Received: 17 November 2022 | Accepted: 15 July 2023

DOI: 10.1111/bjso.12681

ARTICLE

the british psychological society
promoting excellence in psychology

Investigating cycle shifts in women's clothing style and grooming

Julia Stern[1,2] | Sabine Ostermann[2] | Lars Penke[2,3]

[1]Department of Psychology, University of Bremen, Bremen, Germany

[2]Department of Psychology, University of Goettingen, Goettingen, Germany

[3]Leibniz ScienceCampus Primate Cognition, Goettingen, Germany

Correspondence
Julia Stern, University of Bremen, Grazer Straße 2c, Bremen 28359, Germany.
Email: jstern@uni-bremen.de

Funding information
Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Grant/Award Number: 254142454/GRK 2070

Abstract
In contrast to some non-human primate species, human females do not show overt cues to fertility. Previous research argued that women still show systematic changes in their appearance across their ovulatory cycle to enhance their mating success when fertile. We report five studies investigating whether women's clothing style and grooming behaviour change across the ovulatory cycle. All studies were large (with $N = 157$ in Study 1, $N = 109$ in Study 2, $N = 257$ in Studies 3–5), longitudinal studies with four testing sessions per participant. They involved salivary hormone samples and luteinizing hormone tests to validate conception risk estimates. Across all studies, our results suggest no compelling evidence for cycle shifts in clothing style and grooming. Rather, two studies suggest effects in the opposite direction as hypothesized, as women wore more skin-revealing clothes when non-fertile. One study suggests small effects of wearing necklaces more and eyeglasses less often when fertile. However, these effects were not robust across all studies. Our results are in line with other recent null replications and suggest that, if existent, cues to fertility might be even more subtle than previously assumed. We discuss the need for testing competing theories that explain the evolution of concealed ovulation.

KEYWORDS
clothing style, fertility cues, grooming, ovulatory cycle, steroid hormones

INTRODUCTION

In mammalian species, female fertility is usually cyclical and restricted to certain short periods: females typically only engage in sex when conception is possible, and alter their physical appearance and behaviour to signal fertility to potential mates, aiming at increasing their reproductive success. Systematic changes across the estrus cycle are assumed to be regulated by reproductive hormones, mainly estradiol and progesterone (Roney, 2016). Human female ovulatory cycles are on average about 28 days long (Wilcox et al., 2000), with large variations in length, and can be divided into two different phases. The follicular phase begins with menstrual bleeding and ends with ovulation around mid-cycle, which then introduces the start of the luteal phase. Conception is only possible at the day of ovulation and up to five days prior. While estradiol levels are usually higher in the follicular phase and peak around the day of ovulation (with a second, smaller peak mid-luteal), progesterone levels rise in the luteal phase, peaking around mid-luteal. Because of the secretion pattern, the estradiol-to-progesterone ratio (E/P, i.e. estradiol divided by progesterone) is assumed to be a good index of the fertile window timing (Roney, 2019).

Do human females show estrus?

Human females engage in sex throughout the cycle, and also post-menopausal, and do not advertise their fertility by displaying obvious cues, such as sexual swellings (as e.g. chimpanzees, our closest phylogenetic relatives, do, Deschner et al., 2003; though common ancestors of humans and chimpanzees did not have pronounced swellings, Pawłowski, 2015). Thus, ovulation in human females appears concealed, which led to the assumption that (classically defined) estrus was lost over evolutionary time, possibly due to the evolution of pair-bonding (Alexander & Noonan, 1979). However, this claim has been challenged by findings suggesting that there are systematic psychological and behavioural changes across the ovulatory cycle (Thornhill & Gangestad, 2008).

Evidence for subtle cues to fertility

Systematic changes across women's ovulatory cycle seem predominantly related to their mating psychology. For example, women show an increase in sexual desire (Arslan et al., 2021; Jones, Hahn, Fisher, Wang, Kandrik, & DeBruine, 2018; Roney & Simmons, 2013, 2016; van Stein et al., 2019), and self-report feeling more attractive and sexually desirable when fertile (Arslan et al., 2021; Haselton & Gangestad, 2006; Schleifenbaum et al., 2021). These changes across the cycle were reported to be perceived by others and to impact women's social life. For example, men seem to show more mate retention behaviour when their female partner is fertile (Gangestad et al., 2002, 2014; Haselton & Gangestad, 2006; but see Schleifenbaum et al., 2022), and professional lap dancers earn more tips when fertile (Miller et al., 2007). Appearance-related shifts across the cycle that affect women's attractiveness might serve as subtle cues to fertility. For example, it has been reported that fertile women are perceived as having more attractive faces, voices and body scent (Gildersleeve et al., 2012; Lobmaier et al., 2018; Pipitone & Gallup Jr, 2008; Puts et al., 2013; Roberts et al., 2004). However, these findings did not replicate in more recent, higher powered studies (Bleske-Rechek et al., 2011; Catena et al., 2019; Jones, Hahn, Fisher, Wang, Kandrik, Lao, et al., 2018; Mei et al., 2022). Nevertheless, other research suggests cues to fertility that are actively affected by women's behaviour, including changes in clothing style (i.e. women dressing more sexy, provocative or attractive), and in time spent grooming. Both can signal sexual interest, attract potential mates and are detectable by others (Durante et al., 2014; Haselton et al., 2007).

Ovulatory shifts in clothing style and grooming

Evidence for cycle shifts in clothing style and grooming seems compelling. Haselton et al. (2007) conducted the first influential study on this topic. They took pictures of $N=30$ partnered women in their fertile and luteal phase and presented them to $N=42$ judges who had to choose in which picture the displayed woman tried to look more attractive (with faces concealed). In almost 60% of all cases, the judges chose the fertile phase picture, which translates to a large effect size of $d=0.72$ (Haselton & Gildersleeve, 2011). The authors argued that fertile women may 'dress to impress', in that they are trying to attract mates via self-ornamentation. They concluded that changes in women's grooming and clothing style might be the most apparent cues to fertility and responsible for changes in relationship dynamics across the cycle (Haselton et al., 2007). In the following years, several studies reported that when fertile, women wear or aim to wear sexier, more skin revealing clothes (Durante et al., 2008; Saad & Stenstrom, 2012; Schwarz & Hassebrauck, 2008), or prefer to buy sexier clothes (Blake, Bastian, et al., 2017; Durante et al., 2011). Further, women may spend more time grooming when fertile, such as styling their hair, using more make-up, wearing high heels or putting on jewellery (Röder et al., 2009; Saad & Stenstrom, 2012). One very prominent finding is the so-called 'red effect', according to which fertile women wear more red or pink clothes (Beall & Tracy, 2013). These colours are seen as attention grabbing and it has been reported that men perceive women wearing red as more sexually desirable (Elliot & Niesta, 2008; Elliot et al., 2013, but see Peperkoorn et al., 2016) and that women wear red clothes to attract mates (Prokop & Hromada, 2013).

However, research in the recent years has revealed mixed findings concerning cycle shifts in clothing style and grooming. For instance, fertile women might only wear more red clothes on relatively cold days (Tracy & Beall, 2014). Whether the red effect was significant differed also between methods to estimate fertility and hormonal predictors (Eisenbruch et al., 2015). More precisely, the authors found a robust significant red effect predicted by the estradiol-to-progesterone ratio, and a fertile window effect (but not estradiol and progesterone separately) and interpret their results as being in line with the red effect. Two studies failed to find any compelling evidence for the red effect (Blake, Dixson, et al., 2017; Hone & McCullough, 2020), and two pre-registered large-scale diary studies did not find compelling evidence for cycle shifts in self-reported sexy clothing (Arslan et al., 2021) or grooming (Schleifenbaum et al., 2021), casting doubt on previous findings.

Methodological criticism

Over recent years, ovulatory cycle research has been criticized for methodological shortcomings, revealing several reasons to explain mixed findings. As a result, the existence of cycle shifts in women's cognition and behaviour has been hotly debated. First, while validating cycle phase estimates with luteinizing hormone (LH) urine tests is currently seen as the gold standard (Blake et al., 2016; Gangestad et al., 2016), multiple studies relied on comparably invalid counting methods. Second, many studies on cycle shifts were underpowered, which can cause not only false negative, but also false positive results. The median sample size for within-subject studies that reported (at least some) evidence for cycle shifts in clothing style or grooming was $N=43$ ($N=17$ in Saad & Stenstrom, 2012, to $N=96$ in Blake, Bastian, et al., 2017). Some studies even used between-subjects designs to investigate a within-subjects effect, leading to low test power despite generally large sample sizes (between $N=100$ in Hone & McCullough, 2020, Study 1 and $N=617$ in Study 2).

Third, a potential source of bias are subjective self-reports of wearing sexy clothes or spending time grooming. Multiple studies had to rely on self-reports due to their online design (Arslan et al., 2021; Schleifenbaum et al., 2021). A more objective approach is to take standardized pictures of participants' clothes, as done in previous laboratory studies (Eisenbruch et al., 2015; Haselton et al., 2007). Fourth, although reproductive hormones are assumed to regulate cycle shifts, hormone levels have only been assessed in two datasets yielding mixed results (Blake, Bastian, et al., 2017; Blake, Dixson, et al., 2017;

Eisenbruch et al., 2015). Fifth, almost all studies failed to control for plausible confounding variables, such as the weather that likely affects clothes worn. Sixth, researcher degrees of freedom and analytical flexibility, the fact that all researchers across all disciplines have to choose how exactly to conduct their study, formulate their hypotheses, analyse their data and report their results, as well as publication bias in favour of positive findings, attenuates replicability (Wicherts et al., 2016). To solve this issue, pre-registering methods and analysis plans, which are currently recommended as best practice, are highly important.

