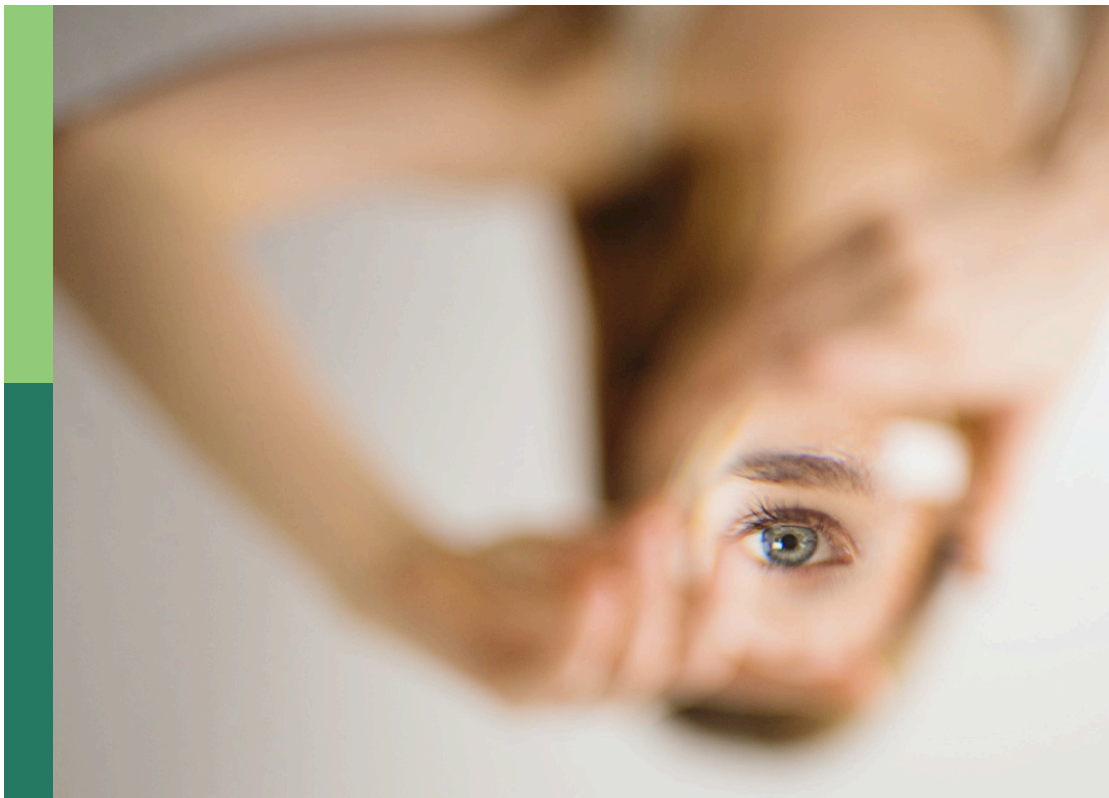
frontiers | Research Topics

Hormones and person perception

Edited by
Lisa L. M. Welling, Amanda Hahn and
Iris J. Holzleitner

Published in
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-6531-5
DOI 10.3389/978-2-8325-6531-5

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Hormones and person perception

Topic editors

Lisa L. M. Welling — Oakland University, United States

Amanda Hahn — Cal Poly Humboldt, United States

Iris J. Holzleitner — University of the West of England, United Kingdom

Citation

Welling, L. L. M., Hahn, A., Holzleitner, I. J., eds. (2025). *Hormones and person perception*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6531-5

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 18 June 2025
DOI 10.3389/fpsyg.2025.1637860

OPEN ACCESS

EDITED AND REVIEWED BY
Barnaby James Wyld Dixson,
The University of Queensland, Australia

*CORRESPONDENCE
Lisa L. M. Welling
welling@oakland.edu

RECEIVED 29 May 2025
ACCEPTED 05 June 2025
PUBLISHED 18 June 2025

CITATION
Welling LLM, Hahn AC and Holzleitner IJ (2025) Editorial: Hormones and person perception. *Front. Psychol.* 16:1637860. doi: 10.3389/fpsyg.2025.1637860

Editorial: Hormones and person perception

Lisa L. M. Welling[1]*, Amanda C. Hahn[2] and Iris J. Holzleitner[3]

[1]Department of Psychology, Oakland University, Rochester, MI, United States, [2]Department of Psychology, Cal Poly Humboldt, Humboldt, CA, United States, [3]School of Social Sciences, University of the West of England Bristol, Bristol, United Kingdom

KEYWORDS

hormones, person perception, testosterone, cortisol, progesterone, estrogen

Editorial on the Research Topic
Hormones and person perception

Recent research bridging behavioral endocrinology and social psychology has refined our understanding of how hormones influence person perception (see Welling and Shackelford, 2019). This Research Topic on *Hormones and Person Perception* brings together empirical studies, methodological insights, and reviews advancing understanding of how hormones can influence perceptions and expressions of traits like attractiveness, health, and social behaviors related to relationship dynamics. The nine accepted papers reflect diverse approaches. Three priorities emerge: hormonal effects involve complex interactions with individual differences and social cues, methodological rigor is vital, and research should integrate genetic, behavioral, and environmental factors to better capture biological foundations of social cognition.

Context-dependent effects

Goetz et al. tested whether exogenous testosterone affects men's perceptions of a woman's sexual interest. The sexual misperception bias (SMB), where men overestimate women's sexual interest, has been interpreted through Error Management Theory (Haselton and Buss, 2000), which proposes that, for men, misperceiving sexual interest carried fewer reproductive costs ancestrally than did missing potential mating opportunities (Buss, 2001). In this placebo-controlled experiment, testosterone administration did not increase SMB, but did heighten sensitivity to affiliative cues, especially among men with average or higher self-perceived attractiveness. Testosterone-treated men who interpreted a woman's behavior as affiliative perceived greater sexual interest, an effect not observed among placebo-treated men. These results suggest testosterone functions as a social hormone shaping perceptual biases in specific contexts.

Donovan and Corpuz explored how testosterone, cortisol, and relationship satisfaction relate in first-time fathers during the postpartum period. Fathers with high testosterone and low cortisol reported higher relationship satisfaction, although this effect was small. The authors suggest that traits often associated with high testosterone, such as dominance and status-seeking (reviewed in Dekkers et al., 2019), may also influence relationship quality beyond mate acquisition. This study suggests complex hormonal influences on relationship dynamics and calls for more precise, multi-method, longitudinal research surrounding the transition to parenthood.

Using an eye-tracking paradigm, Garza and Byrd-Craven investigated women's visual attention to facial masculinity across the menstrual cycle. Contrary to the Ovulatory Shift Hypothesis (Gangestad and Thornhill, 1998), although women spent more time viewing masculine than feminine faces, particularly in a long-term mating context, no hormone measures predicted visual attention to or rated preference for masculine faces across the menstrual cycle. There was partial evidence linking women's hormone levels to visual attention. Higher estradiol to progesterone ratios were associated with shorter first fixations, whereas lower progesterone predicted greater visual attention to male faces in a short-term mating context. Also, higher estradiol levels were related to more overall visual fixations. Findings emphasize that hormonal influences on social cognition may be subtler or more context-dependent than assumed.

Similarly, Lobmaier et al. recorded women during high- and low-fertility phases while reading and reproducing spoken sentences from male and female speakers rated as either attractive or unattractive. Vocal parameters varied by cycle phase, in response to the stimulus speaker's vocal attractiveness and sex, and when reproducing spoken sentences compared to reading written sentences. Women also used breathier, higher-frequency voices when responding to attractive voices, consistent with social mimicry research (Chartrand and Lakin, 2013). In contrast, Friedrich et al. found no cycle phase or hormonal effects on voice-gender categorization, though participants responded faster to feminine voices (see also Lattner et al., 2005). These null results align with emerging reports of weak or inconsistent cycle effects on female social cognition (e.g., Garza and Byrd-Craven, 2019; Jones et al., 2018). Inconsistent findings highlight the need for precise hormone measurement, methodological rigor, and considering other potential sources of variation across studies.

Methodological rigor

Hampson et al. assessed depression in women during the active hormone phase and the hormone-free "washout" week of their contraceptive cycles. The study combined explicit self-reports with implicit measures of depressed affect, finding that implicit measures yielded a pattern of increased depressive affect during active hormone intake, particularly among those who report higher average levels of depressive affect. In contrast, explicit self-reports indicated that participants perceived greater depressive affect when taking inactive pills containing no synthetic hormones. These findings demonstrate the importance of implicit measures for capturing mood effects not detected in self-report measures (e.g., DeCoster et al., 2006) and suggest OC-related mood effects may be most evident in those prone to depression.

Updating earlier work (Grimbos et al., 2010), Swift-Gallant et al. conducted a comprehensive meta-analysis examining second-to-fourth digit ratio (2D:4D), which is thought to be a marker of prenatal androgen exposure (see Swift-Gallant et al., 2020), and sexual orientation. The authors found that homosexual women tend to have lower (more male-typical) digit ratios than heterosexual women, whereas homosexual men exhibited higher (more female-typical) digit ratios than heterosexual men. No significant differences were found for bisexual individuals. These findings add to research on prenatal androgen exposure in sexual orientation development (Swift-Gallant et al., 2021) and suggest future work include multiple biological measures and nuanced sexual orientation categories.

Expanding conceptual frameworks

Gurguis et al. proposed a quantitative genetic framework for studying the evolutionary dynamics of hormone-mediated traits, focusing on person perception and psychiatric conditions using estrogen as an example. The authors argue that person perception is part of broader hormone-regulated suites and should be studied within multivariate evolutionary models that account for genetic correlations among traits (McGlothlin and Ketterson, 2008). Quantitative genetics techniques could test hypotheses about selection on hormone-mediated traits and clarifying trade-offs, like estrogen's dual influence on reproductive fitness (Mittal et al., 2014) and disease risk (Chuffa et al., 2017). This framework offers a promising tool for connecting evolutionary biology, psychology, and psychiatry.

Arnocky and Davis investigated whether male facial attractiveness is related to health through shared hormonal and lifestyle factors, as suggested by Jones et al. (2021), rather than serving as a direct cue to immunocompetence. They measured immunoglobulin A, testosterone, cortisol, lifestyle behaviors, abdominal skinfold, and self-reported health alongside female-rated facial attractiveness. Abdominal skinfold and symptoms of poorer health predicted lower facial attractiveness, and mediated the relationships between exercise, stress, and facial attractiveness. Men with higher testosterone and lower cortisol tended to have more attractive faces, but this was not statistically significant. Results suggest facial attractiveness-immunocompetence associations may partly reflect shared lifestyle and hormonal factors, although inclusion of broader lifestyle measures is warranted.

Conclusion

The above research offers valuable insights into how hormones shape person perception and introduce new priorities for future research. Future work should improve hormone measurement, broaden conceptual frameworks, and examine how individual differences moderate hormonal effects. Together, the published works in this Research Topic underscore the importance of considering hormones in person perception research and advance our understanding of the biological foundations of social cognition.

Author contributions

LW: Writing – review & editing, Conceptualization, Writing – original draft. AH: Writing – review & editing. IH: Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Buss, D. M. (2001). Cognitive biases and emotional wisdom in the evolution of conflict between the sexes. *Curr. Dir. Psychol. Sci.* 10, 219–223. doi: 10.1111/1467-8721.00153

Chartrand, T. L., and Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annu. Rev. Psychol.* 64, 285–308. doi: 10.1146/annurev-psych-113011-143754

Chuffa, L. G., Lupi-Junior, L. A., Costa, A. B., Amorim, J. P., and Seiva, F. R. (2017). The role of sex hormones and steroid receptors on female reproductive cancers. *Steroids* 118, 93–108. doi: 10.1016/j.steroids.2016.12.011

DeCoster, J., Banner, M. J., Smith, E. R., and Semin, G. R. (2006). On the inexplicability of the implicit: Differences in the information provided by implicit and explicit tests. *Soc. Cogn.* 24, 5–21. doi: 10.1521/soco.2006.24.1.5

Dekkers, T. J., van Rentergem, J. A. A., Meijer, B., Popma, A., Wagemaker, E., and Huizenga, H. M. (2019). A meta-analytical evaluation of the dual-hormone hypothesis: does CORT moderate the relationship between testosterone and status, dominance, risk taking, aggression, and psychopathy? *Neurosci. Biobehav. Rev.* 96, 250–271. doi: 10.1016/j.neubiorev.2018.12.004

Gangestad, S. W., and Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proc. Royal Soc. London: Biol. Sci.* 265, 927–933. doi: 10.1098/rspb.1998.0380

Garza, R., and Byrd-Craven, J. (2019). Fertility status in visual processing of men's attractiveness. *Evol. Psychol. Sci.* 5, 328–342. doi: 10.1007/s40806-019-00190-4

Grimbos, T., Dawood, K., Burriss, R. P., Zucker, K. J., and Puts, D. A. (2010). Sexual orientation and the second to fourth finger length ratio: a meta-analysis in men and women. *Behav. Neurosci.* 124:278. doi: 10.1037/a0018764

Haselton, M. G., and Buss, D. M. (2000). Error management theory: a new perspective on biases in cross sex mind reading. *J. Pers. Soc. Psychol.* 78:81. doi: 10.1037/0022-3514.78.1.81

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., et al. (2018). No compelling evidence that preferences for facial masculinity track changes in women's hormonal status. *Psychol. Sci.* 29, 996–1005. doi: 10.1177/0956797618760197

Jones, B. C., Holzleitner, I. J., and Shiramizu, V. (2021). Does facial attractiveness really signal immunocompetence? *Trends in Cognit. Sci.* 25, 1018–1020. doi: 10.1016/j.tics.2021.09.003

Lattner, S., Meyer, M. E., and Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Hum. Brain Mapp.* 24, 11–20. doi: 10.1002/hbm.20065

McGlothlin, J. W., and Ketterson, E. D. (2008). Hormone-mediated suites as adaptations and evolutionary constraints. *Philos. Trans. R. Soc. B, Biol. Sci.* 363, 1611–1620. doi: 10.1098/rstb.2007.0002

Mittal, S., Gupta, P., Malhotra, N., and Singh, N. (2014). Serum estradiol as a predictor of success of in vitro fertilization. *J. Obstet. Gynecol. India* 64, 124–129. doi: 10.1007/s13224-013-0470-7

Swift-Gallant, A., Di Rita, V., Major, C. A., Breedlove, C. J., Jordan, C. L., and Breedlove, S. M. (2021). Differences in digit ratios between gay men who prefer receptive versus insertive sex roles indicate a role for prenatal androgen. *Sci. Rep.* 11:8102. doi: 10.1038/s41598-021-87338-0

Swift-Gallant, A., Johnson, B. A., Di Rita, V., and Breedlove, S. M. (2020). Through a glass, darkly: Human digit ratios reflect prenatal androgens, imperfectly. *Horm. Behav.* 120:104686. doi: 10.1016/j.yhbeh.2020.104686

Welling, L. L. M., and Shackelford, T. K. (2019). *The Oxford Handbook of Evolutionary Psychology and Behavioral Endocrinology*. New York, NY: Oxford University Press.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 February 2023
DOI 10.3389/fpsyg.2023.1067487

OPEN ACCESS

EDITED BY
Darren Burke,
The University of Newcastle, Australia

REVIEWED BY
Maria Alessandra Umilta',
University of Parma,
Italy
Fay A. Guarraci,
Southwestern University,
United States

*CORRESPONDENCE
Ray Garza
✉ ray.garza@tamiu.edu

SPECIALTY SECTION
This article was submitted to Evolutionary Psychology, a section of the journal Frontiers in Psychology

RECEIVED 11 October 2022
ACCEPTED 16 January 2023
PUBLISHED 13 February 2023

CITATION
Garza R and Byrd-Craven J (2023) The role of hormones in attraction and visual attention to facial masculinity.
Front. Psychol. 14:1067487.
doi: 10.3389/fpsyg.2023.1067487

The role of hormones in attraction and visual attention to facial masculinity

Ray Garza[1]* and Jennifer Byrd-Craven[2]

[1]Department of Psychology and Communication, Texas A&M International University, Laredo, TX, United States, [2]Oklahoma Center for Evolutionary Analysis, Department of Psychology, Oklahoma State University, Stillwater, OK, United States

The current study investigated the ovulatory shift hypothesis, which suggests that women prefer more masculine traits when estradiol is high, and progesterone is low (E/P ratio). The current study used an eye tracking paradigm to measure women's visual attention to facial masculinity across the menstrual cycle. Estradiol (E) and progesterone (P) were collected to determine if salivary biomarkers were associated with visual attention to masculine faces in a short- and long-term mating context. Women (N=81) provided saliva samples at three time points throughout their menstrual cycle and were asked to rate and view men's faces that had been manipulated to appear feminine and masculine. Overall, masculine faces were viewed longer compared to feminine faces and this was moderated by mating context, where women viewed masculine faces longer for a long-term relationship. There was not any evidence suggesting that E/P ratio was associated with preferences for facial masculinity, but there was evidence to suggest that hormones were associated with visual attention to men in general. In line with sexual strategies theory, there was evidence to suggest that mating context and facial masculinity are important in mate choice; however, there was no evidence to suggest that women's mate choice was associated with shifts across the menstrual cycle.

KEYWORDS

eye-tracking, mate choice, sexual selection, ovulatory shift hypothesis, attention, attraction

1. Introduction

According to sexual selection theory, members of one sex should be sensitive to cues that advertise important reproductive information about the opposite sex in mate choice (Andersson, 1994). In sexually reproducing species, male ornamentation has evolved due to a variety of factors, such as ecological influences, within-sex competition, and through female choice. In addition, female preferences for exaggerated secondary sexual characteristics have evolved because of indicator mechanisms that signal high heritability and direct phenotypic benefits, such as protection and parental ability (Andersson, 1994). In human males, exaggerated secondary sexual characteristics, such as facial masculinity, are preferred by some women (DeBruine et al., 2006; Boothroyd et al., 2013; Marcinkowska et al., 2018b) perhaps due to its association with immunocompetence (Zahavi, 1975; Folstad and Karter, 1992; Thornhill and Gangestad, 1999a, 1999b; Little et al., 2011a; Rantala et al., 2012). Some suggest that women should have a stronger preference for sexually dimorphic traits in men when conception is optimal (Gangestad and Thornhill, 1998; Gangestad and Haselton, 2015). However, ovulation in human populations is concealed to potential mates and to the female herself (Alexander and Noonan, 1979; Alexander, 1990), which may indicate that preferences across the menstrual cycle may not be under direct conscious awareness. Methods that are sensitive in detecting

subtle behavioral changes, such as eye movements, can demonstrate that women display implicit attentional biases toward masculinity. This study uses an eye tracking paradigm to investigate whether preferences for facial masculinity are influenced by reproductive hormones throughout the menstrual cycle.

Both men and women have faced adaptive challenges in mate selection (Buss and Schmitt, 1993, 2019). Humans have evolved psychological adaptations in response to many facets of mate selection. For women, identifying which men will make good partners has been one of those challenges, as women have had to discern which mates will provide them with indirect benefits in the form of high-quality genes for their offspring, direct benefits in the form of resource protection and status transmission, and/or a mate who is invested in long-term pair bonding. Since women have more to lose in making a poor mate choice, and given the long, intensive parenting effort of human life histories (Flinn et al., 2007), assessing these features in human mating are important. Women have been known to place premiums on characteristics that signal good genes, such as attractiveness and health (Buss, 1989; Cashdan, 1996) and characteristics that signal immediate resource transmission and future resource acquisition (Cashdan, 1996). One principle of sexual strategies theory is that women pursuing men with these features would have obtained benefits from short-term mating, as pursuing men with these features for long-term relationships would put women and their children at risk for abandonment and increased competition from other women (Buss and Schmitt, 1993, 2019). Since short-term mating requires minimal investment, benefits obtained (i.e., superior genes, resources) would have had to outweigh the costs of not pursuing a mate for a long-term commitment.

Facial cues are directly observable characteristics that provide women hormonal information about men. Androgen levels are associated with exaggerations of secondary sexual characteristics, such as the brow ridges and jaw (Thornhill and Gangestad, 1996). Advertisement of these traits should signal indirect benefits (i.e., good genes) to women because of testosterone's immunosuppression. Using this framework, research has demonstrated that women tend to prefer masculine over feminine faces (Johnston et al., 2001; DeBruine et al., 2006; Cornwell and Perrett, 2008; Little et al., 2008; Boothroyd et al., 2013) and find them sexually attractive (Marcinkowska et al., 2018b). Masculine faces have been associated with immunocompetence (Rhodes et al., 2003; Rantala et al., 2012), disease resistance (Thornhill and Gangestad, 2006), and strength (Windhager et al., 2011). Facial masculinity has also been associated with direct benefits, such as protection and resource acquisition (Sell et al., 2009). Men with masculine characteristics display more muscularity, dominance, and physical strength (Fink et al., 2007; Puts, 2010; Windhager et al., 2011). However, there is a tradeoff between securing a mate with good genes or securing a mate that is willing to invest in offspring (Marcinkowska et al., 2018a). Men with masculine features are perceived as less parentally investing (Perrett et al., 1998), and more willing to engage in short-term relationships (Rhodes, 2006; Boothroyd et al., 2008). Therefore, preferences for masculinity may be calibrated to maximize the likelihood and need of obtaining indirect (i.e., good genes) or direct benefits (i.e., resource acquisition, protection).

The ovulatory shift hypothesis (Gangestad and Thornhill, 1998) suggests that shifts in women's hormones across the menstrual cycle are associated with shifting sexual preferences. It asserts that this heightened preference would have been associated with increased reproductive success in ancestral females compared to those who did not exhibit a cyclic shift in preferences (Gildersleeve et al., 2014). It would also indicate that women's preferences would be attenuated outside of the fertile window due to the cost of losing social mates or in choosing the wrong mate (Gangestad et al., 2005). Research into fertility status in women's mate preferences has suggested women prefer men who are symmetrical because of its association with health and immunocompetence (Thornhill and Gangestad, 1999a). Women in the high fertility phase display preferences for men with masculine faces compared to women who are at the low fertile phase of the menstrual cycle (Penton-Voak and Perrett, 2000; Little et al., 2007; Dixson et al., 2018a). However, other studies have not shown an ovulatory shift for preferences to men's masculine characteristics. Cyclic shifts did not play a role in rating men's facial and body masculinity (Marcinkowska et al., 2018c). Women during the fertile phase of the menstrual cycle have shown to rate all men as equally attractive, regardless of masculine traits (i.e., waist to chest/shoulder ratios) (Garza et al., 2017; Jünger et al., 2018a; Garza and Byrd-Craven, 2019). Other research on male traits signaling masculinity (i.e., degree of hair distribution and beardedness) (Gangestad and Thornhill, 2008) have not shown that women shift their preferences for these traits across the fertile period (Rantala et al., 2010; Dixson and Rantala, 2016; Garza et al., 2017; Dixson et al., 2018b). In examining vocal masculinity, women's preferences for masculine voices were not associated with ovulatory status (Jünger et al., 2018b).

One issue in research investigating ovulatory shifts in men's mating preferences has been the accuracy of methods used. Traditionally, self-report methods using the calendar method, where women count the days from their previous menstrual period, have been used to dichotomize women into low or high fertility conceptive probability. The use of precise methods in detecting fertility status have become more common practice. During the pre-ovulatory phase of the menstrual cycle, women experience a rise in estradiol and a decline in progesterone (Roney and Simmons, 2013). By tracking the ratio of estradiol to progesterone, it is possible to determine the increased likelihood of ovulation, and if women's shifts for specific mate preferences were to occur, they should occur during a rise in the E/P ratio. The estradiol to progesterone ratio (i.e., E/P Ratio) is a recommended method in detecting fertility status (Blake et al., 2016; Gangestad et al., 2016) because during the late-follicular phase estradiol is expected to be higher than progesterone than any other point in the menstrual cycle, therefore higher values are an indicator of increased likelihood of ovulation, while lower values are an indicator of post-ovulation. Findings for hormonal shifts in preferences for masculinity have been equivocal. Preferences for masculinity have been associated with increased levels of estradiol (Roney and Simmons, 2008), while others have not found any preferences using luteinizing hormone (Peters et al., 2009). Recent studies on mate preferences for masculine faces using the E/P ratio have not revealed a preference during peak fertility (Marcinkowska et al., 2016, 2018a). Progesterone has shown to be predictive of masculinity preferences as a function of relationship status (Marcinkowska et al., 2018a). Partnered women's progesterone levels were related to a weaker preference for masculinity. Since increased progesterone levels are associated with pregnancy (Gilbert, 2000), when progesterone is high, women may direct their attention to parenting instead of honest signals of good genes (Marcinkowska et al., 2018a). Although these studies have relied extensively on preferences tasks, such as choosing which face (i.e., masculine vs. feminine) is preferred, it is not yet understood if there are any hormonal influences in the way that women process these features visually.

