregnancy, including a downregulation of inflammatory immune responses, which makes females more prone to potential infections (Bouman et al., 2001; Faas et al., 2000; Trzonkowski et al., 2001, cited from Fleischman & Fessler, 2011). Immunosuppression in the luteal phase, most likely elevated by increasing levels of progesterone, presumably happens to prevent the maternal immune system from attacking the half-foreign blastocyst (Fleischman & Fessler, 2011). In

response to a higher vulnerability to diseases due to the down regulated immune system, it is assumed that females engage in disease prophylaxis by showing behavior that helps them to avoid potential sources of infections (e.g. Fessler & Navarette, 2003).

Disgust plays a key role in the avoidance of pathogens (Tybur et al., 2009). In accordance, Fessler and Navarette (2003) hypothesized that women's disgust sensitivity should be adjusted as a function of immunocompetence, especially in phases of reproductive immunomodulation, also known as the Compensatory Prophylaxis Hypothesis (CPH). Various previous studies have already tested the CPH and reported supportive evidence (see Table 1 for a detailed overview of methods used in these studies). For example, the original authors of the Hypothesis reported that sexual disgust is related to women's conception risk (Fessler & Navarette, 2003), although not finding support for the CPH in terms of an association between conception risk and disgust sensitivity. Further, levels of disgust were reported to be highest in the first trimester of pregnancy, when immunosuppression and vulnerability to diseases are highest (Fessler et al., 2005). Probably the most compelling evidence for the CPH so far was reported in a study by Fleischman and Fessler (2011) that involved direct assessments of progesterone. In a sample of $N = 79$ participants, the authors found significant correlations between progesterone levels and different measures of pathogen disgust. However, subsequent studies have reported fairly mixed findings. Being the only study so far that assessed progesterone levels and participants cycle phase at the same time, Zelazniewicz and colleagues (2016) reported no changes of pathogen or moral disgust across the cycle. Nevertheless, in between-subjects analyses, they found that progesterone levels in the luteal phase correlated with different measures of disgust. In a remarkable longitudinal study, Jones and colleagues (2018) directly assessed steroid hormone levels, as well as self-reports of pathogen and moral disgust. They reported no compelling evidence that any of the assessed hormones, their ratios or their interactions are related to pathogen or moral disgust.

As a response, Fleischman and Fessler (2018) concluded that the CPH might be wrong or that there are either methodological reasons for differences in results or that progesterone is not the driving factor for changes in disgust.

Since then, the inconsistent findings regarding the CPH even increased interest in investigating the hypothesis, and studies that were following raised uncertainty in previous results. One study reported findings in the opposite direction of the CPH, in that women in the follicular phase reported being more, rather than less disgusted in response to a video, as compared to women in the luteal phase (Olantunji et al., 2020). Milkowska and colleagues (2019, 2021) conducted two studies on changes in women's levels of disgust across different phases of their ovulatory cycle. Results of their first study suggest that only among women who currently have an infection, pathogen disgust, but not moral disgust, is higher in the luteal phase (Milkowska et al., 2019). In their second study, they report that among all women, pathogen disgust, but not moral disgust, was higher when women were in the luteal phase than when they were in the follicular phase (Milkowska et al., 2021). As all of the above reported studies were observational, Kavaliers and colleagues (2021a) decided to conduct an experimental study investigating the CPH. As ethical constraints make it hard to directly manipulate hormone levels in humans, the authors investigated mice and did not find that progesterone treatments had an impact on the avoidance of pathogens. However, in a reanalysis of the data from this study using more sensitive analytical methods, Bressan and Kramer (2021) argued that progesterone does actually raise disgust. In their answer to the reanalysis, the original authors (Kavaliers et al., 2021b) highlight the need for more research on progesterone and pathogen disgust to answer open research questions. In summary, some studies reported results in line with the CPH that suggest that women with higher levels or progesterone, or women in the luteal phase report higher levels of (pathogen) disgust. One study suggested that the CPH might only be true for women currently feeling sick, whereas

one study reported results in the opposite direction than expected by the CPH. Some studies reported mixed results, depending on the analyses. The largest, methodologically most rigorous study so far rather reports null results for the CPH.

Table 1

Overview of methodological characteristics of previous studies investigating the Compensatory Prophylaxis Hypothesis.

Study	Sample size	Design	Cycle phase estimates?	Hormone assays?
Fessler and Navarette, 2003	*N* = 307	Between-subjects	Conception risk estimates based on forward counting	X
Fleischman and Fessler, 2011	*N* = 79	Between-subjects	X	Progesterone in saliva (immunoassays)
Zelazniewicz and colleagues, 2016	*N* = 30	Within-subjects (two sessions per subject, during menstruation and mid-luteal)	Mid-luteal phase session confirmed by LH tests	Progesterone in blood (immunoassays)
Jones and colleagues, 2018	*N* = 375	Within-subjects (up to 15 sessions per subject)	X	Progesterone, estradiol, testosterone, cortisol in saliva (immunoassays)
Milkowska and colleagues, 2019	*N* = 321	Between-subjects	Cycle phase estimates based on forward counting	X
Olantunji and colleagues, 2020	*N* = 73	Between-subjects	Cycle phase estimates based on forward counting	X
Milkowska and colleagues, 2021	*N* = 93	Within-subjects	Cycle phase estimates based on LH tests	X

Note: LH tests = urine tests measuring the luteinizing hormone.

How can these immense differences in results from previous studies be explained? As Fleischman and Fessler (2018) noted, differences in methods can easily lead to differences in

results. Some of the mentioned studies investigated changes in disgust across cycle phases, others directly assessed progesterone levels. Some employed within-subjects, whereas others used between-subjects designs. Statistical analyses differed. Moreover, studies used different self-report questionnaires to measure levels of disgust, others rather used disgusting pictures or videos as stimuli material.

Overall, studies doing ovulatory cycle research received methodological criticism within the last years. For example, it has been criticized that many studies relied on counting methods to identify participant's current cycle phase, which are probably heavily affected by measurement error (e.g. Blake et al., 2016). Other criticism involved low test power due to small sample sizes and employing between-subjects designs to study a within-subjects effect, which potentially confounds results (e.g. Arslan et al., 2021). Presenting a power analysis, Gangestad and colleagues (2016) suggest that, to be able to find an effect of moderate magnitude (Cohen's $d = 0.5$ with 80% power), between-subjects studies relying on (forward-) counting methods for cycle phase estimate would need a total of 1,213 participants. Within-subjects studies may need significantly smaller sample sizes to detect a medium sized effect (e.g. $N = 190$ for forward-counting estimates, $N = 48$ for more reliable cycle phase estimates that are validated with luteinizing hormone tests). Therefore, all between-subjects studies investigating the CPH so far seem to be heavily underpowered, which can lead to false negative, but also to false positive findings (Button et al., 2013). Inconsistent findings and methodological criticism of previous studies highlight the need for more data investigating the CPH, preferably by employing a large sample size, a longitudinal design, direct hormone measures as well as luteinizing hormone tests that validate cycle phase estimates. This study contributes to the current debate on potential changes of pathogen disgust during women's ovulatory cycle, by employing the mentioned methods to test the CPH. While using questionnaires that were previously used in studies reporting positive or studies reporting null

results, and by assessing progesterone as well as women's cycle phase (validated by luteinizing hormone tests), this study also tests whether differences in these and different analytical methods affect the results.