Current article

Overall, previous mixed findings and methodological shortcomings demonstrate a high demand for studies with more robust designs to clarify whether, how and why women systematically change their clothing style and grooming across their ovulatory cycle. The aim of the current studies is to contribute these research questions. We investigate (a) whether women's clothing style and grooming shift across their ovulatory cycle and with fluctuations in reproductive hormones, (b) whether others perceive fertile women to dress more attractively, (c) whether women aim to wear sexier and more skin revealing clothes to a party when fertile and (d) whether women aim to buy more sexy clothes when fertile. For this purpose, we replicate and widely extend previous findings, while overcoming methodological shortcomings and using methods that are currently considered the gold standard. The majority of the studies reported in this article were pre-registered; we provide open data and analysis scripts for all studies, and open material where possible.

STUDY 1

In the first study, we examined cycle shifts in clothing style and grooming. We pre-registered the following hypotheses: Women wear more body-exposing clothes (H1a), body accentuating/tight clothes (H1c) and will show more skin (H1d) during the fertile than the luteal phase of their ovulatory cycle. Further, women groom themselves more during the high fertility phase (H2a), which is shown by wearing their hair openly more often (H2c), put more effort in their hairstyle (H2d), use more make-up (H2e), wear more nail polish (H2f), more accessories/jewellery (H2g), wear correcting eyeglasses less often (H2h) and wear high heels more often (H2i). Moreover, women are more likely to wear red or pink clothes during the fertile as compared to the luteal phase of their menstrual cycle (H3a), which should be applicable to the upper body part (H3c), the lower body part (H3d) as well as to make-up (H3e) and nail polish (H3f). We further predicted these effects of H1a, H2a and H3a to be mediated by higher E/P (H1b, H2b, H3b), but failed to pre-register hormone effects for the remaining hypotheses. We nevertheless investigate and report hormone effects for all outcome variables (see robustness checks). We repeat all analyses controlling for weather condition, age, relationship status, body mass index (BMI) and self-reported personality traits. Finally, following Blake, Dixson, et al. (2017), we investigated cycle effects of other colours than red or pink in an exploratory manner.

Methods

This study was pre-registered online (as a pdf file; https://osf.io/c7sgv/) before codings or analyses took place. Data collection procedures of the photographs were pre-registered separately, as part of a bigger project on ovulatory cycle effects (https://osf.io/egjwv/). Data, analysis code and other materials for all studies reported in this article are available online (https://osf.io/cfreh/). All participants signed a written consent form and the local ethics committee approved the study protocol.

Participants

A total of $N = 157$ (age: $M = 23.3$ years, $SD = 3.4$) participants finished all sessions and were included in the following analyses. All participants reported to fit the following inclusion criteria: female, between 18 and 30 years old,[1] having regular cycles (25–35 days, and naturally cycling, for details see Supplementary Note 1a). Out of the 180 recruited participants, $n = 23$ dropped out (Supplementary Note 1a). Our sample exceeded the size required to achieve 80% power given a within-subject design and anticipated effects of moderate magnitude (Cohen's $d = 0.5$ with $N = 48$ for LH test validated cycle phases and two testing sessions per participant, suggesting sufficient power to detect much smaller effect sizes in our study), as suggested by a power simulation for ovulatory cycle research (Gangestad et al., 2016).

Procedure

All participants took part in five sessions, one introductory session (for checking inclusion criteria, collecting demographic data and cycle-related information to individually schedule the following sessions) and four testing sessions. Testing sessions took place once during the fertile phase and once during the luteal phase for two (mostly) consecutive cycles. To control for possible effects of diurnal changes in hormone levels, we scheduled all sessions in the second half of the day (11.30 a.m. to 6.00 p.m.). Participants first completed a screening questionnaire that assessed their eligibility and control variables for saliva sampling (Schultheiss & Stanton, 2009). Saliva samples were collected via passive drool. Then, standardized pictures, height and weight were collected. Full-body photographs were taken via a digital camera (Canon® EOS 350D) with standardized angle and distance. Participants were instructed to stand upright on a marked line in front of a white background, next to a Gretag Macbeth Color Checker and to put on a neutral facial expression. One photograph got lost while transferring the photographs, resulting in a total of 627 photographs that were then coded using a standardized coding scheme.

As part of a larger project, participants had to complete other tasks (e.g. attractiveness ratings of men) that are not part of the current article and were published elsewhere (e.g., Stern et al., 2020). Upon completion of all sessions, participants received a payment of 80€ or course credit. Participants were invited to take part in a parallel, but separate online diary study (Arslan, Driebe, et al., 2020) from which we were able to match data for $n = 142$ participants to use measures of self-reported personality traits.

Measures

Ovulatory cycle phase

Cycle phase was determined by the reverse cycle day method, based on the estimated day of the next menstrual onset (Gildersleeve et al., 2012) and confirmed by highly sensitive (10 mIU/mL) urine ovulation tests from Purbay® measuring the luteinizing hormone (LH). LH tests were done at home at the estimated day of ovulation and four days prior. For the cycle phase analyses, we excluded a total of 45 participants due to negative LH tests in both cycles, irregular ovulatory cycles or inappropriate scheduling of testing sessions (see Supplementary Note 1b for details), resulting in $n = 112$ women. Of these participants, 46 started in their luteal phase and 66 started fertile. However, all 157 women were included in the denoted hormone analyses and robustness checks.

[1]One participant later reported to be 35 years old. We included her data as she met all other inclusion criteria, had positive LH tests and excluding her data did not change any results.

Hormone assessments

One saliva sample from each participant was collected each testing session. Contamination of samples was minimized by asking participants to abstain from eating, drinking (except plain water), smoking, chewing gum or brushing teeth for at least 1 hour before each session. Samples were visually inspected for blood contamination and stored at −80°C until shipment on dry ice to the Kirschbaum Lab at Technical University of Dresden, Germany, where estradiol, progesterone, testosterone and cortisol were assessed via liquid chromatography mass spectrometry (LCMS; Gao et al., 2015). In only 22% of the hormone samples, estradiol levels could be detected by LCMS analysis. Therefore, all samples were re-analysed using a highly sensitive 17β-estradiol enzyme immunoassay kit (IBL International). These latter estradiol values were used in subsequent analyses. We centred 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 linear mixed models (Figure S1).[2] We pre-registered excluding outliers >3 *SD*s from the mean. However, we decided to slightly deviate from our pre-registration, as the pre-registered procedure would have led to excluding a number of progesterone levels that are in a plausible range for mid-luteal levels. Thus, we separately excluded outliers >3 *SD*s for the distinct cycle phases (fertile and luteal), which should avoid misidentifying phase-specific peaks as outliers (Roney & Simmons, 2013).

Coding

We developed three different coding sheets (see open material), one for each outcome (body exposure, grooming and clothing colour). Three independent research assistants that were blind to participant's cycle phase coded the photographs with a different order for each coder. They were instructed to code all variables for each photograph before proceeding to the next one. The procedure was practiced in a training session (with five example photographs). The coding procedure is explained in detail in Supplementary Note 1c, coding material is also available (Tables S1–S3, Figure S2). Inter-coder reliabilities were very mixed, from poor for wearing high heels (fleiss $\kappa = .16$) to almost perfect for wearing glasses ($\kappa = .97$); a detailed table can be found in the Supporting Information (Table S4). We nevertheless report all results (in the Supporting Information), but refrain from interpreting results for outcomes with low reliability.

Control variables

Rainfall (in L/m^2) and maximum temperature (in °C) were retrieved for each testing day (from https://www.wetterkontor.de/de/wetter/deutschland/rueckblick.asp[3]). We assessed participant's weight with a scale while wearing standardized sports underwear, and height with a stadiometer to compute BMI. Further, in each session, participant's self-reported relationship status and age. Relationship status was effect coded (1 = partnered, −1 = single). We classified all women who reported to be in an open relationship, in a committed relationship, engaged or married as in a relationship, whereas participants that reported being single were categorized accordingly. At the beginning of the study, 75 of the participants reported to be in a relationship, 82 were single. Participants relationship status changed for $n = 13$ participants during the study, their data were categorized in accordance with their relationship status on the day of the particular testing session. Finally, we assessed self-reported personality traits via an online survey in formr (Arslan, Walther, & Tata, 2020). Relevant for this study are the revised Sociosexual Orientation Inventory (SOI-R; Penke & Asendorpf, 2008), the Narcissistic Admiration and Rivalry Questionnaire (NARQ; Back et al., 2013) and extraversion assessed with the Big Five Inventory (BFI; Lang et al., 2001). Cronbach's

[2]This procedure deviates from our pre-registration, in which we stated that we will log base 10 transform hormone levels. However, we learned that log-transformed hormone data may diminish the real variation between and within ovulatory cycles (Roney, 2019). We still report all analyses with log-transformed hormone levels in the Supporting Information.

[3]We decided to use this database, as it provided more exact data as compared to the pre-registered one.

TABLE 1 Multi-level regression analyses of body exposure scores (including tightness and skin display) as a function of cycle phase or E/P ratio.

	β	95% CI	t	p
H1a: Body exposure composite				
(Intercept)	.00	−0.13 to 0.13	37.20	<.001
Cycle phase	−.03	−0.11 to 0.04	−0.83	.404
H2a: Grooming composite				
(Intercept)	.01	−0.15 to 0.16	24.46	<.001
Cycle phase	.01	−0.05 to 0.07	0.30	.768
H3a: Red composite				
(Intercept)	.04	−0.09 to 0.17	47.33	<.001
Cycle phase	−.08	−0.26 to 0.10	−0.85	.395

alpha for each scale was good (socio-sexual behaviour $\alpha = .86$, socio-sexual attitude $\alpha = .88$, socio-sexual desire $\alpha = .76$, narcissistic admiration $\alpha = .82$, narcissistic rivalry $\alpha = .85$, extraversion $\alpha = .84$).

