A longitudinal evaluation of ovulatory cycle shifts in women's mate attraction and preferences

Julia Stern, Tobias L. Kordsmeyer, & Lars Penke

Department of Psychology & Leibniz ScienceCampus Primate Cognition

University of Goettingen

Gosslerstrasse 14, 37073 Goettingen, Germany

This manuscript has been accepted for publication at *Hormones and Behavior*.

Corresponding author: Julia Stern (julia.stern@psych.uni-goettingen.de)

Declaration of Conflicting Interests

The authors declare that they have no conflicts of interests.

Abstract

Are ovulatory cycle shifts in women's mate attraction and preferences robust? What are underlying mechanisms of potential cycle shifts? These questions are the subject of a current scientific debate surrounding the *good genes ovulatory shift hypothesis*. Here, we report a large, preregistered, within-subjects study, including salivary hormone measures and conception risk estimates based on luteinizing hormone tests. In four sessions across one ovulatory cycle, $N = 257$ women (= 1028 sessions) rated the attractiveness of 40 natural male bodies, 40 natural female bodies and 40 objects. Multilevel analyses yielded weak evidence for ovulatory increases in women's general attraction, specifically to male bodies, though they are not systematically related to changes in steroid hormone levels. Further, we found no compelling robust evidence for mate preference shifts across the cycle, as only one out of many different tests showed some weak evidence for such effects. Mechanisms regulating cycle shifts, the impact of our results on developing and revising cycle shift theories, and influences of different methodologies on results are discussed.

Keywords: ovulatory cycle, mate preferences, mate attraction, steroid hormones, fertility

Introduction

Do women's mate preferences change across the ovulatory cycle? This research question is currently topic of a controversial scientific debate. According to the most prominent hypothesis in the ovulatory cycle literature, the *Good Genes Ovulatory Shift Hypothesis* (GGOSH, Gangestad et al., 2005), women's mate preferences should shift across the cycle. More precisely, when being fertile, women should prefer to mate with men who display putative indicators of genetic fitness, which should aid them in obtaining "good genes" that will be passed on to their offspring. These preferences should only be present when evaluating men for short-term, sexual relationships and be absent when not fertile (Gildersleeve et al., 2014). As the ovulatory cycle is regulated by hormonal changes, steroid hormones are assumed to be the mechanism behind preference shifts across the cycle. Whereas higher levels of estradiol and lower levels of progesterone characterize the fertile (late follicular) phase, levels of estradiol are lower and levels of progesterone much higher during the luteal phase (between ovulation and menstrual onset), except for a second smaller estradiol peak mid-luteal (Roney & Simmons, 2013).

Numerous previous studies have reported evidence in favor of the GGOSH. Cycle shifts in mate preferences were reported for a number of different purported genetic fitness indicators, such as masculine faces (e.g. Penton-Voak et al., 1999) and voices (e.g. Puts, 2005, 2006), masculine or muscular bodies (Gangestad et al., 2007; Little et al., 2007), and dominant behaviors (Gangestad et al., 2004, 2007). However, besides controversies around the claim whether these traits really indicate male genetic fitness (e.g. Lee et al., 2014; Scott et al., 2014), the GGOSH is currently debated because of numerous studies failing to find compelling evidence in line with its predictions regarding preference shifts across the cycle. This controversy became prominent around 2014, when two meta-analyses on cycle shifts in women's mate preferences came to strikingly diverging conclusions about the existence of

such cycle shifts (Gildersleeve et al., 2014; Wood et al., 2014). Furthermore, methods used in previous studies reporting evidence in line with the GGOSH have been extensively criticized (e.g. Arslan et al., in press; Blake et al., 2016; Gangestad et al., 2016; Jones et al., 2019) with regard to insufficently low test power (which could lead to an overestimation of effect sizes and noisy, unreliable results), a high flexibility in defining the fertile window (mostly using rather imprecise counting methods), between-subject designs to test a within-subjects effect, no direct hormone assessments, as well as intransparency and high degrees of flexibility in statistical analyses. Consequently, raised criticism has cast doubt on the reliability and validity of previous findings and highlighted the need for new studies employing more rigorous methods.

During recent years, multiple studies were conducted to address previous criticism and rigorously test the GGOSH. They employed larger sample sizes, more valid fertile window estimates (validated with urine tests of the luteinizing hormone, LH, which peaks shortly before ovulation), within-subjects designs, steroid hormone assays, and employing Open Science practices. Interestingly, the vast majority of these newer studies found no compelling evidence for ovulatory cycle shifts in mate preferences for men‘s faces (Dixson et al., 2018; Jones et al., 2018a; Marcinkowska et al., 2018), bodies (Jünger et al., 2018a; Marcinkowska et al., 2018; van Stein et al., 2019), voices (Jünger et al., 2018b) and behaviors (Stern et al., 2020). However, some of these studies presented evidence for shifts in mate attraction instead of mate preferences, in that all men were evaluated as being a little more attractive when fertile, independent of male characteristics. Such a shift in mate attraction might be linked to an increase in general sexual desire during the fertile phase, indicating an increase in mating effort around ovulation (Arslan et al., in press; Jones et al., 2018b; Jünger et al., 2018a, 2018b; Roney & Simmons, 2013, 2016; Stern et al., 2020). This new data was met with scrutiny by the original proponents of the GGOSH. For example, recently Gangestad and

colleagues (2019) reanalyzed open data from Jünger and colleagues (2018a) and argued that their data actually show cycle shifts in preferences for men's bodies when including women's relationship status in a three-way interaction and changing multiple other analytical decisions. However, this conclusion was challenged by a multiverse analysis showing no robust evidence for cycle shifts in preferences for men's bodies with the same data (Stern et al., 2019), and critiques of Gangestad and colleagues' (2019) analytical decisions and interpretation of results (Higham, 2019; Jones et al., 2019b; Roney, 2019; Stern et al., 2019). Again, all involved researchers highlighted a need for further well-powered studies with rigorous methods to scrutinize potential effects of women's ovulatory cycle and reproductive hormones on their mate preferences.

Alongside this controversy, alternative explanations on cycle effects on female mating psychology were developed. The *Motivational Priority Shifts Hypothesis* (Roney, 2018a) suggests that, rather than their mate preferences, women's motivations change across the cycle: when women can conceive, their mating motivations (e.g. sexual interests) have a greater priority because the probability of conception provides potential fitness benefits that outweighs potential costs of sex (Roney, 2018a; Roney & Simmons, 2017). Other motivations (e.g. motivation to forage and eat) have lower priority during the fertile phase, but higher during cycle phases when women cannot conceive (e.g. the luteal phase). Steroid hormones, especially estradiol and progesterone, should regulate shifts in sexual motivation (Roney, 2018a). Non-hormonal factors, such as women's relationship status, might also affect mating motivation, probably independently of hormonal effects (Roney, 2018a), because women's mating psychology might be sensitive to the presence or absence of a stable investing partner (Pillsworth et al., 2004), as costs of pregnancy might outweight its benefits when a long-term supportive mate is not available. Furthermore, self-reported stress appeared to be another factor that affects reproductive hormones (Roney & Simmons, 2015), which, in turn, might

influence changes in mate attraction (Jünger et al., 2018a). Reported shifts in mate attraction (Jünger et al., 2018a, 2018b; Stern et al., 2020) and general sexual desire (Arslan et al., in press; Jones et al., 2018b; Roney & Simmons, 2013, 2016), as well as effects of relationship status on mate attraction and sexual desire (Jünger et al., 2018a; Roney & Simmons, 2016), are in line with this hypothesis. Shifts in sexual desire and mate attraction might even be connected in that increased fertile phase sexual desire may lead to perceiving other men as being more attractive, thus, mediating a potential increase in mate attraction.

Another hypothesis with some supportive recent evidence is that some of the previously reported effects of changes in women's mating psychology may in fact be between-women, not within-women, hormonal effects (DeBruine et al., 2019; Havlíček et al., 2015; Marcinkowska et al., 2018b). More precisely, mate preferences for masculine faces were predicted by the interaction of between-women progesterone levels and relationship status, rather than by within-women hormonal shifts. This might explain why most of the studies supporting cycle shifts in mate preferences were between-subject studies (e.g. Gangestad et al., 2004, 2007; Little et al., 2007; Puts, 2005; 2006), whereas recent within-subjects studies did not report compelling evidence for this effect (Dixson et al., 2018; Jones et al., 2018a; Jünger et al., 2018a, 2018b; Marcinkowska et al., 2018a; Stern et al., 2020; van Stein et al., 2019).