Methods

Open data, analysis script, and material are provided (https://osf.io/nszpa/). All participants signed a written consent form and the local ethics committee approved the study protocol (no. 225).

Participants and recruitment

A total of 257 heterosexual female participants (aged 18-35 years, $M = 23.2$, $SD = 3.3$), out of 282 recruited, finished all sessions, and were therefore included in further analyses. The 25 dropouts resulted from 16 women who attended only the introductory session and nine women who only completed one or two testing sessions (for the following reasons: not responding to emails anymore (9), decided not to take part without providing further reasons (3), scheduling problems (2), switch to hormonal contraception (2), taking the morning after pill (2), health issues (2), moved to a different city (1), irregular mid-cycle bleeding (1), very long irregular cycle >50 days (1), claimed to not fit into the inclusion criteria anymore (1), or pregnancy (1)). Our participants had to fulfill the following preregistered criteria to take part in the study: female, between 18 and 35 years old, naturally cycling (no hormonal contraception for at least three months, no expected switch to hormonal contraception while in the study, no current pregnancy or breastfeeding, no childbirth or breast-feeding during the previous year, not taking hormone-based medication or anti-depressants, no endocrine disorders). Additionally, included participants reported their ovulatory cycles being of regular length between 25 and 35 days, at least during the last 3 months. Our sample size largely exceeds the size required to achieve 80% power given a within-subjects design and

anticipated effects of moderate magnitude, as suggested by recent guidelines for sample sizes in ovulatory shift research (Gangestad et al., 2016).

Procedure

All participants took part in five individually scheduled sessions. In the first introductory session, participants received detailed information about the general procedure, duration of the study, and compensation. A research assistant explained the luteinizing hormone (LH) ovulation tests and checked the inclusion criteria. Average cycle length as well as the dates of the last, the penultimate and the next menstrual onset were assessed to plan the dates of the next sessions.

Sessions two to five were computer-based testing sessions and took place across different phases of the ovulatory cycle, scheduled based on backward counting and the observed LH test surge. Suitable testing days were computed with the help of an Excel sheet created for that purpose (see open material). All participants completed two sessions in their mid-to-late follicular phase (session A during expected reverse cycle days 19-21, session B during expected reverse cycle days 16-18) and two sessions in their expected luteal phase (session C in the mid luteal phase, expected reverse cycle days 6-11 or 5-10 days after an observed LH surge, session D premenstrual during expected reverse cycle days 1-5 or 11-15 days after an observed LH surge). Scheduling was validated via LH test results and via following up to the day of the next menstrual onset. The starting session for each participant depended on their current cycle phase at the introductory session and their personal schedule. Of all participants who finished all sessions, 134 participants started with the first session in their follicular phase, and 123 started in the luteal phase. Participants who started with their first session being session A usually completed all four sessions within one cycle. Participants who started with their first session being any of the sessions B, C, or D usually completed all four sessions in two consecutive cycles. However, in some cases, we rescheduled one or both

of the fertile testing sessions spontaneously to the next cycle when the LH surge occurred earlier than expected.

To control for possible effects of diurnal changes in hormone levels, all sessions were scheduled in the second half of the day (between 12pm and 6pm). When arriving at the lab, participants first completed a screening questionnaire, assessing their eligibility and some control variables for saliva sampling (Schultheiss & Stanton, 2009). Then, they completed the questionnaires including the items to assess whether they felt sick on that day, pathogen disgust and contamination sensitivity. Next, saliva samples were collected via passive drool.

All data in the lab was conducted using the open source framework Alfred (Treffenstaedt & Wiemann, 2018), which is based on the programming language Python (version 2.7). Besides the tasks described in the current study, participants also had to complete other tasks as part of a larger study (for details on all assessed data, see https://osf.io/th6rf). All different tasks were fully randomized between participants and sessions. Upon completion of all sessions, participants received a payment of 60€ or course credit.

Measures

Pathogen disgust

Pathogen disgust was measured via the pathogen disgust subscale from the Three Domain Disgust Scale (TSSD; Tybur et al., 2009) in each session, following Jones and colleagues (2018), as well as Milkowska and colleagues (2019, 2021). This subscale consists of seven items (e.g. *stepping on dog poop*), which were rated which were rated from 1 = *not at all disgusting* to 7 = *extremely disgusting*. Instructions were directly applied from the TSSD[1]. We added a sentence to the instructions asking people to refer to their feelings today or in the last

[1] Full instructions were: „The following items describe a variety of concepts. Please rate how disgusting you find the concepts described in the items, where 1 means that you do not find the concept disgusting at all and 7 means that you find the concept extremely disgusting. You should particularly focus on your feelings today or in the last 24 hours“.

24 hours to investigate ovulatory cycle phase-dependent changes, and to make sure that they do not report their general feelings or think about their answers from a previous session (adapted from Fleischman & Fessler, 2011). Cronbachs alpha for this scale was α = .70, the mean score was 4.45 (*SD* = 1.07). All items were averaged to form a fullscale score.

Contamination sensitivity

Following Fleischman and Fessler (2011), contamination sensitivity was measured via the eleven contamination-related items (items 1-10 and 39; e.g. "*In the last 24 hours I've felt my hands were dirty when I touched money*") from the Revised Padua Obsessive-Compulsive Inventory (Burns et al., 1996). Participants had to rate the eleven claims on a five-point Likert scale from 1 = *not at all* to 5 = *very much*. Cronbachs alpha for this scale was α = .87, the mean score was 1.44 (*SD* = 0.60). All items were averaged to form a fullscale score.

Feeling sick

Following Milkowska and colleagues (2019), we asked the participants whether they had any form of current infection ("*Are you currently feeling sick (for example strong cold, distinct headache*)?") with a binary "*yes/ no*" item. Participants reported feeling sick in only 55 of all 1028 sessions.