Statistical analyses

All data wrangling and analyses were done with the software R 4.2.0 (R Core Team, 2016), and all used packages can be found in the Supporting Information. All statistical tests were two tailed. All models are linear multi-level models, even if the outcome was binary (following the recommendation of Gomila, 2021). We also report all models with binary outcomes in our open script. Some of these models did not converge and confidence intervals are extremely wide and uninterpretable, potentially because of very little variation in the outcome variables (e.g. wearing eyeglasses). We tested our hypotheses via separate models with participant ID as random intercept. We modelled a random slope for cycle phase in each model, and for hormone predictors in the denoted hormone models.[4]

Results

Main analyses

Body exposure was not significantly related to cycle phase and this association was not mediated by E/P, contradicting Hypotheses 1a, 1b. Cycle phase was significantly negatively related to skin display ($\beta = -.13$, 95% CI $= -0.21$ to -0.06, $t = -3.55$, $p < .001$), indicating a small effect that participants in the luteal phase wear more skin-revealing clothes. This effect is in the opposite direction as predicted by Hypothesis 1d. Results are displayed in Table 1 and in Table S5. Adding the control variables to the model did not change any results (Tables S8–S11). Maximum temperature was significantly related to body exposure and skin display, in that women wore more revealing clothes when the temperature was higher.

None of the grooming variables showed a significant relationship to cycle phase or E/P, contradicting Hypotheses 2a to 2j (Table 1 and Table S6). All standardized effect size estimates were very close to zero. Adding the control variables did not change any of our results (Tables S12–S20). Further, none of the colour variables showed a significant relationship to cycle phase or E/P,

[4]We did not specify random slopes in our pre-registration, but decided to model them to avoid Type 1 error inflation (Barr et al., 2013) and to be in line with our analyses in Studies 2–5. This decision did not change any results.

in contrast to Hypotheses 3a to 3f (Tables S7 and S29). All standardized effect size estimates or their confidence intervals were close to zero. Adding the control variables did not change any of our results (Tables S21–S26). There were small associations between wearing red and some of the personality variables or relationship status, but all other effects were non-significant. None of the other coded clothing colours showed any significant relationship with any of our cycle variables, except for a higher probability to wear upper body clothes coloured in natural colours when fertile (Tables S39–S42).

Robustness checks

To match previous analyses reported in the literature and to test the robustness of our findings, we investigated effects of modelling E/P, or estradiol and progesterone as separate predictors, rather than cycle phase. Furthermore, we repeated all of these analyses with log-transformed hormone levels. In short, almost all results were non-significant and did not change our conclusions. The significant effects were a negative relationship between log-transformed E/P and wearing red clothes ($\beta = -.09$, $p = .039$), as well as between log-transformed estradiol and wearing red clothes and nail polish (βs $= -.10$, ps $= .039$). These effects are in the opposite direction as predicted and would not be significant after controlling for multiple testing. Details can be found in the Supporting Information (Tables S27–S38). We further repeated our analyses (a) without excluding outliers, (b) without excluding women with negative LH tests and (c) repeated all robustness checks when modelling control variables. None of these analyses changed any of conclusions. Details are reported in the open script.

Discussion

Study 1 showed no compelling support for any of our pre-registered hypotheses. Instead, we found that women displayed more skin in the luteal phase of the cycle. However, the effect of wearing more sexy clothes in the luteal phase is in the opposite direction as expected and was not supported by hormonal associations. While the reported null findings are in line with other studies (Arslan et al., 2021; Blake, Bastian, et al., 2017; Hone & McCullough, 2020; Schleifenbaum et al., 2021), they need replication given the limitations of Study 1: First, poor inter-coder reliabilities for some dimensions suggest that they were very hard to detect based on the photographs, and that the measures were thus not reliable (e.g. wearing high heels). Second, there was very little variation in some variables (e.g. wearing red clothes, eyeglasses). Participants might not vary in these variables in our specific sample or changes might rather occur in non-laboratory contexts. Third, women might differ in how they try to look more attractive: some wear more make-up or accessories and others wear more revealing or noticeably coloured clothes. If there is heterogeneity in attractiveness enhancement, the current study design was unable to reveal these effects (though we modelled random slopes to statistically consider individual differences). The following studies address these limitations.

STUDY 2

This study is a conceptual replication of Haselton et al. (2007). The design allows us to investigate whether women shift their overall appearance beyond specifically defined dimensions. This study employs a forced choice design, directly comparing outfits in the fertile versus luteal phase to overcome potential problems of little variation in specific variables. Based on findings of the original study, we hypothesized that the fertile phase photo should be picked more often as looking more attractive than the luteal phase photo (H1a). This effect should be stronger for women in relationships (H1b), and the closer the fertile phase photo was taken to the day of ovulation, or the luteal phase photo was taken to the day of the next menstrual onset (H1c). The latter effects should be stronger for women in relationships (H1d). We predicted effects for E/P to correspond with all predicted cycle phase effects (H2a, H2b, or H2c respectively). In an exploratory manner, we investigated whether effects differ depending on the sex of judges.

Methods

This study was pre-registered online before data collection (https://osf.io/49sp7/). Data, analysis code and material are available at https://osf.io/cfreh/. All participants signed a written consent form and the study protocol was covered by the approved ethics committee proposal for Study 1.

Participants

Target participants were the same as in Study 1. We only used the photographs of the women who showed positive LH tests ($n = 112$). Three more participants were excluded, due to photographs being missing, underexposed or blurred. Photographs were split into two sets to avoid that judges were confronted with the same women twice. Set 1 included both photographs of every target from their Sessions 1 and 2, Set 2 included the remaining photographs of Sessions 3 and 4. Judges were $N = 105$ participants[5] (52 women, 53 men, aged 18–37 years). They were randomly assigned to one of the sets (Set 1 with $n = 25$ women, $n = 27$ men, age $M = 22.85$, $SD = 4.07$; Set 2 with $n = 27$ women, $n = 26$ men, age $M = 23.23$, $SD = 3.98$). The sample size of judges was based on the study by Haselton et al. (2007), who reported that a sample of $N = 42$ judges led to a high inter-judge agreement. Our target sample size exceeds the sample size of the original study ($N = 30$) and should have sufficient power to detect at least the effect sizes reported in the original study (Haselton et al., 2007).

Procedure

After signing an informed consent and entering demographic data, all photographs were presented to the judges on a screen via the open source software Alfred (Treffenstaedt & Wiemann, 2018). Two full-body photographs (with faces concealed) of one woman (one fertile, one luteal) were presented at the same time. Judges were asked 'In which photo is the person trying to look more attractive?' and chose via mouse click (adapted from Haselton et al., 2007). The order of targets and the side the fertile photographed was displayed on (left vs. right) was randomized. Judges showed a fair overall agreement with $\kappa = .23$ for Set 1 and $\kappa = .25$ for Set 2.

Measures

Existing data

Target participant's cycle phase, hormone levels, age, relationship status and weather conditions were described in Study 1. Although we pre-registered not to exclude outliers, we decided to exclude $n = 1$ progesterone value and $n = 1$ E/P value in Set 1, $n = 1$ progesterone, $n = 2$ estradiol and $n = 2$ E/P values in Set 2 for further analyses, as they were extreme outliers and potentially due to measurement error (>8 *SD*s above the mean for Set 1 and >6 *SD*s for Set 2).

[5]We pre-registered that we will collect data from 100 participants, but recruited a few more as we expected more dropouts. We decided to use the data of all participants.

Variables

Days to ovulation/menstruation were measured as the number of days passed between taking the photograph and the day of ovulation or next menstrual onset. Judges self-reported their gender (−1 = male, 1 = female) and chose a picture as trying to look more attractive (coded 0 = *no*, 1 = *yes*).

Statistical analyses

All data wrangling and analyses were done in line with Study 1. We pre-registered to perform all analyses for Set 1 and Set 2 separately for the purpose of an internal replication, but decided to combine both sets for the main analyses to increase statistical power. The separate analyses are reported in the Supporting Information.

Results

The fertile phase photographs were chosen in 46.68%, which suggests a small descriptive preference for the luteal phase photographs. Contradicting Hypothesis 1a, judges descriptively, but not significantly chose the luteal phase photographs more often than the fertile phase photographs ($\beta = -.13$, 95% CI = −0.27 to 0.01, $t = -1.85$, $p = .065$). Choices were not significantly related to an interaction of cycle phase and relationship status, in contrast to Hypothesis 1b. Neither days to ovulation[6] ($\beta = .05$, 95% CI = −0.09 to 0.18, $t = 0.65$, $p = .513$) nor days to menstruation ($\beta = -.03$, 95% CI = −0.23 to 0.17, $t = -0.29$, $p = .771$) or their interaction with relationship status (Table S43) were significantly related to choices. All reported results in this section contradict Hypotheses 1c and 1d. Analyses including E/P did not reveal a significant association of choice and E/P ($\beta = -.09$, 95% CI = −0.43 to 0.25, $t = -0.51$, $p = .607$), contradicting Hypothesis 2a. Choices were not significantly related to an interaction of E/P and relationship status, days to ovulation or menstruation, contradicting Hypothesis 2b and 2c (Table S44).

Exploratory analyses and robustness checks

We investigated whether sex of judge is associated with choices of photographs (Table S45), and repeated our main hormone analyses with estradiol and progesterone as separate predictors (Tables S46, S47). Then, we repeated all analyses with log-transformed hormone levels (Tables S48–S50) and controlled all of our main analyses for weather condition (see open script). We further repeated all analyses for both sets separately (Tables S51–S72). None of these analyses changed our conclusions. If significant results occurred, they were in the opposite direction as predicted (i.e. preference for the luteal phase picture) and not robust across all analyses.

Discussion

The results of Study 2 suggest no compelling evidence that women dress to impress when fertile, as they are not perceived as 'trying to look more attractive'. Thus, we did not replicate the findings reported by Haselton et al. (2007).

[6]Our analyses for days to ovulation and days to menstruation differ from our pre-registered analyses, as we can only analyse the effect of days to ovulation within the fertile days. Thus, we had to compute simple main effects for these variables (not interacting with cycle phase).