Although previous research has depended on women's stated preferences for masculinity, eye tracking has provided researchers with a behavioral measurement of these preferences. Eye tracking is a sensitive gaze contingency technique that provides real-time visual processing, such as implicit and explicit measurements. Eye tracking procedures are advantageous over traditional preference tasks, as they are less susceptible to experimenter expectancy, they correlate with self-reported preferences, and they provide information to smaller features (i.e., regions of interest) that can help investigate the nuances associated with mate preferences (Krupp, 2008). Eye tracking research in women's assessments of sexually dimorphic (i.e., masculine/feminine) faces has been limited. In using both male and female participants in a Chinese sample, Wen and Zuo (2012) showed that participants fixated first at masculine faces, suggesting that masculine features are important in early visual processing. However, since their research combined data from both men and women in tracking their visual movements, it is likely that there were differences in the interpretation of masculine vs. feminine faces (i.e., men may perceive men as dominant/aggressive). In a similar study, albeit using masculinity and attractiveness, Yang et al. (2015) found that masculine faces were preferred for first fixation duration and total visual time, but it was dependent on their level of attractiveness. In using female participants only, Burris et al. (2014) found that women focused their visual attention to feminine rather than masculine faces using visual metrics that account for early stage (i.e., first fixation duration) and late stage (i.e., total dwell time) processing. However, given the differences that exist between Chinese and European populations in masculinity preferences, it can be argued that there are cultural differences in preferences, as both studies using Chinese populations found preferences for masculinity in eye tracking designs, while in a European sample, preferences were for feminine faces. Although not explored, one possible explanation for preferences in the Chinese and European study could be the role of population density. Denser populations may rely on heuristics and exaggerated dimorphic characteristics due to the frequent exchange of social encounters (Scott et al., 2014; Marcinkowska et al., 2018b).

The current study aimed to address issues raised by previous studies investigating the ovulatory shift hypothesis by using an eye tracking paradigm to track women's visual preferences to sexually dimorphic faces. Although recent studies have not found an effect for fertility status using a forced choice task in preferences, the current study addressed whether shifts in hormones affect women's visual preferences by tracking eye movements across different times of the menstrual cycle. This study addressed the following research questions: (1) Do hormones moderate the relationship between sexual dimorphism and attraction/visual attention? It is hypothesized that as estradiol increases and progesterone decreases (i.e., high E/P ratio), women will rate masculine faces more attractive and view them longer, (2) Do hormones affect preferences for masculine features depending on mating context? As a direct test of the sexual strategies theory (Buss and Schmitt, 1993), we investigate if women have evolved psychological mechanisms for short- and long-term mate preferences by tracking visual attention to masculine and feminine faces in these different contexts. Research has suggested that women prefer masculine men for short-term relationships (Little et al., 2002, 2007; Little et al., 2011b), however, tracking reproductive hormones has not been investigated using an eye tracking paradigm. It is hypothesized that women with high E/P ratios will view masculine faces longer when considering men for a short-term relationship.

2. Materials and methods

2.1. Participants

A G*Power analysis in detecting a small to moderate effect size indicated a sample size of approximately 71 participants. Further, we relied on recommendations from Gangestad et al. (2016) in estimating an appropriate sample size for a within-subjects designs testing the ovulatory shift hypothesis. Participants were 81 heterosexual women from Oklahoma State University ($M_{age} = 19.27$, $SD_{age} = 2.83$) who signed up on the university's online participant recruitment system to participate in the three-part study in exchange for course credit. Participants were only allowed to sign up for the study if they were not on any hormonal-based birth control, were not pregnant, did not smoke, and identified primarily as heterosexual. The sample demographics were White ($N = 55$), African-American ($N = 9$), Hispanic ($N = 5$), Native-American ($N = 5$), Asian-American ($N = 4$), and Other ($N = 2$).

2.2. Measures

2.2.1. Sexually dimorphic faces

The sexually dimorphic stimuli used were from the London Face Lab, which include a composite of masculinized and feminized faces of the same individual that have been morphed to indicate −50% femininity and + 50% masculinity (DeBruine and Jones, 2017).

2.2.2. Mating context prompts

Two mating context prompts adopted from Jones et al. (2018) were used to connote information on a short-term and long-term relationship. In the short-term-attractiveness test, women were given the following information: "You are looking for the type of person who would be attractive in a short-term relationship. This implies that the relationship may not last a long time. Examples of this type of relationship would include a single date accepted on the spur of the moment, an affair within a long-term relationship, and possibility of a one-night stand." In the long-term-attractiveness test, women were given the following information: "You are looking for the type of person who would be attractive in a long-term relationship. Examples of this type of relationship would include someone you may want to move in with, someone you may consider leaving a current partner to be with, and someone you may, at some point, wish to marry (or enter into a relationship on similar grounds as marriage)."

2.2.3. Hormonal assays

On the day of salivary analysis, saliva samples were thawed for 1 ½ hrs and then they were centrifuged at 1,500 rpm's for 15 min. Estradiol and Progesterone were assayed using enzyme linked immunosorbent assays (ELISA) following Salimetrics protocols which required a serial dilution for standards and pipetting samples in duplicates (Estradiol: 100 µL, Progesterone: 50 µL). The standards for estradiol were 32, 16, 8, 4, 2, and 1 pg/µL, and for progesterone the standards were 2,430, 810, 270, 90, 30, and pg/µL. After pipetting each plate, they were set aside for incubation for 1 h, then were washed four times using a Bio-Tek plate washer. We added a TMB solution and incubated each assay for 30 min in the dark. This was followed by adding 50 µL of the stop solution to each well and mixed for 3 min before being

read using a Bio-Tek 808 lx plate reader using the Gen5 software. The intra and inter-assay coefficient (CV) for progesterone was 6.68 and 5.64%, and 7.52 and 6.15% for estradiol.

2.2.4. Eye tracking metrics

Eye tracking data were recorded using a Tobii X2-60, which is a non-invasive eye tracking instrument that records eye movements at 60 frames per second (60 hz). We used three eye tracking metrics, first fixation duration, total visit duration, and number of fixations. First fixation duration was defined as the duration of the first fixation on an interest area (i.e., feminine, masculine face) in milliseconds (ms), and it is often used as an eye tracking metric to indicate saliency upon first view or early-onset processing (Conklin et al., 2018). For late-onset processing, total visit duration and number of fixations were used. Total visit duration measures the duration of time spent on an interest area for each trial, and it is often used as a measure of late-onset or effortful processing (Conklin et al., 2018). The number of fixations (i.e., fixation count) is an eye tracking metric often used to complement total visit duration and it is an alternative method of considering attention (Conklin et al., 2018). The number of fixations is defined as the number of times a fixation is made on an interest area (i.e., feminine, masculine faces). Therefore, first fixation duration can be considered a measurement of early- or automatic processing, while total visit duration and number of fixations can be considered late- or effortful levels of processing. A fixation is a period in which a non-moving stimulus is being viewed (Holmqvist et al., 2011), which is then followed by a saccade. Interest areas were created using the Tobii Lab Pro interest area creator, and we created two interest areas per trial, an interest area encompassing a masculine face and an interest area encompassing a feminine face. The regions of interest included the outline of the faces (i.e., ears, forehead, jawline), and excluded the shoulders, hair, and background of the image.

2.3. Procedure

The institutional review board and Oklahoma State University reviewed and approved this study. Women signed up for the 3-part study using the university's SONA online study sign-up system. The study was announced as, "Attention to Male Images," and it included information as to the requirements to be eligible to sign up. A pre-screener was used to only sample participants who were female and primarily identified themselves as heterosexual. Participants were only allowed to sign up for the study if they had typical menstrual cycles and were not on any hormonal-based contraceptives. They were informed that before each session of the 3-part study, they could not have anything to eat 60 min prior to the study, and to choose a timeslot that was consistent throughout the sessions, such as choosing the same time for each of the three sessions. To capture variability in reproductive hormones, women were asked to participate at three different time periods within a menstrual cycle. Following from Marcinkowska et al. (2018a), participants were instructed to choose a time slot that corresponded to the early follicular phase (days 2–8), ovulation (no later than day 20), and luteal phase (last week of their menstrual cycle). As part of the demographics questionnaire, they also indicated their menstrual cycle status that corresponded to their time of visit. Participants had to wait 7-days before signing up for another session in order to prevent them from signing up for consecutive sessions and to maximize variability in their sample.

Upon entering the laboratory, participants were given a consent form which included information that was addressed on SONA, as well as additional information as to what participants could expect (i.e., demographic questions & surveys). Participants were notified that they would viewing images of men, and their task was to view the images and rate them. They were also informed of the different survey instruments that needed to be completed. Upon consent, participants provided saliva sample through passive drool collection using a saliva collection tube into a saliva collection vial (1.5 mL). All participants provided a saliva sample within 1–3 min. No participant took over 3 min in providing a saliva sample. Once saliva collection was complete, it was stored in a −80°C freezer until the day of salivary analysis.

Participants were then instructed to sit in front of a computer screen where an eye tracking device was magnetically connected to the desktop (Tobii X2-60). The eye tracker was situated within 50 cm of the participant to be able to record eye movements adequately. Before beginning the eye tracking study, participants performed a 5-point visual calibration test which consisted of following a red dot randomly across the screen to 5 positions. When complete with the calibration test, their eyes were checked again to ensure that they were centered on the computer screen, and then they were given instructions to the study. For one block, they were given instructions to view the images of men as if they were looking for a short-term relationship, defined as a relationship that consisted of a one-night stand or casual encounter where no commitment was expected. For the other block, they were instructed to view the images of men as if they were looking for a partner for a long-term relationship, defined as a committed partnership, such as marriage or a relationship lasting months. The use of the mating context prompts was adopted from Jones et al. (2018). Participants were instructed that they were to view images of men and to view the images as they would any image on a computer screen. They were presented with 20 pairs of men for two separate blocks randomly ordered to include a masculinized or feminized version of the same male on either side of the presentation (see Figure 1). They viewed each pair for 3,000 ms, followed by a fixation cross 'X' at the center of the screen which was presented for 500 ms before the screen refreshed and randomly presented another pair. The duration for reading each prompt and viewing 20 pairs of images was between 3 and 5 min for each block. For each of the sessions, participants did not view the blocks in the same order, and the images were counterbalanced to appear on different sides of the screen. In total, they viewed 80 images per session. A sample timeline of each session is shown in Figure 2.

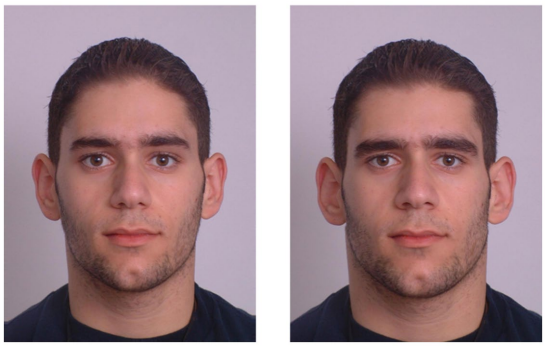


FIGURE 1
Presentation of a male face depicting a femininized (left) and masculinized (right) version. Reproduced with permission from Lisa DeBruine. Available at: https://figshare.com/articles/dataset/Young_Adult_White_Faces_with_Manipulated_Versions/4220517?file=7826521.

Consent	Saliva collection	Demographic forms	Eye-tracking calibration task	Block 1: STM	Block 2: LTM	Rating task
~ 3min	~1-3min	~5min	~1-2 min	~3-5min	~3-5min	~3min

FIGURE 2
Sample timeline for each session. Blocks 1 & 2 were randomized and counterbalanced for each of the three sessions. In total, each session was approximately 25–30 min in duration.

Once complete with the eye tracking portion of the study, participants were instructed to rate the men for their perceived physical attractiveness using a Likert scale where '1 = unattractive' to '7 = extremely attractive.' They provided ratings for men under a short-term and long-term relationship context, in which they viewed each male image sequentially and not in pairs. For each image, the Likert scale rating was presented at the bottom of the screen, and when a rating was given, the screen refreshed another image. Participants were not given a time limit for the attractiveness rating. They then completed a sociodemographic questionnaire which included items about their age, ethnicity, sexual orientation, relationship status, and menstrual cycle status. For the sexual orientation question, participants indicated their sexual orientation by selecting if they identified as heterosexual, homosexual, bisexual, or prefer not to answer. For the menstrual cycle questionnaire, a forward counting method was used to indicate cycle status where participants indicated how many days had passed since the onset of menstrual bleeding. Participants were dismissed from the study and were reminded to return for the subsequent parts (i.e., time 2 & 3). They were sent follow-up emails to remind them of sessions 2 and 3. For time sessions 2, participants were instructed to return no later than the 20th day of their menstrual cycle, and for time session 3, they were instructed to return during the luteal phase of their menstrual cycle or the last week before menstruation. In total, each session was approximately 25–30 min.

3. Results

3.1. Statistical analyses

Data were analyzed using linear mixed-effects models with maximum likelihood using the packages lme4 and lmertest (Bates et al., 2015) in R for dependent variables attractiveness, first fixation duration, total visit duration, and fixation count. Linear mixed-effects models are robust multilevel models that account for variation across subjects, trials, and time varying covariates (i.e., estradiol & progesterone). Since participants were asked to rate and view multiple images across multiple time points, linear mixed effects models are recommended for analyzing designs with multiple repeated observations over time. For each session, participants viewed 80 images (i.e., 40 pairs), which totaled 240 images for all three sessions. Estradiol, progesterone, and E/P ratio were centered on their subject-specific means to interpret within effects (Brauer and Curtin, 2017). Mating context (i.e., STM, LTM), facial masculinity (i.e., feminine, masculine), estradiol, progesterone, estradiol to progesterone ratio (i.e., E/P ratio) were entered as fixed factors, and participants and trials were entered as random effects. All models met the assumptions of normality of residuals, and we performed qq-plots to test those assumptions. Outliers for visual time were screened at ±2.50 z-scores from the mean. For all of our analyses, the R^2 Marginal (i.e., fixed effects) and R^2 Conditional (i.e., random effects: subjects and trials) are reported for our total effect sizes for our models. R^2_{Marginal} and $R^2_{\text{Conditional}}$ effect sizes are to be interpreted as variance accounted for. All post-hoc analyses were conducted using a Bonferroni correction.

3.1.1. Descriptive statistics

Table 1 presents the descriptive statistics for women's hormone levels, estradiol, progesterone, and E/P ratio.

3.1.2. Attractiveness

For attractiveness, the overall variance explained in the model was, $R^2_{\text{Marginal}} = 0.03$, $R^2_{\text{Conditional}} = 0.34$. There was a weak significant effect for progesterone, $b = -0.0007$, $SE = 0.0002$, 95%CI [−0.001, −0.0003], $p < 0.001$. When women's progesterone was higher, ratings of attractiveness were lower. There were no other significant effects or interactions.

3.1.3. First fixation duration

First fixation duration was defined as the average duration of the first fixation to a region of interest (i.e., feminine, masculine face), and it is often used as a measure indicating saliency upon first view. The overall variance explained from the model was, $R^2_{\text{Marginal}} = 0.004$, and $R^2_{\text{Conditional}} = 0.17$. There was a significant effect for women's E/P ratio, $b = -31.55$, $SE = 10.67$, 95%CI [−54.06, −12.21], $p = 0.002$. When women's E/P ratio was higher their first fixation durations were shorter. There were no other significant effects or interactions.

3.1.4. Total visit duration

Total visit duration was defined as the average amount of time spent viewing each region of interest (i.e., feminine, masculine face). The overall variance explained from the model was, $R^2_{\text{Marginal}} = 0.02$, and $R^2_{\text{Conditional}} = 0.24$. There was a significant effect for facial masculinity, $b = 71.44$, $SE = 19.80$, 95%CI [32.64, 110.28], $p < 0.001$. Pairwise comparisons revealed that women viewed masculine faces longer ($M = 881$, $SE = 25$) compared to feminine faces ($M = 843$, $SE = 25$). A main effect for mating context was significant, $b = 85.73$, $SE = 19.91$, 95%CI [43.16, 121.22], $p < 0.001$. Women viewed men longer when considering them for a short-term mating context ($M = 887$, $SE = 25$) compared to a long-term mating context ($M = 838$, $SE = 24.9$). This was further qualified by a facial masculinity by mating context interaction, $b = -65.98$, $SE = 28.10$, 95%CI [−112.11, −10.96], $p = 0.02$. There were significant differences when viewing men's faces as a function of mating context, where women viewed masculine faces longer in a long-term mating context ($M = 873$, $SE = 26.8$) compared to feminine faces ($M = 802$, $SE = 26.8$). When considering a short-term mating context,

TABLE 1 Raw hormone levels for estradiol, progesterone, and E/P ratio across the three sessions.

	Estradiol	Progesterone	E/P Ratio
Session 1	1.64 (1.60)	126 (85.8)	0.01 (0.007)
Session 2	1.45 (0.74)	149 (116)	0.05 (0.34)
Session 3	1.54 (1.05)	145 (93.9)	0.01 (0.008)

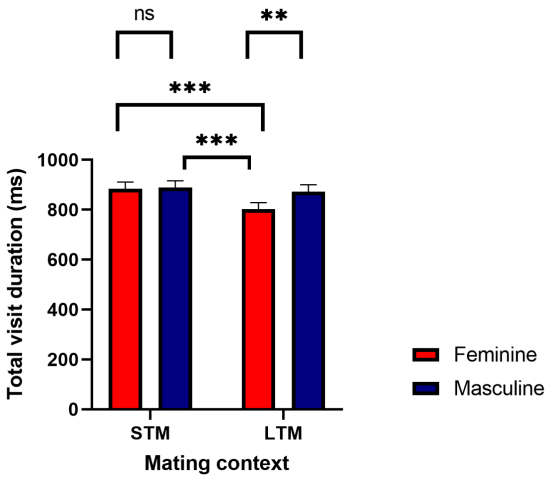


FIGURE 3
Women's total visit duration in milliseconds as a function of facial masculinity and mating context. *p<0.05, **p<0.01, ***p<0.001.

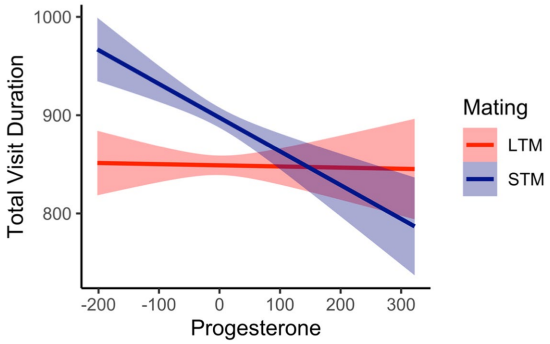


FIGURE 4
Interaction between women's progesterone levels (cluster mean centered) and mating context predicting total visit duration in milliseconds.

women viewed feminine faces (M = 884, SE = 27.2) and masculine faces (M = 889, SE = 27.2) longer compared to viewing feminine faces in a long-term mating context (M = 802, SE = 27.2), see Figure 3. There was a significant interaction between mating context and women's progesterone levels, b = −0.39, SE = 0.08, 95%CI [−0.47, −0.14], p < 0.001. The interaction was probed at −1SD, the mean, and + 1SD from the mean of progesterone. At lower levels (b = 79.42, SE = 16.33, 95%CI [47.14, 111.71], p < 0.001) and at the mean of progesterone (b = 49.17, SE = 14.08, 95%CI [21.11, 77.33], p < 0.001), women's visual attention was higher for a short-term mating context compared to a long-term mating context. At higher levels of progesterone (b = 18.92, SE = 16.33, 95%CI [−13.35, 51.20], p = 0.24), the slopes were not significantly different from each other, see Figure 4. For long-term mating, women's visit duration remained stable in relation to progesterone levels. Overall, women viewed masculine faces longer during a long-term mating context, and visual attention was partly associated with progesterone.

3.1.5. Fixation count

Fixation count was defined as the number of times a visit (i.e., eye movement) was made into a region of interest (i.e., feminine, masculine faces). The overall variance explained from the model was, R^2_{Marginal} = 0.007, and $R^2_{\text{Conditional}}$ = 0.26. There was a significant effect for facial masculinity, b = 0.25, SE = 0.10, 95%CI [0.02, 0.42], p = 0.03. Women made more visual fixations to masculine faces (M = 3.09, SE = 0.10) compared to feminine faces (M = 2.93, SE = 0.09). There was a significant effect for estradiol, b = 0.08, SE = 0.02, 95%CI [0.02, 0.11], p = 0.003. When women's estradiol was higher they made more visual fixations.

4. Discussion

The current study investigated women's visual attention to facial masculinity across their menstrual cycle. Our first hypothesis, that women with high E/P ratios would rate masculine faces more attractive and view them longer compared to feminine faces, was not supported. Masculine faces were not rated more attractive than feminine faces, and there were no interactions with E/P ratio, or estradiol and progesterone individually. Our second hypothesis, that visual attention to masculine faces would be moderated by mating context, where women with high E/P ratios would rate masculine faces as more attractive and view them longer during a short-term mating context, was not supported. Although there were no significant effects for attractiveness, women did view masculine faces longer compared to feminine faces but for a long-term mating context. Overall, feminine faces for a long-term mating context were given the least amount of visual attention, even when compared to masculine and feminine faces in a short-term mating context. In addition, there was no support for women with high E/P ratios viewing masculine men longer under a short-term mating context. Instead, there was partial support for the association between women's hormones (estradiol, progesterone, and E/P ratio) and visual time. When women's E/P ratio was high, they made fewer first fixation durations compared to when their E/P ratio was low. Further, when women's progesterone levels were low, their visual attention was higher when considering men's faces for a short-term mating context compared to a long-term mating context. In addition, there were differences in the number of visual fixations when looking at women's hormones. When women's estradiol was high, they made more visual fixations throughout.

The findings of the current study do not support previous work suggesting that facial masculinity is associated with ratings of attractiveness (Cunningham et al., 1990; Grammer and Thornhill, 1994; DeBruine et al., 2006; Little et al., 2007). This does not imply that facial masculinity is not important in women's mate preferences, as women have shown that their overall preferences to masculinity may be different than what they choose (Flegr et al., 2019). In using a mating context prompt, we did not find any differences in attractiveness for women rating men under different mating contexts. Recent research has reported mixed results in facial masculinity preferences across mating contexts. Women have demonstrated preferences for all levels of facial

masculinity when rating men for a co-parent compared to a short-term relationship (Stower et al., 2020). Others have shown a preference for facial masculinity for rating men for a long-term compared to a short-term mating context (Clarkson et al., 2020), consistent with the present findings.

In addition to ratings of attractiveness, the current study utilized sensitive gaze recording techniques (i.e., eye tracking) that may be able to capture subtle changes in attention to sexually dimorphic faces as a function of mating context and hormones. We did find that women's late-onset visual measures (i.e., total visit duration, fixation count) were longer for masculine faces rather than feminine faces. Further, we found that women's total visit duration was associated with facial masculinity and mating context, where women viewed masculine faces compared to feminine faces longer during a long-term mating context. This finding is surprising, given that we expected women to view masculine faces longer under a short-term mating context, as predicted by sexual strategies theory. This may indicate that when considering a partner under repeated occasions, women prioritize masculine features when considering a long-term relationship. Since short-term mating is meant to indicate a one-time sexual encounter, viewing men's faces repeatedly in three different sessions eliminates the saliency of short-term mating, and long-term mating becomes more salient due to repeated exposure. Sexual strategies theory also suggests that women may prioritize men who are able to acquire resources and invest those resources in parenting, and facial masculinity has been perceived as being associated with the ability to protect (Sell et al., 2009). Further, this finding may be interpreted as women being attentive to features that connote both indirect (i.e., good genes) and direct benefits (i.e., protection, resource acquisition & transmission), which may be ideal attributes in seeking a partner for a long-term commitment. Little et al. (2007) demonstrated similar preferences when using a forced choice technique, where masculine preferences were higher for women considering men for a long-term partnership; however, in their study women were also taking into consideration ecological harshness, a condition not tested in this study. Given that women were simply viewing paired images of men's faces, visual attention may not necessarily indicate that women found masculine faces more attractive, as demonstrated by the numerical ratings of attractiveness provided. Preferences for masculinity has been linked to women relying on heuristics to discern facial profiles, primarily in societies with highly developed health indices, such as the United States (Scott et al., 2014). It is plausible that women were relying on heuristics to discern facial masculinity, as testosterone levels may be higher in industrialized societies (Scott et al., 2014). Considering that facial masculinity has been associated with aggression, visual attention to masculine faces could possibly reflect attention to threatening faces or males that are successful in intrasexual competition (Puts, 2010; Scott et al., 2014; Boothroyd et al., 2015; Mefodeva et al., 2020). Overall, feminine faces for a long-term mating context were given the least amount of visual attention. Perhaps, in the context of long-term mating (i.e., commitment, marriage), facial femininity in men may not be a salient attribute and may not be given a considerable amount of attention compared to facial masculinity in a long-term mating context, and facial femininity and masculinity in a short-term mating context. It could also suggest that there is more attention given overall to faces when considering a short-term mating context, as women may be relying on underlying features associated with the best fit male and spend more time scanning men's faces. Since women may prioritize high-quality genes in a short-term mating context (Buss and Schmitt, 1993), they may expend more cognitive resources in assessing potential mates for a context where immediate sexual access may occur.