Hormone measures

For hormone assays, we collected four saliva samples from each participant (one per testing session). Contamination of saliva samples was minimized by asking participants to abstain from eating, drinking (except plain water), smoking, chewing gum, or brushing teeth for at least one hour before each session. The samples were stored at -80°C directly after collection until shipment on dry ice to the Kirschbaum Lab at Technical University of Dresden, Germany (one freeze-thaw cycle), where progesterone, testosterone and cortisol were assessed via liquid chromatography mass spectrometry (LCMS, Gao et al., 2015)[2]. Since the

[2] Liquid to liquid extraction was carried out by adding 20 µL internal standard and 1 mL ethyl acetate to 400 µL saliva in a 2 mL polypropylene tube. The resulting mixture was subsequently rotated for 1 min on the vortex and

lab had no valid protocol for LCMS analysis of estradiol levels, the samples were reanalyzed for estradiol using the highly sensitive 17β-estradiol enzyme immunoassay kit (IBL International, Hamburg, Germany). Samples were analyzed in singlets, however, the lab reported that their procedure yields CVs < 11% and there was a significant large association between conception risk and E/P (for details see Stern et al., 2021), validating hormone assays and our cycle phase measure. We centered all hormone values on their subject-specific means and scaled them afterwards (i.e. divided them by a constant), so that the majority of the distribution for each hormone varied from -0.5 to 0.5 to facilitate calculations in the linear mixed models (e.g. as in Jones et al., 2018; Stern et al., 2021, amongst others). This is a common procedure to isolate effects of within-subject changes in hormones and dealing with the non-normal distribution of hormone levels. Hormone levels were nearly normally distributed afterwards (see also Stern et al., 2021), the R code for this procedure can be found in the open script.

Cycle phase

Participants' cycle phase for every session was assigned based on highly sensitive (10mIU) LH urine ovulation test strips from MedNet GmbH. Participants started LH-testing after menstruation (around reverse cycle day 21) and continued until a rise of LH (positive tests) was observed and a minimum of two days after the tests were negative again (as suggested by Roney, 2018). The day that we defined as the surge day was mainly based on the first positive test result (although for women with multiple positive tests, we defined the day with the latest positive test result before testing negative again as the day of LH surge). As participants stopped LH testing afterwards, we were not able to detect potentially occurring second LH

then centrifuged for 10 min at 12000 r/min with centrifuge (Hettich, MIKRO 22 R). The ethyl acetate layer was transported to a new glass tube and evaporated to dryness under nitrogen. The residue was resuspended in 120 µL methanol/water in a ratio of 50:50 (v/v), 50 µL of which was injected into the LC-MS/MS system. The LC-MS/MS system consisted of Shimadzu HPLC system, and AB Sciex Triple Quad 6500+ System equipped with the electrospray ionization (ESI) source. See Gao et al. (2015) for more details.

peaks in women with irregularly long cycles (whose first LH peak did likely not lead to ovulation). Participants who had their first session in the luteal phase additionally did five LH tests from the day of the introductory session to their testing day (and beyond) to make sure that we did not accidentally scheduled them in their fertile phase. Participants were provided with a minimum of ten LH tests each and provided daily pictures of the tests to the investigators for confirmation. LH results were used to allow flexible scheduling in case LH test results differed from the scheduling based on counting (for details see Stern et al., 2021).

To determine participants' cycle phase, we first checked how many cycles were reported as being irregular (i.e. > 40 days, < 20 days, or the length deviated more than five days from participant's average cycle length). Even though all participants reported regular cycles in the introductory session, 28 of the 257 women had an irregular cycle (11%). Furthermore, n = 16 participants observed negative LH tests despite having regular cycles, possibly due to non-ovulatory cycles (6%). Nine participants did not do (enough) LH tests to detect a surge or reported invalid results only (4%), and four participants were mis-sampled for other reasons (2%). These participants (n = 57; 22%) were excluded for all cycle phase analyses (as cycle phase cannot be reliably assigned). These numbers are comparable or even lower than in previous cycle studies. We categorized every session between menstrual onset and one day after the LH surge (which we assume to be the day of ovulation) as follicular phase session. Every session between two days after the LH surge and the next menstrual onset were categorized as luteal phase sessions. Of these remaining n = 200 participants for cycle phase analyses, 98 started with the first session in their luteal phase, and 102 started in the follicular phase (to avoid order effects). However, all 257 women were included in the hormone analyses.

Statistical analyses

All analyses were done with the statistic software R 4.0.4 (R Core Team, 2021). The following packages were used: tidyverse 1.3.0 (Wickham, 2019), readxl 1.3.1 (Wickham & Bryan, 2019), psych 2.0.12 (Revelle, 2016), lmerTest 3.1-3 (Kuznetsova et al., 2013), knitr 1.31 (Xie, 2014), kableExtra 1.3.4 (Zhu, 2019), cowplot 1.1.1 (Wilke et al., 2019). All statistical tests were two-tailed. To test the CPH and the robustness of our results, we computed several models for each outcome (pathogen disgust and contamination sensitivity scores served as outcomes). Predictors were either hormone levels or cycle phase. As previous studies used different model specifications for hormone levels (and we wanted to test whether these may lead to differences in results), we specified four different models for each outcome to investigate associations with progesterone: one model including progesterone, estradiol and their interaction, one model including progesterone, estradiol and their ratio, one model with progesterone as a single predictor variable, one model only including women who reported feeling sick with progesterone as the single predictor variable. We further investigated associations of testosterone, cortisol and their interaction with the outcomes in separate models. In all models, we include within-subjects hormone levels, as well as between-subjects (the latter in an exploratory manner as suggested in the review process). Further, all models include a random intercept for each participant to account for the nested data structure, random slopes were specified maximally following Barr and colleagues (2013).

Results

Are steroid hormones associated with pathogen disgust?