Here we aim to contribute to the current debate on the existence of preference shifts across the ovulatory cycle by investigating shifts in female preferences for men's bodies, the dimension for which cycle shifts have been most controversial. For this purpose, we test N = 257 women, each in four individual sessions across their ovulatory cycle (resulting in 1028 sessions), in which they rate the attractiveness of 40 natural 3D male bodies, as well as 40 natural 3D female bodies and 40 3D objects as controls.

To test the Motivational Priorities Hypothesis, we preregistered several hypotheses (https://osf.io/29n5j): Naturally cycling women with higher conception risk (Hypothesis 1a), as well as with a higher estradiol-to-progesterone ratio (henceforth E/P; Hypothesis 1b) should evaluate male bodies as more attractive. We expect both effects to be stronger for women in relationships (Hypothesis 1c), to be stronger when women are less stressed out (Hypothesis 1d) and to be mediated by sexual desire (Hypothesis 1e). We expect the effect of conception risk on attractiveness ratings to be mediated by a higher E/P (Hypothesis 1f)[1].

To test the GGOSH, we preregistered the following hypotheses: Naturally cycling women with higher conception risk should evaluate muscular male bodies (Hypothesis 2a), physically dominant male bodies (Hypothesis 2b), and bodies that are tall, belong to individuals that are physically stronger and have higher baseline testosterone levels, higher upper arm circumference, higher upper torso volume (relative to lower torso volume), higher SHR (shoulder-hip ratio) and SCR (shoulder-chest ratio) (Hypothesis 2c) as more attractive (see Table S1 for a justification for these traits). These effects are expected to be stronger for women in relationships (Hypothesis 2d), to be stronger when women are less stressed out (Hypothesis 2e, see Ditzen et al., 2017), to be significant when conception risk as a predictor is replaced by the E/P (Hypothesis 2f) and to remain stable when controlling for men's body fat mass (Hypothesis 2g).

In an exploratory manner, we also include between-subjects hormone effects and test effects with attractiveness ratings of female bodies and inanimate objects, rather than male bodies, as outcomes. Multiple patterns of results are plausible for female body ratings. Ratings might increase when conception risk is high because women's sexual arousal is non-specific (Chivers, 2005) or other women might be seen as rivals and derogated when

[1] Analyses and results for this hypothesis can be found in the supplementary material (page 9), as we noticed that E/P and conception risk were highly correlated and, thus, having both as predictors in the same model is not as informative as we thought when preregistering the hypothesis.

conception risk is higher, leading to lower attractiveness ratings (Fisher, 2004). However, if participants rate specific female body cues (e.g. muscularity) as being more attractive when conception risk is high, this pattern of results would, in our understanding, contradict the GGOSH. Preference shifts are specifically assumed for men (to enhance reproductive success) and preference shifts for women cannot be seen as an adaption to obtain potential good genes. Object ratings should not be influenced by conception risk or hormone levels, but if they are, this would contradict the Motivational Priority Shifts Hypothesis and might indicate confounding effects of, for example, mood or study design and must be investigated further.

Methods

Our hypotheses, study design, sampling and analysis plan were pre-registered online at the Open Science Framework (https://osf.io/29n5j) before any data on the women have been collected or analyzed. Open data, analysis script, and material are also provided (https://osf.io/4jcuf). All participants signed a written consent form and the local ethics committee approved the study protocol (no. 225).

Participants and recruitment

A total of 257 heterosexual female participants (aged 18-35 years, M = 23.2, SD = 3.3), out of 282 recruited, finished all sessions, and were therefore included in further analyses[2]. The 25 dropouts resulted from 16 women who attended only the introductory session and nine women who only completed one or two testing sessions (for the following reasons: not responding to emails anymore (9), decided not to take part without providing further reasons (3), scheduling problems (2), switch to hormonal contraception (2), taking the morning after

[2] Please note that we have preregistered a sample size of N = 250, to reach a sample of n = 200 participants for our conception risk analyses with all women fulfilling all inclusion criteria. Our actual sample size is seven participants larger than what we have preregistered, because we expected more people to drop out. However, we decided to include all N = 257 participants to reach the preregistered sample of exact n = 200 participants for our conception risk analyses.

pill (2), health issues (2), moved to a different city (1), irregular mid-cycle bleeding (1), very long irregular cycle >50 days (1), claimed to not fit into the inclusion criteria anymore (1), or pregnancy (1)). Our participants had to fulfill the following preregistered criteria to take part in the study: female, between 18 and 35 years old, naturally cycling (no hormonal contraception for at least three months, no expected switch to hormonal contraception while in the study, no current pregnancy or breastfeeding, no childbirth or breast-feeding during the previous year, not taking hormone-based medication or anti-depressants, no endocrine disorders). Additionally, included participants reported their ovulatory cycles being of regular length between 25 and 35 days, at least during the last 3 months. Our sample size largely exceeds the size required to achieve 80% power given a within-subjects design and anticipated effects of moderate magnitude, as suggested by recent guidelines for sample sizes in ovulatory shift research (Gangestad et al., 2016). More precisely, these guidelines suggest a sample size of $N = 48$ participants with two testing sessions each (= 96 sessions) and LH test validated cycle phases, to have sufficient power to detect Cohen's $d = 0.5$. In contrast, the current study has sufficient power to detect much smaller effect sizes with $N = 257$, four testing sessions per participant (= 1028 sessions) and LH validated scheduling. Further, our study exceeds all sample sizes of previous studies that investigated cycle shifts in preferences for male bodies by at least $n = 100$ participants.

Procedure

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

Sessions two to five were computer-based testing sessions and took place across different phases of the ovulatory cycle, scheduled based on backward counting and the

observed LH test surge. Suitable testing days were computed with the help of an Excel sheet created for that purpose (see open material). All participants completed two sessions in their expected fertile phase (mid to late follicular phase, approx. 5-6 days before ovulation) and two sessions in their expected luteal phase (one session in the mid luteal phase, one session in the premenstrual phase). Scheduling was validated via LH test results and via following up to the day of the next menstrual onset. Details can be found in the supplementary material. The starting session for each participant depended on their current cycle phase at the introductory session and their personal schedule. Of all participants who finished all sessions, 134 participants started with the first session in their fertile phase, and 123 started in the luteal phase.

To control for possible effects of diurnal changes in hormone levels, all sessions were scheduled in the second half of the day (between 12pm and 6pm). When arriving at the lab, participants first completed a screening questionnaire, assessing their eligibility and some control variables for saliva sampling (Schultheiss & Stanton, 2009). Next, saliva samples were collected via passive drool. Then, participants started a rating task, rating men's bodies, women's bodies, or objects. Rating tasks and trials were randomized between each participant and session. Instructions, stimuli material (see below) and procedure were similar to the study by Jünger and colleagues (2018a). Participants were instructed to evaluate all men, women, and objects as they perceived them "in that moment", independently of their own current relationship status, sexual orientation, general interest in other men, or ratings in previous sessions. Participants were then presented with the stimuli in a randomized order. Men's bodies, women's bodies, and objects (e.g. chair, candle, clock, bike, table) were not mixed, but shown in blocks of the same kind of stimuli. All stimuli were displayed rotating around their vertical axis, allowing them to be inspected from every side. To avoid the influence of confounding variables like facial attractiveness or skin color, stimuli were consistently colored in grey, without texture or head (see Figure 1), containing information on body

morphology only. Participants rated each stimulus separately after at least one full rotation but were able to inspect them as long as they preferred. Men's and women's bodies were rated for sexual attractiveness on an eleven-point Likert scale from -5 (*extremely unattractive*) to +5 (*extremely attractive*). Objects were rated on comparable Likert scales, but from -5 (*extremely unappealing*) to +5 (*extremely appealing*), as rating objects on sexual attractiveness would have been odd. Definitions of rating categories were provided prior to rating:

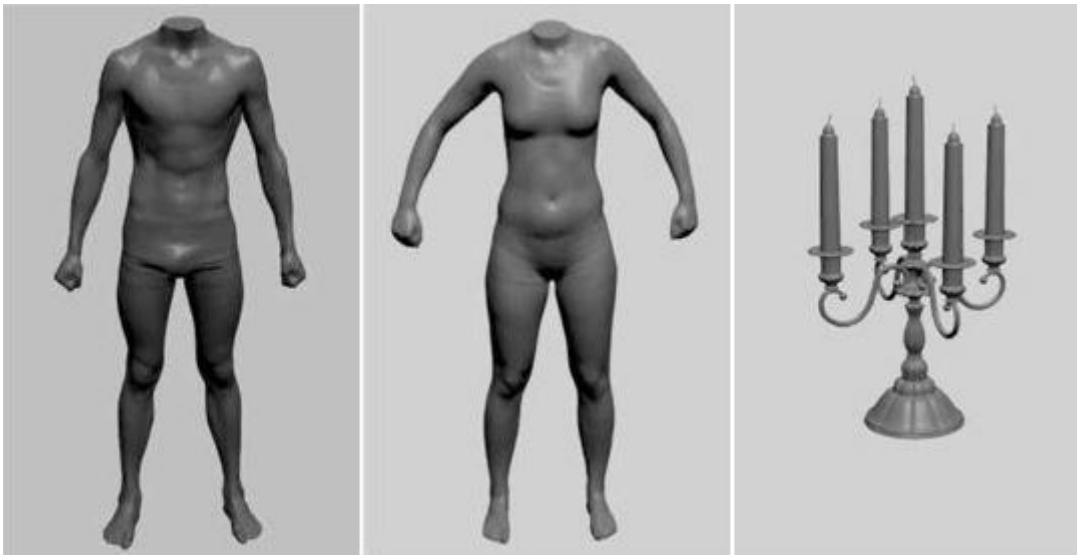
Sexually attractive: Men/women that score high are men/women that you would find very attractive for a sexual relationship that can last for a short time and must not contain any other commitment. Men/women scoring low are men/women that you would find very unattractive for a sexual relationship.

Appealing: Objects that score high are objects that you find aesthetic, beautiful, and comparable to what would be attractive in humans. Objects scoring low are objects that you find are very unaesthetic, not beautiful, and comparable to what you would see as being unattractive in humans.

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

Figure 1

Static examples of 3D stimuli.



Measures

Conception risk

Participants' conception risk was assigned based on highly sensitive (10mIU) LH urine ovulation test strips from MedNet GmbH, following the procedure of Study 1 in Jünger et al. (2018b) and Shirazi et al. (2019), which is based on previous studies (see Table 1)[3]. Participants started LH-testing after menstruation (around reverse cycle day 21) and continued until a rise of LH (positive tests) was observed and a minimum of two days after the tests were negative again (as suggested by Roney, 2018b). Participants were provided with a minimum of ten LH tests each and provided daily pictures of the tests to the investigators for confirmation. LH results were used to allow flexible scheduling in case LH test results differed from the scheduling based on counting (see supplementary material for details).

[3] Newer conception risk data has been published recently (Faust et al., 2019). However, we decided not to include this data in our conception risk measure because we preregistered our measure as displayed in Table 1 and wanted to follow our preregistered plan.

Table 1

Conception risk estimates relative to the day of ovulation from previous studies and the weighted average value used in the current study.

Day relative to ovulation	Day relative to LH peak	Schwartz et al., (1980)	Wilcox et al., (1998)	Colombo & Masarotto (2000)	Weighted average
-8	**-7**	0	0	0.00	**0.00**
-7	**-6**	0	0	0.01	**0.01**
-6	**-5**	0	0	0.03	**0.02**
-5	**-4**	0.04	0.04	0.07	**0.06**
-4	**-3**	0.14	0.13	0.18	**0.16**
-3	**-2**	0.2	0.08	0.24	**0.20**
-2	**-1**	0.2	0.29	0.26	**0.25**
-1	**0**	0.34	0.27	0.21	**0.24**
0	**1**	0.14	0.08	0.10	**0.10**
1	**2**	0.07	0	0.01	**0.02**
2	**3**	0	0	0.04	**0.02**

Note: All other days (not mentioned here) were assigned a conception risk of 0.00

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

Hormone measures

For hormone assays, we collected four saliva samples from each participant (one per testing session). Contamination of saliva samples was minimized by asking participants to abstain

from eating, drinking (except plain water), smoking, chewing gum, or brushing teeth for at least one hour before each session. The samples were stored at -80°C directly after collection until shipment on dry ice to the Kirschbaum Lab at Technical University of Dresden, Germany (one freeze-thaw cycle), where progesterone was assessed via liquid chromatography mass spectrometry (LCMS, Gao et al., 2015)[4]. Since the lab had no valid protocol for LCMS analysis of estradiol levels, the samples were reanalyzed for estradiol using the highly sensitive 17β-estradiol enzyme immunoassay kit (IBL International, Hamburg, Germany). Samples were analyzed in singlets, however, the lab reported that their procedure yields CVs < 11%. There was a significant large association (β = 0.71, *95%CI* [0.41; 1.01]) between conception risk and E/P (γ = 3.95, *SE* = 0.93, *95%CI* = [2.12; 5.77], t = 4.24, p <.001), validating our conception risk measure. Table 2 displays hormone concentrations as a function of conception risk. We centered all hormone values on their subject-specific means and scaled them afterwards (i.e. divided them by a constant), so that the majority of the distribution for each hormone varied from -0.5 to 0.5 to facilitate calculations in the linear mixed models (e.g. as in Jones et al., 2018; Jünger et al., 2018a, 2018b). This is a common procedure to isolate effects of within-subject changes in hormones, avoiding the influence of outliers on results and dealing with the non-normal distribution of hormone levels. Hormone levels were nearly normally distributed afterwards, a figure showing the distribution of hormone levels after this procedure can be found in the supplement (Figure S1). Importantly, this procedure did not change any findings compared to

[4] Liquid to liquid extraction was carried out by adding 20 µL internal standard and 1 mL ethyl acetate to 400 µL saliva in a 2 mL polypropylene tube. The resulting mixture was subsequently rotated for 1 min on the vortex and then centrifuged for 10 min at 12000 r/min with centrifuge (Hettich, MIKRO 22 R). The ethyl acetate layer was transported to a new glass tube and evaporated to dryness under nitrogen. The residue was resuspended in 120 µL methanol/water in a ratio of 50:50 (v/v), 50 µL of which was injected into the LC-MS/MS system. The LC-MS/MS system consisted of Shimadzu HPLC system, and AB Sciex Triple Quad 6500+ System equipped with the electrospray ionization (ESI) source. See Gao et al. (2015) for more details.

analyses with untransformed hormone values. The R code for this procedure can be found in the open script.

Table 2

Hormone concentrations as a function of conception risk

Conception risk	Estradiol pg/ml mean (SD)	Progesterone pg/ml mean (SD)	E/P mean (SD)
0.00	3.95 (1.85)	42.60 (56.54)	0.54 (0.99)
0.01	3.86 (1.85)	7.56 (11.99)	1.39 (0.99)
0.02	3.74 (1.51)	9.13 (11.79)	1.31 (1.39)
0.06	4.30 (2.55)	4.30 (8.79)	2.27 (2.29)
0.1	4.66 (1.75)	8.80 (6.81)	0.98 (0.96)
0.16	4.06 (1.48)	6.51 (9.75)	1.41 (0.96)
0.2	3.82 (1.67)	5.78 (7.58)	1.39 (1.31)
0.24	4.34 (1.65)	7.65 (8.33)	1.54 (2.90)
0.25	4.11 (2.01)	5.88 (6.28)	1.39 (1.65)

Stimuli and masculinity measures

Fourty male bodies and 40 female bodies, collected in independent studies (men: Kordsmeyer & Penke, 2019; women: Jünger et al., 2018a), and 40 objects, randomly selected from a database (https://archive3d.net) and converted to fit to the presentation of the bodies, were presented. All bodies were natural bodies of men or women in standardized underwear (tight shorts or sports underwear), captured with a high-resolution 3D body scanner (Vitus Smart XXL by Human Solutions). Men and women were instructed to stand upright with legs hip-wide apart, arms extended and held slightly away from the body, making a fist with thumbs showing forward, the head positioned in accordance with the Frankfort Horizontal, and to breathe normally during the scanning process. Body models were scaled so that they retained original height differences. Out of 165 available male and 157 female bodies, we preselected stimuli based on adequate scan quality and no missing values on target's data, and randomly selected 40 bodies for males and females each out of the remaining ones.