STUDY 3

This study is a conceptual replication of Study 1 in an independent sample with several improvements. First, the sample size is larger, we assessed conception risk as a continuous, more valid indicator of fertility (Gangestad et al., 2016) and employed a tighter sampling schedule to capture more distinct hormonal profiles. Second, participants did more LH tests and results were sent to the study team via pictures, thereby reducing the chance of misinterpretation. Third, we trained our coders more thoroughly and improved coding instructions to increase inter-rater reliability. We decided to avoid coding categories that are difficult to code and show little variation,[7] and added more sensitive categories of wearing red clothes[8] to employ a continuous predictor. We also collected self-reports of clothing style. We decided to drop the unnecessary large amount of control variables that did not have a significant association with clothing style or grooming in Study 1. Our hypotheses are in line with those of Study 1 (though we acknowledge not finding much support for them in Study 1) and can be found in Supplementary Note 3.

Methods

This study was pre-registered online (https://osf.io/s2jgt/) before codings or analyses took place. Data collection procedures of the photographs were pre-registered separately, as part of a bigger project (https://osf.io/dwcsm/). Data, analysis code and material are available at https://osf.io/cfreh/. All participants signed a written consent form and the local ethics committee approved the study protocol.

Participants

A total of 257 heterosexual female participants (aged 18–35 years, $M = 23.2$, $SD = 3.3$), out of 282 recruited (for dropouts, see Supplementary Note 3), finished all sessions, and were therefore included for further analyses.[9] Pre-registered eligibility criteria were the same as in Study 1. Again, this sample size largely exceeds the size required to achieve 80% power as suggested by a power simulation for cycle research (Gangestad et al., 2016).

Procedure

Again, all participants took part in five individually scheduled sessions with procedures being similar to Study 1, except that these sessions took part across (mostly) one ovulatory cycle. All participants completed two sessions in their expected fertile phase and two sessions in their expected luteal phase (one session in the mid-luteal phase, one in the pre-menstrual phase). Scheduling was validated via LH test results and following up to the next menstrual onset. Details can be found in Supplementary Note 3. After completing a screening questionnaire for saliva sampling, participants self-reported grooming. Next, saliva samples and standardized pictures were taken as reported in Study 1. Again, as part of a

[7]We did not code transparency, as transparent clothes were extremely rare in Study 1. We also dropped variables that had very low reliabilities, including assess accentuation/tightness of clothes, effort of hairstyle, make-up usage, nail polish and heels.

[8]We focused on red/pink clothes, rather than other colours, make-up or nail polish. We added proportion of red/pink in clothing and intensity of red/pink as new categories and dropped the very complex coding scheme of Study 1 (separately coding multiple pieces of clothes).

[9]We have pre-registered a sample size of $N = 250$, to reach a sample of $n = 200$ participants for our conception risk analyses with women fulfilling all inclusion criteria. We decided to include all $N = 257$ participants to reach the pre-registered sample of exact $n = 200$ participants for our conception risk analyses.

larger project, participants completed other tasks that are not relevant for the current article and were published elsewhere (Stern et al., 2021). All tasks were randomized between participants and sessions. Upon completion of all sessions, participants received a payment of 60€ or course credit. Coding of photographs took place afterwards.

Measures

Conception risk

Participants' conception risk was assigned based on highly sensitive (10 mIU) LH test strips from MedNet GmbH. Participants started LH testing after menstruation and continued until a rise of LH (positive tests) was observed, and for a minimum of two days after the tests were negative again (as suggested by Roney, 2018). Participants were provided with at least 10 LH tests each and provided daily pictures of the tests to the investigators for confirmation. Results were used to allow flexible scheduling (see Supplementary Note 3). Following our pre-registration, we checked how many cycles were reported as irregular (i.e. >40 days, <20 days or a deviation of more than five days from participant's average cycle length) as these participants ($n = 57$; 22%) had to be excluded for conception risk analyses (see Supplementary Note 3). These numbers are comparable to or even lower than in previous cycle studies. Of the remaining $n = 200$ participants for conception risk analyses, 98 started testing in their luteal phase and 102 started fertile. All 257 women were included in the hormone analyses.

Hormone measures

Sampling methods, hormone transformations and analyses were identical to Study 1. Following Roney and Simmons (2013), we checked outliers >3 *SD*s from the mean before data transformation. Hence, we divided the cycle into three categories (the day of ovulation and up to 9 days before; the 10 days after the day of ovulation; all other days). This procedure helps to avoid the misidentification of phase-specific peaks as outliers (as progesterone levels are much higher in the mid-luteal phase). Then, we excluded outliers >3 *SD*s from the mean values of all phases ($n = 6$ progesterone levels >233.96 pg/mL = 0.7%; $n = 12$ estradiol levels >9.94 pg/mL = 1.17%; $n = 11$ E/P > 6.88 pg/mL = 1.27%).

Self-reported grooming

We assessed self-reported grooming via one item 'Today, I put effort into my outfit (clothes, make-up)' on a 5-point scale from 1 = *not at all applicable* to 5 = *very applicable*.

Control variables

Participant's relationship status was assessed as reported in Study 1. At the beginning of the study, 121 of the participants reported being partnered, 136 were single. Weight and height were assessed via self-report before participants took part in any testing session. Maximum temperature was retrieved as in Study 1. The weekday on which each testing session took place and was coded (1 = Monday, 2 = Tuesday, 3 = Wednesday, 4 = Thursday, 5 = Friday, 6 = Saturday, 7 = Sunday).

Coding

In line with Study 1, we developed three different coding sheets (see open material) to code skin display, grooming and clothing colour. Three independent research assistants coded the photographs. The procedure was generally similar to Study 1, but we increased the number of practice trials in the training session and gave more detailed instructions (see open material), which clearly improved inter-rater reliability to satisfying values for all categories (see Table S80). Further, we adapted the coding sheets for the category earrings, as well as the clothing colour coding sheet. All coding sheets are described in detail in the Supporting Information (Tables S74–S79).

Results

Main analyses

Data wrangling and analyses were done in line with Study 1. Neither conception risk, nor E/P was significantly related to skin display, in contrast to Hypotheses 1a and 1b (Table S81). Adding control variables to the model did not change the results (Table S85). There was a strong significant relationship between skin display and maximum temperature, in that participants wore more skin-revealing clothes when temperatures were higher. All other effects remained non-significant. Further, none of the outcomes of wearing red/pink were related to conception risk (Table S83), in contrast to all concerning hypotheses. Results remained virtually identical when adding control variables (Tables S89–S92).

Conception risk was significantly positively related to the grooming composite ($\beta = .08$, 95% CI = 0.03–0.12, $t = 3.22$, $p = .001$), supporting Hypothesis 2a, and suggesting more grooming when conception risk was higher, though the effect size was small. We found significant relationships between conception risk and wearing necklaces ($\beta = .08$, 95% CI = 0.02–0.13, $t = 2.78$, $p = .006$) and eyeglasses ($\beta = .03$, 95% CI = 0.00–0.06, $t = 2.28$, $p = .023$), suggesting that with higher conception risk, women are more likely to wear necklaces and less likely to wear eyeglasses,[10] in line with Hypotheses 2f and 2g, though the effect sizes were small and the link for eyeglasses did not remain significant after controlling for multiple testing. None of the other hypotheses were supported; results are displayed in Table S82. Results remained virtually identical when adding control variables (Tables S86–S88).

Exploratory analyses and robustness checks

Next, we investigated whether participant's self-reported grooming was associated with conception risk or ovarian hormone levels. Results suggest no significant relationships between self-reported grooming and either of our fertility measures (Table S84). Finally, we investigated whether effects reported earlier change when modelling E/P or estradiol and progesterone as separate predictors. Then, we repeated all of these analyses with log-transformed hormone levels. All results are reported in the Supporting Information (Tables S93–S102). Overall, the results reported earlier were supported by our robustness checks. There was one small significant association between skin display and log-transformed estradiol levels (Table S98). Log-transformed E/P was positively associated with the grooming composite and negatively with wearing eyeglasses (Table S99). Raw and log-transformed progesterone levels were negatively associated with the grooming composite and with wearing eyeglasses (Tables S95, S100). However, all effects were very small and not significant after controlling for multiple testing.

Discussion

In Study 3, we found slight evidence that higher conception risk is related to higher levels of grooming: participants wore necklaces more and eyeglasses less often when conception risk was higher. However, effects were not supported by hormonal links and not significant after controlling for multiple testing. Self-reported grooming was not related to any of the fertility variables. All reported associations between wearing red/pink or display of skin and any of the fertility variables were non-significant. While the null results of wearing red clothes are in line with results of Study 1 and previous studies (Blake,

[10] Please note that eyeglasses were coded as 1 = *yes* and 2 = *no*.

Dixson, et al., 2017; Hone & McCullough, 2020), this study did not replicate that women show less skin when fertile, as reported in Study 1. Neither did it replicate previous studies reporting that women wear sexier or more skin revealing clothes when fertile (in line with Arslan et al., 2021). The reported significant effects of the grooming variables were not found in Study 1. Given that all significant effect sizes were very small, not stable across all robustness checks and that the vast majority of reported findings were null findings, in our judgement, the current study provides no compelling evidence that women's clothing style and grooming shifts systematically across the ovulatory cycle. Further, effects of more grooming when conception risk was higher were not reflected by participant's self-reported grooming, in line with the results of Schleifenbaum et al. (2021).

The three studies reported so far overcame methodological limitations criticized in prior studies. However, one remaining limitation is that these studies involved pictures taken in a laboratory context. Women might be more likely to dress to impress across other contexts, when more mating opportunities are apparent, or might be restricted in their clothing choices by societal norms (e.g. not dressing too sexy when attending a lecture).

STUDY 4

This study focuses on whether women's clothing style changes across the ovulatory cycle when assessing clothing choices at social gatherings. We conducted a conceptual replication of the influential study by Durante et al. (2008), in which women drew outfits that they would like to wear to a party. Our pre-registered hypotheses closely follow the reported effects of the original study: naturally cycling women with higher conception risk (H1a) and with higher estradiol and lower progesterone levels (H1b) prefer to wear more *revealing* clothes to a social event. Associations should be stronger for singles (H1c). We predict the exact same effects for wearing more *sexy* clothes (H2a, H2b, H2c respectively). In addition, we investigated whether cycle variables are related to preferring to wear more red clothes in an exploratory manner.