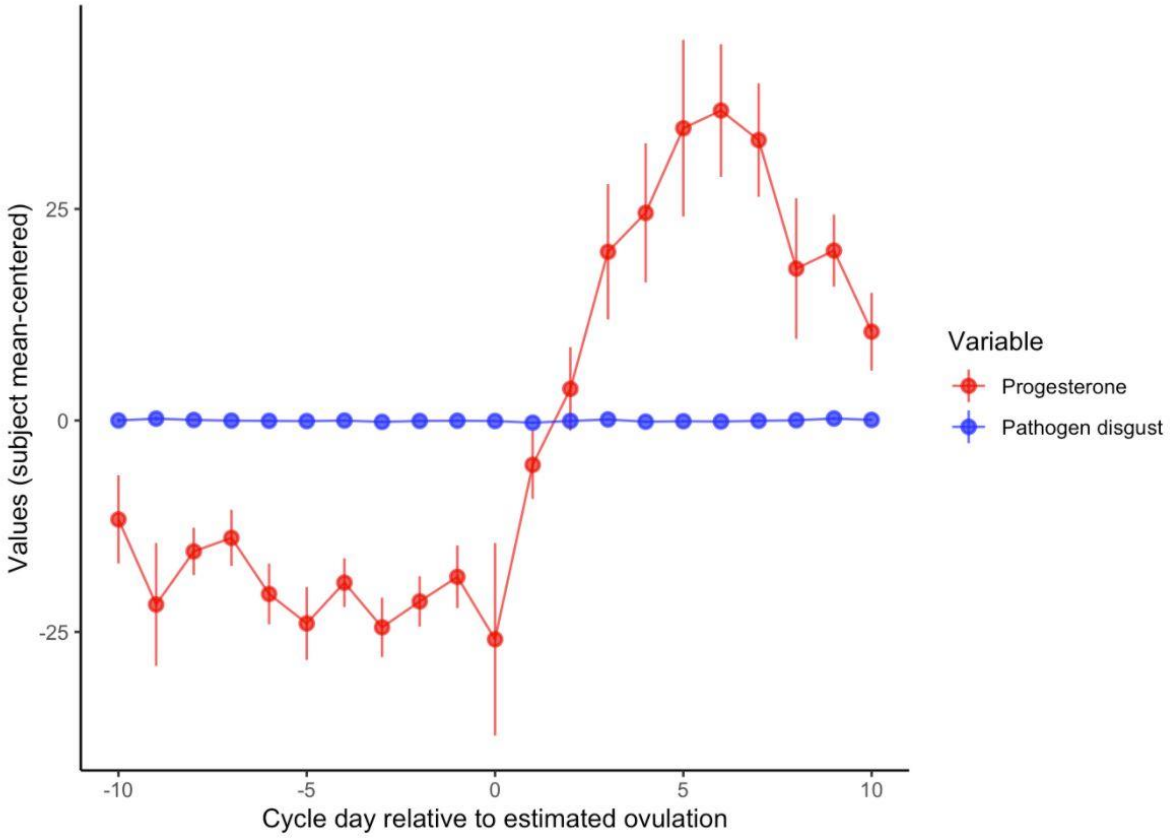
Our Model 1 showed that neither within-subjects nor between-subjects progesterone, estradiol or their interaction were significantly associated with pathogen disgust (Table 2). Model 2 also showed no significant relationship of within-subjects progesterone (estimate = 0.11, 95%CI = [-0.06; 0.29]; $t = 1.27$, $p = .205$), within-subjects estradiol (estimate = -0.10, 95%CI

= [-0.39; 0.19]; t = -0.69, p = .495), or the within-subjects estradiol-to-progesterone ratio (estimate = 0.03, 95%CI = [-0.07; 0.13], t = 0.51, p = .611) with pathogen disgust. All between-subjects results were non-significant as well (progesterone: estimate = -0.02, 95%CI = [-0.15; 0.11]; t = -0.30, p = .763; estradiol: estimate = -0.06, 95%CI = [-0.19; 0.07]; t = -0.92, p = .362; estradiol-to-progesterone ratio: estimate = -0.05, 95%CI = [-0.18; 0.08]; t = -0.73 p = .490).

Model 3 revealed no significant association of within-subjects or between-subjects progesterone and pathogen disgust as well (within-subjects: estimate = 0.10, 95%CI = [-0.07; 0.27], t = 1.19, p = .236; between-subjects: estimate = -0.01, 95%CI = [-0.13; 0.11], t = -0.20, p = .845), even when only including women who reported feeling sick (within-subjects: estimate = 0.01, 95%CI = [-1.38; 1.40], t = 0.01, p = .990; between-subjects: estimate = 0.37, 95%CI = [-0.31; 1.04], t = 1.07, p = .293). Figure 1 shows no strong variation of pathogen disgust in relation to within-subjects progesterone across the cycle. Finally, Model 4 showed no significant relationship between within-subjects testosterone (estimate = 0.03, 95%CI = [-0.12; 0.19], t = 0.42, p = .677), within-subjects cortisol (estimate = -0.08, 95%CI = [-0.27; 0.12], t = -0.74, p = .460), or the interaction between within-subjects testosterone and cortisol (estimate = -0.42, 95%CI = [-1.27; 0.43], t = -0.97, p = .349) and pathogen disgust. All between-subjects results were non-significant as well (testosterone: estimate = 0.003, 95%CI = [-0.13; 0.14]; t = 0.04, p = .967; cortisol: estimate = -0.01, 95%CI = [-0.15; 0.13]; t = -0.15, p = .881; testosterone x cortisol interaction: estimate = 0.01, 95%CI = [-0.15; 0.17]; t = 0.16, p = .878).

Figure 1

Association of within-subjects progesterone values and pathogen disgust scores across the cycle



Note: Progesterone and pathogen disgust values were subject mean-centered (subtracting each subject's mean). The x-axis shows the estimated day of the cycle centered on the estimated day of ovulation (one day after the LH-surge was detected). Displayed are means (dots) and standard errors (lines) for each variable. Subject mean-centered values for pathogen disgust vary between -1 and 1, a more detailed graph displaying this variable can be found in the supplementary material.

Are steroid hormones associated with contamination sensitivity?