The results of the study did not find any evidence for the ovulatory shift hypothesis, that women shift their preferences to men with putative markers of genetic quality (i.e., masculinity) across the menstrual cycle, supporting recent studies (Jones et al., 2018; reviewed in Marcinkowska et al., 2018a; Jones et al., 2019). The lack of evidence may have to do with the limited number of salivary samples (e.g., 3 in this study) provided and not confirming ovulation. Using an eye-tracking paradigm, which is meant to capture subtle movements in visual attention, we did not find evidence that women's preferences to facial masculinity, either by attractiveness ratings or tracking eye movements, were associated with high E/P ratios.

The study includes limitations that future studies can investigate and build upon using an eye tracking paradigm. First, although we did collect three time periods across the menstrual cycle, in order to capture true conception probability, research should investigate fertility status by collecting daily salivary samples in addition to luteinizing hormone kits to confirm ovulation in women. In this study, women's E/P ratio lacked heterogeneity, only showing slight changes across the three sessions. This impacted our ability to test if changes in E/P ratio was associated with shifting preferences and visual attention to facial masculinity, therefore, our results in reference to E/P ratio should be taken with caution. By testing women daily or including more than three samples, researchers may be able to capture heterogeneity in women's reproductive hormones and be able to test accurately the ovulatory shift hypothesis using an eye-tracking paradigm. One possible explanation of the lack of heterogeneity in our sample is due to the study's requirements and restrictions. We only recruited participants who had normal menstrual cycles and did not specify a particular range of cycle length (e.g., 23–35 day cycle length), nor did we assess average cycle length. Since "normal menstrual cycle" could have been broadly construed, future work should account for collecting women's information of cycle length. Second, relying on a non-diverse sample limits what can be tested according to theoretical frameworks proposed in trade-offs made by women. That is, it is unclear if college students are making similar trade-offs in choosing a masculine partner for a short or a long-term relationship. Ecological constraints, such as diverse socioeconomic statuses and different life histories make more salient trade-offs compared to women attending a university. University women have a dense mating pool and are surrounded by cues of safety, where seeking a mate that may demonstrate physical features associated with protection (i.e., masculinity) may not be a priority according to their local ecology. Women have shown to calibrate their preference to facial masculinity dependent upon whether they are experimental (Little et al., 2007) or actual constraints (Marcinkowska et al., 2018b). Given these differences in life histories, it is unclear if overall preferences for facial masculinity are due to their indications of putative indicators of high-quality genes or successful intrasexual competition (Puts, 2010; Little et al., 2013). Masculine traits, such as beardedness, are preferred by women in populations where beards are frequent (Dixson et al., 2017) and where the sex ratio is male-biased (Dixson et al., 2019a). Bearded men are also considered as having higher parenting abilities (Dixson and Brooks, 2013), primarily among women with young children (Dixson et al., 2019b). Women have also shown preferences to bearded men because it may signal their ability to provide direct benefits in the form of immediate resources and protection. Further research may consider the overall perceptions of facial masculinity to indicate inter- or intrasexual selection traits. Lastly, we only included women who self-identified as exclusively heterosexual, and did not sample

women with other sexual orientations (i.e., homosexual, bisexual). It is possible that women from this heterosexual sample could have a more nuanced sexual orientation, that could have been measured more precisely using a valid sexual orientation instrument. One study showed that women differ in their visual assessments of female bodies according to their sexual fluidity, showing similar patterns on how men view women (Widman et al., 2021). It is recommended future work incorporate a valid measure of sexual orientation in assessing attractiveness and visual assessments.

This study contributes to the overall literature on human mate choice by providing direct tests of sexual strategies theory and the ovulatory shift hypothesis. It introduces a different approach in studying the nuanced factors associated with women's cyclic shifts and preferences for sexually dimorphic features. By using an eye tracking paradigm, the overall goal was to determine if the ovulatory shift hypothesis could be supported using real-time, implicit measures, by tracking eye movements to men's facial masculinity. Although there was no strong evidence to suggest that women's visual attention shifted when E/P ratios were higher, there were indications that women's hormone levels (i.e., estradiol, progesterone, E/P ratio) predicted visual time to faces. This suggests these biomarkers play a supportive role in evaluating men's faces under a particular mating context. However, it is important to interpret the findings using the E/P ratio with caution, as the study did not have enough variability in E/P ratio to make solid predictions about changes in that hormone measurement and its association with attractiveness and visual attention. Although, it is very unlikely that women will view the same male with feminine and masculine facial traits in an actual mating context, women do make quick decisions when considering potential mates in a real-world setting, such as in social gatherings, using mobile base dating apps (e.g., Tinder), and in speed dating (Todd et al., 2007).

In summary, women's visual attention to facial masculinity was longer compared to feminine faces for a long-term mating context. We did not find any evidence to suggest that hormones were associated with visual attention to facial masculinity, but there was support that hormones are associated with overall visual attention. This indicates that there are biological underpinnings to the way that women strategically view men when considering them for a potential mate. Further work is needed to disentangle the role of hormones and their role in the cognitive processes of mate choice. Expanding work on the cognitive processes in mate choice can further the field by understanding the attentional processes in facial evaluations and by considering the biological underpinnings associated in mate preferences.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Oklahoma State University, Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

RG and JB-C conceived and designed the study and wrote the first draft of the manuscript. RG collected the data and performed the statistical analysis. All authors contributed to manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alexander, R. D. (1990). *How Did Humans Evolve? Reflections on the Uniquely Unique Species*. Ann Arbor, MI: University of Michigan Museum of Zoology Special Publication, 1–36.

Alexander, R. D., and Noonan, K. M. (1979). "Concealment of ovulation, parental care, and human social evolution" in *Evolutionary Biology and Human Social Behavior, an Anthropological Perspective*. eds. N. A. Chagnon and W. Irons (Belmont, CA: Duxbury Press), 436–453.

Andersson, M. B. (1994). *Sexual Selection*. Princeton University Press, Princeton.

Bates, D., Maechler, M., Bolker, B., and Walker, B. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Blake, K., Dixson, B. J. W., O'Dean, S. M., and Denson, T. F. (2016). Standardized methodological protocols for measuring the effects of fertility on women's behavior: a data-driven approach contrasting counting and hormonal methods. *Horm. Behav.* 81, 74–83. doi: 10.1016/j.yhbeh.2016.03.004

Boothroyd, L. G., Denson, T. F., Dixson, B. J., Carre, J. M., and McCormick, C. M. (2015). Evidence from meta-analysis of the facial width-to-height ratio as an evolved cue of threat. *PLoS One* 10:e0132726. doi: 10.1371/journal.pone.0132726

Boothroyd, L. G., Jones, B. C., Burt, D. M., DeBruine, L. M., and Perrett, D. I. (2008). Facial correlates of sociosexuality. *Evol. Hum. Behav.* 29, 211–218. doi: 10.1016/j.evolhumbehav.2007.12.009

Boothroyd, L., Scott, I., Gray, A. W., Coombes, C. I., and Pound, N. (2013). Male facial masculinity as a cue to health outcomes. *Evol. Psychol.* 11, 1044–1058. doi: 10.1177/147470491301100508

Brauer, M., and Curtin, J. J. (2017). Linear mixed-effects models and the analysis of non-independent data: a unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychol. Methods* 23, 389–411. doi: 10.1037/met0000159

Burris, R. P., Marcinkowska, U. M., and Lysons, M. T. (2014). Gaze properties of women judging the attractiveness of masculine and feminine male faces. *Evol. Psychol.* 12, 19–35. doi: 10.1556/jep.12.2014.1.2

Buss, D. M. (1989). Sex difference in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behav. Brain Scie.* 12, 1–14. doi: 10.1017/S0140525X00023992

Buss, D. M., and Schmitt, D. P. (1993). Sexual strategies theory: an evolutionary perspective on human mating. *Psychol. Rev.* 100, 204–232. doi: 10.1037/0033-295X.100.2.204

Buss, D. M., and Schmitt, D. P. (2019). Mate preferences and their behavioral manifestations. *Annu. Rev. Psychol.* 70, 77–110. doi: 10.1146/annurev-psych-010418-103408

Cashdan, E. (1996). Women's mating strategies. *Evol. Anthropol.* 5, 134–143. doi: 10.1002/(SICI)1520-6505(1996)5:4<134::AID-EVAN3>3.0.CO;2-G

Clarkson, T. R., Sidari, M. J., Sains, R., Alexander, M., Harrison, M., Mefodeva, V., et al. (2020). A multivariate analysis of women's mating strategies and sexual selection on men's facial morphology. *R. Soc. Open Sci.* 7:191209. doi: 10.1098/rsos.191209

Conklin, K., Pellicer-Sanchez, A., and Carrol, G. (2018). *Eye-Tracking: A Guide for Applied Linguistics Research*. Cambridge Press, New York, NY.

Cornwell, R. E., and Perrett, D. I. (2008). Sexy sons and sexy daughters: the influence of parents' facial characteristics on offspring. *Anim. Behav.* 76, 1843–1853. doi: 10.1016/j.anbehav.2008.07.031

Cunningham, M. R., Barbee, A. P., and Pike, C. L. (1990). What do women want? Facial metric assessment of multiple motives in the perception of male facial physical attractiveness. *J. Pers. Soc. Psychol.* 59, 61–72. doi: 10.1037/0022-3514.59.1.61

DeBruine, L., and Jones, B. (2017). Young Adult White Faces With Manipulated Versions. figshare. *Dataset*. doi: 10.6084/m9.figshare.4220517.v1

DeBruine, L. M., Jones, B. C., Little, A. C., Boothroyd, L. G., Perrett, D. I., Penton-Voak, I. S., et al. (2006). Correlated preferences for facial masculinity and ideal or actual partner's masculinity. *Proc. R. Soc. B Biol. Sci.* 273, 1355–1360. doi: 10.1098/rspb.2005.3445

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

Dixson, B. J., and Brooks, R. C. (2013). The role of facial hair in women's perceptions of men's attractiveness, health, masculinity and parenting abilities. *Evol. Hum. Behav.* 34, 236–241. doi: 10.1016/j.evolhumbehav.2013.02.003

Dixson, B. J. W., and Rantala, M. J. (2016). The role of facial and body hair distribution in women's judgments of men's sexual attractiveness. *Arch. Sex. Behav.* 45, 877–889. doi: 10.1007/s10508-015-0588-z

Dixson, B. J., Kennedy-Costantini, S., Lee, A. J., and Nelson, N. L. (2019b). Mothers are sensitive to men's beards as a potential cue of paternal investment. *Horm. Behav.* 113, 55–66. doi: 10.1016/j.yhbeh.2019.04.005

Dixson, B. J., Lee, A. J., Blake, K. R., Jasienska, G., and Marcinkowska, U. M. (2018b). Women's preferences for men's beards show no relation to their ovarian cycle phase and sex hormone levels. *Horm. Behav.* 97, 137–144. doi: 10.1016/j.yhbeh.2017.11.006

Dixson, B. J., Rantala, M. J., and Brooks, R. C. (2019a). Cross-cultural variation in women's preferences for men's body hair. *Adapt. Hum. Behav. Physiol.* 5, 131–147. doi: 10.1007/s40750-019-0107-x

Dixson, B. J., Rantala, M. J., Melo, E., and Brooks, R. C. (2017). Beards and the big city: displays of masculinity may be amplified under crowded conditions. *Evol. Hum. Behav.* 38, 259–264. doi: 10.1016/j.evolhumbehav.2016.10.009

Fink, B., Neave, N., and Seydel, H. (2007). Male facial appearance signals physical strength to women. *Am. J. Hum. Biol.* 19, 82–87. doi: 10.1002/ajhb.20583

Flegr, J., Blum, A. E., Nekola, O., and Kroupa, S. (2019). What people prefer and what they think they prefer in short- and long-term partners. The effects of the phase of the menstrual cycle, hormonal, contraception, pregnancy, and the marital and the parenthood status on partner preferences. *Evol. Hum. Behav.* 40, 112–125. doi: 10.1016/j.evolhumbehav.2018.09.003

Flinn, M. V., Quinlan, R. J., Coe, K., and Ward, C. V. (2007). "Evolution of the human family: Cooperative males, long social childhoods, smart mothers, and extended kin networks" in *Family Relationships: An Evolutionary Perspective*. eds. C. A. Salmon and T. K. Shackelford (Oxford: Oxford University Press)

Folstad, I., and Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *Am. Nat.* 139, 603–622.

Gangestad, S. W., and Haselton, M. G. (2015). Human estrus: implications for relationship science. *Curr. Opin. Psychol.* 1, 45–51. doi: 10.1016/j.copsyc.2014.12.007

Gangestad, S. W., Haselton, M. G., Welling, L. M., Gildersleeve, K., Pillsworth, E. G., Burris, R. P., et al. (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evol. Hum. Behav.* 37, 85–96. doi: 10.1016/j.evolhumbehav.2015.09.001

Gangestad, S. W., and Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proc. R. Soc. B Biol. Sci.* 265, 927–933.

Gangestad, S. W., and Thornhill, R. (2008). Human oestrus. *Proc. Biol. Sci.* 275, 991–1000. doi: 10.1098/rspb.2007.1425

Gangestad, S. W., Thornhill, R., and Garver-Apgar, C. E. (2005). Adaptations to ovulation: implications for sexual and social behavior. *Curr. Dir. Psychol. Sci.* 14, 312–316. doi: 10.1111/j.0963-7214.2005.00388.x

Garza, R., and Byrd-Craven, J. (2019). Fertility status in visual processing of men's attractiveness. *Evol. Psychol. Sci.* 5, 328–342. doi: 10.1007/s40806-019-00190-4

Garza, R., Heredia, R. R., and Cieslicka, A. B. (2017). An eye tracking examination of men's attractiveness by conceptive risk women. *Evol. Psychol.* 15, 1–11. doi: 10.1177/1474704917690741

Gilbert, S.F. (2000). *Developmental Biology*. Sinauer; Sunderland, MA.

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

Grammer, K., and Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: the role of symmetry and averageness. *J. Comp. Psychol.* 108, 233–242.

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., and Grammer, K. (2001). Male facial attractiveness: evidence for a hormone-mediated adaptive design. *Evol. Hum. Behav.* 22, 251–267. doi: 10.1016/S1090-5138(01)00066-6

Jones, B. C., Hahn, A. C., and DeBruine, L. M. (2019). Ovulation, sex hormones, and women's mating psychology. *Trends Cogn. Sci.* 23, 51–62. doi: 10.1016/j.tics.2018.10.008

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., et al. (2018). No compelling evidence that preferences for facial masculinity track changes in women's hormonal status. *Psychol. Sci.* 29, 996–1005. doi: 10.1177/0956797618760197

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

Jünger, J., Motta-Mena, N. V., Cardenas, R., Bailey, D., Rosenfield, K. A., Schild, C., et al. (2018b). Do women's preferences for masculine voices shift across the ovulatory cycle? *Horm. Behav.* 106, 122–134. doi: 10.1016/j.yhbeh.2018.10.008

Krupp, D. B. (2008). Through evolution's eyes: extracting mate preferences by linking visual attention to adaptive design. *Arch. Sex. Behav.* 37, 57–63. doi: 10.1007/s10508-007-9273-1

Little, A. C., Cohen, D. L., Jones, B. C., and Belsky, J. (2007). Human preferences for facial masculinity change with relationship type and environmental harshness. *Behav. Ecol. Sociobiol.* 61, 967–973. doi: 10.1007/s00265-006-0325-7

Little, A. C., Connely, J., Feinberg, D. R., Jones, B. C., and Roberts, S. C. (2011b). Human preferences for masculinity differs according to context in faces, bodies, voices, and smell. *Behav. Ecol.* 22, 862–868. doi: 10.1093/beheco/arr061

Little, A. C., DeBruine, L. M., and Jones, B. C. (2013). Environment contingent preferences: exposure to visual cues of direct male-male competition and wealth increase women's preferences for masculinity in male faces. *Evol. Hum. Behav.* 34, 193–200. doi: 10.1016/j.evolhumbehav.2012.11.008

Little, A. C., Jones, B. C., and DeBruine, L. M. (2011a). Facial attractiveness: evolutionary based research. *Phil. Trans. R. Soc. B* 366, 628–1659. doi: 10.1098/rstb.2010.0404

Little, A. C., Jones, B. C., DeBruine, L. M., and Feinberg, D. R. (2008). Symmetry and sexual dimorphism in human faces: interrelated preferences suggest both signal quality. *Behav. Ecol.* 19, 902–908. doi: 10.1093/beheco/arn049

Little, A. C., Jones, B. C., Penton-Voak, I. S., Burt, D. M., and Perrett, D. I. (2002). Partnership status and the temporal context of relationships influence human female preferences of sexual dimorphism in male face shape. *Proc. R. Soc. B Biol. Sci.* 269, 1095–1193. doi: 10.1098/rspb.2002.1984

Marcinkowska, U. M., Galbarczyk, A., and Jasienska, G. (2018c). 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. doi: 10.1016/j.psyneuen.2017.11.007

Marcinkowska, U. M., Ellison, P. T., Galbarczyk, A., Milkowska, K., Pawlowski, B., Thune, I., et al. (2016). Lack of support for relation between woman's masculinity preferences, estradiol level, and mating context. *Horm. Behav.* 78, 1–8. doi: 10.1016/j.yhbeh.2015.10.012

Marcinkowska, U. M., Kaminski, G., Little, A. C., and Jasienska, G. (2018a). Average ovarian hormones, rather than daily values and their fluctuations, are related to facial preferences among women. *Horm. Behav.* 102, 114–118. doi: 10.1016/j.yhbeh.2018.05.013

Marcinkowska, U. M., Rantala, M. J., Lee, A. J., Kozlov, M. V., Aavik, T., Cai, H., et al. (2018b). Women's preferences for men's masculinity are strongest under favorable ecological conditions. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-39350-8

Mefodeva, V., Sidari, M. J., Chau, H., Fitzsimmons, B., Antoine, G., Clarkson, T. R., et al. (2020). Multivariate intra-sexual selection on men's perceptions of male facial morphology. *Adapt. Hum. Behav. Physiol.* 6, 143–169. doi: 10.1007/s40750-020-00128-2

Penton-Voak, I. S., and Perrett, D. I. (2000). Female preferences for male faces changes cyclically: further evidence. *Evol. Hum. Behav.* 21, 39–48. doi: 10.1016/j.psyneuen.2011.11.007

Perrett, D. I., Lee, K., Penton-Voak, I. S., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature* 394, 884–887.

Peters, M., Simmons, L. W., and Rhodes, G. (2009). Preferences across the menstrual cycle for masculinity and symmetry in photographs of male faces and bodies. *PLoS One* 4, 1–7. doi: 10.1371/journal.pone.0004138

Puts, D. A. (2010). Beauty and the beast: mechanisms of sexual selection in humans. *Evol. Hum. Behav.* 31, 157–175. doi: 10.1016/j.evolhumbehav.2010.02.005

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., et al. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. *Nat. Commun.* 278, 774, –780. doi: 10.1098/rspb.2010.1678

Rantala, M. J., Polkki, M., and Rantala, L. M. (2010). Preference for human male body hair changes across the menstrual cycle and menopause. *Behav. Ecol.* 21, 419–423. doi: 10.1093/beheco/arp206

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.* 57, 199–226. doi: 10.1146/annurev.psych.57.102904.190208

Rhodes, G., Chan, J., Zebrowitz, I. A., and Simmons, L. W. (2003). Does sexual dimorphism in human faces signal health? *Proc. R. Soc. B* 270, 93–95. doi: 10.1146/annurev.psych.57.102904.190208

Roney, J. R., and Simmons, Z. I. (2008). Women's estradiol predicts preferences for facial cues of men's testosterone. *Horm. Behav.* 63, 636–645. doi: 10.1016/j.yhbeh.2007.09.008

Roney, J. R., and Simmons, Z. I. (2013). Hormonal predictors of sexual motivation in natural menstrual cycles. *Horm. Behav.* 53, 14–19. doi: 10.1016/j.yhbeh.2013.02.013

Scott, I. M. L., Clark, A. P., Josephson, S. C., Boyette, A. H., Cuthill, I. C., Fried, R. L., et al. (2014). Human preferences for sexually dimorphic faces may be evolutionary novel. *Proc. Natl. Acad. Sci.* 111, 14388–14393. doi: 10.1073/pnas.1409643111

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., and Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proc. R. Soc. Lond. B* 276, 575–584. doi: 10.1098/rspb.2008.1177

Stower, R. E., Lee, A. J., McIntosh, T. L., Sidari, M. J., Sherlock, J. M., and Dixson, B. J. (2020). Mating strategies and the masculinity paradox: how relationship context, relationship status, and sociosexuality shape women's preferences for facial masculinity and beardedness. *Arch. Sex. Behav.* 49, 809–820. doi: 10.1007/s10508-019-1437-2

Thornhill, R., and Gangestad, S. W. (1996). The evolution of human sexuality. *Trends Ecol. Evol.* 11, 98–102.

Thornhill, R., and Gangestad, S. W. (1999a). Facial attractiveness. *Trends Cogn. Sci.* 3, 452–460. doi: 10.1371/journal.pone.0245557

Thornhill, R., and Gangestad, S. W. (1999b). The scent of symmetry: a human sex pheromone that signals fitness? *Evol. Hum. Behav.* 20, 175–201. doi: 10.1159/000265559

Thornhill, R., and Gangestad, S. W. (2006). Facial sexual dimorphism, developmental stability, and susceptibility to disease in men and women. *Evol. Hum. Behav.* 27, 131–144. doi: 10.1016/j.evolhumbehav.2005.06.001

Todd, P. M., Penke, L., Fasolo, B., and Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preference. *PNAS* 104, 1500–15016. doi: 10.1073/pnas.0705290104

Wen, F. F., and Zuo, B. (2012). The effects of transformed gender facial features on face preference of college students: based on the test of computer graphics and eye movement tracks. *Acta Psychol. Sin.* 44, 14–29.

Widman, D. R., Bennetti, M. K., and Anglemyer, R. (2021). Gaze patterns of sexually fluid women and men at nude females and males. *Evol. Behav. Sci.* 15, 315–325. doi: 10.1037/ebs0000183

Windhager, S., Schaefer, K., and Fink, B. (2011). Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness, dominance, and masculinity. *Am. J. Hum. Biol.* 23, 805–814. doi: 10.1002/ajhb.21219

Yang, T., Chen, H., Hu, Y., Zheng, Y., and Wang, W. (2015). Preferences for sexual dimorphism on attractiveness levels: an eye-tracking study. *Personal. Individ. Differ.* 77, 179–185. doi: 10.1027/1618-3169/a000248

Zahavi, A. (1975). Mate selection: a selection for a handicap. *J. Theor. Biol.* 53, 205–214. doi: 10.1016/0022-5193(75)90111-3

frontiers | Frontiers in Psychology

TYPE Perspective
PUBLISHED 17 June 2024
DOI 10.3389/fpsyg.2024.1395974

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Joelson Moreno Brito De Moura,
Federal Rural University of Pernambuco, Brazil

*CORRESPONDENCE
Christopher I. Gurguis
✉ Christopher.Gurguis@uth.tmc.edu

RECEIVED 04 March 2024
ACCEPTED 03 June 2024
PUBLISHED 17 June 2024

CITATION
Gurguis CI, Kimm TS and Pigott TA (2024) Perspective: the evolution of hormones and person perception—a quantitative genetic framework.
Front. Psychol. 15:1395974.
doi: 10.3389/fpsyg.2024.1395974

Perspective: the evolution of hormones and person perception—a quantitative genetic framework

Christopher I. Gurguis*, Tyler S. Kimm and Teresa A. Pigott

Department of Psychiatry and Behavioral Sciences, McGovern Medical School at UTHealth, Houston, TX, United States

Evolutionary biology provides a unifying theory for testing hypotheses about the relationship between hormones and person perception. Person perception usually receives attention from the perspective of sexual selection. However, because person perception is one trait in a suite regulated by hormones, univariate approaches are insufficient. In this Perspectives article, quantitative genetics is presented as an important but underutilized framework for testing evolutionary hypotheses within this literature. We note tacit assumptions within the current literature on psychiatric genetics, which imperil the interpretation of findings thus far. As regulators of a diverse manifold of traits, hormones mediate tradeoffs among an array of functions. Hormonal pleiotropy also provides the basis of correlational selection, a process whereby selection on one trait in a hormone-mediated suite generates selection on the others. This architecture provides the basis for conflicts between sexual and natural selection within hormone-mediated suites. Due to its role in person perception, psychiatric disorders, and reproductive physiology, the sex hormone estrogen is highlighted as an exemplar here. The implications of this framework for the evolution of person perception are discussed. Empirical quantification of selection on traits within hormone-mediated suites remains an important gap in this literature with great potential to illuminate the fundamental nature of psychiatric disorders.