Visual cues of upper body strength were directly measured from the body scans using the automatic measures of the software Anthroscan (all according to ISO 20685:2005),

including the following parameters relevant to this study: bust-chest girth (Anthroscan measure 4510), hip girth (7520), and upper arm girth (8520). In addition to automatic measurements, biacromial shoulder width was measured manually (on screen) as the direct distance between the left and right acromion processes. The volume (in liters) of upper torso and lower torso was also measured from scans. We calculated shoulder-chest ratio, shoulder-hip ratio, and the relative volume of upper torso to lower torso. Physical strength was operationalized as the aggregated mean of the dominant hand grip and upper body strength, measured with a hand dynamometer (Saehan SH5001), following the procedure described in Sell and colleagues (2009). The maximum strength of three trials for each measurement was used. Height was measured with a stadiometer. To measure testosterone levels, saliva samples were taken (for details for male stimuli see Kordsmeyer & Penke, 2019, and for female stimuli Jünger et al., 2018a). Stimuli muscularity was rated separately for male and female stimuli by 20 independent raters on seven-point Likert scales from 1 = "*not at all muscular*" to 7 = "*very muscular*". Cronbach's alphas for ratings were high (α = .97 for male bodies, α = .91 for female bodies), thus, ratings were averaged. Men's bodily dominance ratings were collected in a previous study (Kordsmeyer et al., 2018), by asking participants "*How likely is it that this man would win a physical fight with another man?*" on an eleven-point Likert scale from -5 = "*extremely unlikely*" to +5 = "*extremely likely*".

Relationship status

Women's self-reported relationship status was assessed in every testing session using a single item "What is your current relationship status? (*Single – in a committed relationship – in an open relationship – engaged - married)*". In case the relationship status changed across the different sessions, their data were categorized in accordance with their relationship status on the particular testing day. Relationship status was effect-coded with -1 = single and 1 = in a relationship (including all categories other than single). At the beginning of the study, 121 of the participants reported being partnered, 136 reported being single.

Self-reported stress

Participant's self-reported stress levels were assessed via one item ("*Today I am stressed out*" on a 5-point Likert scale from "*not at all applicable*" to "*completely applicable*") at the beginning of each session.

Sexual desire

Women's sexual desire was assessed in each testing session with the item[5] "*How much do you desire sexual contact today*?" on a seven-point Likert scale from 1 = "*not at all"* to 7 = "*very much"*. This item was taken from Roney and Simmons (2013).

Statistical analyses

All analyses were calculated with the statistic software R 3.6.1 (R Core Team, 2016). The following packages were used: lme4 1.1-21 (Bates et al., 2014), lmerTest 3.1-0 (Kuznetsova et al., 2013), sjPlot 2.8.3 (Lüdecke, 2018), psych 1.8.12 (Revelle, 2016), dplyr 0.8.3 (Wickham, 2011), tidyverse 1.2.1 (Wickham, 2019), GPArotation 2014.11-1 (Bernaards & Jennrich, 2015).

As we study a large number of different effects, we preregistered to adjust our alpha level to a significance threshold of $p < .01$. A simple Bonferroni correction would be overly conservative. The same significance threshold is applied to all reported effects. All statistical tests are two-tailed.

Results

Cycle shifts in mate attraction

For our main analyses, we first tested possible ovulatory cycle shifts in mate attraction (Hypotheses 1a-f). For this purpose, we computed multilevel regression models for all

[5] Participants also responded to the Sexual Desire Inventory 2 (*SDI-2*, Spector et al., 1996) and the Sociosexual Orientation Inventory Revised (*SOI-R*, Penke & Asendorpf, 2008) for other purposes. As preregistered, both are not part of the current study.

analyses, all random slopes were specified maximally[6] (including interactions) following Barr and colleagues (2013). In addition, for testing Hypothesis 1a, we included attractiveness ratings of men's bodies as outcome variable, women's conception risk as the predictor variable and random intercepts for female participants and male stimuli. This model showed a very small significant cycle shift (β = 0.02, *95%CI* [0.005; 0.03]) in women's attraction: with higher conception risk, women's ratings of men's bodies slightly increased (γ = 0.54, *SE* = 0.21, *95%CI* = [0.13; 0.95], t = 2.62, p = .009), supporting Hypothesis 1a. Further, when modelling E/P rather than conception risk, no significant association of E/P with men's body attractiveness was detected (γ = -0.04, *SE* = 0.08, *95%CI* = [-0.20; 0.13], t = -0.43, p = .671), not supporting Hypothesis 1b. Further, we found no compelling evidence that the effects of conception risk or E/P were moderated by women's relationship status (Table 3) or self-reported stress (Table S4), in contrast to Hypotheses 1c and 1d.

Table 3

Multilevel regression analyses of male body attractiveness rating as a function of conception risk or E/P and women's relationship status

	Estimates	*SE*	*t*	*p*	*95% CI*
Conception risk model					
Conception risk	0.70	0.26	2.72	.006	[0.20; 1.21]
Relationship status	-0.15	0.12	-1.42	.211	[-0.38; 0.08]
Conception risk x Relationship status	-0.29	0.43	-0.69	.489	[-1.13; 0.54]
E/P model					
E/P	-0.01	0.07	-0.08	.935	[-0.14; 0.13]
Relationship status	-0.19	0.11	-1.71	.087	[-0.40; 0.03]
E/P x Relationship status	-0.09	0.19	-0.48	.633	[-0.45; 0.28]

Note. The outcome in the conception risk model had 32,000 observations (200 participants x 4 test sessions x 40 stimuli), the outcome in the E/P model had 34,680 observations (257 participants x 4 test sessions x 40 stimuli – missing values). We effect-coded relationship status with -1 = single, 1 = in a relationship.

[6] We described random slopes for every predictor variable in our preregistration. However, we are slightly deviating from the pseudo code in our preregistration (but not from the text), as it accidentally included copy and paste errors. More precisely, the pseudocode models did not include slopes, but rather reduced scalar random effects. Hence, the pseudo code is in contradiction to the preregistered text and our intentions.

Second, we investigated the relationship between male attractiveness ratings, sexual desire, and conception risk or E/P (Hypothesis 1e). General sexual desire was positively associated with attractiveness ratings of male bodies (γ = 0.08, *SE* = 0.02, *95%CI* = [0.03; 0.13], t = 3.23, p =.001), indicating that men's bodies were evaluated as being more attractive when participants' sexual desire was high, but the effect size was close to zero (β = 0.03, *95%CI* [0.01; 0.05]). Further, sexual desire increased with higher conception risk (γ = 4.36, *SE* = 1.26, *95%CI* = [1.90; 6.83], t = 3.46, p <.001; β = 0.36, *95%CI* [0.16; 0.57]). When modeling conception risk and sexual desire as predictors for male attractiveness ratings in the same model, both previously reported effects on attractiveness ratings decreased (conception risk: γ = 0.50, *SE* = 0.22, *95%CI* = [0.07; 0.93], t = 2.30, p = .022; sexual desire: γ = 0.04, *SE* = 0.03, *95%CI* = [-0.01; 0.10], t = 1.44, p = .149). However, there was no significant association of E/P and sexual desire (γ = 0.45, *SE* = 0.98, *95%CI* = [-1.47; 2.37], t = 0.45, p =.650). When E/P and sexual desire were included as predictors in the same model, previously reported effects remained stable (E/P: γ = -0.04, *SE* = 0.09, *95%CI* = [-0.21; 0.13], t = -0.44, p = .658; sexual desire: γ = 0.12, *SE* = 0.03, *95%CI* = [0.06; 0.17], t = 4.26, p <.001).

Additional exploratory analyses and robustness checks

In an exploratory manner, we repeated all analyses reported above with estradiol and progesterone as separate predictors. Results revealed a positive main effect of estradiol on ratings in only one model, which was not robust in other models and did not reach p <.01[7]. No further significant effects were detected; details can be found in the supplementary material (Tables S5 – S11). Next, we investigated whether ratings of female bodies or objects were related to conception risk or E/P. There was no significant effect of conception risk (γ = -0.16, *SE* = 0.24, *95%CI* = [-0.63; 0.30], t = -0.69, p = .493) or sexual desire (γ = 0.03, *SE* = 0.03,

[7] We specified our significance level as p <.01 in our preregistration to adjust for multiple testing.