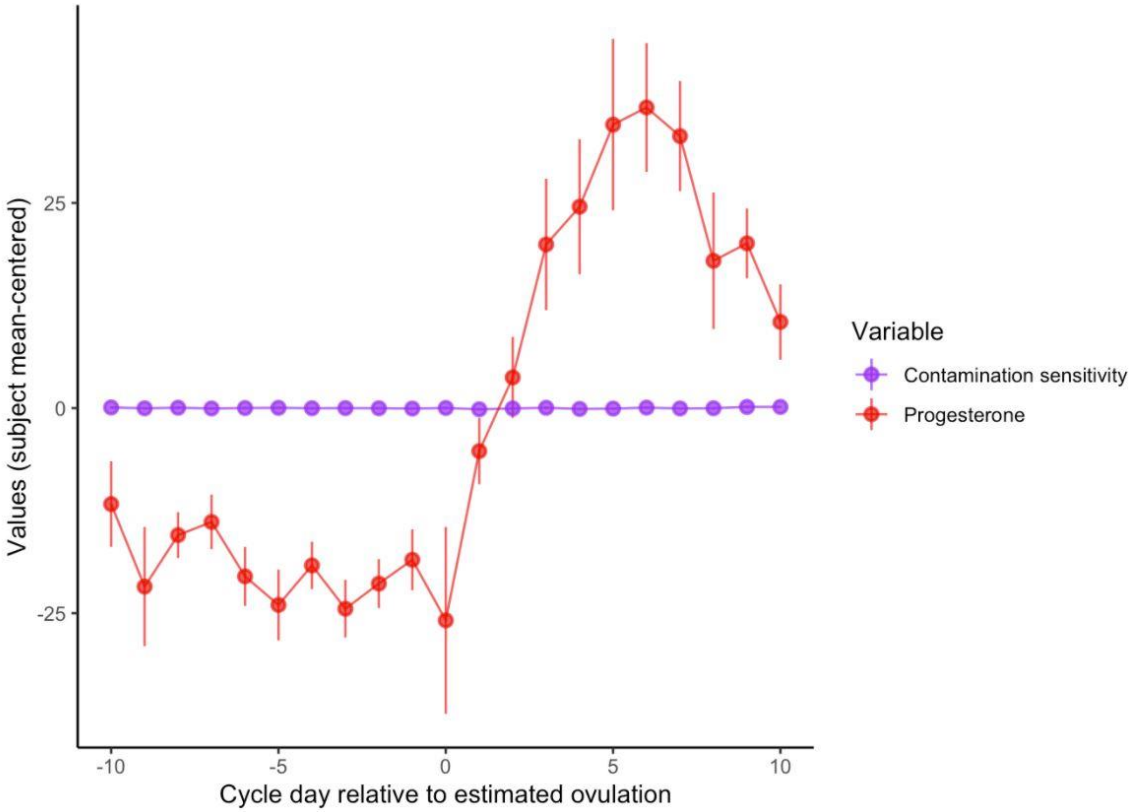
Model 5 showed that neither within-subjects nor between-subjects progesterone, estradiol or their interaction were significantly associated with contamination sensitivity (Table 2). Model 6 also showed that neither within-subjects progesterone (estimate = 0.04, 95%CI = [-0.10; 0.18], t = 0.56, p = .577), within-subjects estradiol (estimate = 0.02, 95%CI = [-0.14; 0.18], t = 0.23 p = .819), or within-subjects estradiol-to-progesterone ratio (estimate = 0.001, 95%CI = [-0.05; 0.06], t = 0.05, p = .960) were significantly associated with contamination sensitivity. All between-subjects results were non-significant as well (progesterone: estimate = -0.02, 95%CI = [-0.09; 0.05]; t = -0.54, p = .590; estradiol: estimate = -0.01, 95%CI = [-0.08; 0.07]; t = -0.13, p = .894; estradiol-to-progesterone ratio: estimate = -0.05, 95%CI = [-0.16; 0.05]; t = -0.97 p = .340).

Model 7 showed no significant association of within-subjects progesterone (estimate = 0.04, 95%CI = [-0.09; 0.16], t = 0.63, p = .537) or between-subjects progesterone (estimate = -0.01, 95%CI = [-0.08; 0.06], t = -0.40, p = .687) and contamination sensitivity and results remained virtually identical when including only women who reported feeling sick (within-subjects: estimate = 0.29, 95%CI = [-0.60; 1.18], t = 0.63, p = .535; between-subjects: estimate = 0.08, 95%CI = [-0.48; 0.63], t = 0.27, p = .790). The association between within-subjects progesterone and contamination sensitivity across the cycle is displayed in Figure 2. Lastly, Model 8 revealed no significant association of contamination sensitivity and within-subjects testosterone (estimate = 0.01, 95%CI = [-0.07; 0.09], t = 0.32, p = .749), within-subjects cortisol (estimate = 0.03, 95%CI = [-0.08; 0.13], t = 0.51, p = .609), or the interaction between within-subjects testosterone and cortisol (estimate = -0.23, 95%CI = [-0.53; 0.08], t = -1.44, p = .151). All between-subjects results were non-significant as well (testosterone: estimate = 0.01, 95%CI = [-0.06; 0.07]; t = 0.14, p = .892; cortisol: estimate = -0.02, 95%CI =

[-0.09; 0.05]; t = -0.51, p = .609; testosterone x cortisol interaction: estimate = 0.004, 95%CI = [-0.07; 0.08]; t = 0.11, p = .913).

Figure 2

Association of within-subjects progesterone values and contamination sensitivity scores across the cycle



Note: Progesterone and contamination sensitivity values were subject mean-centered (subtracting each subject's mean). The x-axis shows the estimated day of the cycle centered on the estimated day of ovulation (one day after the LH-surge was detected). Displayed are means (dots) and standard errors (lines) for each variable. Subject mean-centered values for contamination sensitivity vary between -1 and 1, a more detailed graph displaying this variable can be found in the supplementary material.

Table 2

Associations between pathogen disgust or contamination sensitivity and within-subjects as well as between-subjects progesterone, estradiol and their interaction.

	Estimate	*95%CI*	*t*	*p*
Pathogen disgust				
Progesterone (within-subjects)	0.10	[-0.06; 0.27]	1.23	.221
Estradiol (within-subjects)	-0.08	[-0.34; 0.18]	-0.58	.564
Progesterone (between-subjects)	-0.01	[-0.13; 0.12]	-0.11	.917
Estradiol (between-subjects)	-0.07	[-0.19; 0.06]	-1.02	.313
Progesterone * Estradiol (within-subjects)	0.88	[-0.76; 2.52]	1.05	.292
Progesterone * Estradiol (between-subjects)	-0.04	[-0.13; 0.06]	-0.72	.477
Contamination sensitivity				
Progesterone (within-subjects)	0.04	[-0.06; 0.13]	0.77	.444
Estradiol (within-subjects)	0.05	[-0.10; 0.19]	0.65	.514
Progesterone (between-subjects)	-0.01	[-0.08; 0.06]	-0.31	.757
Estradiol (between-subjects)	-0.02	[-0.09; 0.05]	-0.57	.573
Progesterone * Estradiol (within-subjects)	0.24	[-0.67; 1.14]	0.51	.608
Progesterone * Estradiol (between-subjects)	-0.01	[-0.07; 0.04]	-0.52	.609