KEYWORDS

evolutionary psychiatry, hormones, quantitative genetics, person perception, evolutionary theory

1 Introduction

Hormones integrate traits into adaptive suites (McGlothlin and Ketterson, 2008). When several traits are regulated by one hormone, their response to evolutionary processes will be linked. Under these circumstances, hormonal mediation may facilitate or restrain phenotypic responses to selection, depending on the form of selection and underlying genetic architecture. Though hormones have adaptive functions, their actions can also predispose individuals to certain conditions, such as cancers and infections (Klein, 2000; Chuffa et al., 2017). Many psychiatric disorders are also linked with variation in hormone expression (Jacobson, 2014; Naughton et al., 2014; Iovino et al., 2018; Gogos et al., 2019). The evolution of hormone-mediated traits involves tradeoffs between their beneficial effects (e.g., maintaining reproduction) and deleterious effects (e.g., susceptibility to cancer).

As an example, estrogen regulates reproductive physiology in women, but a rapid drop in estrogen following parturition leaves women susceptible to depressive episodes (Schiller et al., 2015; Yim et al., 2015). Estrogen's involvement in both phenotypes provides the basis for a tradeoff between the two. In the presence of tradeoffs, predicting the evolutionary response of traits can be complex and requires a rigorous framework for analysis. The direction of selection on traits involved must be measured empirically because it may differ between traits within the same hormone-mediated suite (McGlothlin and Ketterson, 2008).

Prior work also implicates hormones in the modulation of person perception (Gangestad and Thornhill, 2008; Romero-Martinez et al., 2021). As part of hormone-mediated suites, person perception is unlikely to evolve independently of other traits regulated by the same hormones. In this Perspectives article, we focus on estrogen because it directly influences fitness components and is a well-studied influence on person perception and psychiatric disorders. We introduce evolutionary quantitative genetics to provide a framework for discussion. We then review known and emerging functions of estrogen. Finally, we discuss the evolutionary dynamics of sex hormone-mediated suites and implications for the evolution of psychiatric disorders and person perception. Our hope is that the framework discussed here could similarly be applied to other traits of interest to evolutionary psychology as has been accomplished robustly in behavioral quantitative genetics (Boake, 1994).

1.1 Quantitative genetics – a framework for testing evolutionary hypotheses

Darwin's theory of evolution by natural selection can be distilled to the following syllogism: for a given population, if more offspring are produced than survive to reproduce, and if individuals vary in their traits (including fitness), and if some of that variation is heritable, then that population will evolve (Darwin, 1859). Darwin recognized a second mechanism, sexual selection, which hinges on variation in the ability to obtain mates (via intrasexual competition or mate choice) rather than variation in survival. This theory was formalized beginning in the 20th century by Ronald Fisher and Sewell Wright, marking the birth of evolutionary quantitative genetics (Lynch and Walsh, 1998; Mayr and Provine, 1998). This approach focuses on phenotypic variation associated with multiple genetic loci, which is the case for most behaviors, psychological traits, and psychiatric disorders (Geschwind and Flint, 2015). A deceptively simple set of equations may be used to predict how variation in quantitative traits is expected to respond to natural or sexual selection. Quantitative genetics thus provides an analytically rigorous framework for applying evolutionary theory to and empirically testing hypotheses posited by evolutionary psychology.

First, phenotypic variance can be decomposed into multiple sources. Twin studies are the most commonly used quantitative genetic study design for psychiatric disorders and psychological traits (Kendler, 1993; Kendler, 2001; Vukasovic and Bratko, 2015). Modern quantitative genetics benefits from pedigree-based studies, including the use of the "animal model," which permits the use of pedigrees that are missing measurement of traits in some individuals and allows for more precise estimation of variance components (Kruuk, 2004; Kruuk and Hadfield, 2007). When combined with adoption studies, common environmental effects can be estimated (Lynch and Walsh, 1998). If pedigrees are sufficiently large and complex (e.g., contain mothers who have offspring from different fathers) or contain repeated measurements, other parameters such as maternal effects and permanent environmental effects may be estimated. The basic equation describing decomposition of phenotypic variance is given as:

$$P = G + E + G \times E + R$$

In this equation, P refers to phenotypic variance, G to genetic variance, E to environmental variance, GxE to variance from gene x environment effects, and R to residual variance (Falconer and Mackay, 1996). Genetic variance may arise from additive genetic variation, dominance, or epistasis. The proportion of phenotypic variance accounted for by genetic variance is called broad-sense heritability. Sources of environmental variance include parental effects, common environmental effects, or permanent environmental effects. GxE effects occur when different genotypes respond to environmental change in nonparallel ways. For all psychiatric disorders studied, phenotypic variance results from both environmental and genetic sources (Sullivan and Geschwind, 2019). Face perception likewise has both environmental and heritable sources of variance (Zhu et al., 2010).

For predicting the response to selection, two key parameters are important: narrow-sense heritability and the strength of selection (Lush, 1937; Walsh and Lynch, 2018). Narrow-sense heritability (h^2) is defined as the proportion of phenotypic variance (P) due to additive genetic effects (A) (Falconer and Mackay, 1996; Lynch and Walsh, 1998):

$$h^2 = \frac{A}{P}$$

Of note, narrow-sense heritability describes only part of the resemblance between relatives' phenotypes. Other sources of genetic variation will also cause resemblance among relatives. Quantitative genetics focuses on narrow-sense heritability, however, because this component responds to selection. Common environmental and parental effects may also cause resemblance between siblings and can inflate estimates of heritability if not measured (Lynch and Walsh, 1998). Some variance due to epigenetics may be heritable; recent extensions of quantitative genetic models incorporate this (Jablonka and Raz, 2009; Franklin et al., 2010; Stopher et al., 2012; Thomson et al., 2018). In humans, cultural inheritance is especially important and has also recently received attention within quantitative genetics (Danchin et al., 2011, 2013). Any trait with non-zero heritability has the potential to respond to selection.

When examining suites of traits, however, heritability must be extended to account for genetic correlations between those traits. The multivariate extension of heritability is defined by a matrix of additive genetic variances and covariances among traits called the G-matrix. From the G-matrix, one can calculate genetic correlations (r_G):

$$r_G = \frac{\mathrm{cov}\left(A_1, A_2\right)}{\sqrt{A_1 \cdot A_2}}$$

The genetic correlation is simply the Pearson correlation between the additive genetic components of two traits (A_1 and A_2). Recent studies suggest genetic correlations among many psychiatric disorders (Brainstorm et al., 2018; Grotzinger et al., 2022). Interestingly, another recent study suggested that psychiatric disorders are genetically correlated most strongly with pulmonary, gastrointestinal, and neurological disorders (Athanasiadis et al., 2022). Collectively, these studies imply that evolution of traits involved in one psychiatric disorder will depend on evolution of traits involved not only in other psychiatric disorders, but also in disorders involving organs beyond the brain. Similarly, genetic correlations among scores in the Minnesota Multiphasic Personality Inventory would suggest that personality traits will not evolve entirely independently of each other – a pattern of great import for the evolutionary psychology of personality (Viken and Rose, 2007).

The second key parameter, the strength of selection, was formalized by George Price (Price, 1970, 1972). The "Price Equation" defines selection on a phenotype as the covariance between that phenotype (P) and fitness (ω):

$$S = \mathrm{cov}(P, \omega)$$

Fitness in evolutionary genetics is defined as differential reproductive success—an individual's lifetime number of pregnancies relative to the population mean. The average number of children varies widely among human populations (United Nations, 2022). Because usually $S < 1$, variance in relative reproductive success sets the maximum potential response of a trait to selection, known as the "opportunity for selection" (Crow, 1958). According to Fisher, the strength of selection is more accurately defined by the genetic correlation between a phenotype and fitness, though this is rarely empirically measured (Fisher, 1958). Of note for studies of sexual selection, three covariances are important: between a trait and relative mating success, between a trait and relative reproductive success, and between relative mating success and relative reproductive success (Arnold and Wade, 1984). When there is no relationship between relative mating success and relative reproductive success, no sexual selection can occur. Sexual and natural selection may oppose each other in direction.

A few studies have examined fitness consequences of psychiatric disorders. Bipolar disorder appears to reduce fecundity, yet may be associated with increased fertility at younger ages (Power et al., 2013; Jacobson, 2016; Grover et al., 2019; Hope et al., 2020). Compared to women without psychiatric disorders, affected women are at risk for several negative fitness outcomes, including recurrent miscarriage, sexually transmitted infections, and reproductive cancer (Hope et al., 2022). Higher than average anxiety, on the other hand, was associated with quadratic (U-shaped) increases in fitness with individuals showing lower than average or higher than average anxiety having more children (Jacobson and Roche, 2018). For major depressive disorder, one study showed affected individuals do not have decreased fecundity when compared to their siblings, but another showed that affected individuals have lower fecundity compared with the general population (Tondo et al., 2011; Power et al., 2013). Although the relationship between psychiatric disorders and fitness outcomes has been preliminarily examined, the association between other psychological traits and fitness outcomes warrants further study, as this relationship is key to understanding their contemporary evolution.

The "Breeder's Equation" describes the expected response to selection (Lush, 1937):

$$R = h^2 S$$

The response to selection for any given trait (R) is that trait's heritability multiplied by the strength of selection. Whenever $h^2 < 1$, the effect of selection on a trait is proportionally diminished.

The multivariate extension of the Breeder's Equation allows for estimation of response of a suite of traits to selection (Lande, 1979). This equation is especially important for hormone-mediated suites of traits:

$$\Delta \bar{z} = \mathrm{GP}^{-1}\mathrm{S}$$

Here, $\Delta\bar{z}$ is the vector of responses in a suite of phenotypes, G is the G-matrix, P-1 is the phenotypic variance–covariance matrix for the traits, and S is the vector of selection differentials on those traits. The pertinent consequence of this equation for hormone-mediated suites is that the response of one trait to selection depends on selection directly on that trait in addition to selection on every other trait with which it is genetically correlated (Lande and Arnold, 1983). When one trait responds to selection on another trait with which it is genetically correlated, the process is called correlational selection. Within a suite of genetically correlated traits, direct selection on each trait may differ in strength, form, or direction. The overall direction of change for a trait in response to selection, thus, does not only depend on selection directly on that trait, but on correlational selection through other traits as well (Lande and Arnold, 1983; Arnold, 1992). To our knowledge, no studies have used the quantitative genetic framework to predict response to selection for psychiatric disorders or other psychological traits.

Hormonal pleiotropy is an important source of genetic covariance (Wittman et al., 2021). For traits involved in hormone-mediated suites, evolution of each of those traits will potentially be dependent on selection acting on multiple others. This may cause the trait of interest to respond to selection in ways not predicted by univariate models (Arnold, 1992; McGlothlin and Ketterson, 2008).

Some properties of quantitative genetic parameters are largely overlooked, but have important implications for evolutionary psychology. First, these parameters are properties of populations, not individuals. Second, all quantitative genetic parameters are specific to the age during which they are measured and can change over the life of an organism. Prior empirical work demonstrates that heritability and other quantitative genetic parameters can change significantly with age (Wilson et al., 2005). Selection may favor increases in a phenotype in young individuals and decreases in that phenotype in older individuals (Roff, 1992; Stearns, 1992). Third, these parameters are specific to the population in which they are measured. One ought not assume that heritability measured in one population will be the same as heritability measured in another. Finally, quantitative genetic parameters are specific to the generation in which they are measured and may change in response to different processes, including selection (Athanasiadis et al., 2022). Selection may change strength, form, or direction from one generation to the next and in turn alter trait

heritability. For these reasons, the adaptive value of a trait in the past or future should be distinguished from current selection on that trait. This discordance is called "mismatch" (Corbett et al., 2018). For example, the neurobiological systems regulating response to reward are not adapted to stimuli from recreational drugs, and those with substance use disorders may have lower lifetime reproductive success (Nesse and Berridge, 1997; Troisi, 2001; Jacobson, 2016). Other examples of quantitative genetic predictions of hypotheses from the evolutionary psychology literature are given in Table 1. These examples illustrate how quantitative genetics may analytically bridge evolutionary theory with evolutionary psychology by allowing for empirical tests of hypotheses about how evolutionary processes shape variation in psychological traits.

1.2 The estrogen-mediated suite of traits

Sex hormones are privileged with regard to selection because they regulate reproductive traits and thus generate strong potential for correlational selection among other traits in their suite (McGlothlin and Ketterson, 2008). In humans, sex hormones are well-studied for their effects on the development of sexual characteristics. Here, we focus on estrogen, but the principles we elucidate should be considered for other sex hormones.

Estrogen signaling, in addition to regulating growth, development, and physiology of female reproduction, is crucial for the timing of life history transitions (e.g., menarche and menopause), metabolism, immune function, adipogenesis, skeletal modeling, cardiovascular system functioning, and mood regulation (Dluzen, 2005; Kovats, 2015). Estrogen may also regulate mate preference (Gangestad and Thornhill, 2008; Jünger et al., 2018). Additionally, estrogen is a major factor in carcinogenesis, especially in cancers of the breast and female reproductive tract (Grady et al., 1995; Clemons and Goss, 2001; Kaaks et al., 2002; Rossouw et al., 2002; Yager and Davidson, 2006; Reid et al., 2017). This network of traits influenced by estrogen provides the architecture for correlational selection (Figure 1).

Estrogen signaling accomplishes this diversity of functions by acting through several molecular mechanisms, including binding to receptors or directly to DNA to alter transcription (Bjornstrom and Sjoberg, 2005). Serum estradiol concentration is regulated by sex hormone binding globulin (Siiteri et al., 1982; Arathimos et al., 2020). Free estradiol may be important for some phenotypes, such as carcinogenesis, while fluctuations in hormone levels are important for others, such as menstrual cycles, maintenance and progression of pregnancies, and mood regulation (Deroo and Korach, 2006; Burns and Korach, 2012; Christensen et al., 2012; Hamilton et al., 2017; Gordon et al., 2019). Nuances in the mechanisms by which a suite of traits is regulated allow for partial independence in the evolution of those traits when $r_G < 1$ (Wagner and Lynch, 2008).

Though some of the functions of estrogen are limited to women, estrogen is important in regulating several functions in men as well (Kousteni et al., 2001). This scenario is the basis for cross-sex genetic correlations (r_{MF}), whereby evolution of trait expression in men will be linked to the evolution of those same traits in women. Cross-sex genetic correlations have been the target of intensive work in evolutionary biology due to the role that this genetic architecture can play in sexual selection (Badyaev, 2002). The major consequence of r_{MF} is that selection on one sex will generate selection in the other. If r_{MF} between the sexes is negative, and selection on the sexes is in opposite directions, the evolution of sex differences is expected (Lande, 1980). On the other hand, if r_{MF} is positive and selection on the sexes is in opposite directions, the evolution of sex differences will be constrained. The long-term outcomes of sex differences depend on the strength, direction, and consistency of selection and the strength and direction of r_{MF} (Badyaev, 2002; Walsh and Lynch, 2018).

Estrogen is implicated in several psychiatric disorders and may be partially responsible for observed sex differences. Sex differences in risk of depression, for example, emerge in puberty and continue throughout life (Kessler, 2003). This risk is mechanistically linked with fluctuations in estrogen levels, for example during pregnancy, peripartum periods, and menopause (Payne, 2003; Bennett et al., 2004; Freeman et al., 2006; O'Hara and Swain, 2009; O'Hara and McCabe, 2013). Estrogen, by regulating mood, is associated with both major depressive disorder and bipolar disorder (Halbreich and Kahn, 2001; Borrow and Cameron, 2014; Frey and Dias, 2014). Cycle-related changes in mood symptoms in both these disorders have also been reported (Payne et al., 2007). Estrogen and progesterone jointly increase vulnerability for developing anxiety disorders and influence the presentation, course, and treatment response of anxiety disorders,

TABLE 1 Examples of hypotheses from evolutionary psychology and their concordant quantitative genetic predictions.

Evolutionary psychology hypothesis	Quantitative genetic predictions
Trait A is adaptive	Positive covariance between Trait A and fitness outcomes
Trait B is deleterious	Negative covariance between Trait B and fitness outcomes
Trait C is the result of a historical process of strong directional selection	Low heritability of Trait C
Trait D is an adaptation to environment E, which is deleterious in environment F ("mismatch")	Positive covariance between Trait D and fitness in environment E, but negative covariance between Trait D and fitness in environment F, e.g., Hereford (2009)
Patterns of personality covariance are adaptive strategies	Genetic integration among personality traits with concordant fitness benefits, e.g., Duckworth and Kruuk (2009)
The mind is organized into adaptive modules	Pattern of genetic covariance between traits within the same posited modules and weak or limited genetic covariance between traits within different posited modules, e.g., Drake and Klingenberg (2010)
Emotion G has a specific function, H	Positive covariance between performance of the posited function and fitness, e.g., Arnold (1983)
Person perception is an adaptation that facilitates mate choice	Covariance between person perception traits and mate choice AND covariance between mate choice and fitness outcomes, e.g., Brooks and Endler (2001)

Empirical work investigating some of these examples in non-human taxa is provided.

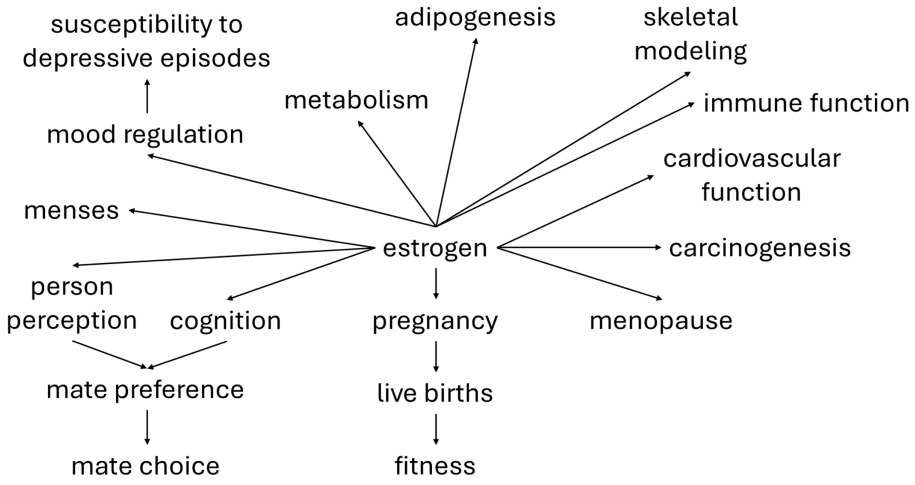


FIGURE 1
The pleiotropic effects of estrogen. Due to the reproductive effects of estrogen, its correlation with fitness should be quite strong, providing the basis for correlational selection on other traits that it regulates. The strength and direction of genetic correlations between traits within this estrogen-mediated suite partially determines whether the evolutionary response of individual traits is facilitated or constrained. A thorough understanding of the function of hormone-regulated traits must include consideration of this underlying genetic architecture.

especially in women (Pigott et al., 2019). In schizophrenia, several possible mechanisms also associate estrogen signaling with cognitive function (McGregor et al., 2017). Estrogen receptor expression in the frontal cortex and hippocampus shows sex differences in major depressive disorder, but not in schizophrenia and bipolar disorder (Perlman et al., 2005). Sex differences in susceptibility to neurodegenerative diseases may also be related to neuroprotective effects of estrogen (Vegeto et al., 2020).

2 Discussion

When an array of phenotypes is regulated by one hormone, predicting the response to selection of a single phenotype is difficult without large sampling and detailed phenotyping efforts. Within the field of hormones and person perception, most efforts have focused on relating estrogen function individually to various phenotypes, such as mate choice or preference (Feinberg et al., 2006; Garver-Apgar et al., 2008; Lukaszewski and Roney, 2009). Large quantitative genetic studies within these fields could provide important empirical tests of theory, especially considering existing methods to construct pedigrees from already collected genetic data (Staples et al., 2014). For clinical fields, evolutionary theory can shed light on unresolved fundamental questions about the nature of psychiatric disorders (Nesse, 2023).

Hormonal regulation of person perception is a key area for understanding how psychiatric disorders evolve due to its role in sexual selection. Sex hormones are associated with variation in perception of faces, morphology, and emotions (Gangestad and Thornhill, 2008; Romero-Martinez et al., 2021). Person perception varies between populations and biological sexes, is influenced by psychiatric disorders, and changes with reproductive physiology, season, and age (Gangestad and Thornhill, 2008; Pawlowski and Sorokowski, 2008; Kohler et al., 2010, 2011; Boothroyd and Vukovic, 2019; Olderbak et al., 2019). This abundant phenotypic variation suggests ample opportunity for evolutionary processes to change the distribution of these traits from one generation to the next.

Crucially, hormonal regulation of person perception modifies the sensory processes necessary for expressing mating preference and choice, but is not equivalent to these. Mating preference and mate choice must be measured empirically, a problem which presents unique methodological challenges (Andersson and Simmons, 2006; Dougherty, 2020; Clancey et al., 2022). Variation in the perception of mates over time (e.g., across the menstrual cycle) allows for changes in mate preference, but these changes do not necessarily entail changes in mate choice, especially given variation in human mating systems (Todd et al., 2007; Schacht and Kramer, 2019). In turn, changes variation in mate choice may or may not be under natural or sexual selection depending on the relationship between mate choice and fitness (Shuster and Wade, 2003; Walsh and Lynch, 2018). For example, in many human societies, marriage is primarily an economic transaction involving the influence of an individual's parents or family (Ingoldsby, 2006; Buunk et al., 2009). In these societies, the expression of individual preference may be facilitated or opposed by familial influence. The fitness consequences (e.g., frequency of extrapair mating) can depend on alignment of individual and familial preferences (Scelza, 2011). Quantitative genetics provides a robust framework for studying the role of hormonal regulation of person perception in sexual selection as well as the evolution of psychiatric disorders in the setting of hormone-mediated suites of traits.

More empirical work is needed to clarify several evolutionary quantitative genetic questions about hormones and person perception, such as quantification of univariate and multivariate heritability of hormone-mediated suites. For reasons outlined above, this work must include studies of different populations, ages, and generations. Most studies of heritability of clinical phenotypes do not examine these differences. The unwarranted, tacit assumption is that risky genetic loci are stable across these contexts or change on longer timescales than are clinically relevant. Given that sex-hormone mediated traits undergo large regulatory changes during puberty, these knowledge gaps are especially important targets for empirical work. Few studies have quantified contemporary selection on hormones and person perception or examined mechanisms of selection [though see (Hill

et al., 2013)], which means the processes possibly driving changes in heritability are almost completely overlooked.

Overall, we have discussed beneficial and adverse consequences of sex hormone signaling, highlighting the fact that tradeoffs among various functions of hormonal suites are common. This discussion emphasizes the importance of caution in drawing conclusions about the evolution of psychological traits and psychiatric disorders from single phenotype studies. This caveat is especially important for the field of hormones and person perception due to the possibility of conflicts between sexual and natural selection. Our hope is that future work will incorporate evolutionary quantitative genetic approaches to studying adaptive hormone-mediated suites of traits.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

CG: Writing – review & editing, Writing – original draft, Visualization, Conceptualization. TK: Writing – review & editing, Supervision, Conceptualization. TP: Writing – review & editing, Supervision, Conceptualization.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank J. M. West for fruitful discussion during the writing and revision of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Andersson, M., and Simmons, L. W. (2006). Sexual selection and mate choice. *Trends Ecol. Evol.* 21, 296–302. doi: 10.1016/j.tree.2006.03.015

Arathimos, R., Millard, L. A. C., Bell, J. A., Relton, C. L., and Suderman, M. (2020). Impact of sex hormone-binding globulin on the human phenome. *Hum. Mol. Genet.* 29, 1824–1832. doi: 10.1093/hmg/ddz269

Arnold, S. J. (1983). Morphology, performance and fitness. *Am. Zool.* 23, 347–361. doi: 10.1093/icb/23.2.347

Arnold, S. J. (1992). Constraints on phenotypic evolution. *Am. Nat.* 140, S85–S107. doi: 10.1086/285398

Arnold, S. J., and Wade, M. J. (1984). On the measurement of natural and sexual selection: theory. *Evolution* 38, 709–719. doi: 10.2307/2408383

Athanasiadis, G., Meijsen, J. J., Helenius, D., Schork, A. J., Ingason, A., Thompson, W. K., et al. (2022). A comprehensive map of genetic relationships among diagnostic categories based on 48.6 million relative pairs from the Danish genealogy. *Proc. Natl. Acad. Sci. USA* 119, 1–9.

Badyaev, A. V. (2002). Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends Ecol. Evol.* 17, 369–378. doi: 10.1016/S0169-5347(02)02569-7

Bennett, H. A., Einarson, A., Taddio, A., Koren, G., and Einarson, T. R. (2004). Prevalence of depression during pregnancy: systematic review. *Obstet. Gynecol.* 103, 698–709. doi: 10.1097/01.AOG.0000116689.75396.5f

Bjornstrom, L., and Sjoberg, M. (2005). Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Mol. Endocrinol.* 19, 833–842. doi: 10.1210/me.2004-0486

Boake, C. R. B. (1994). *Quantitative Genetic Studies of Behavioral Evolution*. Chicago: University of Chicago Press.

Boothroyd, L. G., and Vukovic, J. (2019). "Mate preferences across the lifespan" in *The Oxford handbook of evolutionary psychology and behavioral endocrinology*. eds. L. L. M. Welling and T. K. Shackelford (New York, NY, USA: Oxford University Press).