Methods

This study was pre-registered at the Open Science Framework (https://osf.io/p89ty/). Again data, analysis code and other materials are also available online (https://osf.io/cfreh/).

Participants and procedure

Participants were the same as in Study 3. This sample size greatly exceeds the sample size of Durante et al. (2008, with $N=88$ across two sessions). In each testing session, participants were given coloured pencils and a sheet of paper displaying an outline of a woman's figure, on which the participants were asked to draw outfits they would wear to a social gathering that same night (a party with a lot of attractive singles; Figure S4). This procedure was directly adapted from previous studies (Durante et al., 2008; Hone & McCullough, 2020). We excluded illustrations from one participant because of insufficient quality (drawing was so light that clothes were hardly recognizable), resulting in a total of 1024 drawings from 256 participants. Figure 1 shows examples of drawings from two participants.

Measures

Hormone levels, conception risk and relationship status were assessed as described in Study 3.

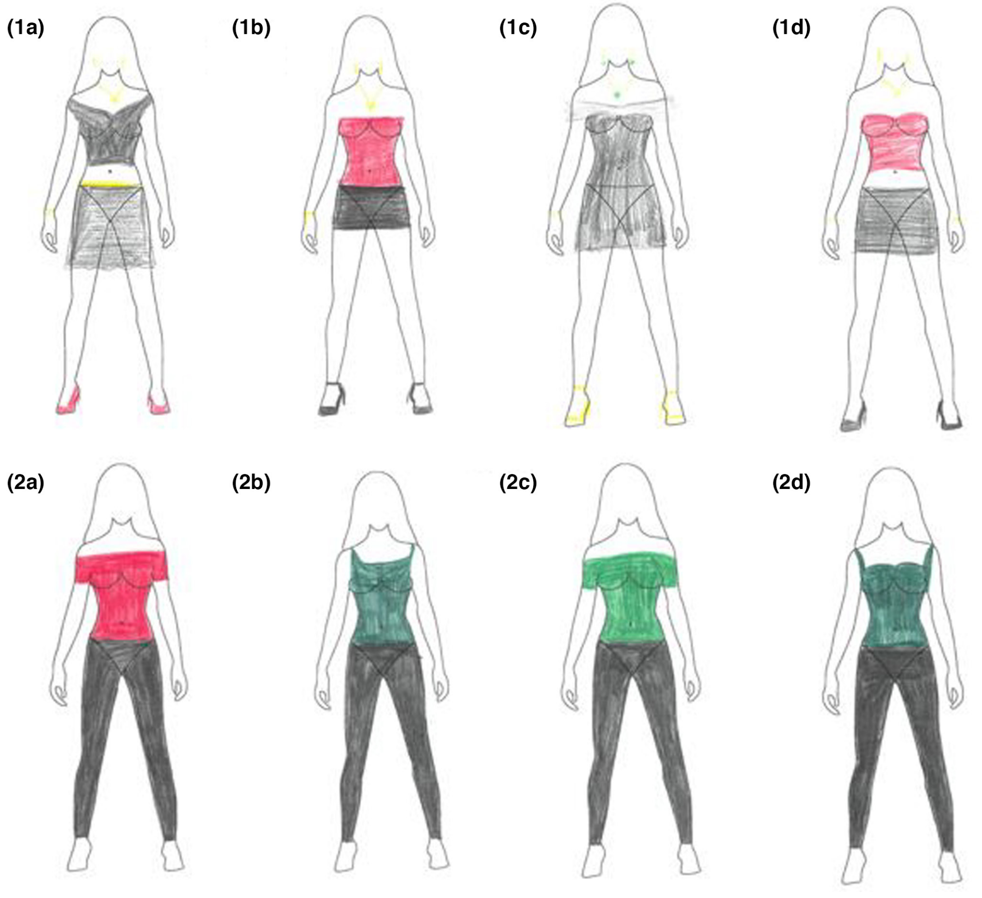

FIGURE 1 Examples of drawings from two participants. *Note*: All four drawings by one participant are displayed next to each other in the same row. For both participants, session (a) was pre-menstrual, sessions (b) and (c) were in the fertile phase, session (d) was in the mid-luteal phase.

Ratings

We collected ratings on how revealing and how sexy the illustrated clothes were. For this purpose, we divided the 1024 drawings into four ratings sets, each containing 256 illustrations, one from each participant. Raters were randomly assigned to one of the rating sets, each rater was presented 128 randomly selected illustrations of the respective set in a randomized order via Alfred3 (Treffenstaedt et al., 2021). Due to technical errors, we collected data from a total of $N = 215$ raters (78 men, 137 women; mean age = 23.28 years, $SD = 4.01$), in contrast to the pre-registered $N = 160$ raters. To still follow our pre-registered plan, we decided to randomly select $n = 20$ raters for each illustration and disregard the remaining ratings. After reading the instructions and giving consent to the GDPR-based data protection information, ratings were collected on 9-point scales (1 = *not at all revealing/sexy*, 9 = *extremely revealing/sexy*). Intra-class correlations (1, k; Shrout & Fleiss, 1979) were good (sexiness ICC = .80, revealing ICC = .91), ratings were averaged to form a revealing and a sexiness score respectively.

Automated analyses

Durante et al. (2008) analysed the amount of skin revealed by counting the number of squares on a 1-mm sheet that contained exposed skin. We decided to run automated analyses as a more objective measure, to avoid mistakes, and for economic reasons. We scanned all illustrations and edited them so that all drawn clothes were completely coloured black, to make sure that the computer program does not falsely identify white clothes or extremely light drawings as showing a large amount of skin (see Figure S5 for an example). Then, automated analyses were done using the software R, details can be found in Supplementary Note 4. We also ran automated analyses using the unedited versions, to extract whether women drew red/pink clothes.

Results

In contrast to Hypotheses 1a, 1b and 1c, we did not observe a significant association of conception risk, E/P, estradiol or progesterone or their interaction with relationship status and drawing revealing or sexy clothes (Tables S103–S105). Next, we investigated whether participant's conception risk and hormone levels were associated with drawing more skin revealing or more red clothes (automatically extracted). Again, none of the analyses revealed any significant results (Tables S106–S109). As skin-revealing and sexiness ratings were substantially correlated ($r = .73$), we repeated our main analyses with a composite score of both ratings to test the robustness of our results. All results were non-significant (Tables S110, S111). Finally, we repeated all hormone models with log-transformed hormone levels. Results were virtually identical (Tables S112–S121).

Discussion

We did not replicate the results reported by Durante et al. (2008). We found no compelling evidence that women with higher conception risk (or higher estradiol and lower progesterone levels) aim to wear sexier or more revealing clothes to a social event. Rather, as can be seen in Figure 1, participants seemed to show little variation from their personally preferred clothing style, as outfits drawn by one participant often showed high amounts of similarity. The results of this study suggest that the null results reported in Study 3 cannot be explained by restrictions in clothing choice due to the laboratory context. Nevertheless, another explanation remains, in that some previous studies suggest that cycle shifts in sexy and revealing clothes might be implicitly measurable as changes in the *desire* for wearing or spending money on sexier and revealing clothes when fertile (Blake, Bastian, et al., 2017; Durante et al., 2011).

STUDY 5

This study examines whether women's willingness to spend money for sexy clothes changes across the ovulatory cycle as a conceptual replication of the study by Blake, Bastian, et al. (2017). Our pre-registration of a larger data collection covers the sample characteristics and the procedure (https://osf.io/dwcsm/), but not hypotheses or analyses for this study. Thus, this study should be interpreted as exploratory. Our analyses follow the logic of our previous studies, and the analyses reported by Blake, Bastian, et al. (2017). Data, analysis code and material are available at https://osf.io/cfreh/.

Methods

Participants and procedure

The sample was the same as in Studies 3 and 4. This sample size greatly exceeds the size of previous studies investigating cycle shifts in interest in buying sexy clothes (Blake, Bastian, et al., 2017: $N = 98$

participants; Durante et al., 2011: *N*s = 60 and 48 in their Studies 1 and 2 respectively). In each of the testing sessions, participants rated pictures of 40 outfits in a randomized order. The pictures were selected in a pre-test, in which *n* = 14 raters rated 100 modern outfits from a shopping website (faces of the models concealed) on the item 'How sexy is this outfit?' (1 = *very unsexy*, 7 = *very sexy*). Intra-class correlations (1, k) were good (ICCs = .86), ratings were averaged and we chose the 20 pictures with the highest (i.e. *sexy*, coded = 1), and 20 pictures with the lowest (i.e. *non-sexy*, coded = 0) mean ratings.

Measures

Hormone levels, conception risk and relationship status were assessed as described in Study 3.

Interest in buying sexy clothes

We assessed how much money participants would be willing to spend on each outfit on a 10-point scale (from 1 = *20€* to 10 = *200€*, *M* = 3.32, *SD* = 1.98).

Control variables

Following Blake, Bastian, et al. (2017), we assessed women's self-perceived mate value and socio-sexual orientation as control variables via an online survey (presented with formr; Arslan, Walther, & Tata, 2020) that participants filled out before their first testing session. Self-perceived mate value was assessed via three items adapted from Landolt et al. (1995) that were answered on a 5-point scale (1 = *strongly disagree*, 5 = *strongly agree*). Cronbach's alpha was good ($\alpha = .88$), all three items were averaged and *z*-scored. Women's socio-sexual orientation was assessed via the SOI-R (Penke & Asendorpf, 2008). All nine items were averaged and *z*-scored ($\alpha = .85$).

Results

We modelled a multi-level model with willingness to spend as an outcome variable, conception risk, sexy clothing and their interaction as fixed effects and random slopes. There was a small significant main effect for sexiness of clothes, suggesting that participants were willing to spend more money on sexy clothes as compared to non-sexy clothes. Neither the main effect of conception risk nor the interaction between conception risk and sexiness of clothes were significant, and effect sizes were very close to zero (Table S122). Results remained virtually identical when adding control variables (Table S123). Next, we tested women's willingness to spend in relation to ovarian hormone levels. None of the models revealed a significant interaction effect. There were no significant main effects of either hormone, the main effect of sexiness was only significant in one model. Results remained virtually identical when adding the control variables (Tables S124–S126). We repeated all hormone models with log-transformed hormone levels (Tables S127–S129).