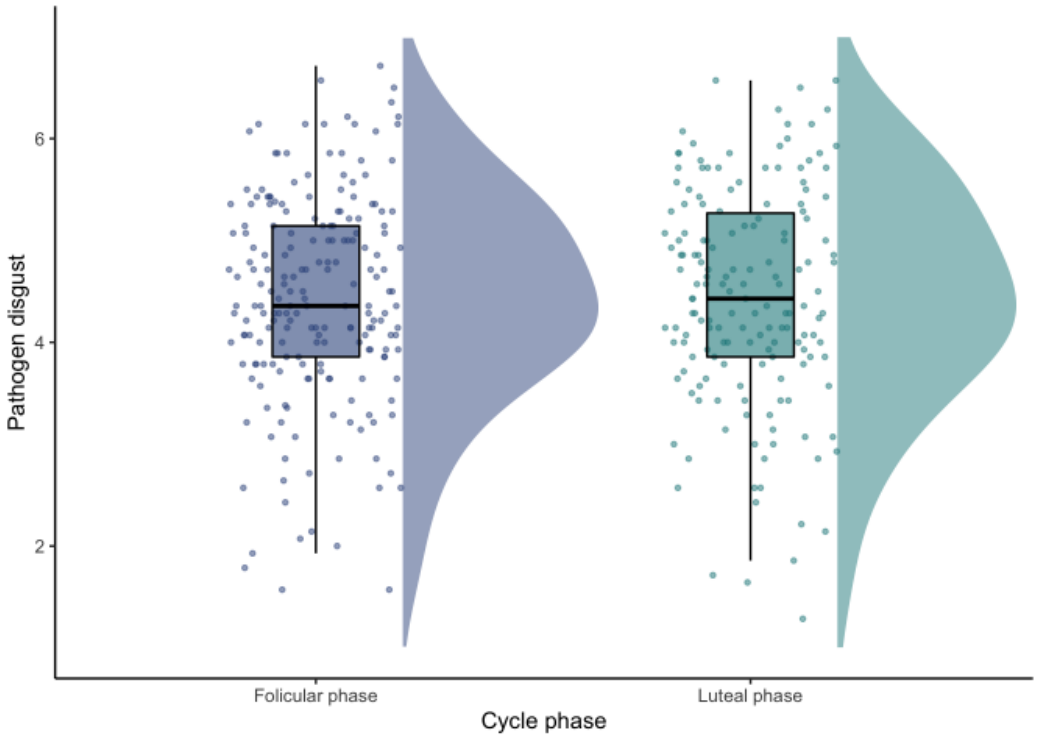
Note: Estimate = unstandardized effect size estimate, 95% CI = 95% confidence interval, t = t-value, p = p-value.

Does pathogen disgust change across the ovulatory cycle?

Our model showed no significant association of ovulatory cycle phase (estimate = -0.01, 95%CI = [-0.09; 0.07], t = -0.21, p = .832) and pathogen disgust scores (Figure 3). Robust analysis including only women who reported feeling sick remained virtually identical to the main analysis (estimate = -0.02, 95%CI = [-0.36; 0.33], t = -0.09, p = .936).

Figure 3

Association of ovulatory cycle phase and pathogen disgust scores



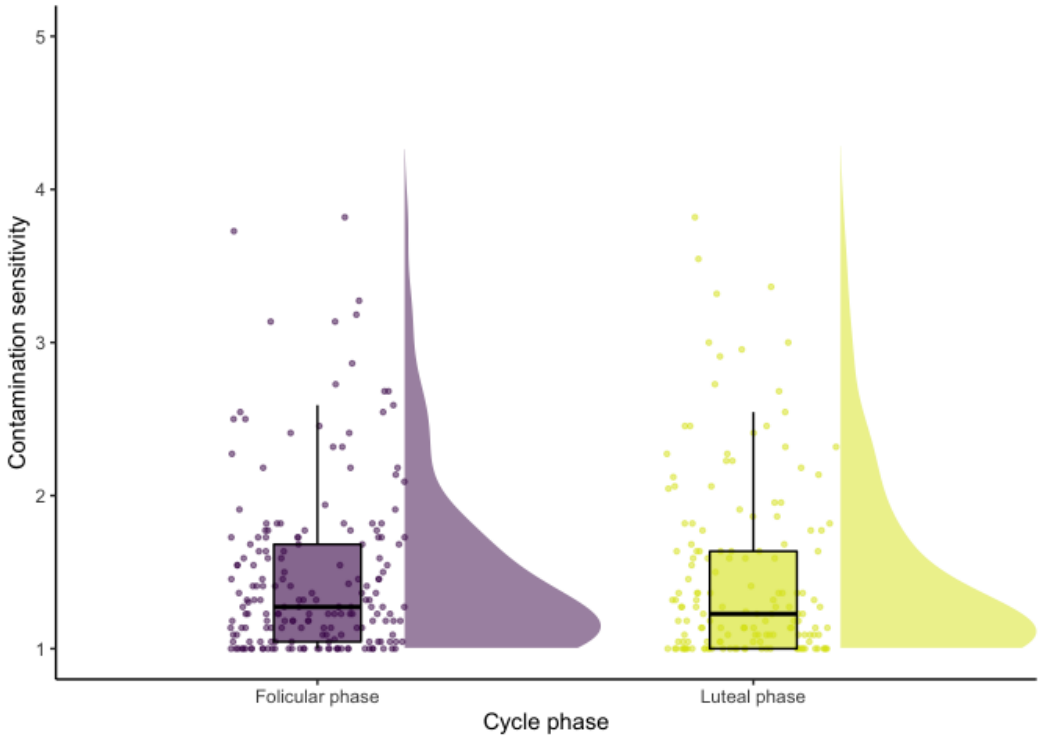
Note: The box plots, points, and distributions represent the average pathogen disgust score for each participant. The box plots are showing the median, first and third quartile, and the minimum and maximum pathogen disgust score for follicular (purple) and luteal (green) phase.

Does contamination sensitivity change across the ovulatory cycle?

Ovulatory cycle phase (estimate = -0.000, 95%CI = [-0.06; 0.06], t = -0.002, p = .998) was not significantly related to contamination sensitivity (Figure 4). Results remained virtually identical to the main analysis when included only women who reported (estimate = 0.08, 95%CI = [-0.09; 0.24], t = 0.935, p = .377).

Figure 4

Association of ovulatory cycle phase and contamination sensitivity scores



Note: The box plots, points, and distributions represent the average contamination sensitivity score for each participant. The box plots are showing the median, first and third quartile, and the minimum and maximum contamination sensitivity score for follicular (purple) and luteal (yellow) phase.

Robustness checks

We computed additional analyses to test the robustness of our results. As some models (especially models only including women feeling sick) showed convergence warnings, we

decided to repeat all models when not including random slopes. Results remained virtually identical, details can be found in the supplement and the open script file. Further, we tested whether our effects are robust when controlling for testing session (to see whether order effects may affect our results). Session was (descriptively) negatively associated with pathogen disgust and contamination sensitivity (estimates between -0.01 and -0.06, *p*s between .018 and .190), suggesting that participants self-reported slightly more pathogen disgust and contamination sensitivity in earlier sessions. However, these effects would not remain significant when controlling for multiple testing. All of our reported results regarding hormone and cycle phase effects remained virtually identical. Details on all results can be found in the supplement and the open script file.