Borrow, A. P., and Cameron, N. M. (2014). Estrogenic mediation of serotonergic and neurotrophic systems: implications for female mood disorders. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* 54, 13–25. doi: 10.1016/j.pnpbp.2014.05.009

Brainstorm, C., Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., et al. (2018). Analysis of shared heritability in common disorders of the brain. *Science* 360, 1–12.

Brooks, R., and Endler, J. A. (2001). Female guppies agree to differ: phenotypic and genetic variation in mate-choice behavior and the consequences for sexual selection. *Evolution* 55, 1644–1655

Burns, K. A., and Korach, K. S. (2012). Estrogen receptors and human disease: an update. *Arch. Toxicol.* 86, 1491–1504. doi: 10.1007/s00204-012-0868-5

Buunk, A. P., Park, J. H., and Duncan, L. A. (2009). Cultural variation in parental influence on mate choice. *Cross-Cult. Res.* 44, 23–40.

Christensen, A., Bentley, G. E., Cabrera, R., Ortega, H. H., Perfito, N., Wu, T. J., et al. (2012). Hormone regulation of female reproduction. *Horm. Metab. Res.* 44, 587–591. doi: 10.1055/s-0032-1306301

Chuffa, L. G., Lupi-Junior, L. A., Costa, A. B., Amorim, J. P., and Seiva, F. R. (2017). The role of sex hormones and steroid receptors on female reproductive cancers. *Steroids* 118, 93–108. doi: 10.1016/j.steroids.2016.12.011

Clancey, E., Johnson, T. R., Harmon, L. J., and Hohenlohe, P. A. (2022). Estimation of the strength of mate preference from mated pairs observed in the wild. *Evolution* 76, 29–41. doi: 10.1111/evo.14397

Clemons, M., and Goss, P. (2001). Estrogen and the risk of breast cancer. *N. Engl. J. Med.* 344, 276–285. doi: 10.1056/NEJM200101253440407

Corbett, S., Courtiol, A., Lummaa, V., Moorad, J., and Stearns, S. (2018). The transition to modernity and chronic disease: mismatch and natural selection. *Nat. Rev. Genet.* 19, 419–430. doi: 10.1038/s41576-018-0012-3

Crow, J. F. (1958). Some possibilities for measuring selection intensities in man. *Hum. Biol.* 61, 763–775.

Danchin, E., Charmantier, A., Champagne, F. A., Mesoudi, A., Pujol, B., and Blanchet, S. (2011). Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat. Rev. Genet.* 12, 475–486. doi: 10.1038/nrg3028

Danchin, E., Pujol, B., and Wagner, R. H. (2013). The double pedigree: a method for studying culturally and genetically inherited behavior in tandem. *PLoS One* 8:e61254. doi: 10.1371/journal.pone.0061254

Darwin, C. (1859). *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. Edison, NJ, USA: Castle Books.

Deroo, B. J., and Korach, K. S. (2006). Estrogen receptors and human disease. *J. Clin. Invest.* 116, 561–570. doi: 10.1172/JCI27987

Dluzen, D. E. (2005). Unconventional effects of estrogen uncovered. *Trends Pharmacol. Sci.* 26, 485–487. doi: 10.1016/j.tips.2005.08.001

Dougherty, L. R. (2020). Designing mate choice experiments. *Biol. Rev. Camb. Philos. Soc.* 95, 759–781. doi: 10.1111/brv.12586

Drake, A. G., and Klingenberg, C. P. (2010). Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *Am. Nat.* 175, 289–301. doi: 10.1086/650372

Duckworth, R. A., and Kruuk, L. E. (2009). Evolution of genetic integration between dispersal and colonization ability in a bird. *Evolution* 63, 968–977. doi: 10.1111/j.1558-5646.2009.00625.x

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. Essex, U.K.: Longman.

Feinberg, D. R., Jones, B. C., Law Smith, M. J., Moore, F. R., DeBruine, L. M., Cornwell, R. E., et al. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Horm. Behav.* 49, 215–222. doi: 10.1016/j.yhbeh.2005.07.004

Fisher, R. A. (1958). *The Genetical theory of natural selection*. Toronto: Dover.

Franklin, T. B., Russig, H., Weiss, I. C., Graff, J., Linder, N., Michalon, A., et al. (2010). Epigenetic transmission of the impact of early stress across generations. *Biol. Psychiatry* 68, 408–415. doi: 10.1016/j.biopsych.2010.05.036

Freeman, E. W., Sammel, M. D., Lin, H., and Nelson, D. B. (2006). Associations of hormones and menopausal status with depressed mood in women with no history of depression. *Arch. Gen. Psychiatry* 63, 375–382. doi: 10.1001/archpsyc.63.4.375

Frey, B. N., and Dias, R. S. (2014). Sex hormones and biomarkers of neuroprotection and neurodegeneration: implications for female reproductive events in bipolar disorder. *Bipolar Disord.* 16, 48–57. doi: 10.1111/bdi.12151

Gangestad, S. W., and Thornhill, R. (2008). Human oestrus. *Proc. Biol. Sci.* 275, 991–1000.

Garver-Apgar, C. E., Gangestad, S. W., and Thornhill, R. (2008). Hormonal correlates of women's mid-cycle preference for the scent of symmetry. *Evol. Hum. Behav.* 29, 223–232. doi: 10.1016/j.evolhumbehav.2007.12.007

Geschwind, D. H., and Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science* 349, 1489–1494. doi: 10.1126/science.aaa8954

Gogos, A., Ney, L. J., Seymour, N., Van Rheenen, T. E., and Felmingham, K. L. (2019). Sex differences in schizophrenia, bipolar disorder, and post-traumatic stress disorder: are gonadal hormones the link? *Br. J. Pharmacol.* 176, 4119–4135. doi: 10.1111/bph.14584

Gordon, J. L., Peltier, A., Grummisch, J. A., and Sykes Tottenham, L. (2019). Estradiol fluctuation, sensitivity to stress, and depressive symptoms in the menopause transition: a pilot study. *Front. Psychol.* 10:1319. doi: 10.3389/fpsyg.2019.01319

Grady, D., Gebretsadik, T., Kerlikowske, K., Ernster, V., and Petitti, D. (1995). Hormone replacement therapy and endometrial cancer risk: a Meta-analysis. *Obstet. Gynecol.* 85, 304–313. doi: 10.1016/0029-7844(94)00383-O

Grotzinger, A. D., Mallard, T. T., Akingbuwa, W. A., Ip, H. F., Adams, M. J., Lewis, C. M., et al. (2022). Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat. Genet.* 54, 548–559.

Grover, S., Sahoo, S., Chaudhary, S., Chakrabarti, S., Nehra, R., and Avasthi, A. (2019). Gender differences, family size and fertility rate among patients with bipolar disorder: a study from India. *Psychiatry Res.* 272, 562–568. doi: 10.1016/j.psychres.2018.12.156

Halbreich, U., and Kahn, L. S. (2001). Role of estrogen in the Aetiology and treatment of mood disorders. *CNS Drugs* 15, 797–817. doi: 10.2165/00023210-200115100-00005

Hamilton, K. J., Hewitt, S. C., Arao, Y., and Korach, K. S. (2017). Estrogen Hormone Biology. *Curr. Top. Dev. Biol.* 125, 109–146. doi: 10.1016/bs.ctdb.2016.12.005

Hereford, J. (2009). A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* 173, 579–588. doi: 10.1086/597611

Hill, A. K., Hunt, J., Welling, L. L. M., Cárdenas, R. A., Rotella, M. A., Wheatley, J. R., et al. (2013). Quantifying the strength and form of sexual selection on men's traits. *Evol. Hum. Behav.* 34, 334–341. doi: 10.1016/j.evolhumbehav.2013.05.004

Hope, H., Parisi, R., Ashcroft, D. M., Williams, R., Coton, S., Kosidou, K., et al. (2020). Fertility trends of women with serious mental illness in the United Kingdom 1992-2017: a primary care cohort study using the clinical practice research datalink. *J. Affect. Disord.* 269, 141–147. doi: 10.1016/j.jad.2020.03.037

Hope, H., Pierce, M., Johnstone, E. D., Myers, J., and Abel, K. M. (2022). The sexual and reproductive health of women with mental illness: a primary care registry study. *Arch. Womens Ment. Health* 25, 585–593. doi: 10.1007/s00737-022-01214-y

Ingoldsby, B. B. (2006). "Mate selection and marriage" in *Families in global and multicultural perspective*. eds. B. B. Ingoldsby and S. D. Smith (Thousand Oaks, CA, USA: Sage Publications).

Iovino, M., Messana, T., De Pergola, G., Iovino, E., Dicuonzo, F., Guastamacchia, E., et al. (2018). The role of Neurohypophyseal hormones vasopressin and oxytocin in neuropsychiatric disorders. *Endocr. Metab. Immune Disord. Drug Targets* 18, 341–347. doi: 10.2174/1871530318666180220104900

Jablonka, E., and Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q. Rev. Biol.* 84, 131–176. doi: 10.1086/598822

Jacobson, L. (2014). Hypothalamic-pituitary-adrenocortical axis: neuropsychiatric aspects. *Compr. Physiol.* 4, 715–738. doi: 10.1002/cphy.c130036

Jacobson, N. C. (2016). Current evolutionary adaptiveness of psychiatric disorders: fertility rates, parent-child relationship quality, and psychiatric disorders across the lifespan. *J. Abnorm. Psychol.* 125, 824–839. doi: 10.1037/abn0000185

Jacobson, N. C., and Roche, M. J. (2018). Current evolutionary adaptiveness of anxiety: extreme phenotypes of anxiety predict increased fertility across multiple generations. *J. Psychiatr. Res.* 106, 82–90. doi: 10.1016/j.jpsychires.2018.10.002

Jünger, J., Kordsmeyer, T. L., Gerlach, T. M., and Penke, L. (2018). Fertile women evaluate male bodies as more attractive, regardless of masculinity. *Evol. Hum. Behav.* 39, 412–423. doi: 10.1016/j.evolhumbehav.2018.03.007

Kaaks, R., Lukanova, A., and Kurzer, M. S. (2002). Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. *Am. Soc. Prevent. Oncol.* 11, 1531–1543

Kendler, K. S. (1993). Twin studies of psychiatric illness: current status and future directions. *Arch. Gen. Psychiatry* 50, 905–915. doi: 10.1001/archpsyc.1993.01820230075007

Kendler, K. S. (2001). Twin studies of psychiatric illness: an update. *Arch. Gen. Psychiatry* 58, 1005–1014. doi: 10.1001/archpsyc.58.11.1005

Kessler, R. (2003). Epidemiology of women and depression. *J. Affect. Disord.* 74, 5–13. doi: 10.1016/S0165-0327(02)00426-3

Klein, S. L. (2000). The effects of hormones on sex differences in infection: from genes to behavior. *Neurosci. Biobehav. Rev.* 24, 627–638. doi: 10.1016/S0149-7634(00)00027-0

Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K., and Moberg, P. J. (2011). Facial emotion perception in depression and bipolar disorder: a quantitative review. *Psychiatry Res.* 188, 303–309. doi: 10.1016/j.psychres.2011.04.019

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., and Moberg, P. J. (2010). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophr. Bull.* 36, 1009–1019. doi: 10.1093/schbul/sbn192

Kousteni, S., Bellido, T., Plotkin, L. I., O'Brien, C. A., Bodenner, D. L., Han, L., et al. (2001). Nongenotropic, sex-nonspecific signaling through the estrogen or androgen receptors: dissociation from transcriptional activity. *Cell* 104, 719–730

Kovats, S. (2015). Estrogen receptors regulate innate immune cells and signaling pathways. *Cell. Immunol.* 294, 63–69. doi: 10.1016/j.cellimm.2015.01.018

Kruuk, L. E. (2004). Estimating genetic parameters in natural populations using the "animal model". *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359, 873–890. doi: 10.1098/rstb.2003.1437

Kruuk, L. E., and Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity between relatives. *J. Evol. Biol.* 20, 1890–1903. doi: 10.1111/j.1420-9101.2007.01377.x

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33, 402–416. doi: 10.1111/j.1558-5646.1979.tb04694.x

Lande, R. (1980). Sexual dimorphism, sexual selection and adaptation in polygenic characters. *Evolution* 34, 292–305. doi: 10.1111/j.1558-5646.1980.tb04817.x

Lande, R., and Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution* 37, 1210–1226. doi: 10.2307/2408842

Lukaszewski, A. W., and Roney, J. R. (2009). Estimated hormones predict women's mate preferences for dominant personality traits. *Personal. Individ. Differ.* 47, 191–196. doi: 10.1016/j.paid.2009.02.019

Lush, J. L. (1937). *Animal breeding plans*. Ames, IA, USA: Iowa State Press.

Lynch, M., and Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA, USA: Sinauer Associates, Inc.

Mayr, E., and Provine, W. (1998). *The evolutionary synthesis: Perspectives on the unification of biology*. Cambridge, MA, USA: Harvard University Press.

McGlothlin, J. W., and Ketterson, E. D. (2008). Hormone-mediated suites as adaptations and evolutionary constraints. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 363, 1611–1620. doi: 10.1098/rstb.2007.0002

McGregor, C., Riordan, A., and Thornton, J. (2017). Estrogens and the cognitive symptoms of schizophrenia: possible neuroprotective mechanisms. *Front. Neuroendocrinol.* 47, 19–33. doi: 10.1016/j.yfrne.2017.06.003

Naughton, M., Dinan, T. G., and Scott, L. V. (2014). Corticotropin-releasing hormone and the hypothalamic-pituitary-adrenal axis in psychiatric disease. *Handb. Clin. Neurol.* 124, 69–91. doi: 10.1016/B978-0-444-59602-4.00005-8

Nesse, R. M. (2023). Evolutionary psychiatry: foundations, progress and challenges. *World Psychiatry* 22, 177–202. doi: 10.1002/wps.21072

Nesse, R. M., and Berridge, K. C. (1997). Psychoactive drug use in evolutionary perspective. *Science* 278, 63–66. doi: 10.1126/science.278.5335.63

O'Hara, M. W., and McCabe, J. E. (2013). Postpartum depression: current status and future directions. *Annu. Rev. Clin. Psychol.* 9, 379–407. doi: 10.1146/annurev-clinpsy-050212-185612

O'Hara, M. W., and Swain, A. M. (2009). Rates and risk of postpartum depression—a meta-analysis. *Int. Rev. Psychiatry* 8, 37–54.

Olderbak, S., Wilhelm, O., Hildebrandt, A., and Quoidbach, J. (2019). Sex differences in facial emotion perception ability across the lifespan. *Cogn. Emot.* 33, 579–588. doi: 10.1080/02699931.2018.1454403

Pawlowski, B., and Sorokowski, P. (2008). Men's attraction to women's bodies changes seasonally. *Perception* 37, 1079–1085. doi: 10.1068/p5715

Payne, J. L. (2003). The role of estrogen in mood disorders in women. *Int. Rev. Psychiatry* 15, 280–290. doi: 10.1080/0954026031000136893

Payne, J. L., Roy, P. S., Murphy-Eberenz, K., Weismann, M. M., Swartz, K. L., McInnis, M. G., et al. (2007). Reproductive cycle-associated mood symptoms in women with major depression and bipolar disorder. *J. Affect. Disord.* 99, 221–229. doi: 10.1016/j.jad.2006.08.013

Perlman, W. R., Tomaskovic-Crook, E., Montague, D. M., Webster, M. J., Rubinow, D. R., Kleinman, J. E., et al. (2005). Alteration in estrogen receptor alpha mRNA levels in frontal cortex and hippocampus of patients with major mental illness. *Biol. Psychiatry* 58, 812–824. doi: 10.1016/j.biopsych.2005.04.047

Pigott, T. A., Duran, A. N., Jalnapurkar, I., Kimm, T., Linscheid, S., and Allen, M. K. (2019). "Sex differences in anxiety disorders" in *The Oxford handbook of evolutionary psychology and behavioral endocrinology*. eds. L. L. M. Welling and T. K. Shackelford (New York, NY, USA: Oxford University Press).

Power, R. A., Kyaga, S., Uher, R., MacCabe, J. H., Langstrom, N., Landen, M., et al. (2013). Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry* 70, 22–30. doi: 10.1001/jamapsychiatry.2013.268

Price, G. R. (1970). Selection and covariance. *Nature* 227, 520–521. doi: 10.1038/227520a0

Price, G. R. (1972). Extension of covariance selection mathematics. *Ann. Hum. Genet.* 35, 485–490. doi: 10.1111/j.1469-1809.1957.tb01874.x

Reid, B. M., Permuth, J. B., and Sellers, T. A. (2017). Epidemiology of ovarian cancer: a review. *Cancer Biol. Med.* 14, 9–32. doi: 10.20892/j.issn.2095-3941.2016.0084

Roff, D. A. (1992). *Life history evolution*. Sinauer, MA, USA: Sinauer Associates, Inc.

Romero-Martinez, A., Sarrate-Costa, C., and Moya-Albiol, L. (2021). A systematic review of the role of oxytocin, cortisol, and testosterone in facial emotional processing. *Biology (Basel)* 10, 1–34. doi: 10.3390/biology10121334

Rossouw, J. E., Anderson, G. L., and Prentice, R. L. (2002). Investing in healthy postmenopausal women: principal results from the Women's health initiative randomized controlled trial. *JAMA* 288, 321–333.

Scelza, B. A. (2011). Female choice and extra-pair paternity in a traditional human population. *Biol. Lett.* 7, 889–891. doi: 10.1098/rsbl.2011.0478

Schacht, R., and Kramer, K. L. (2019). Are we monogamous? A review of the evolution of pair-bonding in humans and its contemporary variation cross-culturally. *Front. Ecol. Evol.* 7:230. doi: 10.3389/fevo.2019.00230

Schiller, C. E., Meltzer-Brody, S., and Rubinow, D. R. (2015). The role of reproductive hormones in postpartum depression. *CNS Spectr.* 20, 48–59. doi: 10.1017/S1092852914000480

Shuster, S. M., and Wade, M. J. (2003). *Mating systems and strategies*. Princeton, NJ, USA: Princeton University Press.

Siiteri, P. K., Murai, J. T., Hammond, G. L., Nisker, J. A., Raymoure, W. J., and Kuhn, R. W. (1982). The serum transport of steroid hormones. *Recent Prog. Horm. Res.* 38, 457–510

Staples, J., Qiao, D., Cho, M. H., and Silverman, E. K.G. University of Washington Center for Mendelian, D.A. Nickerson, and J.E. Below (2014). PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am. J. Hum. Genet.* 95, 553–564.

Stearns, S. C. (1992). *The evolution of life histories*. New York, NY, USA: Oxford University Press.

Stopher, K. V., Walling, C. A., Morris, A., Guinness, F. E., Clutton-Brock, T. H., Pemberton, J. M., et al. (2012). Shared spatial effects on quantitative genetic parameters: accounting for spatial autocorrelation and home range overlap reduces estimates of heritability in wild red deer. *Evolution* 66, 2411–2426. doi: 10.1111/j.1558-5646.2012.01620.x

Sullivan, P. F., and Geschwind, D. H. (2019). Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* 177, 162–183. doi: 10.1016/j.cell.2019.01.015

Thomson, C. E., Winney, I. S., Salles, O. C., and Pujol, B. (2018). A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PLoS One* 13:e0197720. doi: 10.1371/journal.pone.0197720

Todd, P. M., Penke, L., Fasolo, B., and Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proc. Natl. Acad. Sci. USA* 10, 15011–15016.

Tondo, L., Lepri, B., and Baldessarini, R. J. (2011). Reproduction among 1975 Sardinian women and men diagnosed with major mood disorders. *Acta Psychiatr. Scand.* 123, 283–289.

Troisi, A. (2001). Harmful effects of substance abuse: a Darwinian perspective. *Funct. Neurol.* 16, 237–243

United Nations (2022). *Department of economic and social affairs, world population prospects 2022: Summary of results*. New York, United Nations Publication: UN DESA.

Vegeto, E., Villa, A., Della Torre, S., Crippa, V., Rusmini, P., Cristofani, R., et al. (2020). The role of sex and sex hormones in neurodegenerative diseases. *Endocr. Rev.* 41, 273–319. doi: 10.1210/endrev/bnz005

Viken, R. J., and Rose, R. J. (2007). Genetic variation and covariation in the original and restructured clinical scales of the MMPI. *J. Abnorm. Psychol.* 116, 842–847. doi: 10.1037/0021-843X.116.4.842

Vukasovic, T., and Bratko, D. (2015). Heritability of personality: a meta-analysis of behavior genetic studies. *Psychol. Bull.* 141, 769–785. doi: 10.1037/bul0000017

Wagner, G. P., and Lynch, V. J. (2008). The gene regulatory logic of transcription factor evolution. *Trends Ecol. Evol.* 23, 377–385. doi: 10.1016/j.tree.2008.03.006

Walsh, B., and Lynch, M. (2018). *Evolution and selection of quantitative traits*. New York: Oxford University Press.

Wilson, A. J., Kruuk, L. E., and Coltman, D. W. (2005). Ontogenetic patterns in heritable variation for Body size: using random regression models in a wild ungulate population. *Am. Nat.* 166, e177–e192. doi: 10.1086/497441

Wittman, T. N., Robinson, C. D., McGlothlin, J. W., and Cox, R. M. (2021). Hormonal pleiotropy structures genetic covariance. *Evol Lett* 5, 397–407. doi: 10.1002/evl3.240

Yager, J. D., and Davidson, N. E. (2006). Estrogen carcinogenesis in breast Cancer. *N. Engl. J. Med.* 354, 270–282. doi: 10.1056/NEJMra050776

Yim, I. S., Tanner Stapleton, L. R., Guardino, C. M., Hahn-Holbrook, J., and Dunkel Schetter, C. (2015). Biological and psychosocial predictors of postpartum depression: systematic review and call for integration. *Annu. Rev. Clin. Psychol.* 11, 99–137. doi: 10.1146/annurev-clinpsy-101414-020426

Zhu, Q., Song, Y., Hu, S., Li, X., Tian, M., Zhen, Z., et al. (2010). Heritability of the specific cognitive ability of face perception. *Curr. Biol.* 20, 137–142. doi: 10.1016/j.cub.2009.11.067

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 14 August 2024
DOI 10.3389/fpsyg.2024.1404387

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Neil Robert Caton,
The University of Queensland, Australia
Christopher Watkins,
Abertay University, United Kingdom
James Rutter,
Durham University, United Kingdom

*CORRESPONDENCE
Steven Arnocky
✉ stevena@nipissingu.ca

RECEIVED 21 March 2024
ACCEPTED 09 July 2024
PUBLISHED 14 August 2024

CITATION
Arnocky S and Davis AC (2024) Do lifestyle and hormonal variables explain links between health and facial attractiveness?
Front. Psychol. 15:1404387.
doi: 10.3389/fpsyg.2024.1404387

Do lifestyle and hormonal variables explain links between health and facial attractiveness?

Steven Arnocky[1]* and Adam C. Davis[2]

[1]Human Evolution Laboratory, Department of Psychology, Nipissing University, North Bay, ON, Canada, [2]Department of Social Sciences, Canadore College, North Bay, ON, Canada

Introduction: Facial attractiveness has recently been considered an indicator of underlying immunocompetence. However, studies examining this relationship have yielded mixed findings. Previous research suggested that these discrepant findings could be due to the common influence of lifestyle factors upon both rated facial attractiveness and health.

Methods: Young men ($N = 162$) provided standardized facial photos with a neutral expression subsequently rated by eight women for overall attractiveness. Saliva was assayed for immunoglobulin A, testosterone (T) and cortisol (C), and body fat was measured using a skinfold caliper. Self-reports of poor health, and lifestyle factors that could influence health status (age, sleep habits, smoking, drinking alcohol, family stress, and exercising) were collected.

Results: Results showed that symptoms of poor health and skinfold negatively predicted facial attractiveness. There was a modest but statistically non-significant T x C interaction where higher T lower C men trended toward having more attractive faces. A sequential mediation model examining the influence of lifestyle showed support for an indirect effect on facial attractiveness. Specifically, skinfold and poor health symptoms mediated the links between exercise, stress, and facial attractiveness.

Discussion: These findings suggest links between facial attractiveness and immunocompetence could be linked to some common lifestyle and hormonal variables, but that more comprehensive research involving lifestyle indicators (such as nutrition) are necessary.

KEYWORDS

facial attractiveness, immunocompetence, good genes sexual selection, unhealthy lifestyle, skinfold, secondary sexual characteristics

1 Introduction

Humans are remarkably consistent in their assessment of what constitutes an attractive face (Langlois et al., 1991, 2000; Cunningham et al., 1995). Symmetry, averageness, skin quality, and hormone-linked sexually dimorphic features together form a facial structure that may be considered along a spectrum of attractiveness to the opposite sex (Arnocky et al., 2014). Faces may be an important source of reproductively relevant information accessible within a small amount of space (see Arnocky et al., 2014 for review). Specifically, attractive faces are believed to serve as a cue to an individual's genotypic quality (e.g., Hume and Montgomerie, 2001). Indeed, some previous research has reported positive links between facial attractiveness and health (see Arnocky et al., 2014; Jones et al., 2021 for review). From this perspective, ancestors who happened to prefer immunocompetence-linked morphological traits, such as those contributing to facial attractiveness, would have mated with partners who

were better able to survive, accrue resources, and successfully rear offspring, and to have produced healthier offspring who themselves would be more likely to survive and reproduce.