95%CI = [-0.03; 0.09], t = 1.06, p = .288) on female body attractiveness ratings, in contrast to both the hypothesis that fertile women rate other women as less attractive because of increased intrasexual competitiveness in the fertile phase (Fisher, 2004) and the hypothesis that women's sexual desire might be non-specific to men (Chivers, 2005). Moreover, there was no significant effect of conception risk on ratings of objects (γ = 0.24, *SE* = 0.19, *95%CI* = [-0.13; 0.61], t = 1.27, p = .206), in contrast to the alternative explanations that the increase in attractiveness ratings of male bodies with higher conception risk is due to generally milder rating behavior when fertile or other methodological confounds that might artificially create attraction effects. No other significant effects were detected; details can be found in the supplementary material (Tables S12 – S17). Finally, we controlled our main analyses for a possible effect of testing session. All results remained virtually identical (for details see our open script PDF file, pp. 142-148).

Cycle shifts in mate preferences

Next, we investigated whether women's mate preferences for masculine bodies shift across the cycle. For this purpose, we first ran a factor analysis for the seven body characteristics mentioned in Hypothesis 2c, which yielded two factors: Factor 1 (labelled "shoulder factor") and Factor 2 (labelled "strength factor"). Details can be found in the supplementary material (Table S3). Second, we ran four models with male body attractiveness as the outcome, and conception risk and one male body cue at a time (rated muscularity, rated physical dominance, shoulder factor, or strength factor) as predictor variables. As described above, we modelled random intercepts for female participants and male stimuli and maximal random slopes. Results can be found in Table 4. In all four models, there was a significant but small (close to zero) main effect (β = 0.02, *95%CI* [0.005; 0.03]) of conception risk, indicating that attractiveness ratings increased with higher conception risk, again supporting Hypothesis 1a. Further, there were significant medium-sized main effects for rated muscularity (β = 0.60,

95%CI [0.48; 0.72]) and rated physical dominance (β = 0.40, *95%CI* [0.21; 0.58]), indicating that women tend to find more muscular men and men with higher physical dominance as being more sexually attractive in general. The main effect for the shoulder factor was significant according to conventional standards, but did not reach our preregistered conservative significance level of p <.01. However, the confidence interval included small to medium effect size estimates (β = 0.25, *95%CI* [0.04; 0.46]). None of the interaction effects between conception risk and body cues were significant according to our predefined significance level, thus showing no compelling evidence for the GGOSH, in contrast to Hypotheses 2a, 2b, and 2c. The interaction of conception risk with muscularity (β = 0.01, *95%CI* [0.001; 0.02) was descriptively in line with the GGOSH (and significant according to conventional standards), but had a standardized effect size and confidence interval close to zero.

Table 4

Multilevel regression analyses of male body attractiveness rating as a function of conception risk interacting with masculinity cues

	Estimates	*SE*	*t*	*P*	*95% CI*
Muscularity model					
Conception risk	0.54	0.21	2.62	.009	[0.13; 0.95]
Muscularity	1.57	0.16	10.15	<.001	[1.26; 1.88]
Conception risk x Muscularity	0.24	0.10	2.33	.020	[0.04; 0.45]
Dominance model					
Conception risk	0.54	0.21	2.62	.009	[0.13; 0.95]
Physical dominance	1.04	0.25	4.25	<.001	[0.56; 1.53]
Conception risk x Physical dominance	0.10	0.11	0.91	.364	[-0.11; 0.31]
Shoulder factor model					
Conception risk	0.54	0.21	2.62	.009	[0.13; 0.95]
Shoulder factor	0.73	0.31	2.37	.020	[0.12; 1.34]
Conception risk x Shoulder factor	0.19	0.12	1.61	.108	[-0.04; 0.42]
Strength factor model					
Conception risk	0.54	0.21	2.60	.009	[0.13; 0.95]
Strength factor	0.56	0.29	1.93	.054	[-0.01; 1.13]
Conception risk x Strength factor	0.09	0.11	0.81	.417	[-0.13; 0.30]

Note. The outcome variable had 32,000 observations (200 participants x 4 test sessions x 40 stimuli). Estimates are unstandardized effect size estimates.

For testing Hypothesis 2f, we ran the same models as described above, but with E/P as a predictor. We also included between-women hormone levels in an exploratory manner. Results are displayed in Table 5. Neither within- nor between-women E/P showed significant associations with attractiveness ratings. The main effects for the body cues were virtually identical to the effects reported above. We found no compelling evidence for interaction effects of E/P and body cues, contradicting Hypothesis 2f. Two between-women E/P interactions showed descriptive effects in that women with a higher average E/P, relative to women with lower E/P, tended to rate less physically dominant men (β = -0.02, *95%CI* [-0.03; 0.001]) and men who scored lower on the strength factor (β = -0.01, *95%CI* [-0.02; -0.001]) as more attractive. However, confidence intervals for all interaction effects were very narrow around zero.

Table 5

Multilevel regression analyses of male body attractiveness rating as a function of within- and between-women estradiol-to-progesterone ratio interacting with masculinity cues

	γ	*SE*	*t*	*p*	*95% CI*
Muscularity model					
E/P_ww	-0.03	0.09	-0.44	.663	[-0.21; 0.13]
Muscularity	1.59	0.15	10.29	<.001	[1.29; 1.90]
E/P_bw	0.09	0.06	1.49	.138	[-0.03; 0.20]
E/P_ww x Muscularity	0.00	0.02	0.25	.805	[-0.03; 0.04]
E/P_bw x Muscularity	-0.03	0.03	-0.88	.379	[-0.10; 0.04]
Dominance model					
E/P_ww	-0.04	0.09	-0.43	.666	[-0.21; 0.13]
Physical dominance	1.04	0.25	4.19	<.001	[0.55; 1.52]
E/P_bw	0.09	0.06	1.49	.139	[-0.03; 0.20]
E/P_ww x Physical dominance	0.02	0.02	0.94	.346	[-0.02; 0.06]
E/P_bw x Physical dominance	-0.05	0.03	-1.87	.063	[-0.11; 0.00]
Shoulder factor model					
E/P_ww	-0.04	0.09	-0.43	.666	[-0.21; 0.13]
Shoulder factor	0.75	0.31	2.43	.020	[0.15; 1.34]
E/P_bw	0.09	0.06	1.49	.139	[-0.03; 0.20]
E/P_ww x Shoulder factor	-0.02	0.02	-0.86	.391	[-0.06; 0.03]
E/P_bw x Shoulder factor	0.02	0.03	0.74	.459	[-0.04; 0.08]
Strength factor model					
E/P_ww	-0.04	0.08	-0.43	.669	[-0.20; 0.13]
Strength factor	0.56	0.29	1.94	.060	[-0.01; 1.13]
E/P_bw	0.86	0.06	1.48	.139	[-0.03; 0.20]
E/P_ww x Strength factor	0.03	0.02	1.57	.116	[-0.01; 0.07]
E/P_bw x Strength factor	-0.04	0.02	-2.08	.039	[-0.08; -0.00]

Note. The outcome variable had 34,680 observations (257 participants x 4 test sessions x 40 stimuli – missing values). E/P_ww reflects within-woman effects of E/P ratio, E/P_bw reflects between-women effects of the E/P ratio. Estimates are unstandardized effect size estimates.

As preregistered, we repeated all models described above with all seven masculinity indicators separately instead of their factors. None of the models showed compelling evidence for preference shifts for any of the masculine cues; details can be found in the supplementary material (Tables S19 – S23). Further, we tested whether shifts in women's mate preferences might be moderated by their relationship status or self-reported stress. For this purpose, we repeated all models reported in Tables 3 and 4, additionally including an interaction effect with either relationship status or self-reported stress. None of these three-way interactions

revealed a significant effect, contradicting Hypotheses 2d and 2e. Main effects for conception risk, muscularity, physical dominance, and the shoulder and strength factors remained virtually identical. Details can be found in the supplementary material (Tables S24 – S35). For testing Hypothesis 2g, we repeated all main models for Hypotheses 2a-2f, additionally controlling for men's body fat mass, in that we included an interaction effect with weight-to-height ratio[8] (WHtR). All previously reported effects remained virtually identical. In addition, all models showed a small to medium negative main effect of WHtR (β = -0.28, *95%CI* [-0.40; -0.16]), indicating that bodies with higher fat mass were rated as being less attractive. Overall, the models did not show compelling evidence for preference shifts across the ovulatory cycle. Details can be found in the supplementary material (Tables S36 – S55).