Discussion

In the current study, we aimed to test the CPH by employing a longitudinal design, a large sample size, direct hormone assessments and LH-test validated cycle phase estimates. Across a variety of different analyses, we found no compelling evidence that pathogen disgust or contamination sensitivity are related to LH-validated cycle phase or different hormones levels (within-subjects and between-subjects). We further found no significant effects for women who reported feeling sick in our data.

Previous studies testing the CPH yield mixed findings. Our results are in line with the results by Jones and colleagues (2018) who also reported no compelling evidence that pathogen disgust tracks changes in women's salivary progesterone, estradiol, testosterone, or cortisol. They are further in line with studies not reporting compelling evidence for an association of different cycle phases and increased pathogen disgust (e.g. Fessler & Navarette, 2003; Zelazniewicz et al., 2016), but in contrast to previous studies reporting significant effects for either (within-subjects or between-subjects) hormone levels or different cycle phases.

Fleischman and Fessler (2018) published three possible explanations for the null results reported by Jones and colleagues (2018): a) that the CPH might be entirely wrong, b) measurement issues might explain differences in findings, c) progesterone might not be the driving factor. We argue that these three explanations might also pertain to our findings, as they were virtually identical to the findings published by Jones and colleagues (2018), although Jones and colleagues (2018) did not specifically assess women's cycle phases. Of course, it is possible that the CPH is wrong and that "changes in immune functioning are too small or not consistent enough to exert selective pressure on mechanisms governing behavior" (Fleischman & Fessler, 2018, p. 468). However, we refrain from such strong conclusions based on our findings, given that there are always limitations in single datasets, that absence of evidence does not equal evidence of absence and that our results might not be generalizable to other contexts or samples (e.g. pregnant women). Nevertheless, we think that our findings further challenge the CPH. We agree that differences in used measures might at least partly explain differences in findings between studies. For example, we did not investigate disgust responses to pictures depicting disease cues, for which some previous studies reported findings in line with the CPH (Fleischman & Fessler, 2011; Milkowska et al., 2021). However, not using pictures in the current study does not explain differences in findings of self-reported pathogen disgust or contamination sensitivity via questionnaires also used by Fleischman and Fessler (2011) or Milkowska and colleagues (2019; 2021). Other differences in methods (besides used stimuli) might be more likely to explain differences in findings. For example, the study by Jones and colleagues (2018) and the current study are the studies with the largest sample sizes so far, and also the only large-scale within-subject studies with direct hormone assessments. Further, our study used randomized sessions (e.g. not all participants started testing in the same cycle phase), whereas every participant had her first testing session in the fertile phase in the studies by Milkowska and colleagues (2019, 2021). Interestingly, Fleischman and Fessler (2018) also stated that the higher test power in the study by Jones and

colleagues (2018) suggests that the CPH might rather be wrong than differences in results could be explained by measurement issues. Regarding the potential explanation that progesterone is not the driving factor that regulates disgust sensitivity, we agree that other factors related to pregnancy might lead to upregulated disgust sensitivity. To investigate which factors are actually responsible for fluctuations in disgust sensitivity (and higher disgust sensitivity when pregnant), we suggest that future studies should collect data from pregnant women, as only testing pregnant women can answer this research question properly.

Limitations

We note several limitations regarding our study that might be addressed in subsequent research. First, whereas we investigated whether different self-report questionnaires or different analyses yield different findings regarding our research question, we did not investigate disgust responses to pictures or videos depicting disease cues. Second, we cannot draw strong conclusions on whether having an infection might moderate shifts in disgust sensitivity across the cycle. Although we did not find compelling evidence for this claim, only a very small number of our participants reported feeling sick (potentially as, if they were truly sick, they would not have attended the session). Third, due to ethical constraints, our study was observational, not experimental. Hence, it remains unclear whether progesterone administration might raise disgust, comparably to mice (Bressan & Kramer, 2021).

Conclusion

In conclusion, our results further challenge the CPH, as we report no compelling evidence for associations of steroid hormones or ovulatory cycle phase with pathogen disgust or contamination sensitivity. Future studies might further investigate the influence of current sickness on the CPH, rather test pregnant women, or investigate whether different stimuli (e.g. pictures) might uncover a link between progesterone or the luteal phase and pathogen disgust or contamination sensitivity. Our study further underlines the importance of large-scale

replication studies and direct hormone assessments to test links between hormones, perceptions and behaviors.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 254142454 / GRK 2070, awarded to the first author. We thank our research assistants Paula Bange, Laura Botzet, Sarah Forsthoff, Isa Garbisch, Kim Gloystein, Salome Johannigmann, Indra Kirchberg, Laila Knapp and Sabine Ostermann for data collection. We further thank Wei Gao for providing detailed information on the LCMS analyses.

References

Arslan, R.C., Schilling, K.M., Gerlach, T. M., & Penke, L. (2021). Using 26 thousand diary entries to show ovulatory changes in sexual desire and behavior. *Journal of Personality and Social Psychology*, *121*, 410-431. https://doi.org/10.1037/pspp0000208

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Blake, K. R., Dixson, B. J., O'Dean, S. M., & Denson, T. F. (2016). Standardized protocols for characterizing women's fertility: A data-driven approach. *Hormones and Behavior*, *81*, 74–83. https://doi.org/10.1016/j.yhbeh.2016.03.004

Bouman, A., Moes, H., Heineman, M. J., de Leij, L. F., & Faas, M. M. (2001). The immune response during the luteal phase of the ovarian cycle: increasing sensitivity of human monocytes to endotoxin. *Fertility and Sterility*, *76*, 555-559. https://doi.org/10.1016/S0015-0282(01)01971-9

Bressan, P., & Kramer, P. (2021). Progesterone does raise disgust. *Hormones and Behavior*, 104937. https://doi.org/10.1016/j.yhbeh.2021.104937

Burns, G. L., Keortge, S. G., Formea, G. M., & Sternberger, L. G. (1996). Revision of the Padua Inventory of obsessive compulsive disorder symptoms: Distinctions between worry, obsessions, and compulsions. *Behaviour Research and Therapy*, *34*, 163-173. https://doi.org/10.1016/0005-7967(95)00035-6

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365-376. https://doi.org/10.1038/nrn3475