In support of the facial attractiveness immunocompetence hypothesis, Shackelford and Larsen (1997) found that facial asymmetry was linked with poorer health among men and women. Hume and Montgomerie (2001) found that female facial attractiveness was tied to their body mass index (BMI) and health history, whereas facial attractiveness in men was linked to their childhood socioeconomic standing, which could indicate a role of environmental or lifestyle factors affecting facial development. There is also circumstantial evidence suggesting a putative link between facial attractiveness and immunocompetence. Humans tend to reliably rate more attractive faces as being healthier. For instance, Fink et al. (2006) found that female faces that were more symmetrical were also more attractive, but also were perceived as healthier. Foo et al. (2020) found that facial traits contributing to overall attractiveness, such as averageness, symmetry, skin yellowness, and adiposity in men, predicted raters' perceptions of the health of those faces. Facial attractiveness is also tied to mating success in some studies: men with attractive faces have more short-term sex partners, and women with attractive faces start having sex at an earlier age and have more long-term sex partners (Rhodes et al., 2005). Some research suggests that there may be a sex difference in the link between facial attractiveness and immunocompetence. For example, men (Rantala et al., 2012), but not women (Rantala et al., 2013), with attractive faces have a stronger immune response to a hepatitis vaccine.

Still, other research has found null links between facial attractiveness and health. Kalick et al. (1998) examined the relationship between adult health and their rated facial attractiveness at late adolescence. They found no links across the lifespan. Nevertheless, raters inaccurately perceived attractive faces as being healthier within the sample. Similar findings were observed by Foo et al. (2020), where (as described earlier), attractive faces were viewed as healthier by raters, yet facial attractiveness was nevertheless unrelated to markers of immunocompetence including oxidative stress and lysozyme activity. Other research using a large (> 4,000 participant) sample found no links between longitudinal measures of childhood health and facial asymmetry (Pound et al., 2014). More recently, Cai et al. (2019) found that neither female facial attractiveness, sexual dimorphism, averageness, or coloration predicted self-reported health or salivary immunoglobulin-A (sIgA). Similarly, other work found that male facial attractiveness did not predict antibody levels following vaccination (Pátková et al., 2022).

1.1 Considering potentially important covariates

Jones et al. (2021) recently suggested that the discordant findings pertaining to the link between facial attractiveness and health might be due to covariates that could impact both variables. Specifically, they proposed that "rather than reflecting immunocompetence, facial attractiveness is instead more closely linked to aspects of lifestyle that produce health benefits" (pp. 3). The researchers argued that lifestyle factors, which can vary intra-individually over time, might explain changes in individuals' facial attractiveness over time. Which lifestyle factors are relevant to facial attractiveness and health? Jones et al. (2021) focused on the examples of diet and body fat, which certainly have implications for health status and may have a stronger link to facial attractiveness than do markers of immunocompetence (Cai et al., 2019).

Rantala et al. (2013) found that body fat was curvilinearly related to facial attractiveness: Women with low or high body fat were rated as less attractive than those having intermediate body fat. Exercise also has well-established links to health (e.g., Akimoto et al., 2003; Murphy et al., 2009). Diets rich in highly processed and refined foods, typical of Western populations, have been linked to a range of physical and mental health problems (Cordain et al., 2005). Focusing on unhealthy dietary habits, Visine et al. (2024) found that consumption of food high in refined carbohydrates with a high glycemic load was associated with reduced facial attractiveness (rated by opposite-sex others) in both women and men. These effects remained after controlling for potential confounds, including age, sexual dimorphism, BMI, physical activity, smoking, and relationship status. Despite well-established links between exercise and health (e.g., Murphy et al., 2009), the link between exercise and facial attractiveness is less clear. Hönekopp et al. (2010) found that a composite measure of physical fitness predicted rated body but not facial attractiveness. Yet other research has shown that higher performance athletes are rated as being more facially attractive (e.g., Bagozzi et al., 2018). Moreover, men with stronger grip strength are rated as being more facially attractive (Fink et al., 2007).

Other candidates include exposure to smoke and alcohol, which when used in excess are known to have widespread negative health consequences (see Hurley et al., 2012 for review). Prototype faces of identical twins who smoke are rated less attractive than the non-smoking twin images (Skinner et al., 2017). Likely mechanisms of smoking-related change in attractiveness include skin wrinkling, pale-yellow (i.e., sallow) complexion, and gaunt facial structure (reviewed in Doshi et al., 2007). Some studies show that acute alcohol use can increase others' ratings of the drinker's facial attractiveness (e.g., Van Den Abbeele et al., 2015). Nevertheless, excessive alcohol use can lead to psoriasis, eczema, and skin infections (Higgins and Du Vivier, 1992) as well as jaundice, hyperpigmentation, and vascular issues including spider telangiectasias and angiomas (Liu et al., 2010).

Stress has also been implicated in both features influencing facial attractiveness (such as skin quality; see Koizumi et al., 2023 for review) and a diverse range of negative health consequences (see Apanius, 1998). Finally, sleep might also affect both facial attractiveness and health. Individuals photographed following 2 days of sleep restriction were rated as less attractive than when they had appropriate sleep. The researchers reasoned that aversion to mating with a sleep disturbed partner could help avoid sleep-related health issues (Sundelin et al., 2017).

Besides the study by Visine et al. (2024) described above, one other study to consider lifestyle factors in relation to facial attractiveness and health was conducted by Mengelkoch et al. (2022). These researchers examined rated facial attractiveness and various markers of health along with covariates, including BMI, adult socioeconomic status (SES), exercise, smoking behavior, and recent stress. However, given that these variables were not the primary focal point of the study, only those that were significantly related to rated attractiveness (BMI and age) were retained in their models. Their findings suggested that facial attractiveness was related with higher rates of phagocytosis and lower rates of bacterial growth in plasma, along with lower neutrophil counts, together suggesting better anti-bacterial immunity, but not with cellular proliferation or cytokine production.

1.2 Hormones

Hormones play an important role in coordinating phenotypic development (Roney, 2016) and therefore might also serve as important covariates when examining links between facial attractiveness and health. In a sample of young Latvian women, Rantala et al. (2013) found that facial attractiveness (as rated by men) was unrelated to the production of anti-hepatitis B surface antigen following a hepatitis B vaccination. However, they did find that (in addition to the body fat finding described earlier), women with high cortisol (C) had faces that were rated as less attractive. The authors considered that perhaps facial attractiveness serves as a cue to one's exposure to, or ability to cope with, life stressors, or that low C also signals health in humans. Other research has found either null links between women's facial attractiveness and C (Gonzalez-Santoyo et al., 2015, Study 1) or mixed results whereby some samples rate women with low C as having either more attractive (US raters) or less attractive (Mexican raters) faces (Gonzalez-Santoyo et al., 2015). Meta-analysis shows that flatter diurnal Cortisol slopes are associated with diverse negative health markers (Adam et al., 2017; see also: Knack et al., 2013). Nevertheless, comparatively less work has considered the role between men's cortisol and their facial attractiveness.

Testosterone (T) is another hormone that may be complicit in both men's health and facial attractiveness. Some research has shown that men with higher T are rated by women as having more attractive faces (e.g., Roney et al., 2006; Rantala et al., 2012), and T also has implications for immune functioning. For example, T is positively associated with sIgA in men (Arnocky et al., 2018; Hodges-Simeon et al., 2020). Yet other studies have found null links between T and rated male facial attractiveness (e.g., Swaddle and Reierson, 2002; Neave et al., 2003; Penton-Voak and Chen, 2004), and others have found null links between T, C, and both facial attractiveness and other-rated perceptions of health (Kandrik et al., 2017). Some researchers have suggested that relying on baseline T or C levels may be insufficient, and that the dual hormone hypothesis involving an interaction between high T and low C might be complicit in phenotypic masculinization. Indeed, Rantala et al. (2012) found that high T low C men's facial photos were rated as most attractive by women. However, other research has failed to observe these effects (Kordsmeyer et al., 2019). For example, T, C, and percentage of adipose tissue were unrelated to ratings of men's facial attractiveness (Pátková et al., 2022). Thus, more research is needed regarding the potential impacts of immunocompetence by hormone interactions on the development of phenotypic characteristics and the perceived attractiveness of those traits (Davis and Arnocky, 2022).

2 The present study

The goal of this research was to examine whether individual lifestyle factors, as well as abdominal skinfold measurements, that are theoretically common to both facial attractiveness and health might eliminate these links when controlled for in a regression analysis. In so doing, this research also aimed to examine whether previously reported links between facial attractiveness and biological and self-report markers of health are broadly replicable, given previously inconsistent findings.

Two indices of health were examined in the current study: Self-reports of poor health symptom frequency and severity and, following Cai et al. (2019), salivary immunoglobulin-A (sIgA). sIgA is a potentially important marker of underlying immunocompetence that is produced by plasma B cells, comprising over 70% of our mucosal antibodies (Macpherson et al., 2008) that provide an initial defense against pathogens (Macpherson et al., 2008; Brandtzaeg, 2009). Low levels of sIgA have been linked to increased infection (Fahlman and Engels, 2005; Nakamura et al., 2006; Volkmann and Weekes, 2006) as well as to self-reported severity and frequency of poor health symptoms in otherwise healthy young adult university students (Arnocky et al., 2023) and, longitudinally, to death in older adults (Phillips et al., 2015). sIgA has previously been linked to other apparently sexually selected phenotypic traits that may serve as cues to underlying immunocompetence, including the deep male voice (Arnocky et al., 2018) and female breast morphology (Locke and Arnocky, 2021).

We expected an initial negative bivariate correlation between facial attractiveness and self-reported symptoms of poor health (Hypothesis 1A), and a positive correlation between facial attractiveness and sIgA (Hypothesis 1B). We then examined whether controlling for age (in years), lifestyle variables (sleep, familial stress, alcohol and tobacco use, exercise) and skinfold, along with hormones that have been linked to both health and facial attractiveness (T, C, and a T x C interaction), would weaken any observed links between symptoms of poor health, sIgA, and facial attractiveness (Hypothesis 2). Finally, we considered a sequential mediation model whereby unhealthy lifestyle habits have an indirect effect on facial attractiveness. Specifically, we expected that unhealthy lifestyle variables would predict a thicker abdominal skinfold, which has been identified as a predictor of future health problems in previous research (Loh et al., 2018). Therefore, in our model, skinfold was entered as a predictor of poor health symptoms, which in turn would predict lower rated facial attractiveness (Hypothesis 3, see Figure 1).

3 Materials and methods

3.1 Participants and procedure

As a part of a larger study on immune function and phenotypic development, males from a small Canadian University and the community were recruited through the institutional research participation system and posters in local businesses around town. A sample size calculation was performed using G*Power (3.1.9.7) with an expected medium effect size ($F^2 = 0.15$), 80% power, $\alpha = 0.05$, and 12 predictors, which yielded a sample size of 127. The total sample in the existing data set was 162 young adult men, aged 18–39 years ($M_{age} = 22.7$, SD = 4.7; 91.4% were students). The ethnic composition of the sample was Caucasian (90%), Black (4%), East Asian (3%), South Asian (2%), and Indigenous/First peoples (1%). Participants received either $50 CAD remuneration or partial course credit and $10. This research received approval by the Nipissing University Research Ethics Board (protocol # 100770–26,667).

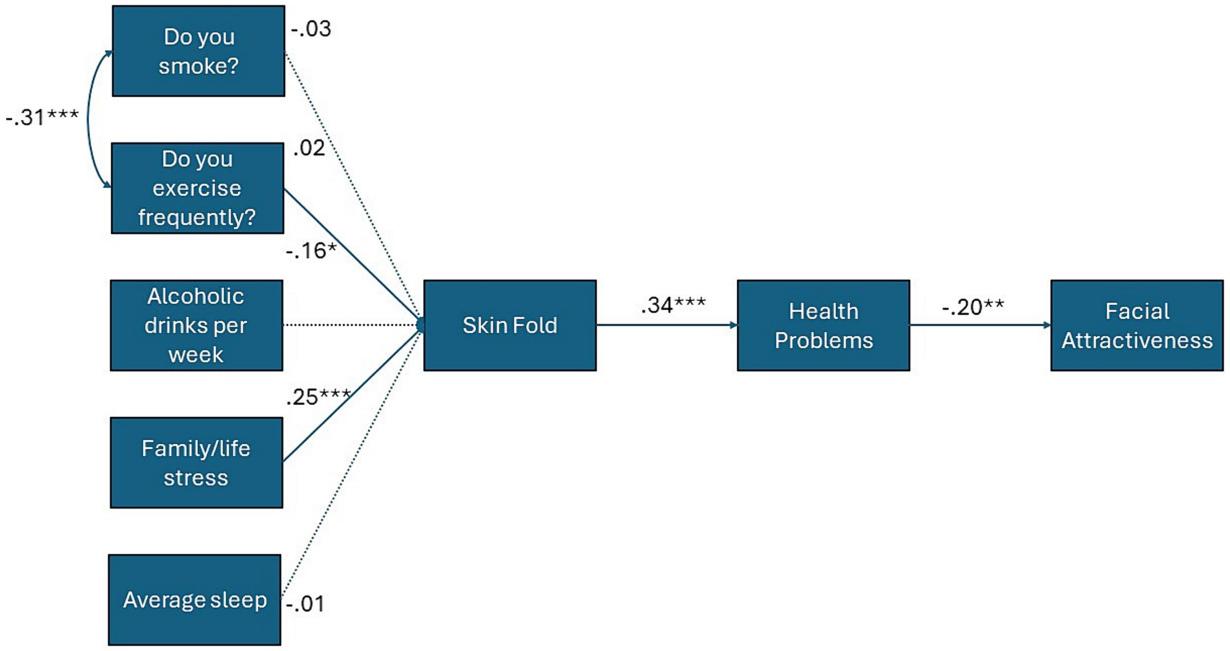


FIGURE 1
Path model for indirect effects of lifestyle variables upon facial attractiveness via skinfold and symptoms of poor health. Solid lines depict statistically significant paths, dashed lines depict statistically non-significant paths. Standardized coefficients shown. * = $p < .05$; ** = $p < 0.01$; *** = $p < 0.001$.

3.2 Measures

3.2.1 Hormones and immunoglobulin-a

Participants were asked not to eat, drink (except water), brush their teeth, or exercise 2 hrs prior to their testing session, and were rescheduled if they reported any current or recent acute symptoms of illness during their telephone screener prior to their session. Saliva samples were collected in 5 mL polystyrene culture tubes and stored at −80°C until assayed in duplicate via enzyme immunoassay kits (DRG International, NJ, United States) in the Principal Investigator's lab. Sample provision time ranged between 8:30 AM and 5:00 PM. For sIgA, intra- and inter-assay CVs were below 6%. For T (pg/mL), intra-assay CV was below 4% and the inter-assay CV was below 8%. For C (ng/mL), the intra-assay CV was below 2% and the inter-assay CV was below 11%. To account for the typical non-normal distribution of these markers, the average of the duplicates was log-transformed. Given that salivary flow rate affects sIgA levels, we corrected the concentration value to reflect a flow (mL/s)-corrected µg/mL score (log-transformed). Sample provision time was related to C, and participant age was related to both C ($r = -0.21$, $p = 0.007$) and T ($r = -0.24$, $p = 0.002$), whereas sIgA was unrelated to either age or sample provision time.

3.2.2 Unhealthy lifestyle

As part of a self-reported health screener, participants then completed items which addressed unhealthy lifestyle factors, including alcohol consumption ("*How many alcoholic beverages do you drink [on average] per week?*"), smoking/tobacco exposure ("*Do you smoke?*" [binary]), life stress ("*Have you or your family recently experienced any life changes or unusual psychological stress?*" [binary]), exercise ("*Do you exercise regularly?*" [binary]), along with sleep ("*How many hours do you sleep on average at night?*" [continuous]). A second indicator of unhealthy lifestyle was individuals' skinfold measurement obtained using a digital body fat caliper. Thicker skinfold is associated with unhealthy eating habits from an early age (e.g., Dalrymple et al., 2019), including being linked to consumption of ultra-processed foods (Rohatgi et al., 2017), and is associated with a host of cardiometabolic risks in adulthood (González-Torres et al., 2023). Accordingly, skinfold has been used by researchers as an indicator of nutritional status (e.g., Bernstein et al., 2002). Body fat is highly correlated with facial adiposity (see Sierra-Johnson and Johnson, 2004) and plays an important role in determining male facial attractiveness (Windhager et al., 2011). The participant's suprailliac skinfold (approximately one inch about the right hipbone) was measured three times and then averaged, ($\alpha = 0.99$, 95% LLCI = 0.991, 95% ULCI = 0.995).

3.2.3 Self-reported health

Self-reported health was assessed using The Health Symptoms Survey (Knack et al., 2011, 2012), which records both the frequency and severity of physical health problems. The measure demonstrates good construct validity, correlating with health-linked personality factors and behavioral issues (Knack et al., 2012), altered hypothalamic–pituitary–adrenal axis functioning (Knack et al., 2011), and sIgA as a biological marker of immunocompetence (Arnocky et al., 2023). The measure includes 56 items ranging from 1 (*Not at All*/*Does not Hurt at All*) to 4 (*All the Time*/*Unbearable Pain*) to determine the frequency and severity (28 items each) of symptoms, including stomach aches, flu, mouth sores, fatigue, chest pain, diarrhea, muscle aches and pains, headache or migraine, coughing, and fever experienced over the past year. A mean score was created with the measure demonstrating good internal consistency ($\alpha = 0.91$, 95% LLCI = 0.77, 95% ULCI = 0.94).

TABLE 1 Descriptive statistics for study variable.

		N	*M*	SD	Min	Max
1. Age		160	22.71	4.71	18.00	39.00
2. Facial Attractiveness		162	4.00	1.28	1.25	7.75
3. Health Problems		162	1.35	0.25	1.00	2.29
4. sIgA		161	111.33	84.74	20.80	702.20
5. Skin Fold (mm)		162	14.26	8.44	3.30	41.23
6. Sleep		162	7.31	0.96	3.50	9.50
7. Exercise	Yes	133				
	No	29				
8. Alcohol (drink/week)		162	4.96	5.19	0.00	24.00
9. Smoking	Yes	9				
	No	153				
10. Life Stress	Yes	33				
	No	129				
11. Testosterone (T)		162	132.60	101.29	11.28	1155.50
12. Cortisol (C)		162	6.09	4.86	0.07	20.36

Biomarker (sIgA, T, C) concentration values are reported prior to Log transformation. sIgA concentration is reported in μg/mL, T is reported in pg/mL, and C is reported in ng/mL.

3.2.4 Facial attractiveness

Each male participant provided a standardized color photograph with a neutral facial expression. Photos were taken from a stationary camera (Canon EOS Rebel T6) in a well-lit room with no windows. Photos were in color and were 4,608 pixels wide by 3,456 pixels high. The facial stimuli took up most of the photo area, with only a small portion of the neck and shoulders visible. Photos were not edited in any manner, with the intention of having the ratings being made on naturalistic stimuli. These photos were rated by eight Caucasian women (M_{age} = 21, SD = 1.70) who were asked to report the level of facial attractiveness of each photo, presented in random order, using a Likert-type scale ranging from 1 = *Very unattractive*, to 10 = *Very attractive*. The raters were reliably consistent in their ratings for facial attractiveness (α = 0.82, 95% LLCI = 0.77, 95% ULCI = 0.86). Previous studies have demonstrated that researchers can obtain reliable attractiveness ratings using a small number of raters (e.g., Buss and Shackelford, 2008; Kordsmeyer et al., 2019).

4 Results

Descriptive statistics are presented in Table 1. Analyses were preformed using SPSS (29.0.1.0; IBM Corp, 2023). First, a bivariate correlation analysis (Table 2) was conducted to determine whether facial attractiveness correlated with the control variables (age, lifestyle factors, skinfold, and hormones) and the two health indicators (symptoms of poor health, sIgA). Age was negatively correlated with T, C, and their interaction, but was otherwise unrelated to lifestyle, health, and facial attractiveness. T and C were positively correlated. C was correlated negatively with exercise and positively with smoking, whereas T was unrelated to all lifestyle variables. Results showed that those with more attractive faces had lower abdominal skin fold values, fewer health problems, exercised more, and were modestly higher in sIgA.

Second, regression analysis was conducted with specific lifestyle indicators, skinfold, age, and hormones, entered as predictors of facial attractiveness simultaneously (Table 3) using Model 1 of the PROCESS macro for SPSS (Hayes, 2013). Results showed that average skinfold and symptoms of poor health[1] were the only statistically significant predictors of facial attractiveness, such that poorer health and more body fat was linked to lower facial attractiveness. Although T, C, and the T x C interaction were not statistically significant predictors of facial attractiveness, these variables trended toward being statistically significant (e.g., p's < 0.10).[2] Visual examination of the interaction suggests a trend toward men with high T and low C being rated as more facially attractive (Figure 2).[3]

Third, we considered the possibility that lifestyle factors might instead have an indirect effect on facial attractiveness, specifically via increased skinfold and associated health problems. To test this prediction, we used AMOS (version 29; Arbuckle, 2019) to create an observed variable path model, with facial attractiveness entered as the dependent variable, unhealthy lifestyle variables (family stress, smoking, drinking alcohol, exercise, and sleep) as the predictors, and abdominal skinfold and poor health symptoms as the sequential mediators. The chi-square test of significance (relative χ^2 index values <3.00), Comparative Fit Index (CFI; values >0.90), and the root mean square error of approximation (RMSEA; values <0.08, Kline, 2016) were used to determine model fit. Indirect (mediation) effects were examined using 1,000 bootstrap samples and bias-corrected 95% confidence intervals. Results showed that, of the unhealthy lifestyle

1 Given past research linking facial attractiveness to less respiratory illness, less antibiotic use, but not fewer stomach bugs in male (Boothroyd et al., 2013) and female participants (Gray and Boothroyd, 2012). We reran the model with only respiratory symptoms comprising the health problems variable. Results did not meaningfully change from those reported.

2 Including saliva sample time of day as a covariate did not meaningfully alter the results reported herein.

3 Excluding the participant case with an outlying (very high) T concentration did not meaningfully alter the results of the interaction results reported herein.

TABLE 2 Bivariate correlations for variables.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Age	-----											
2. Facial attractiveness	−0.01	-----										
3. Health problems	−0.08	−0.28***	-----									
4. sIgA	−0.02	0.14†	−0.14†	-----								
5. Skin Fold	0.10	−0.28***	0.34***	−0.01	-----							
6. Sleep	−0.03	0.04	−0.09	−0.05	−0.05	-----						
7. Exercise	−0.01	0.20*	−0.24***	0.27***	−0.13†	0.05	-----					
8. Alcohol (drink/week)	−0.15†	0.08	0.09	0.01	−0.03	0.08	0.04	-----				
9. Smoking	0.01	−0.02	0.14†	−0.05	−0.01	0.01	−0.31***	0.02	-----			
10. Life Stress	0.01	−0.10	0.12	0.01	0.25***	−0.15†	0.08	0.03	−0.12	-----		
11. Testosterone (T)	−0.24***	0.10	−0.01	−0.02	−0.03	0.05	0.05	0.09	0.06	−0.03	-----	
12. Cortisol (C)	−0.21***	0.04	0.10	0.12	−0.08	−0.06	−0.16*	0.07	0.20*	0.05	0.33***	-----
13. T x C	−0.22***	0.02	0.11	0.11	−0.07	−0.05	−0.15†	0.09	0.20*	0.03	0.41***	0.98***

† = $p < 0.01$; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

TABLE 3 Bootstrapped regression analyses (Model 1, PROCESS Macro) for lifestyle factors, skinfold, sIgA, testosterone and cortisol as predictors of men's facial attractiveness (as rated by women).