Discussion

In this study, we did not find supporting evidence that fertile women are willing to spend more money for sexy versus non-sexy clothes. While in line with the overall pattern of results reported in this article, the results are only partly in line with those reported by Blake, Bastian, et al. (2017), who found that interest in buying sexy clothes was negatively associated with progesterone, but not associated with estradiol or estimated cycle phase. They are further in contrast to findings by Durante et al. (2011), who reported that women chose to buy a greater percentage of sexy clothes near ovulation (though this effect was only significant when women were primed with an attractive woman).

GENERAL DISCUSSION

Across five studies, we aimed to investigate whether women's clothing style or grooming change across the ovulatory cycle as a function of fluctuating levels in ovarian hormones. Our studies addressed different contexts, assessed self-reports, other perceptions as well as objective codings, and took a number of potentially confounding variables into account. In addition to testing our pre-registered hypotheses, we ran a large number of exploratory analyses and robustness checks to scrutinize our results. The overall pattern of results suggests no compelling evidence for cycle shifts in or ovarian hormonal associations with changes in women's clothing style or grooming. There were some exceptions that are discussed next.

Wearing sexy/revealing clothes

We found slight evidence for an effect in the opposite direction as expected: women wore more skin-revealing clothes in the luteal phase. There are several explanations for this finding: higher basal temperature when in the luteal phase (Lee et al., 2014), less risky behaviour when fertile (Bröder & Hohmann, 2003; Chavanne & Gallup Jr, 1998) and stronger protection against intra-sexual aggression (Krems et al., 2021). Given that these effects were not robust and not replicated in the other studies, we refrain from further interpretation before replication in independent samples. Nevertheless, our results are in contrast to previous studies reporting that fertile women wear more, aim to wear more or spend more money on sexy and skin-revealing clothes (Blake, Bastian, et al., 2017; Durante et al., 2008, 2011; Haselton et al., 2007; Saad & Stenstrom, 2012; Schwarz & Hassebrauck, 2008). Rather, they are in line with a recent study also reporting no compelling evidence for these cycle shifts (Arslan et al., 2021). Our findings suggest that other variables are more important predictors of wearing sexier clothes, as the strongest indicator for displaying more skin was the outside temperature.

Wearing red/pink clothes

We did not find any support that fertile women wear more red or pink clothes. Previous research on the red effect provided fairly mixed findings. Given that we controlled for potential confounding variables, and investigated a wide number of different fertility indicators, we think that the explanation that cycle shifts in red clothes can only be found under specific conditions (e.g. only on cold days, Tracy & Beall, 2014) is unlikely. Further, it has been reported that women wearing red are perceived as being more sexually receptive by other women and are thus more often subject to intra-sexual aggression (Pazda et al., 2014). Female intra-sexual aggression can be costly, negatively affecting the ability to conceive (Huchard & Cowlishaw, 2011), and has even been discussed as a major reason for the evolution of concealed ovulation in human females (Krems et al., 2021). Thus, costs of potentially higher intra-sexual aggression may also prevent women from wearing more red or pink clothes when fertile.

Grooming

Women often use cosmetics or jewellery, or spend more time and effort on grooming to enhance their attractiveness for desirable mates (Durante et al., 2011), especially when conception is possible (Haselton et al., 2007). Previous studies reported evidence in line with this assumption (Röder et al., 2009; Saad & Stenstrom, 2012). However, a recent large-scale study did not replicate this effect (Schleifenbaum et al., 2021). Yet, previous studies have not investigated cycle shifts in grooming beyond self-reports. The evidence for enhanced grooming when fertile was mixed across the studies reported in the current article. We found evidence that women wear necklaces more but eyeglasses less often when conception risk was higher in Study 3. While necklaces might be one strategy to enhance attractiveness, a well-known stereotype is that wearing eyeglasses decreases facial attractiveness (Leder et al., 2011). However, the latter effect was not significant after controlling for multiple testing. Both effects were not found in Study 1, not significantly linked to fluctuating hormone levels across most analyses and effect sizes were very small. All other analyses involving efforts of grooming were non-significant with effect sizes very close to zero. We consider it unlikely that wearing necklaces more but eyeglasses less often evolved as strategies for attracting mates, while other grooming behaviour did not. Wearing jewellery (including necklaces) was not found among the grooming behaviours that increased in the fertile phase in a previous study (Saad & Stenstrom, 2012). In our human history, eye-sight correcting glasses are very novel and a large number of women does not own any eyeglasses. The functional mechanism that would lead women to pursue exactly these two strategies to enhance their own attractiveness when fertile remains unclear. Overall, we do not interpret the evidence for enhanced grooming when fertile as compelling.

Is ovulation concealed in human females?

Previously, it was argued that ovulation in human females is relatively concealed or non-advertised, as they do not display overt cues to fertility. Nevertheless, subtle cues to ovulation, albeit small and maybe even only non-adaptive side effects of cycle changes (Gangestad & Haselton, 2015), might still exist, potentially detectable by others and influencing social behaviour (Haselton & Gildersleeve, 2011). Here, in line with the most recent studies investigating cycle shifts in clothing style and grooming (Arslan et al., 2021; Blake, Bastian, et al., 2017; Hone & McCullough, 2020; Schleifenbaum et al., 2021), we did not find supporting evidence for detectable cues to fertility. However, it might be premature to conclude that 'leaky cues' do not exist. All studies entail limitations and we welcome replication studies. Nevertheless, we think that the overall pattern of published results within recent years casts doubt on previous evidence that was in favour of cycle shifts in clothing style and grooming, especially given that newer studies had higher test power and superior methods. At least, we can conclude that cycle shifts in clothing style and grooming, if existent, are much smaller and harder to detect than previously assumed. Effects for other, more passively occurring shifts in, for example, body odour, voice pitch or facial cues could exist independently, but recent evidence also casts doubt on strong, perceivable cues to fertility in other traits (Bleske-Rechek et al., 2011; Catena et al., 2019; Jones, Hahn, Fisher, Wang, Kandrik, & DeBruine, 2018; Jones, Hahn, Fisher, Wang, Kandrik, Lao, et al., 2018 Mei et al., 2022; Schleifenbaum et al., 2022). Overall, it currently seems as if cues to fertility might be too subtle to be perceivable (see also Burriss et al., 2015).

Explanations for mixed findings

Overall, there has been large inconsistency in findings supporting cycle shifts in clothing style and grooming, or not finding compelling evidence for these shifts. One likely explanation is the large differences in used methods and designs between different studies, including but not restricted to cycle phase validation methods, hormone measures, self-reports versus codings, within- versus between-subject designs, test power, analytical choices and operationalizations of outcomes measured. Another explanation might be that effects are very small and hard to detect, and at the same time local and specific. More precisely, what is perceived as more attractive (and also women's tactics to enhance their attractiveness) may be influenced by local cultural norms and vary between different social groups. If this is true, it might explain why some studies find effects that others cannot replicate. A third explanation, partly related to the previous one, is that if the effect exists, we (but also other studies) did probably not measure it properly. Rather than affecting specific tactics to enhance one's own attractiveness that might vary within and between women (e.g. wearing more jewellery or specific clothing colours), the effect may rather be in the overall resulting appearance, that is global attractiveness.

Thus, if cultural norms of attractiveness are not well reflected in the measures of specific attractiveness enhancement tactics, focusing on specific tactics rather than the perceived overall attractiveness might lead to replicability issues.

Implications and directions for future research

Future research will further clarify if there are any cues to fertility in human females and, if existing, whether they are perceivable by others, whether they reflect specific tactics or rather global attractiveness and whether there are local or cross-cultural differences in women displaying them. Every new dataset contributes valuable new information and should be used to re-evaluate the overall picture. Meta-analyses on this topic might be especially insightful. Further, we think it is crucial to empirically test different theories explaining why concealed ovulation evolved. While it has previously been assumed that human females benefit from concealed ovulation to increase the opportunity to obtain resources and investment from male partners, a recent study suggests that concealed ovulation might have evolved to avoid aggression from other females when fertile (Krems et al., 2021). To be able to contribute to an understanding of human female reproductive physiology and psychology, future studies should be designed and powered to discriminate between these and other theories.

We found support for environmental circumstances that influence clothing and grooming behaviour (especially outside temperature). It seems plausible that these variables are influenced by other social and environmental circumstances that might have a larger effect than hormonal changes across the cycle. For example, we tend to wear different clothes when we stay at home, compared to attending a job interview or a wedding. What we wear is restricted by our wardrobe and what is sold in shops is further influenced by current fashion and trends. Future research may investigate factors that explain individual differences in clothing choice and grooming.

Limitations

Although current studies were able to overcome most of the methodological criticism raised against previous studies, there are limitations that need to be addressed in future studies. First, some of the inter-rater reliabilities in the first two studies were very low, raising concerns regarding the measure and indicating caution for the interpretation of the related findings. We tried to overcome this issue by not interpreting coded categories with low inter-rater agreement (e.g. wearing heels) and replicated our null results in Study 3 with higher inter-rater reliability. Still, these specific findings are in need for replication. Second, the samples of the five studies were not fully independent and consisted of only two independent target samples. Thus, peculiarities of each sample (e.g. sample specific measurement error) potentially affect the results of multiple studies. Third, the item used in Study 2 measured the effort women spend in looking more attractive, but an item assessing global attractiveness, rather than the effort that leaves more room for judge's individual interpretations, would have been preferable. Fourth, our Studies 4 and 5 may still lack ecological validity as compared to real field studies. Fifth, although we did not inform the participants about the purpose of our studies, they might have guessed the reasons why they were photographed or had to draw outfits, potentially leading to a reactivity bias influencing the measured outcome variables. Note that this limitation also pertains to earlier studies that found significant results (Durante et al., 2008; Eisenbruch et al., 2015; Haselton et al., 2007). Future studies may employ cover stories to solve this issue. Sixth, the results of the drawing task in Study 4 might be dependent on participant's ability to draw. Future studies may employ different tasks to overcome this issue. Seventh, recent research recommends daily hormone assays to validate fertile window estimates as more reliable than LH tests alone (Marcinkowska, 2020) to increase the validity of conception risk measures. Eighth, recent work has pointed out that estradiol levels assessed with salivary immunoassays may not correspond to conception risk or show the expected peak in the fertile phase (Arslan et al., 2023). While our measure of progesterone performed better as compared to our estradiol measure, E/P was also strongly related to our conception risk measure which was validated with highly sensitive LH tests (Stern et al., 2021). However, we cannot rule out that analysing estradiol with a different method (e.g. analytic protocols for LCMS that were not available when our hormone assays were analysed), would have led to different results. Finally, we collected our data in a WEIRD country (Henrich et al., 2010) and are not able to generalize our results cross-culturally.