Additional exploratory analyses and robustness checks

For robustness checks, we first tested for an interaction effect of conception risk and rated muscularity on female body attractiveness ratings, comparing it to the interaction effect of the analyses for male bodies. Analyses revealed that more muscular women were rated as being more attractive ($p < .001$; β = 0.55, *95%CI* [0.44; 0.67]). A main effect of conception risk was trending negative ($p = .023$; β = -0.01, *95%CI* [-0.02; -0.01]). Interestingly, the interaction between conception risk and muscularity was significant by conventional standards (but did not reach $p < .01$), was trending in the same direction and had a similar effect size as the interaction for male attractiveness ($p = .048$; β = 0.01, *95%CI* [0.0002; 0.02]), which, to our understanding of the hypothesis, contradicts the GGOSH. For details see Table S18.

Second, we controlled our main analyses for potential confounding effects of testing session, which might for example capture practice effects. All results remained virtually

[8] As WHtR is a superior index of body fat than waist-to-hip ratio (WHR) or BMI (Gelber et al., 2008). However, all effects were virtually identical when controlling for WHR or BMI rather than WHtR. See our open script PDF file (pp. 169-213) for details.

identical (for details see our open script PDF file, pp. 148-167). Third, to better match analyses from some previous studies and to further contribute to the cycle shifts discussion, we repeated all main models reported above with estradiol and progesterone separately as predictors, rather than conception risk or E/P. All previously reported main effects were robust. None of the two-way interaction effects between estradiol or progesterone and the body cues reached our predefined level of significance. However, a negative interaction between progesterone and the strength factor occurred (p = .023; Table S57), that reached p = .006 in a model focusing on a non-significant three-way interaction with relationship status (Table S73), but again effect size and confidence intervals were virtually zero (β = -0.01, *95%CI* [-0.01; -0.003]). None of the three-way interaction effects between hormone levels, body cues and relationship status or self-reported stress were significant (nor descriptively trending towards a significant effect). Details can be found in the supplementary material (Tables S56 – S77). Thus, there was no compelling evidence for within- or between-women hormone effects on preferences for masculine male bodies.

Discussion

This study aimed to contribute to the current scientific discourse on mate attraction and preference shifts across the ovulatory cycle in that it replicated previous studies, added potential moderator and mediator variables, and employed strong methods. Our results suggest that conception risk is weakly, but significantly positively, associated with sexual desire and women's mate attraction to men's bodies, but show no strong evidence for any steroid hormonal effects on mate attraction. Further, one interaction effect between conception risk and muscularity shows some evidence for preference shifts across the cycle, but the vast majority of our models show no compelling evidence for the GGOSH. Most effects do not reach our conservative level of significance, predefined in the preregistration to control for multiple testing. All observed effect sizes for mate attraction and preference shifts

are very small. Women's relationship status and self-reported stress did not influence mate attraction or preference shifts in a noteworthy manner.

Mate attraction shifts

The reported increase in women's mate attraction with higher conception risk is in line with previous studies reporting generally higher female attractiveness ratings for men's faces, bodies, voices, and behaviors when fertile (Dixson et al., 2018; Jünger et al., 2018a, 2018b; Stern et al., 2020), as well as higher visual attention to male faces and bodies when fertile (Garza & Byrd-Craven, 2019). These shifts seemed to be robust across a variety of checks and to be exclusively evident when evaluating male bodies, not female bodies or objects, supporting the idea that they reflect a higher mating motivation. However, the standardized effect sizes were very small and close to zero (in line with what has been reported in Jünger et al., 2018a). Still, an increase in general sexual desire with higher conception risk is in line with other studies reporting cycle shifts in general sexual desire (Arslan et al., in press; Jones et al., 2018b; Roney & Simmons, 2013; 2016; Shirazi et al., 2019; van Stein et al., 2019). These results provide weak support for the Motivational Priority Shifts Hypothesis (Roney, 2018a). However, this hypothesis suggests that an increase in mate attraction and sexual desire should be regulated by hormonal shifts across the cycle, a claim for which we did not find compelling evidence, even though conception risk was related to E/P. This pattern of results is comparable to previous studies (Jünger et al., 2018b; Stern et al., 2020). Furthermore, we did not find strong evidence for an association of women's relationship status or self-reported stress with mate attraction, in contrast to previous findings (Jünger et al., 2018a) and contrary to the Motivational Priority Shifts Hypothesis.

Comparing methods and results on ovulatory cycle shifts in humans to the literature on non-human primates, Higham (2019) suggests that whether cycle shifts are related to cycle phase estimates or to hormone levels should be seen as two separate, potentially not

converging, yet equally important questions. For example, sexually proceptive behavior in non-human primates shows inter-specific variation, in that it seems to be linked to changes in hormone levels (e.g. E/P) in some species (e.g. long-tailed macaques), but to fertile phase timing in other species (e.g. olive baboons). This is interesting, as it suggests that mechanisms that regulate cycle shifts might vary across species, possibly explaining that our reported shifts in conception risk were not regulated by hormonal changes. Further, reactions to changes in conception risk and hormone levels might vary inter- and intraindividually. Previous studies often came to diverging conclusions regarding cycle shifts depending on whether cycle phase estimates or hormone levels were assessed (Higham, 2019). Other potential explanations for differences in results for conception risk and steroid hormones might be lagged effects, in that for example estradiol affects behavior two days later (Roney & Simmons, 2013), or perimenstrual symptoms (e.g. cramps/headaches) mediate steroid hormonal changes and social behavior (for details see Kiesner et al., 2020). We highly encourage future research to take a closer look at proximate mechanisms regulating cycle shifts and how they might lead to differences in results.

Mate preference shifts

Our results regarding mate preference shifts across the cycle are overall not supporting the GGOSH. We note that some models suggest weak supportive evidence for preference shifts across the cycle, but effects did not reach our conservative levels of significance, preregistered to control for multiple testing. Further, although confidence intervals exceeded zero, standardized effect size estimates and confidence intervals were very small (standardized βs between -.01 and .01), and the ovulatory shift effect for muscularity preferences had a comparable *p*-value and effect size for women's bodies. Interestingly, these results are mostly in line with other recent work challenging previous evidence for the GGOSH (as reviewed in Jones et al., 2019a). Why do some studies, but not others, report

evidence in favor of the GGOSH? A recent study suggests that different methods might have an impact on the results, in that studies using rating vs. forced-choice designs might come to different conclusions (Lewis, 2020). We agree that methods may have very important implications on the results, and argue that there might be other, probably more important factors beyond stimuli evaluation designs. Taking a closer look at differences between studies reporting different results regarding the GGOSH, we notice that studies reporting no compelling evidence for the GGOSH were published rather recently, were more highly powered, and used more rigorous methods and designs than previous studies in support of the GGOSH (for an overview see Jones et al., 2019a). Some of the former even employed Open Science methods, such as open data or preregistration. However, in a reanalysis of Jünger and colleagues (2018a) data, Gangestad and colleagues (2019) argue that there is a three-way interaction between progesterone levels, muscularity indicators or rated physical dominance, and women's relationship status that supports the GGOSH. Stern and colleagues (2019) already showed that this effect is not robust in a multiverse analysis of the same dataset. Here, we do not find evidence for such a three-way interaction effect across a great number of different models and robustness checks. Further, whereas Gangestad and colleagues (2019) argue that effect sizes for preference shifts are meaningful (of medium to large magnitude), the confidence intervals we report here based on new data suggest that effect sizes are close to zero for the three-way interaction emphasized by Gangestad et al. (2019). Thus, we did not replicate Gangestad and colleagues' (2019) reanalysis of Jünger and colleague's (2018a) data. To conclude, more data is always helpful, but it seems as if preference shifts across the cycle are much smaller and potentially not evident for all characteristics previously expected in line with the GGOSH.