	B	Std. Error	*t*	*p*	R^2
DV: Facial attractiveness					0.18
Age	0.01	0.02	0.66	0.51	
Health Problems	−0.89	0.44	−0.20	0.04*	
Log sIgA	0.27	0.28	0.97	0.34	
Average skinfold (mm)	−0.03	0.01	−2.29	0.02*	
Sleep	−0.01	0.10	−0.01	0.99	
Exercise	0.42	0.29	1.45	0.15	
Alcohol (drinks/week)	0.02	0.02	1.26	0.21	
Smoking	0.11	0.47	0.23	0.82	
Life Stress	−0.16	0.25	0.63	0.53	
Testosterone (T)	0.21	0.52	0.40	0.69	
Cortisol (C)	−0.07	0.29	−0.23	0.82	
T x C	−1.18	0.70	−1.70	0.09†	

DV, Dependent Variable; B, Unstandardized regression coefficient. † = $p < 0.10$; * = $p < 0.05$.

habits, life stress ($B = 5.33$, SE = 1.56, $\beta = 0.25$, $p < 0.001$) and exercise ($B = -3.49$, SE = 1.74, $\beta = -0.16$, $p = 0.045$) predicted abdominal skin fold, whereas smoking ($B = -0.84$, SE = 2.91, $\beta = 0.02$, $p = 0.77$), drinking alcohol ($B = 0.04$, *SE = 0.12*, $\beta = -0.03$, $p = 0.74$), and sleep ($B = -0.06$, SE = 0.66, $\beta = -0.01$, $p = 0.93$) did not. Skinfold, in turn, directly predicted the severity and frequency of poor health symptoms ($B = 0.01$, SE = 0.002, $\beta = 0.34$, $p < 0.001$). Poor health symptoms, in turn, predicted lower facial attractiveness ($B = -1.05$, SE = 0.41, $\beta = -0.20$, $p = 0.01$). Skinfold directly negatively predicted facial attractiveness ($B = -0.03$, SE = 0.01, $\beta = -0.21$, $p = 0.007$), and had an indirect effect through the mediator of poor health symptoms ($B = -0.01$, SE = 0.004, $p = 0.015$, 95% LLCI = −0.02, 95% ULCI = −0.003). Exercise ($B = 0.15$, SE = 0.09, $p = 0.03$, 95% LLCI = 0.02, 95% ULCI = 0.36) and life stress ($B = -0.23$, SE = 0.10, $p = 0.02$, 95% LLCI = −0.46, 95% ULCI = −0.06) also showed statistically significant indirect effects through the sequentially mediated pathway of skinfold and poor health symptoms upon facial attractiveness. The sequential mediation model fit the data well, relative χ^2 index = 1.31 ($df = 19$, $p = 0.16$), RMSEA = 0.01 (95% CI = 0.00–0.08), CFI = 0.91 (Figure 1).

5 Discussion

Tests of the links between facial attractiveness and health have yielded mixed results, with some researchers suggesting that lifestyle factors common to both facial attractiveness and health might account for these links (Jones et al., 2021). Accordingly, the present study examined whether indicators of immunocompetence (self-reported poor health symptoms and sIgA), unhealthy lifestyle (smoking, alcohol consumption, poor sleep, lack of exercise, and

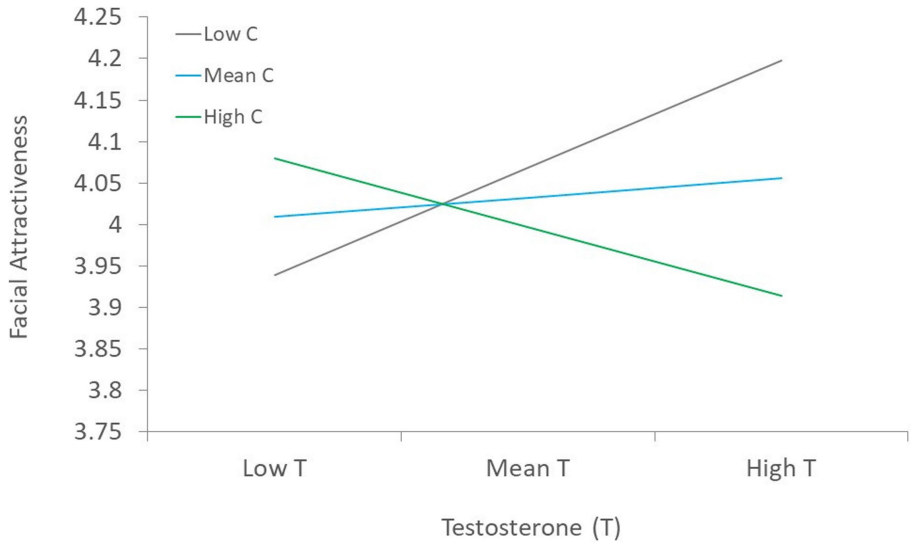


FIGURE 2
Visual depiction of the T x C interaction in a model predicting facial attractiveness.

family stress), age, along with skinfold (as an index of body fat) and hormones (testosterone and cortisol) predicted facial attractiveness. Initially, results showed a significant bivariate link between facial attractiveness and self-reported health symptoms, and this relationship remained statistically significant when including age, unhealthy lifestyle habits, along with skinfold, T, C, and the T x C interaction in the model. However, none of the lifestyle factors themselves predicted facial attractiveness, whereas skinfold did. Previous work has shown that abdominal skinfold is a strong predictor of both facial adiposity and overall facial attractiveness (Sierra-Johnson and Johnson, 2004; Windhager et al., 2011). Similar findings have been observed in women, where BMI predicts facial attractiveness (Han et al., 2016). These links are likely due to related changes to facial morphology that are associated with visceral body fat (Lee and Kim, 2014). Abdominal skinfold is strongly associated with diverse indices of poor health (see Lee and Kim, 2014 for review) and future all-cause mortality in white males (Loh et al., 2018). Skinfold is influenced by lifestyle factors, including nutrition and exercise (e.g., Kwak et al., 2010). We therefore considered whether lifestyle factors instead had an indirect effect upon facial attractiveness via abdominal skinfold and subsequent poor health. Results of a mediation analysis supported this for two of the lifestyle variables: Exercise and life stress. This finding suggests that exercise and life stress have an indirect effect on facial attractiveness via changes to body fat and related poor health symptoms. Some of the lifestyle factors examined here do not necessarily increase body fat. For example, although smoking has been linked to long-term weight gain (Carrasquilla et al., 2024), it may have more meaningful effects in young adulthood upon skin quality and specific health problems (e.g., lung disease). Given that only 6% of our sample smoked, we were likely unable to appropriately assess the potential indirect effects of smoking on facial attractiveness. Future research using a broader community-based sample could address this limitation.

There was also a modest positive correlation between female-rated facial attractiveness and men's sIgA, but this effect was eliminated in the regression equation that included the control variables. This finding corresponds with that of Cai et al. (2019) who found that sIgA was broadly unrelated to female facial appearance. Unlike other sexually dimorphic features that have been linked to sIgA, such as male voice pitch (Arnocky et al., 2018; Hodges-Simeon et al., 2024) and female breast symmetry (Locke and Arnocky, 2021), this null finding could mean that links between health and facial attractiveness are weaker than with other attractive secondary sex characteristics, or perhaps are more strongly driven by lifestyle influences. Future work involving a broader range of immunological markers in relation to facial appearance is therefore encouraged.

Both T and C were also uncorrelated with men's facial attractiveness. However, the regression equation controlling for other variables led to a modestly significant positive link between the T x C interaction and facial attractiveness. Specifically, men with higher T and lower C were rated as most attractive, but this effect did not reach the conventional benchmark for statistical significance. However, it is noteworthy that this finding does conform to that of Rantala et al. (2012), who found the same effect. The overall weak association between hormones and facial attractiveness diverges from a study of women which showed a link between high C and lower facial attractiveness (Rantala et al., 2013), but corresponds with others of male facial attractiveness showing no links with either hormone (Swaddle and Reierson, 2002; Neave et al., 2003; Penton-Voak and Chen, 2004; Kandrik et al., 2017; Kordsmeyer et al., 2019). It has long been assumed in evolutionary psychology that male facial attractiveness is an honest cue of an individual's health and immunocompetence (see Jones et al., 2021 for discussion). Some work does support links between certain immune markers (e.g., high functioning natural killer cells) being associated with female perceptions of male facial attractiveness (Mengelkoch et al., 2022).

T and C did not correlate with symptoms of poor health or sIgA. These findings contrast with previous work on similar samples of young adult men from Northern Ontario that have shown positive links between single samples of sIgA and T (Arnocky et al., 2018). There is a need for more comprehensive assessments of hormonal

markers in relation with health variables, perhaps by assessing 'trait' levels of these hormones across multiple timepoints and days (discussed by Davis and Arnocky, 2022).

5.1 Limitations

One limitation of this work is the use of a homogenous sample of young, primarily Caucasian, undergraduates. This segment of the population tends to be particularly healthy, relative to the broader population. This likely limited variability in lifestyle, which might partly account for the relatively weak predictive role of most lifestyle factors. For instance, College graduates eat healthier, smoke less, and exercise more (see Lawrence, 2017 for review). Similarly, the brief measurement of each lifestyle factor was also limiting. There exist longer form measures of diet quality (Warren-Findlow et al., 2017), drinking behavior (The Drinking Styles Questionnaire DSQ; Smith et al., 1995), tobacco use (e.g., Fagerström Test for Nicotine Dependence FTND; Heatherton et al., 1991), physical activity (Healey et al., 2020), and sleep quality (Yi et al., 2006). The measures used in the present study also asked participants to self-report their own health symptoms and behaviors, and it is important to consider the various sources of self-report bias that can influence this kind of data (e.g., recall bias; Van den Bergh and Walentynowicz, 2016). Moreover, young adult men were the target population in the current study. Therefore, we cannot say that the same results would apply to different age groups, such as older adult men (e.g., Ponholzer et al., 2005). Although this study was sufficiently powered, the sample size was also a limitation, with some researchers suggesting that stability of estimates requires a larger sample than what was achieved in this study (Schönbrodt and Perugini, 2013). Future work should therefore consider these links in larger and more heterogenous samples. Reliance on statistical non-significance may be limited when examining control variables in a regression model to determine whether health remains a meaningful predictor of facial attractiveness. Another limitation lies in the reliance upon assessments of overall facial attractiveness. Although ecologically valid, this measure does not identify specific phenotypic structures of the face that might be tied to either immunocompetence or the effects of an unhealthy lifestyle. Using explicitly facial-oriented variables (e.g., geometric morphometric analyses, GMM) could help to determine how specific facial features contribute proportionally to explained variance in attractiveness, health, and lifestyle. For example, GMM has recently been used to examine facial features in relation with men's and women's sociosexual orientation (Antar and Stephen, 2021).

Future research could examine the impact of both lifestyle factors and hormones during development (adolescence) on adult facial attractiveness. Indeed, some aspects of facial attractiveness are relatively changeable (e.g., such as those affected by current health), whereas other aspects are more stable, such as facial masculinity, which is heavily influenced by steroids and some aspects of immune function during early adolescence (see Foo et al., 2020). Measuring these relationships during adolescence and again during adulthood might help to clarify their unique contributions to facial attractiveness. Finally, it may be useful for future work to consider including a measure of lean muscle mass, such as flexed bicep circumference (see, e.g., Holzleitner et al., 2014) as it may be related both to lifestyle and hormonal factors and has been tied to women's ratings of men's attractiveness via modified facial stimuli (e.g., Lei et al., 2019).

6 Conclusion

Mixed findings characterize the research on the links between facial attractiveness, health, and immunocompetence in men (Jones et al., 2021), which has significant implications for evolutionary theories dealing with the purported ultimate explanations for attractive phenotypic traits (e.g., immunocompetence handicap hypothesis; Nowak et al., 2018). We add to this growing literature to help make sense of the equivocal findings by considering lifestyle and hormonal factors that might influence the links between facial attractiveness and immune function. As suggested by previous authors (Jones et al., 2021), we did find evidence that lifestyle habits (indirectly) and hormones appear to matter when studying the relations between facial attractiveness and immunocompetence. These insights help to advance our understanding of why certain phenotypic traits (e.g., facial characteristics) are regarded as attractive and what kind of information these attractive traits communicate to others, such as health status and lifestyle habits.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://osf.io/adxj3.

Ethics statement

The studies involving humans were approved by Nipissing University Research Ethics Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SA: Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. AD: Writing – review & editing, Writing – original draft, Formal analysis.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding for this research provided by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (DG) (file # RGPIN-2019-05988) awarded to SA.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adam, E. K., Quinn, M. E., Tavernier, R., McQuillan, M. T., Dahlke, K. A., and Gilbert, K. E. (2017). Diurnal cortisol slopes and mental and physical health outcomes: a systematic review and meta-analysis. *Psychoneuroendocrinology* 83, 25–41. doi: 10.1016/j.psyneuen.2017.05.018

Akimoto, T., Kumai, Y., Akama, T., Hayashi, E., Murakami, H., Soma, R., et al. (2003). Effects of 12 months of exercise training on salivary secretory IgA levels in elderly subjects. *Br. J. Sports Med.* 37, 76–79. doi: 10.1136/bjsm.37.1.76

Antar, J. C., and Stephen, I. D. (2021). Facial shape provides a valid cue to sociosexuality in men but not women. *Evol. Hum. Behav.* 42, 361–370. doi: 10.1016/j.evolhumbehav.2021.02.001

Apanius, V. (1998). "Stress and immune response" in *Stress and behavior*. eds. A. P. Møller, M. Milinski and P. J. B. Slater (San Diego, CA: Academic Press), 133–154.

Arbuckle, J. L. (2019). *Amos 26.0 User's Guide*. Chicago: IBM SPSS.

Arnocky, S., Bird, B. M., and Perilloux, C. (2014). "An evolutionary perspective on characteristics of physical attractiveness in humans" in *Psychology of interpersonal perception and relationships*. ed. A. Rennolds (New York: NOVA publishers), 115–155.

Arnocky, S., Denomme, B., Hodges-Simeon, C. R., Hlay, J. K., Davis, A. C., and Brennan, H. (2023). Self-perceived mate value is predicted by biological and self-reported indices of health in young adults. *Adapt. Hum. Behav. Physiol.* 9, 54–71. doi: 10.1007/s40750-022-00209-4

Arnocky, S., Hodges-Simeon, C., Ouellette, D., and Albert, G. (2018). Do men with more masculine voices have better immunocompetence? *Evol. Hum. Behav.* 39, 602–610. doi: 10.1016/j.evolhumbehav.2018.06.003

Bagozzi, R. P., Verbeke, W. J. M. I., Belschak, F., and van Poele, M. (2018). Facial attractiveness as a function of athletic prowess. *Evol. Psychol.* 16:147470491880136. doi: 10.1177/1474704918801369

Bernstein, M. A., Tucker, K. L., Ryan, N. D., O'Neill, E. F., Clements, K. M., Nelson, M. E., et al. (2002). Higher dietary variety is associated with better nutritional status in frail elderly people. *J. Am. Diet. Assoc.* 102, 1096–1104. doi: 10.1016/S0002-8223(02)90246-4

Boothroyd, L. G., Scott, I., Gray, A. W., Coombes, C. I., and Pound, N. (2013). Male facial masculinity as a cue to health outcomes. *Evol. Psychol.* 11, 1044–1058. doi: 10.1177/147470491301100508

Brandtzaeg, P. (2009). Mucosal immunity: Induction, dissemination, and effector functions. *Scand. J. Immunol.* 70, 505–515. doi: 10.1111/j.1365-3083.2009.02319.x

Buss, D. M., and Shackelford, T. K. (2008). Attractive women want it all: good genes, economic investment, parenting proclivities, and emotional commitment. *Evol. Psychol.* 6:600. doi: 10.1177/147470490800600116

Cai, Z., Hahn, A. C., Zhang, W., Holzleitner, I. J., Lee, A. J., DeBruine, L. M., et al. (2019). No evidence that facial attractiveness, femininity, averageness, or coloration are cues to susceptibility to infectious illnesses in a university sample of young adult women. *Evol. Hum. Behav.* 40, 156–159. doi: 10.1016/j.evolhumbehav.2018.10.002

Carrasquilla, G. D., García-Ureña, M., Romero-Lado, M. J., and Kilpeläinen, T. O. (2024). Estimating causality between smoking and abdominal obesity by Mendelian randomization. *Addiction* 119, 1024–1034. doi: 10.1111/add.16454

Cordain, L., Eaton, S. B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B. A., et al. (2005). Origins and evolution of the Western diet: health implications for the 21st century. *Am. J. Clin. Nutr.* 81, 341–354. doi: 10.1093/ajcn.81.2.341

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., and Wu, C. (1995). "Their ideas of beauty are, on the whole, the same as ours": consistency and variability in the cross-cultural perception of female physical attractiveness. *J. Pers. Soc. Psychol.* 68, 261–279. doi: 10.1037/0022-3514.68.2.261

Dalrymple, K. V., Flynn, A. C., Seed, P. T., Briley, A. L., O'Keeffe, M., Godfrey, K. M., et al. (2019). Associations between dietary patterns, eating behaviours and body composition and adiposity in 3-year old children of mothers with obesity. Pediatric. *Obesity* 15:e12608. doi: 10.1111/ijpo.12608

Davis, A. C., and Arnocky, S. (2022). Darwin versus Wallace: esthetic evolution and preferential mate choice. *Front. Psychol.* 13:385. doi: 10.3389/fpsyg.2022.862385

Doshi, D. N., Hanneman, K. K., and Cooper, K. D. (2007). Smoking and skin aging in identical twins. *Arch. Dermatol.* 143, 1543–1546. doi: 10.1001/archderm.143.12.1543

Fahlman, M. M., and Engels, H. J. (2005). Mucosal IgA and URTI in American college football players: A year longitudinal study. *Med Sci Sports Exerc*, 37, 374–380. doi: 10.1249/01.mss.0000155432.67020.88

Fink, B., Neave, N., Manning, J. T., and Grammer, K. (2006). Facial symmetry and judgements of attractiveness, health and personality. *Personal. Individ. Differ.* 41, 491–499. doi: 10.1016/j.paid.2006.01.017

Fink, B., Neave, N., and Seydel, H. (2007). Male facial appearance signals physical strength to women. *Am. J. Hum. Biol.* 19, 82–87. doi: 10.1002/ajhb.20583

Foo, Y. Z., Simmons, L. W., Perrett, D. I., Holt, P. G., Eastwood, P. R., and Rhodes, G. (2020). Immune function during early adolescence positively predicts adult facial sexual dimorphism in both men and women. *Evol. Hum. Behav.* 41, 199–209. doi: 10.1016/j.evolhumbehav.2020.02.002

Gonzalez-Santoyo, I., Wheatley, J. R., Welling, L. L. M., Cárdenas, R. A., Jimenez-Trejo, F., Dawood, K., et al. (2015). The face of female dominance: women with dominant faces have lower cortisol. *Horm. Behav.* 71, 16–21. doi: 10.1016/j.yhbeh.2015.03.006

González-Torres, S., Anaya-Esparza, L. M., Trigueros del Valle, G. F., Rivera-León, E. A., Villagrán, Z., and Sánchez-Enríquez, S. (2023). Skinfold thickness as a cardiometabolic risk predictor in sedentary and active adult populations. *J. Pers. Med.* 13:1326. doi: 10.3390/jpm13091326

Gray, A. W., and Boothroyd, L. G. (2012). Female facial appearance and health. *Evol. Psychol.* 10, 66–77. doi: 10.1177/147470491201000108

Han, C., Hahn, A. C., Fisher, C. I., Debruine, L. M., and Jones, B. C. (2016). Women's facial attractiveness is related to their body mass index but not their salivary cortisol. *Am. J. Hum. Biol.* 28, 352–355. doi: 10.1002/ajhb.22792

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis. A regression-based approach*. New York: Guilford Press.

Healey, E. L., Allen, K. D., Bennell, K., Bowden, J. L., Quicke, J. G., and Smith, R. (2020). Self-report measures of physical activity. *Arthritis Care Res.* 72, 717–730. doi: 10.1002/acr.24211

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., and Fagerström, K. O. (1991). The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. *Br. J. Addict.* 86, 1119–1127. doi: 10.1111/j.1360-0443.1991.tb01879.x

Higgins, E. U., and Du Vivier, W. P. (1992). Invited review: alcohol and the skin. *Alcohol Alcohol.* 27, 595–602. doi: 10.1093/oxfordjournals.alcalc.a045308

Hodges-Simeon, C. R., Albert, G., McHale, T., Gaulin, S. J. C., Gurven, M., Landry, N., et al. (2024). Masculine voice is associated with better mucosal immune defense in adolescent and adult males. *Evol. Hum. Behav.* 45:106590. doi: 10.1016/j.evolhumbehav.2024.05.004

Hodges-Simeon, C. R., Grail, G. P. O., Albert, G., Landry, N., Ortiz, T. L., Carré, J. M., et al. (2020). Testosterone, cortisol, and secretory immunoglobulin-a within a single day and across two sequential days among trans- and cis-gender men. *Steroids* 160:108640. doi: 10.1016/j.steroids.2020.108640

Holzleitner, I. J., Hunter, D. W., Tiddeman, B. P., Seck, A., Re, D. E., and Perrett, D. I. (2014). Men's facial masculinity: when (body) size matters. *Perception* 43, 1191–1202. doi: 10.1068/p7673

Hönekopp, J., Rudolph, U., Beier, L., Liebert, A., and Müller, C. (2010). Physical attractiveness of face and body as indicators of physical fitness in men. *Evol. Hum. Behav.* 28, 106–111. doi: 10.1016/j.evolhumbehav.2006.09.001

Hume, D. K., and Montgomerie, R. (2001). Facial attractiveness signals different aspects of "quality" in women and men. *Evol. Hum. Behav.* 22, 93–112. doi: 10.1016/S1090-5138(00)00065-9

Hurley, L. L., Taylor, R. E., and Tizabi, Y. (2012). Positive and negative effects of alcohol and nicotine and their interactions: a mechanistic review. *Neurotox. Res.* 21, 57–69. doi: 10.1007/s12640-011-9275-6

IBM Corp (2023). *IBM SPSS statistics for windows, version 29.0.1.0*. Armonk, NY: IBM Corp.

Jones, B. C., Holzleitner, I. J., and Shiramizu, V. (2021). Does facial attractiveness really signal immunocompetence? *Trends Cogn. Sci.* 25, 1018–1020. doi: 10.1016/j.tics.2021.09.003

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., and Johnson, R. M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychol. Sci.* 9, 8–13. doi: 10.1111/1467-9280.00002

Kandrik, M., Hahn, A. C., Han, C., Wincenciak, J., Fisher, C. I., DeBruine, L. M., et al. (2017). Does the interaction between cortisol and testosterone predict men's facial attractiveness? *Adapt. Hum. Behav. Physiol.* 3, 275–281. doi: 10.1007/s40750-017-0064-1

Kline, R. B. (2016). *Principles and practice of structural equation modeling. 4th* Edn. New York: Guilford Publications.

Knack, J. M., Arnocky, S., Lawton, C., Wallace, M., and Sussman, J. (2013). "School bullying and health problems: a developmental examination of predictive and protective factors and coping strategies" in *School bullying: Predictive factors, coping strategies, and*

effects on mental health. eds. K. Dekker and M. Dijkstra (New York: NOVA Publishers), 147–167.