CONCLUSION

Across five studies, we aimed to investigate whether women show systematic changes in their clothing style and grooming across their ovulatory cycle. Overall, we found no compelling evidence for links between clothing style or grooming and conception risk or ovarian hormone levels. These results are in line with other recent non-replications, but in contrast to earlier studies reporting that women try to enhance their appearance when fertile. While we cannot answer the question whether ovulation is fully concealed in human females with the studies reported in this article alone, we contribute to an overall picture: the current evidence is insufficient to support the conclusion that there are subtle, perceivable cues to fertility that could lead to changes in social behaviour. Future studies should test competing theories that explain why concealed or non-advertised ovulation in human females evolved.

AUTHOR CONTRIBUTIONS

Julia Stern: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; software; validation; visualization; writing – original draft; writing – review and editing. **Sabine Ostermann:** Data curation; formal analysis; validation; writing – review and editing. **Lars Penke:** Conceptualization; funding acquisition; methodology; resources; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 254142454/GRK 2070. We thank our research assistants and other students who worked in this project: Paula Bange, Laura Botzet, Silvia Bradatsch, Sarah Forsthoff, Isa Garbisch, Kim Gloystein, Salome Johannigmann, Indra Kirchberg, Laila Knapp, Paulina Lauer, Laura Specker-Mattißen, Stephanie Rudolph and Annelie Strachotta for data collection, helping with the pre-registrations or literature research. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interests.

OPEN RESEARCH BADGES

This article has earned Open Data, Open Materials and Preregistered Research Design badges. Data, materials and the preregistered design and analysis plan are available at https://osf.io/cfreh/; Study 1: https://osf.io/c7sgv/?view_only=50bb191885124db9b0047e0b59d9c500. Study 2: https://osf.io/49sp7/?view_only=1029973328ef4521b5aac832c38ad276. Study 3: https://osf.io/s2jgt/?view_only=1fcc37e678d54896bb4841236bc1188d. Study 4: https://osf.io/p89ty/?view_only=8e42030e32894de3b46558f0531c06c4. Study 5: https://osf.io/dwcsm/?view_only=e042065672934440950259c86a68eb1b.

DATA AVAILABILITY STATEMENT

All data, analysis code and material (where possible) have been made publicly available at the OSF and can be accessed at https://osf.io/cfreh/. Links to the pre-registrations for the studies can be found in the main text.

ORCID

Julia Stern https://orcid.org/0000-0001-8749-6392

REFERENCES

Alexander, R. D., & Noonan, K. M. (1979). Concealment of ovulation, paternal care, and human social evolution. In N. A. Chagnon & W. G. Irons (Eds.), *Evolutionary biology and human social behavior: An anthropological perspective* (pp. 436–453). Duxbury.

Arslan, R. C., Blake, K., Botzet, L. J., Bürkner, P.-C., DeBruine, L., Fiers, T., Grebe, N., Hahn, A., Jones, B. C., Marcinkowska, U. M., Mumford, S. L., Penke, L., Roney, J., Schisterman, E., & Stern, J. (2023). Not within spitting distance: Salivary immunoassays of estradiol have subpar validity for predicting cycle phase. *Psychoneuroendocrinology*, *149*, 105994. https://doi.org/10.1016/j.psyneuen.2022.105994

Arslan, R. C., Driebe, J. C., Stern, J., Gerlach, T. M., Ostner, J., & Penke, L. (2020). Goettingen Ovulatory Cycle Diaries 2. https://doi.org/10.17605/OSF.IO/D3AVF

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

Arslan, R. C., Walther, M. P., & Tata, C. S. (2020). Formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*, *52*, 376–387. https://doi.org/10.3758/s13428-019-01236-y

Back, M. D., Küfner, A. C., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*, 1013–1037. https://doi.org/10.1037/a0034431

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

Beall, A. T., & Tracy, J. L. (2013). Women are more likely to wear red or pink at peak fertility. *Psychological Science*, *24*, 1837–1841.

Blake, K. R., Bastian, B., O'Dean, S. M., & Denson, T. F. (2017). High estradiol and low progesterone are associated with high assertiveness in women. *Psychoneuroendocrinology*, *75*, 91–99.

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

Blake, K. R., Dixson, B. J., O'Dean, S. M., & Denson, T. F. (2017). No compelling positive association between ovarian hormones and wearing red clothing when using multinomial analyses. *Hormones and Behavior*, *90*, 129–135. https://doi.org/10.1016/j.yhbeh.2017.03.005

Bleske-Rechek, A., Harris, H. D., Denkinger, K., Webb, R. M., Erickson, L., & Nelson, L. A. (2011). Physical cues of ovulatory status: A failure to replicate enhanced facial attractiveness and reduced waist-to-hip ratio at high fertility. *Evolutionary Psychology*, *9*, 336–353. https://doi.org/10.1177/147470491100900306

Bröder, A., & Hohmann, N. (2003). Variations in risk taking behavior over the menstrual cycle: An improved replication. *Evolution and Human Behavior*, *24*(6), 391–398. https://doi.org/10.1016/S1090-5138(03)00055-2

Burriss, R. P., Troscianko, J., Lovell, P. G., Fulford, A. J., Stevens, M., Quigley, R., Payne, J., Saxton, T. K., & Rowland, H. M. (2015). Changes in women's facial skin color over the ovulatory cycle are not detectable by the human visual system. *PLoS One*, *10*, e0130093. https://doi.org/10.1371/journal.pone.0130093

Catena, T. M., Simmons, Z. L., & Roney, J. R. (2019). Do women's faces become more attractive near ovulation? *Hormones and Behavior*, *115*, 104560. https://doi.org/10.1016/j.yhbeh.2019.07.008

Chavanne, T. J., & Gallup, G. G., Jr. (1998). Variation in risk taking behavior among female college students as a function of the menstrual cycle. *Evolution and Human Behavior*, *19*(1), 27–32. https://doi.org/10.1016/S1090-5138(98)00016-6

Deschner, T., Heistermann, M., Hodges, K., & Boesch, C. (2003). Timing and probability of ovulation in relation to sex skin swelling in wild West African chimpanzees, *Pan troglodytes verus*. *Animal Behaviour*, *66*, 551–560. https://doi.org/10.1006/anbe.2003.2210

Durante, K. M., Griskevicius, V., Cantu, S. M., & Simpson, J. A. (2014). Money, status, and the ovulatory cycle. *Journal of Marketing Research*, *51*(1), 27–39. https://doi.org/10.1509/2Fjmr.11.0327

Durante, K. M., Griskevicius, V., Hill, S. E., Perilloux, C., & Li, N. P. (2011). Ovulation, female competition, and product choice: Hormonal influences on consumer behavior. *Journal of Consumer Research*, *37*, 921–934. https://doi.org/10.1086/656575

Durante, K. M., Li, N. P., & Haselton, M. G. (2008). Changes in women's choice of dress across the ovulatory cycle: Naturalistic and laboratory task-based evidence. *Personality and Social Psychology Bulletin*, *34*, 1451–1460. https://doi.org/10.1177/0146167208323103

Eisenbruch, A. B., Simmons, Z. L., & Roney, J. R. (2015). Lady in red: Hormonal predictors of women's clothing choices. *Psychological Science*, *26*, 1332–1338. https://doi.org/10.1177/0956797615586403

Elliot, A. J., & Niesta, D. (2008). Romantic red: Red enhances men's attraction to women. *Journal of Personality and Social Psychology*, *95*, 1150–1164. https://psycnet.apa.org/doi/10.1037/0022-3514.95.5.1150

Elliot, A. J., Tracy, J. L., Pazda, A. D., & Beall, A. T. (2013). Red enhances women's attractiveness to men: First evidence suggesting universality. *Journal of Experimental Social Psychology*, *49*, 165–168. https://doi.org/10.1016/j.jesp.2012.07.017

Gangestad, S. W., Garver-Apgar, C. E., Cousins, A. J., & Thornhill, R. (2014). Intersexual conflict across women's ovulatory cycle. *Evolution and Human Behavior*, *35*, 302–308. https://doi.org/10.1016/j.evolhumbehav.2014.02.012

Gangestad, S. W., & Haselton, M. G. (2015). Human estrus: Implications for relationship science. *Current Opinion in Psychology*, *1*, 45–51. https://doi.org/10.1016/j.copsyc.2014.12.007

Gangestad, S. W., Haselton, M. G., Welling, L. L., Gildersleeve, K., Pillsworth, E. G., Burriss, R. P., Larson, C. M., & 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

Gangestad, S. W., Thornhill, R., & Garver, C. E. (2002). Changes in women's sexual interests and their partner's mate–retention tactics across the menstrual cycle: Evidence for shifting conflicts of interest. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*, 975–982. https://doi.org/10.1098/rspb.2001.1952

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.