Addressing test power and validity of measures

When reporting a number of non-significant effects, it is always up for debate whether the pattern of results might also be due to low test power or invalid measures. However, it is unlikely that our results are impacted by these issues for several reasons. First, our reported small effect size estimates and narrow confidence intervals suggest that our study had sufficient power to detect even effects of an extremely small magnitude (β = .01). Second, our study has the largest sample of all studies investigating cycle phase or conception risk shifts in mate preferences, exceeding previous sample sizes by at least n = 100 (Jones et al., 2018a had a larger sample size for hormone analyses only and did not find compelling evidence for mate preference shifts either). Third, according to the most recent power simulation on cycle studies, our sample exceeds the estimated sample size of N = 48 required to achieve 80% power to detect a within-subjects effect of moderate magnitude (Gangestad et al., 2016). Fourth, we used methods that were recommended as gold standard recently (Blake et al., 2016; Gangestad et al., 2016) and are highly superior to previous studies, including conception risk as a continuous measure rather than cycle phase, LH tests to validate our scheduling, direct hormone assays analyzed with LCMS, a large sample size, a within-subjects design, and employing Open Science practices, including preregistration.

However, we cannot rule out that difficulties with measuring hormones, especially estradiol, might account for some null effects. It has been reported that it is difficult to validly assess estradiol levels, especially in saliva, as estradiol levels in saliva are in very small ranges, thus often undetectable or overestimated by the assay method (Rosner et al., 2013; Schultheiss et al., 2019). Further, intra-assay CVs of estradiol are often reported to be high (>10%), which contributes to measurement error. High CVs and other problems of salivary estradiol assays are potential sources of null results when trying to link hormones to behavior (possibly in the current study as well, see also the limitations mentioned below), but also have the potential to produce non-replicable positive results. Importantly, these problems are

evident for all studies involving estradiol measures and should be kept in mind when interpreting results. As it is very common to draw conclusions of hormone-behavior links on salivary estradiol, not only in ovulatory cycle research, we encourage researchers to develop and improve methods that analyze salivary hormones more precisely, rendering results more trustworthy (see also Newman & Handelsman, 2014; Rosner et al., 2013).

Limitations

We also note limitations of the current study. First, recent research recommends daily hormone assays to validate fertile window estimates as more reliable than LH tests alone (Marcinkowska, 2020). Although we assessed salivary hormones four times during a cycle, daily hormone assays would have helped to identify false positive or negative LH test results. Further, daily hormone assays allow the investigation of potential lagged effects of steroid hormones (especially estradiol) on social behavior, such as mate choice (Roney & Simmons, 2013), which might explain the lack of significant hormone effects in the current study, that we were not able to investigate with our design. Second, we did not include an early follicular phase measurement and were, thus, not able to test effects of low estradiol (when progesterone is almost absent). This issue might explain why hormonal effects do not eventuate, which we recommend to be investigated in future studies. Third, our methods for analyzing hormone samples might not have been optimal. Although salivary progesterone levels were analyzed with LCMS, estradiol levels had to be analyzed with immunoassay kits, which have been criticized for potentially overestimating salivary hormone concentrations (e.g. Schultheiss et al., 2019). Therefore, analyses involving estradiol should be interpreted cautiously. However, given that to our knowledge no reliable protocol for LCMS estradiol analyses exists so far, we saw no alternative to immunoassays and we encourage endocrinologists to develop a reliable protocol for estradiol LCMS analyses. Moreover, hormone analyses should be done in duplicates rather than singlets and control samples

should be added to obtain reliable intra-assay and inter-assay CVs, as well as to reduce measurement error.

Conclusion

In conclusion, our study provides some evidence for a very small association of conception risk with both women's mate attraction to male bodies and sexual desire. Evidence for mate preference shifts across the ovulatory cycle was at least mixed. We found one very small effect for muscular male bodies that was also evident for female bodies. All other tests of the GGOSH showed no compelling evidence. Further, we report no compelling evidence for effects of hormones on mate attraction, or moderating effects of relationship status or reported stress on either mate attraction or preferences. These results are predominantly in line with other recent large-scale studies, but also provide insight on other aspects, such as moderation by relationship status or self-reported stress, between-women hormone effects, and ratings for women's bodies and objects as controls.

Acknowledgments

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

References

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

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

Bernaards, C., Jennrich, R., & Gilbert, M. P. (2015). Package 'GPArotation'. *Retrieved from http://www.stat.ucla.edu/research/gpa*.

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

Colombo, B., Masarotto, G., & Menstrual Cycle Fecundability Study Group. (2000). Daily fecundability: first results from a new data base. *Demographic research*, *3*.

Chivers, M. L. (2005). A brief review and discussion of sex differences in the specificity of sexual arousal. *Sexual and Relationship Therapy*, *20*, 377 – 390.

DeBruine, L. M., Hahn, A. C., & Jones, B. C. (2019). Does the interaction between partnership status and average progesterone level predict women's preferences for facial masculinity? *Hormones and Behavior*, *107*, 80-82. https://doi.org/10.1016/j.yhbeh.2018.12.004

Ditzen, B., Palm-Fischbacher, S., Gossweiler, L., Stucky, L., & Ehlert, U. (2017). Effects of stress on women's preference for male facial masculinity and their endocrine correlates. *Psychoneuroendocrinology*, *82*, 67-74. https://doi.org/10.1016/j.psyneuen.2017.05.006

Dixson, B. J., Blake, K. R., Denson, T. F., Gooda-Vossos, A., O'Dean, S. M., Sulikowski, D., ... & Brooks, R. C. (2018). The role of mating context and fecundability in women's preferences for men's facial masculinity and beardedness. *Psychoneuroendocrinology*, *93*, 90-102. https://doi.org/10.1016/j.psyneuen.2018.04.007

Faust, L., Bradley, D., Landau, E., Noddin, K., Farland, L. V., Baron, A., & Wolfberg, A. (2019). Findings from a mobile application-based cohort are consistent with established knowledge of the menstrual cycle, fertile window, and conception. *Fertility and Sterility*, *112*, 450-457. https://doi.org/10.1016/j.fertnstert.2019.05.008

Fisher, M. L. (2004). Female intrasexual competition decreases female facial attractiveness. *Proceedings of the Royal Society B:*, *271*, S283-S285.

Gangestad, S. W., Dinh, T., Grebe, N. M., Del Giudice, M., & Emery Thompson, M. (2019). Psychological cycle shifts redux: Revisiting a preregistered study examining preferences for muscularity. *Evolution and Human Behavior*, *40*, 501-516. https://doi.org/10.1016/j.evolhumbehav.2019.05.005

Gangestad, S. W., Garver-Apgar, C. E., Simpson, J. A., & Cousins, A. J. (2007). Changes in women's mate preferences across the ovulatory cycle. *Journal of Personality and Social Psychology, 92,* 151–163. https://doi.org/10.1037/0022-3514.92.1.151

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

Gangestad, S. W., Simpson, J. A., Cousins, A. J., Garver-Apgar, C. E., & Christensen, P. N. (2004). Women's preferences for male behavioral displays change across the menstrual cycle. *Psychological Science, 15,* 203–207. https://doi.org/10.1111/j.0956-7976.2004.01503010.x

Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2005). Adaptations to ovulation: Implications for sexual and social behavior. *Current Directions in Psychological Science, 14,* 312–316. https://doi.org/10.1111/j.0963-7214.2005.00388.x

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.

Garza, R., & Byrd-Craven, J. (2019). Fertility status in visual processing of men's attractiveness. *Evolutionary Psychological Science, 5,* 328-342. https://doi.org/10.1007/s40806-019-00190-4

Gelber, R. P., Gaziano, J. M., Orav, E. J., Manson, J. E., Buring, J. E., & Kurth, T. (2008). Measures of obesity and cardiovascular risk among men and women. *Journal of the American College of Cardiology, 52,* 605-615. https://doi.org/10.1016/j.jacc.2008.03.066

Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014). Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. *Psychological Bulletin, 140,* 1205–1259. https://doi.org/10.1037/a0035438

Havlíček, J., Cobey, K. D., Barrett, L., Klapilová, K., & Roberts, S. C. (2015). The spandrels of Santa Barbara? A new perspective on the peri-ovulation paradigm. *Behavioral Ecology, 26,* 1249-1260. https://doi.org/10.1093/beheco/arv064

Higham, J. (2019). A comparative perspective on measures of cycle phase, and how they relate to cues, signals and mating behavior: A commentary on Gangestad, Dinh, Grebe, Del Giudice, and Emery Thompson (2019). *Evolution and Human Behavior, 40,* 533-536.