Faas, M., Bouman, A., Moesa, H., Heineman, M. J., de Leij, L., & Schuiling, G. (2000). The immune response during the luteal phase of the ovarian cycle: a Th2-type response?. *Fertility and Sterility*, *74*, 1008-1013. https://doi.org/10.1016/S0015-0282(00)01553-3

Fessler, D. M., & Navarrete, C. D. (2003). Domain-specific variation in disgust sensitivity across the ovulatory cycle. *Evolution and Human Behavior*, *24*, 406-417. https://doi.org/10.1016/S1090-5138(03)00054-0

Fessler, D. M., Eng, S. J., & Navarrete, C. D. (2005). Elevated disgust sensitivity in the first trimester of pregnancy: Evidence supporting the compensatory prophylaxis hypothesis. *Evolution and Human Behavior*, *26*, 344-351. https://doi.org/10.1016/j.evolhumbehav.2004.12.001

Fleischman, D. S., & Fessler, D. M. (2011). Progesterone's effects on the psychology of disease avoidance: Support for the compensatory behavioral prophylaxis hypothesis. *Hormones and Behavior*, *59*, 271-275. https://doi.org/10.1016/j.yhbeh.2010.11.014

Fleischman, D. S., & Fessler, D. M. (2018). Response to "Hormonal correlates of pathogen disgust: Testing the compensatory prophylaxis hypothesis.". *Evolution and Human Behavior*, *39*, 468-469. https://doi.org/10.1016/j.evolhumbehav.2018.03.006

Gangestad, S. W., Haselton, M. G., Welling, L. L., Gildersleeve, K., Pillsworth, E. G., Burriss, R. P., …& Puts, D. A. (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evolution and Human Behavior, 37,* 85–96. https://doi.org/10.1016/j.evolhumbehav.2015.09.001

Gao, W., Stalder, T., & Kirschbaum, C. (2015). Quantitative analysis of estradiol and six other steroid hormones in human saliva using a high throughput liquid chromatography–tandem mass spectrometry assay. *Talanta*, *143*, 353-358.

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Lee, A. J., Tybur, J. M., & DeBruine, L. M. (2018). Hormonal correlates of pathogen disgust: Testing the Compensatory Prophylaxis Hypothesis. *Evolution and Human Behavior*, *39*, 166-169. https://doi.org/10.1016/j.evolhumbehav.2017.12.004

Kavaliers, M., Bishnoi, I. R., Ossenkopp, K. P., & Choleris, E. (2021a). Differential effects of progesterone on social recognition and the avoidance of pathogen threat by female mice. *Hormones and Behavior*, *127*, 104873. https://doi.org/10.1016/j.yhbeh.2020.104873

Kavaliers, M., Bishnoi, I. R., Ossenkopp, K. P., & Choleris, E. (2021b). Progesterone and disgust: A response to" progesterone does raise disgust". *Hormones and Behavior*, 104936. https://doi.org/10.1016/j.yhbeh.2021.104936

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, *2*.0-33.

Milkowska, K., Galbarczyk, A., & Jasienska, G. (2019). Disgust sensitivity in relation to ovulatory cycle phase in women with and without an infection. *American Journal of Human Biology*, *31*, e23233. https://doi.org/10.1002/ajhb.23233

Miłkowska, K., Galbarczyk, A., Klimek, M., Zabłocka-Słowińska, K., & Jasienska, G. (2021). Pathogen disgust, but not moral disgust, changes across the ovulatory cycle. *Evolution and Human Behavior*, *42*, 402-408. https://doi.org/10.1016/j.evolhumbehav.2021.03.002

Olatunji, B. O., Cox, R. C., & Li, I. (2020). Disgust regulation between ovulatory cycle phases: Differential effects of emotional suppression and reappraisal. *Journal of Behavior Therapy and Experimental Psychiatry*, *68*, 101543. https://doi.org/10.1016/j.jbtep.2019.101543

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Revelle, W. (2016). *Psych: Procedures for personality and psychological research* (Version 1.6.12) [Computer software]. Retrieved from https:// CRAN.R-project.org/packagepsych

Roney, J. R. (2018). Hormonal mechanisms and the optimal use of luteinizing hormone tests in human ovulatory cycle research. *Hormones and Behavior*, *106*, A7-A9.

Roney, J. R., & Simmons, Z. L. (2013). Hormonal predictors of sexual motivation in natural ovulatory cycles. *Hormones and Behavior*, *63*, 636-645. https://doi.org/10.1016/j.yhbeh.2013.02.013

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J.S. Beer (Eds.), *Methods in Social Neuroscience* (pp. 17-44). New York, NY: Guilford Press.

Stern, J., Kordsmeyer, T. L., & Penke, L. (2021). A longitudinal evaluation of ovulatory cycle shifts in women's mate attraction and preferences. *Hormones and Behavior*, *128*, 104916. https://doi.org/10.1016/j.yhbeh.2020.104916

Treffenstaedt, C., & Wiemann, P. (2018). Alfred - A library for rapid experiment development (Version 0.2b5). https://doi.org/10.5281/zenodo.1437220

Trzonkowski, P., Myśliwska, J., Łukaszuk, K., Szmit, E., Bryl, E., & Myśliwski, A. (2001). Luteal phase of the menstrual cycle in young healthy women is associated with decline in interleukin 2 levels. *Hormone and Metabolic Research*, *33*, 348-353. https://doi.org/10.1055/s-2001-15420

Tybur, J. M., Lieberman, D., & Griskevicius, V. (2009). Microbes, mating, and morality: individual differences in three functional domains of disgust. *Journal of Personality and Social Psychology*, *97*, 103 - 122. https://doi.org/10.1037/a0015474

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., ... & Kuhn, M. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*, 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2019). readxl: Read excel files. *R package version*, *1*(1).

Wilke, C. O., Wickham, H., & Wilke, M. C. O. (2019). Package 'cowplot'. *Streamlined Plot Theme and Plot Annotations for 'ggplot2*. R package version, 1.1.1 [Computer software]. Retrieved from https://cran.r-project.org/web/packages/cowplot/index.html

Xie, Y. (2014). knitr: a comprehensive tool for reproducible research in R. In Stodden, V. Leisch, F., & Peng, R. D. (Eds.). *Implementing Reproducible Research* (pp. 3-29). Boca Raton, FL: CRC Press.

Żelaźniewicz, A., Borkowska, B., Nowak, J., & Pawłowski, B. (2016). The progesterone level, leukocyte count and disgust sensitivity across the ovulatory cycle. *Physiology & Behavior*, *161*, 60-65. https://doi.org/10.1016/j.physbeh.2016.04.002

Zhu, H. (2019). KableExtra: Construct complex table with 'kable' and pipe syntax. R package version, 1(0) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/kableExtra/index.html