Gildersleeve, K. A., Haselton, M. G., Larson, C. M., & Pillsworth, E. G. (2012). Body odor attractiveness as a cue of impending ovulation in women: Evidence from a study using hormone-confirmed ovulation. *Hormones and Behavior*, *61*, 157–166. https://doi.org/10.1016/j.yhbeh.2011.11.005

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, *150*(4), 700–709. https://psycnet.apa.org/doi/10.1037/xge0000920

Haselton, M. G., & Gangestad, S. W. (2006). Conditional expression of women's desires and men's mate guarding across the ovulatory cycle. *Hormones and Behavior*, *49*, 509–518. https://doi.org/10.1016/j.yhbeh.2005.10.006

Haselton, M. G., & Gildersleeve, K. (2011). Can men detect ovulation? *Current Directions in Psychological Science*, *20*, 87–92. https://doi.org/10.1177/2F0963721411402668

Haselton, M. G., Mortezaie, M., Pillsworth, E. G., Bleske-Rechek, A., & Frederick, D. A. (2007). Ovulatory shifts in human female ornamentation: Near ovulation, women dress to impress. *Hormones and Behavior*, *51*, 40–45. https://doi.org/10.1016/j.yhbeh.2006.07.007

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29. https://doi.org/10.1038/466029a

Hone, L. S., & McCullough, M. E. (2020). Are women more likely to wear red and pink at peak fertility? What about on cold days? Conceptual, close, and extended replications with novel clothing colour measures. *British Journal of Social Psychology*, *59*, 945–964. https://doi.org/10.1111/bjso.12371

Huchard, E., & Cowlishaw, G. (2011). Female–female aggression around mating: An extra cost of sociality in a multimale primate society. *Behavioral Ecology*, *22*(5), 1003–1011. https://doi.org/10.1093/beheco/arr083

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., & DeBruine, L. M. (2018). General sexual desire, but not desire for uncommitted sexual relationships, tracks changes in women's hormonal status. *Psychoneuroendocrinology*, *88*, 153–157. https://doi.org/10.1101/155788

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Lao, J., Han, C., Lee, A. J., Holzleitner, I. J., & DeBruine, L. M. (2018). No compelling evidence that more physically attractive young adult women have higher estradiol or progesterone. *Psychoneuroendocrinology*, *98*, 1–5. https://doi.org/10.1016/j.psyneuen.2018.07.026

Krems, J. A., Claessens, S., Fales, M. R., Campenni, M., Haselton, M. G., & Aktipis, A. (2021). An agent-based model of the female rivalry hypothesis for concealed ovulation in humans. *Nature Human Behaviour*, *5*, 726–735. https://doi.org/10.1038/s41562-020-01038-9

Landolt, M. A., Lalumière, M. L., & Quinsey, V. L. (1995). Sex differences in intra-sex variations in human mating tactics: An evolutionary approach. *Ethology and Sociobiology*, *16*, 3–23. https://doi.org/10.1016/0162-3095(94)00012-V

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Validity and psychometric equivalence of the German version of the big five inventory in young, middle-aged and old adults. *Diagnostica*, *73*, 111–121.

Leder, H., Forster, M., & Gerger, G. (2011). The eyeglasses stereotype revisited: Effects of eyeglasses on perception, recognition, and impression of faces. *Swiss Journal of Psychology*, *70*(4), 211–222. https://doi.org/10.1024/1421-0185/a000059

Lee, H., Petrofsky, J., Shah, N., Awali, A., Shah, K., Alotaibi, M., & Yim, J. (2014). Higher sweating rate and skin blood flow during the luteal phase of the menstrual cycle. *The Tohoku Journal of Experimental Medicine*, *234*(2), 117–122. https://doi.org/10.1620/tjem.234.117

Lobmaier, J. S., Fischbacher, U., Wirthmüller, U., & Knoch, D. (2018). The scent of attractiveness: Levels of reproductive hormones explain individual differences in women's body odour. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20181520. https://doi.org/10.1098/rspb.2018.1520

Marcinkowska, U. M. (2020). Importance of daily sex hormone measurements within the menstrual cycle for fertility estimates in cyclical shifts studies. *Evolutionary Psychology*, *18*, 1474704919897913. https://doi.org/10.1177/1474704919897913

Mei, M., Grillot, R. L., Abbey, C. K., Emery Thompson, M., & Roney, J. R. (2022). Does scent attractiveness reveal women's ovulatory timing? Evidence from signal detection analyses and endocrine predictors of odour attractiveness. *Proceedings of the Royal Society B*, *289*(1970), 20220026. https://doi.org/10.1098/rspb.2022.0026

Miller, G., Tybur, J. M., & Jordan, B. D. (2007). Ovulatory cycle effects on tip earnings by lap dancers: Economic evidence for human estrus? *Evolution and Human Behavior*, *28*, 375–381. https://doi.org/10.1016/j.evolhumbehav.2007.06.002

Pawlowski, B. (2015). Ovulation, concealed. In P. Whelehan & A. Bolin (Eds.), *The international encyclopedia of human sexuality* (pp. 831–860). Wiley-Blackwell.

Pazda, A. D., Prokop, P., & Elliot, A. J. (2014). Red and romantic rivalry: Viewing another woman in red increases perceptions of sexual receptivity, derogation, and intentions to mate-guard. *Personality and Social Psychology Bulletin*, *40*, 1260–1269. https://doi.org/10.1177/2F0146167214539709

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*, 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Peperkoorn, L. S., Roberts, S. C., & Pollet, T. V. (2016). Revisiting the red effect on attractiveness and sexual receptivity: No effect of the color red on human mate preferences. *Evolutionary Psychology*, *14*, 1474704916673841. https://doi.org/10.1177/1474704916673841

Pipitone, R. N., & Gallup, G. G., Jr. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*, *29*, 268–274. https://doi.org/10.1016/j.evolhumbehav.2008.02.001

Prokop, P., & Hromada, M. (2013). Women use red in order to attract mates. *Ethology*, *119*, 605–613. https://doi.org/10.1111/eth.12102

Puts, D. A., Bailey, D. H., Cárdenas, R. A., Burriss, R. P., Welling, L. L., Wheatley, J. R., & Dawood, K. (2013). Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle. *Hormones and Behavior*, *63*, 13–19. https://doi.org/10.1016/j.yhbeh.2012.11.007

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C., Perrett, D. I., & Petrie, M. (2004). Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*, S270–S272. https://doi.org/10.1098/rsbl.2004.0174

Röder, S., Brewer, G., & Fink, B. (2009). Menstrual cycle shifts in women's self-perception and motivation: A daily report method. *Personality and Individual Differences*, *47*, 616–619. https://doi.org/10.1016/j.paid.2009.05.019

Roney, J. R. (2016). Theoretical frameworks for human behavioral endocrinology. *Hormones and Behavior*, *84*, 97–110. https://doi.org/10.1016/j.yhbeh.2016.06.004

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

Roney, J. R. (2019). On the use of log transformations when testing hormonal predictors of cycle phase shifts: Commentary on Gangestad, Dinh, Grebe, Del Giudice and Emery Thompson (2019). *Evolution and Human Behavior*, *40*, 526–530.

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

Roney, J. R., & Simmons, Z. L. (2016). Within-cycle fluctuations in progesterone negatively predict changes in both in-pair and extra-pair desire among partnered women. *Hormones and Behavior*, *81*, 45–52. https://doi.org/10.1016/j.yhbeh.2016.03.008

Saad, G., & Stenstrom, E. (2012). Calories, beauty, and ovulation: The effects of the menstrual cycle on food and appearance-related consumption. *Journal of Consumer Psychology*, *22*, 102–113. https://doi.org/10.1016/j.jcps.2011.10.001

Schleifenbaum, L., Driebe, J. C., Gerlach, T. M., Penke, L., & Arslan, R. C. (2021). Women feel more attractive before ovulation: Evidence from a large-scale online diary study. *Evolutionary Human Sciences*, *3*, e47. https://doi.org/10.1017/ehs.2021.44

Schleifenbaum, L., Stern, J., Driebe, J. C., Wieczorek, L. L., Gerlach, T. M., Arslan, R. C., & Penke, L. (2022). Men are not aware of and do not respond to their female partner's fertility status: Evidence from a dyadic diary study of 384 couples. *Hormones and Behavior*, *143*, 105202. https://doi.org/10.1016/j.yhbeh.2022.105202

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). Guilford Press.

Schwarz, S., & Hassebrauck, M. (2008). Self-perceived and observed variations in women's attractiveness throughout the menstrual cycle—a diary study. *Evolution and Human Behavior*, *29*(4), 282–288. https://doi.org/10.1016/j.evolhumbehav.2008.02.003

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Stern, J., Gerlach, T. M., & Penke, L. (2020). Probing ovulatory cycle shifts in women's mate preferences for men's behaviors. *Psychological Science*, *31*, 424–436. https://doi.org/10.1177/0956797619882022

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

Thornhill, R., & Gangestad, S. W. (2008). *The evolutionary biology in human female sexuality*. Oxford University Press.

Tracy, J. L., & Beall, A. T. (2014). The impact of weather on women's tendency to wear red or pink when at high risk for conception. *PLoS One*, *9*, e88852. https://doi.org/10.1371/journal.pone.0088852

Treffenstaedt, C., Brachem, J., & Wiemann, P. (2021). Alfred3 – A library for rapid experiment development (Version 2.1.6). Göttingen, Germany. https://doi.org/10.5281/zenodo.1437219

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

van Stein, K. R., Strauß, B., & Brenk-Franz, K. (2019). Ovulatory shifts in sexual desire but not mate preferences: An LH-test-confirmed, longitudinal study. *Evolutionary Psychology*, *17*, 1–10. https://doi.org/10.1177/1474704919848116

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, 1832. https://doi.org/10.3389/fpsyg.2016.01832

Wilcox, A. J., Dunson, D., & Baird, D. D. (2000). The timing of the "fertile window" in the menstrual cycle: Day specific estimates from a prospective study. *BMJ*, *321*, 1259–1262. https://doi.org/10.1136/bmj.321.7271.1259

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Stern, J., Ostermann, S., & Penke, L. (2023). Investigating cycle shifts in women's clothing style and grooming. *British Journal of Social Psychology*, *00*, 1–25. https://doi.org/10.1111/bjso.12681