Jones, B. C., Hahn, A. C., & DeBruine, L. M. (2019a). Ovulation, Sex Hormones and Women's Mating Psychology. *Trends in Cognitive Sciences,* 23, 51-62. https://doi.org/10.1016/j.tics.2018.10.008

Jones, B. C., Marcinkowska, U. M., & DeBruine, L. M. (2019b). Assessing the evidentiary value of secondary data analyses: A commentary on Gangestad, Dinh, Grebe, Del Giudice and Emery Thompson (2019). *Evolution and Human Behavior, 40,* 531-532.

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., … DeBruine, L. M. (2018a). No compelling evidence that preferences for facial masculinity track changes in women's hormonal status. *Psychological Science*, *29*, 996-1005. https://doi.org/10.1177/0956797618760197

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., & DeBruine, L. M. (2018b). 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

Jünger, J., Kordsmeyer, T., Gerlach, T. M., & Penke, L. (2018a). Fertile women evaluate male bodies as more attractive, regardless of masculinity. *Evolution and Human Behavior*, *39*, 412–423. https://doi.org/10.1016/j.evolhumbehav.2018.03.007

Jünger, J., Motta-Mena, N. V., Cardenas, R., Bailey, D., Rosenfield, K. A., Schild, C., Penke, L., & Puts, D. A. (2018b). Do women's preferences for masculine voices shift across the ovulatory cycle? *Hormones and Behavior*, *106*, 122-134. https://doi.org/10.1016/j.yhbeh.2018.10.008

Kiesner, J., Eisenlohr-Moul, T., & Mendle, J. (2020). Evolution, the Menstrual Cycle, and Theoretical Overreach. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691620906440

Kordsmeyer, T. L., Hunt, J., Puts, D. A., Ostner, J., & Penke, L. (2018). The relative importance of intra-and intersexual selection on human male sexually dimorphic traits. *Evolution and Human Behavior*, *39*, 424-436.

Kordsmeyer, T. L., & Penke, L. (2019). Effects of male testosterone and its interaction with cortisol on self-and observer-rated personality states in a competitive mating context. *Journal of Research in Personality*, *78*, 76-92. https://doi.org/10.1016/j.jrp.2018.11.001

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

Lee, A. J., Mitchem, D. G., Wright, M. J., Martin, N. G., Keller, M. C., & Zietsch, B. P. (2014). Genetic factors that increase male facial masculinity decrease facial attractiveness of female relatives. *Psychological Science*, *25*, 476-484.

Lewis, M. B. (2020). Challenges to both reliability and validity of masculinity-preference measures in menstrual-cycle-effects research. *Cognition*, *197*, 104201.

Little, A. C., Jones, B. C., & Burriss, R. P. (2007). Preferences for masculinity in male bodies change across the menstrual cycle. *Hormones and Behavior*, *51*, 633–639. https://doi.org/10.1016/j.yhbeh.2007.03.006

Lüdecke, D. (2018). sjPlot: Data visualization for statistics in social science. *R package version*, *2*(1).

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

Marcinkowska, U. M., Galbarczyk, A., & Jasienska, G. (2018a). La donna è mobile? Lack of cyclical shifts in facial symmetry, and facial and body masculinity preferences – A hormone based study. *Psychoneuroendocrinology, 88,* 47-53.

Marcinkowska, U. M., Kaminski, G., Little, A. C., & Jasienska, G. (2018b). Average ovarian hormone levels, rather than daily values and their fluctuations, are related to facial preferences among women. *Hormones and Behavior, 102,* 114-119. https://doi.org/10.1016/j.yhbeh.2018.05.013

Newman, J. D. & Handelsman, D. J. (2014). Challenges to the measurement of oestradiol: comments on an endocrine society position statement. *The Clinical Biochemist Reviews, 35,* 75.

Pillsworth, E. G., Haselton, M. G., & Buss, D. M. (2004). Ovulatory shifts in female sexual desire. *Journal of Sex Research, 41,* 55-65. https://doi.org/10.1080/00224490409552213

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.

Penton-Voak, I. S., Perrett, D. I., Castles, D. L., Kobayashi, T., Burt, D. M., Murray, L. K., & Minamisawa, R. (1999). Menstrual cycle alters face preference. *Nature, 399,* 741–742. https://doi.org/10.1038/21557

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior, 26,* 388–397. https://doi.org/10.1016/j.evolhumbehav.2005.03.001

Puts, D. A. (2006). Cyclic variation in women's preferences for masculine traits: Potential hormonal causes. *Human Nature, 17,* 114–127. https://doi.org/10.1007/s12110-006-1023-x

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

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

Roney, J.R. (2018a). Functional roles of gonadal hormones in human pair bonding and sexuality. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge International Handbook of Social Neuroendocrinology* (pp. 239–255). New York, NY: Routledge

Roney, J. R. (2018b). 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. (2015). Elevated psychological stress predicts reduced Estradiol concentrations in young women. *Adaptive Human Behavior and Physiology, 1,* 30-40. https://doi.org/10.1007/s40750-014-0004-2

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

Roney, J. R., & Simmons, Z. L. (2017). Ovarian hormone fluctuations predict within-cycle shifts in women's food intake. *Hormones and Behavior, 90,* 8–14. https://doi.org/10.1016/j.yhbeh.2017.01.009

Rosner, W., Hankinson, S. E., Sluss, P. M., Vesper, H. W., & Wierman, M. E. (2013). Challenges to the measurement of estradiol: an endocrine society position statement. *The Journal of Clinical Endocrinology & Metabolism, 98,* 1376-1387. https://doi.org/10.1210/jc.2012-3780

Schultheiss, O. C., Dlugash, G., & Mehta, P. H. (2019). Hormone measurement in social neuroendocrinology: A comparison of immunoassays and mass spectroscopy methods. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge International Handbook of Social Neuroendocrinology* (pp. 26-40). New York, NY: Routledge

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

Schwartz, D., MacDonald, P. D. M., & Heuchel, V. (1980). Fecundability, coital frequency and the viability of ova. *Population Studies, 34,* 397-400.

Scott, I. M., Clark, A. P., Josephson, S. C., Boyette, A. H., Cuthill, I. C., Fried, R. L., ... & Honey, P. L. (2014). Human preferences for sexually dimorphic faces may be evolutionarily novel. *Proceedings of the National Academy of Sciences, 111,* 14388-14393.

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B, 276,* 575-584. https://doi.org/10.1098/rspb.2008.1177

Shirazi, T., Jones, B. C., Roney, J., DeBruine, L., & Puts, D. (2019). Conception risk affects in-and extra-pair desire similarly: a comment on Shimoda et al. (2018). *Behavioral Ecology, 30,* e6-e7. https://doi.org/10.1093/beheco/arz063

Spector, I. P., Carey, M. P., & Steinberg, L. (1996). The Sexual Desire Inventory: Development, factor structure, and evidence of reliability. *Journal of Sex & Marital Therapy, 22,* 175-190.

Stern, J., Arslan, R. C., Gerlach, T. M., & Penke, L. (2019). No robust evidence for cycle shifts in preferences for men's bodies in a multiverse analysis: Commentary on

Gangestad, Dinh, Grebe, Del Giudice and Emery Thompson (2019). *Evolution and Human Behavior*, *40*, 517-525. https://doi.org/10.1016/j.evolhumbehav.2019.08.005

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

Treffenstaedt, C., & Wiemann, P. (2018). Alfred - A library for rapid experiment development (Version 0.2b5).

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

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*, 1-29. doi:10.18637/jss.v040.i01

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

Wilcox, A. J., Weinberg, C. R., & Baird, D. D. (1998). Post-ovulatory ageing of the human oocyte and embryo failure. *Human Reproduction*, *13*, 394-397. https://doi.org/10.1093/humrep/13.2.394

Wood, W., Kressel, L., Joshi, P. D., & Louie, B. (2014). Meta-analysis of menstrual cycle effects on women's mate preferences. *Emotion Review*, *6*, 229–249. https://doi.org/10.1177/1754073914523073