Knack, J. M., Iyer, P. A., and Jensen-Ccampbell, L. A. (2012). Not simply "in their heads": individual differences associated with victimization and health1. *J. Appl. Soc. Psychol.* 42, 1625–1650. doi: 10.1111/j.1559-1816.2012.00898.x

Knack, J. M., Jensen-Campbell, L. A., and Baum, A. (2011). Worse than sticks and stones? Bullying is associated with altered HPA axis functioning and poorer health. *Brain Cogn.* 77, 183–190. doi: 10.1016/j.bandc.2011.06.011

Koizumi, K., Hirao, N., Yamanami, H., and Ohira, H. (2023). Effects of mild psychological stress on facial impressions. *Front. Psychol.* 14:1186046. doi: 10.3389/fpsyg.2023.1186046

Kordsmeyer, T. L., Lohöfener, M., and Penke, L. (2019). Male facial attractiveness, dominance, and health and the interaction between cortisol and testosterone. *Adapt. Hum. Behav. Physiol.* 5, 1–12. doi: 10.1007/s40750-018-0098-z

Kwak, L., Kremers, S. P. J., Candel, M. J. J. M., Visscher, T. L. S., Brug, J., and van Baak, M. A. (2010). Changes in skinfold thickness and waist circumference after 12 and 24 months resulting from the NHF-NRG in balance-project. *Int. J. Behav. Nutr. Phys. Act.* 7:26. doi: 10.1186/1479-5868-7-26

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., and Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychol. Bull.* 126, 390–423. doi: 10.1037/0033-2909.126.3.390

Langlois, J. H., Ritter, J. M., Roggman, L. A., and Vaughn, L. S. (1991). Facial diversity and infant preferences for attractive faces. *Dev. Psychol.* 27, 79–84. doi: 10.1037/0012-1649.27.1.79

Lawrence, E. M. (2017). Why do College graduates behave more healthfully than those who are less educated? *J. Health Soc. Behav.* 58, 291–306. doi: 10.1177/0022146517715671

Lee, B. J., and Kim, J. Y. (2014). Predicting visceral obesity based on facial characteristics. *BMC Complement. Altern. Med.* 14:248. doi: 10.1186/1472-6882-14-248

Lei, X., Holzleitner, I. J., and Perrett, D. I. (2019). The influence of body composition effects on male facial masculinity and attractiveness. *Front. Psychol.* 9:2658. doi: 10.3389/fpsyg.2018.02658

Liu, S. W., Lien, M. H., and Fenske, N. A. (2010). The effects of alcohol and drug abuse on the skin. *Clin. Dermatol.* 28, 391–399. doi: 10.1016/j.clindermatol.2010.03.024

Locke, A., and Arnocky, S. (2021). Breast symmetry, but not size or volume, predicts salivary immunoglobulin-a (sIgA) in women. *Evol. Hum. Behav.* 42, 517–523. doi: 10.1016/j.evolhumbehav.2021.05.001

Loh, W. J., Johnston, D. G., Oliver, N., and Godsland, I. F. (2018). Skinfold thickness measurements and mortality in white males during 27.7 years of follow-up. *Int. J. Obes.* 42, 1939–1945. doi: 10.1038/s41366-018-0034-0

Macpherson, A. J., McCoy, K. D., Johansen, F. E., and Brandtzaeg, P. (2008). The immune geography of IgA induction and function. *Mucosal Immunol.* 1, 11–22. doi: doi: 10.1038/mi.2007.6

Mengelkoch, S., Gassen, J., Prokosch, M. L., Boehm, G. W., and Hill, S. E. (2022). More than just a pretty face? The relationship between immune function and perceived facial attractiveness. *Proc. R. Soc. B* 289:20212476. doi: 10.1098/rspb.2021.2476

Murphy, M. H., Blair, S. N., and Murtagh, E. M. (2009). Accumulated versus continuous exercise for health benefit. *Sports Med.* 39, 29–43. doi: 10.2165/00007256-200939010-00003

Nakamura, D., Akimoto, T., Suzuki, S., and Kono, I. (2006). Daily changes of salivary secretory immunoglobulin A and appearance of upper respiratory symptoms during physical training. *J Sports Med Phys Fitness.* 46, 152–157. Available at: https://pubmed.ncbi.nlm.nih.gov/16596115/

Neave, N., Laing, S., Fink, B., and Manning, J. T. (2003). Second to fourth digit ratio, testosterone and perceived male dominance. *Proc. R. Soc. B Biol. Sci.* 270, 2167–2172. doi: 10.1098/rspb.2003.2502

Nowak, J., Pawłowski, B., Borkowska, B., Augustyniak, D., and Drulis-Kawa, Z. (2018). No evidence for the immunocompetence handicap hypothesis in male humans. *Sci. Rep.* 8:7392. doi: 10.1038/s41598-018-25694-0

Pátková, Ž., Schwambergová, D., Třebická Fialová, J., Třebický, V., Stella, D., Kleisner, K., et al. (2022). Attractive and healthy-looking male faces do not show higher immunoreactivity. *Sci. Rep.* 12:18432. doi: 10.1038/s41598-022-22866-x

Penton-Voak, I. S., and Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evol. Hum. Behav.* 25, 229–241. doi: 10.1016/j.evolhumbehav.2004.04.003

Phillips, A. C., Carroll, D., Drayson, M. T., and Der, G. (2015). Salivary immunoglobulin a secretion rate is negatively associated with cancer mortality: The West of Scotland Twenty-07 Study. *PLoS ONE*, 10, Article e0145083. doi: 10.1371/journal.pone.0145083

Ponholzer, A., Plas, E., Schatzl, G., Struhal, G., Brössner, C., Mock, K., et al. (2005). Relationship between testosterone serum levels and lifestyle in aging men. *Aging Male* 8, 190–193. doi: 10.1080/13685530500298154

Pound, N., Lawson, D. W., Toma, A. M., Richmond, S., Zhurov, A. I., and Penton-Voak, I. S. (2014). Facial fluctuating asymmetry is not associated with childhood ill-health in a large British cohort study. *Proc. R. Soc. B Biol. Sci.* 281:20141639. doi: 10.1098/rspb.2014.1639

Rantala, M. J., Coetzee, V., Moore, F. R., Skrinda, I., Kecko, S., Krama, T., et al. (2013). Facial attractiveness is related to women's cortisol and body fat, but not with immune responsiveness. *Biol. Lett.* 9:20130255. doi: 10.1098/rsbl.2013.0255

Rantala, M. J., Moore, F. R., Skrinda, I., Krama, T., Kivleniece, I., Kecko, S., et al. (2012). Evidence for the stress-linked immunocompetence handicap hypothesis in humans. Nature. *Communications* 3:694. doi: 10.1038/ncomms1696

Rhodes, G., Simmons, L. W., and Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evol Hum Behav.* 26, 186–201. doi: 10.1016/j.evolhumbehav.2004.08.014

Rohatgi, K. W., Tinius, R. A., Cade, W. T., Martínez Steele, E., Cahill, A. G., and Parra, D. C. (2017). Relationships between consumption of ultra-processed foods, gestational weight gain and neonatal outcomes in a sample of US pregnant women. *PeerJ* 5:e4091. doi: 10.7717/peerj.4091

Roney, J. R. (2016). Theoretical frameworks for human behavioral endocrinology. *Horm. Behav.* 84, 97–110. doi: 10.1016/j.yhbeh.2016.06.004

Roney, J. R., Hanson, K. N., Durante, K. M., and Maestripieri, D. (2006). Reading men's faces: Women's mate attractiveness judgments track men's testosterone and interest in infants. *Proc. R. Soc. Lond. B Biol. Sci.* 273, 2169–2175. doi: 10.1098/rspb.2006.3569

Schönbrodt, F. D., and Perugini, M. (2013). At what sample size do correlations stabilize? *J. Res. Pers.* 47, 609–612. doi: 10.1016/j.jrp.2013.05.009

Shackelford, T. K., and Larsen, R. J. (1997). Facial asymmetry as an indicator of psychological, emotional, and physiological distress. *J. Pers. Soc. Psychol.* 72, 456–466. doi: 10.1037/0022-3514.72.2.456

Sierra-Johnson, J., and Johnson, B. D. (2004). Facial fat and its relationship to abdominal fat: a marker for insulin resistance? *Med. Hypotheses* 63, 783–786. doi: 10.1016/j.mehy.2004.06.020

Skinner, A. L., Woods, A., Stone, C. J., Penton-Voak, I., and Munafò, M. R. (2017). Smoking status and attractiveness among exemplar and prototypical identical twins discordant for smoking. Royal Society open. *Science* 4:1076. doi: 10.1098/rsos.161076

Smith, G. T., McCarthy, D. M., and Goldman, M. S. (1995). Self-reported drinking and alcohol-related problems among early adolescents: dimensionality and validity over 24 months. *J. Stud. Alcohol* 56, 383–394. doi: 10.15288/jsa.1995.56.383

Sundelin, T., Lekander, M., Sorjonen, K., and Axelsson, J. (2017). Negative effects of restricted sleep on facial appearance and social appeal. *R. Soc. Open Sci.* 4:918. doi: 10.1098/rsos.160918

Swaddle, J. P., and Reierson, G. W. (2002). Testosterone increases perceived dominance but not attractiveness in human males. *Proc. R. Soc. Lond. B Biol. Sci.* 269, 2285–2289. doi: 10.1098/rspb.2002.2165

Van Den Abbeele, J., Penton-Voak, I. S., Attwood, A. S., Stephen, I. D., and Munafò, M. R. (2015). Increased facial attractiveness following moderate, but not high, alcohol consumption. *Alcohol Alcohol.* 50, 296–301. doi: 10.1093/alcalc/agv010

Van den Bergh, O., and Walentynowicz, M. (2016). Accuracy and bias in retrospective symptom reporting. *Curr. Opin. Psychiatry* 29, 302–308. doi: 10.1097/YCO.0000000000000267

Visine, A., Durand, V., Guillou, L., Raymond, M., and Berticat, C. (2024). Chronic and immediate refined carbohydrate consumption and facial attractiveness. *PLoS One* 19:e0298984. doi: 10.1371/journal.pone.0298984

Volkmann, E. R., and Weekes, N. Y. (2006). Basal SIgA and cortisol levels predict stress-related health outcomes.*Stress and Health: Journal of the International Society for the Investigation of Stress.* 22: 11–23. doi: 10.1002/smi.1077

Warren-Findlow, J., Reeve, C. L., and Racine, E. F. (2017). Psychometric validation of a brief self-report measure of diet quality: the DASH-Q. *J. Nutr. Educ. Behav.* 49, 92–99.e1. doi: 10.1016/j.jneb.2016.09.004

Windhager, S., Schaefer, K., and Fink, B. (2011). Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness, dominance, and masculinity. *Am. J. Hum. Biol.* 23, 805–814. doi: 10.1002/ajhb.21219

Yi, H., Shin, K., and Shin, C. (2006). Development of the sleep quality scale. *J. Sleep Res.* 15, 309–316. doi: 10.1111/j.1365-2869.2006.00544.x

frontiers | Frontiers in Psychology

TYPE Brief Research Report
PUBLISHED 09 September 2024
DOI 10.3389/fpsyg.2024.1425389

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Ray Garza,
Texas A&M International University, United States
Hajime Fukui,
Nara University of Education, Japan
Jan Havlicek,
Charles University, Czechia

*CORRESPONDENCE
Stefan M. M. Goetz
✉ stefanmmgoetz@gmail.com;
✉ goetzste@msu.edu

RECEIVED 29 April 2024
ACCEPTED 26 August 2024
PUBLISHED 09 September 2024

CITATION
Goetz SMM, Lucas T and Carré JM (2024) Under the influence: exogenous testosterone influences men's cross-sex perceptions of sexual interest.
Front. Psychol. 15:1425389.
doi: 10.3389/fpsyg.2024.1425389

Under the influence: exogenous testosterone influences men's cross-sex perceptions of sexual interest

Stefan M. M. Goetz[1]*, Todd Lucas[1] and Justin M. Carré[2]

[1]Charles Stewart Mott Department of Public Health, College of Human Medicine, Michigan State University, Flint, MI, United States, [2]Laboratory of Social Neuroendocrinology, Department of Psychology, Nipissing University, North Bay, ON, Canada

The sexual misperception bias is a cognitive bias in which men tend to overestimate sexual interest from women, potentially shaped by evolutionary mating strategies. Testosterone, often linked to mating behaviors, might play a role in sustaining sexual overperceptions. To explore this possibility, we conducted a placebo-controlled study with 190 heterosexual men, administering either 11 mg of testosterone or a placebo. Participants interacted with an attractive female confederate, while naïve raters assessed the confederate's affiliative behaviors. Our findings suggest that exogenous testosterone did not broadly impact sexual overperception. However, we found that affiliative behavior from the confederate was positively correlated with perceived sexual interest among testosterone-treated, but not placebo-treated men. In addition, we found that this effect among testosterone-treated men was contingent on their self-perceived attractiveness. Specifically, the confederate's affiliative behaviors were positively correlated with perceived sexual interest, but only for testosterone-treated men with average or above average self-perceived attractiveness. Furthermore, our data revealed that men's tendency to project their own short-term and long-term mating interests increases as a function of self-perceived attractiveness, and this coupling is enhanced by testosterone for long-term interest. Taken together, these results suggest that testosterone may potentiate existing biases, particularly when sexual motivation is high, and bias perceptions of friendly behavior when engaging in cross-sex mindreading. This study adds to the understanding of the neuroendocrine bases of social cognition, suggesting that testosterone can affect men's perceptions of potential mates.

KEYWORDS

exogenous testosterone, individual differences, social perception, sexual misperception, error management theory, attractiveness projection bias

1 Introduction

In the context of mating intelligence, cross-sex "mindreading"—the cognitive representation of the desires of a potential mate of the opposite sex—has long intrigued evolutionary psychologists. Cross-culturally, compared to women, men tend to perceive higher levels of flirtatiousness, seductiveness, and promiscuousness (La France et al., 2009). The sexual misperception bias (SMB) describes the tendency of men to overperceive sexual interest (Abbey, 1982). At the ultimate level, the sex difference in SMB has been proposed to function as a means for males to promote mating by minimizing missed opportunities (Buss, 2001).

Error Management Theory (EMT) posits that inferring sexual intentions under conditions of uncertainty was a recurrent adaptive challenge over evolutionary history (Haselton and Buss, 2000). According to EMT, under such circumstances where there are either cost asymmetries between errors and/or benefit asymmetries between hits, bias will evolve to reduce rates of the higher-cost error and maximize rates of hits to maximize benefits (Haselton and Buss, 2000; Brandner et al., 2021). In fitness terms, false positives (overperception) were less costly to men than false negatives (underperception), and correctly identifying sexual opportunities more beneficial than correctly identifying disinterest. Due to sex differences in obligatory parental investment, the asymmetries in the costs/benefits of errors/hits were not analogous for women, leading to a sex difference in SMB.

Beyond global sex differences, research has increasingly focused on individual differences that contribute to SMB (e.g., Perilloux et al., 2012). Factors such as the tendency to project one's own desires onto others, self-perceived attractiveness, sex drive, and sociosexuality—one's openness to uncommitted sex (Penke and Asendorpf, 2008)—have been identified as significant contributors (Shotland and Craig, 1988; Koenig et al., 2007; Perilloux et al., 2012; Lee et al., 2020; Samara et al., 2021). Several of these mechanisms are associated with androgens and have themselves been posited to be the result of sexual selection on males for pluralistic mating (Baumeister et al., 2001; Schmitt, 2005; Lippa, 2009; Howell et al., 2012; Roth et al., 2021), and in the context of SMB, have been variably invoked as a potential proximate explanation for the observed sex difference (Koenig et al., 2007; Roth et al., 2021; Samara et al., 2021).

To the best of our knowledge, the only study probing the role of androgens in individual differences in SMB, examined endogenous testosterone, pre-post competition and social interaction in a sample of 57 undergraduate men (Perilloux, 2011). Participants engaged in a 20-min online game, ostensibly against another male student, either winning or losing in the final 2 min before interacting on a cooperative puzzle task with an attractive female confederate instructed to behave in a "friendly but not flirtatious" manner. Although the competition failed to induce a 'winner effect' (Geniole et al., 2017), increases in testosterone both across the competition and social interaction with the woman were associated with greater SMB. Given these results, and that many evolved psychological sex differences and individual differences are potentiated by sex hormones (Hooven, 2021), testosterone is a plausible candidate for explaining individual differences in susceptibility to SMB.

In vertebrates, testosterone functions as an endocrinological mechanism supporting reproductive physiology and behavior (Hau, 2007; Fuxjager and Schuppe, 2018). It acts both throughout development and in a trait and state-like manner, exerting both organizational and activational effects[1] (Phoenix et al., 1959; Sellers et al., 2007; Van der Meij et al., 2012). In humans, basal testosterone is associated with sexual function and activity, mating success, and relationship status (for review, see Luberti and Carré, 2023), and numerous studies have documented associations between social challenges pertaining to reproductive behavior and endogenous surges of testosterone (Roney et al., 2007; Ronay and Hippel, 2010; Zilioli and Bird, 2017). Thus, it is evident that testosterone is highly responsive to various reproductive challenges hinting at its direct involvement.

Nevertheless, after two decades of research, few studies have applied single-dose pharmacological challenge paradigms to more firmly establish whether these are causal effects (see Carré et al., 2023). Those that have examined questions pertaining to mating psychology have produced mixed evidence in favor of testosterone's causal role. For instance, testosterone increases impulsivity for sexual rewards (Wu et al., 2022), shifts men's preferences towards more feminized faces (Han et al., 2020), and differentially influences men's perceptions of female facial attractiveness across relationship status such that among single men, testosterone increases attraction to relatively unattractive faces, while among partnered men, testosterone increases attraction to relatively attractive faces (Geniole et al., 2022). Others have found that testosterone modulates facial femininity preferences across short- and long-term mating contexts, but the effect is driven by a decrease in preference for facial femininity among long-term mating preferences (Bird et al., 2016), and one recent study failed to find an effect of exogenous testosterone on an attitudinal measure of sociosexuality (Polo et al., 2024). Yet no studies have explored the role of exogenous testosterone on the capacity to infer mental states in a mating domain. Given testosterone's role in the development, maintenance, and individual variation of various psychological sex differences, and the—albeit mixed—empirical evidence connecting testosterone with SMB directly (Perilloux, 2011) and with several established psychological mediators of SMB, in the present exploratory study we investigated the impact of a single-dose of exogenous testosterone on men's perceptions of a woman's sexual interest upon their initial encounter during a brief interaction.

We employed a naturalistic zero-acquaintance paradigm to investigate the effects of a single-dose of exogenous testosterone, sexual interest, and self-perceived attractiveness on SMB and perceptions of interest (PSI). We also considered the interaction between these variables on PSI, as well as whether an attractive female confederate's affiliative behavior further influenced perceptions. Based on previous evidence, we expected that testosterone, sexual interest, and self-perceived attractiveness would be positively linked to SMB and PSI. Furthermore, we expected that the degree to which the female confederate engaged in affiliative behavior would be positively linked to SMB and PSI.

2 Methods

2.1 Participants

Participants were recruited from a larger study on testosterone and decision-making that was being conducted in the laboratory on the same day. The participant pool comprised 322 healthy heterosexual men, aged 18–40, recruited from local media sites, medical databases, and colleges and universities in Ontario, Canada. All procedures were approved by the university ethics board. After the decision-making study, participants were given the option to complete a second short study for an additional $5 CAD. Of the original 322 participants, 212 opted to participate. Drug-treatment was not related to opt-in rates [$X^2(1) = 0.642$, $p = 0.423$], nor was basal testosterone [$X^2(1) = 0.002$, $p = 0.960$].

1 Organizational effects refer to the largely permanent influences on traits resulting from sex hormone exposure during sensitive periods of the mediating tissues' development, primarily occurring *in utero* (Phoenix et al., 1959), but also during perinatal (Lanciotti et al., 2018) and pubertal stages (Sisk and Zehr, 2005). In contrast, activational effects describe the influence of sex hormones, typically post-pubertally, which are transient and often contingent upon earlier organizational effects (for a review, see Arnold, 2009).

Of those who opted to participate, 20 indicated a sexual orientation other than exclusively heterosexual[2], and two participants failed to record their perception of her sexual interest, leaving a final sample of 190 (M_{age} = 23, SD_{age} = 5.19; range = 25) exclusively heterosexual men, of whom 73% identified as White, 7% as multiracial ancestry, 6% Asian, 6% Black, 4% Aboriginal, 1% Hispanic, and 1% as 'other'. For the analyses of SMB, due to an error in instruction the confederate failed to include her actual interest in the participant, reducing the sample to 175 for these analyses.

Participants also indicated their relationship status, choosing between single (44%), non-exclusively dating (5%), exclusively dating one person (38%), common law (1%), engaged (2%), married (10%), and in an open marriage (1%); those indicating that they were in an exclusive relationship, married or common law, or engaged were recoded as 'paired' (n = 97), whereas those indicating that they were single, dating but not committed, dating multiple, or being in an open marriage were recoded as 'single' (n = 93) as these relationship status entail being active on the 'mating market' and testosterone may function to serve continued mating-seeking effort (e.g., van Anders and Watson, 2007). Indeed, basal testosterone was significantly higher among single [M = 65.84 pg./mL, SD = 33.10; M(age adjusted) = 79.19] versus paired men [M = 54.41 pg./mL, SD = 29.29; M(age adjusted) = 69.02; $t(188) = 2.525$, $p = 0.012$, $d = 0.37$], as was age [$M_{paired} = 24.09$, $SD = 5.93$; $M_{single} = 22.01$, $SD = 4.06$; $t(188) = -2.812$, $p = 0.003$, $d = -0.41$]. Relationship status was both independent of opting into the study [$X^2(1) = 0.003$, $p = 0.959$] and drug-treatment [$X^2(1) = 1.020$, $p = 0.312$].

2.2 Task procedures

Participants arrived at the laboratory for the economic decision-making study between 9:30AM and 5:30PM for a 2-h study. The protocol involved completing a battery of questionnaires and computer based neuroeconomic decision-making tasks.

Thirty minutes after arriving, participants were administered a single dose of either 5.5 mg of testosterone gel to each nostril (11 mg in total) or placebo gel. Both the participants and researchers were blind to the drug-treatment status. The dosage used rapidly increases testosterone concentrations to the high-normal physiological range within 15 min and remaining elevated up until 180 min post administration (Geniole et al., 2019). An additional 90 min elapsed before the participants were invited to participate in a second study (the current study) on impression formation and personality judgments. After agreeing to participate, they were told that the computer was currently in use and asked to wait in a conference room equipped with audio-video devices.

There, participants were seated across from an attractive female confederate,[3] there presumably as a recruiter for another study. The confederate was instructed to be friendly and warm and to initiate a scripted conversation if the participant failed to do so after 60 s had elapsed.

After 3 min had passed, the research assistant escorted the participant to another room where they completed a short questionnaire (items described below). The participant was then debriefed and dismissed.

2.3 Measures

2.3.1 Self-perceived attractiveness

Participants rated their own overall attractiveness ("*How attractive do you consider yourself?*"), using a 10-point scale (1 = not at all, 10 = very much so; $M = 6.33$, $SD = 1.82$).

2.3.2 Sexual misperception bias

For each participant, we calculated SMB by subtracting the participant's estimate of the confederate's short-term mating interest ($M = 3.20$, $SD = 2.35$) from the confederate's actual short-term mating interest ($M = 2.51$, $SD = 2.45$). SMB ranged from −9 to 8 ($M = -0.73$, $SD = 3.30$). Thus, negative values indicated overperception, while a value of zero indicated accurate perception.

2.3.3 Perceived sexual interest

PSI was calculated by adding the participant's perceptions of her short-term and long-term interest (LT: $M = 2.97$, $SD = 2.33$; PSI: $M = 6.16$, $SD = 4.23$). Perception of short- and long-term mating orientation were not significantly different [$t(189) = 1.55$, $p = 0.122$] and were strongly correlated [$r(188) = 0.631$, $p < 0.001$].

2.3.4 Sexual interest

The participants also reported their interest in the confederate as a short-term ($M = 5$, $SD = 3.24$) and long-term partner ($M = 3.98$, $SD = 2.86$). A paired sample t-test revealed a greater interest in the participant as a short-term partner than a long-term partner ($M_{diff} = 1.02$, $d = 0.34$, $p < 0.001$).

2.3.5 Affiliative behaviors

Two trained male judges blind to the hypotheses and drug-treatment[4] rated the confederate's behaviors from the audio-video recordings, across nine affiliative behaviors (see Van der Meij et al., 2012)[5]. Inter-rater reliability for the full scale across the two raters was adequate (Cronbach's $\alpha = 0.852$). The nine items were then averaged across raters and a composite affiliation measure was computed by weighting the items by their factor loadings using a single-factor principal axis analysis ($M = 23.26$, $SD = 3.66$). See Table 1 for the zero-order correlations and means and standard deviations of the study variables.

2 Sexual orientation was assessed by asking participants to select any of the following that applied: heterosexual, homosexual, bisexual (clarified to participants as sexual attraction and desire toward both men and women), or asexual (clarified to the participants as having little/no sexual attraction or desire toward others).

3 Independent raters rated her as 8/10 whereby 1 – not at all attractive and 10 = very attractive.

4 A binomial test indicated that the judges were no better than chance at guessing drug-treatment (p's > 0.314).

5 Note, affiliative behaviors were also measured in men by two trained female judges blind to the hypotheses. However, as the focus of the current paper is on men's perceptions, we did not include his affiliative behaviors in the present set of analyses. See Goetz (2020) for a description of the main study.

2.3.6 Basal Testosterone

Basal testosterone was collected via Salivette® swabs and immediately stored at −20°C until hormone analysis. Saliva was assayed in duplicate via enzyme immunoassay kits from DRG International, Inc. The average intra-assay coefficient of variation (CV) was 9.80% and the average inter-assay CV was 13.13%. The average value across duplicates was used in our analysis after outliers were Winsorized to ±3 *SD*s. Basal testosterone was marginally higher in the testosterone relative to the placebo group [$M_{testo} = 63.81$ pg./mL ($SD = 31.65$) versus $M_{placebo} = 55.77$ pg./mL ($SD = 31.04$); $t(190) = -1.774$, $p = 0.078$, $d = -0.26$].

2.4 Analytical approach

Due to violations of the normality assumption, nonparametric tests, Wilcoxon rank tests, were used to examine the presence of SMB in the overall sample and separately for testosterone and placebo groups. A Mann–Whitney test was used to test for mean differences in SMB between groups. A series of GLMs were used to probe the effects of drug-treatment, affiliation behavior, self-perceived attractiveness, and short- and long-term interest on PSI. Benjamini-Hochberg procedure was used to control false discovery rate (FDR). All predictors were entered simultaneously in the models. Simple slope analyses were used to characterize the nature of any significant interactions. Robust SE estimation (HC1) was used to compensate for violations of parametric tests assumptions (MacKinnon and White, 1985). All analyses were conducted using Jamovi (v2.4.11) using the GAMLj3 module.

3 Results

3.1 Sexual perception bias

The Shapiro–Wilk test of normality was violated ($p < 0.001$); as such, a Wilcoxon rank test was applied which confirmed the presence of the SMB in our sample, testing whether SMB differed from zero ($M = -0.731$, $SD = 3.198$; $W(174) = 2666.5$, $p = 0.005$, rank biserial = −0.289 [$d = -0.22$]). A negative nonparametric (Spearman's rho) correlation was observed between basal testosterone and SMB ($\rho(175) = -0.208$, $p = 0.006$ [$r(175) = -0.153$, $p = 0.043$]). When analyzing SMB partial residuals controlling for basal testosterone, SMB was not present ($M = 0.229$, $SD = 3.259$; $W(174) = 8179.0$, $p = 0.476$, rank biserial = 0.062 [$d = 0.070$]). A Mann–Whitney *U*, indicated the lack of difference in SMB between treatment groups ($U = 3436.0$, $p = 0.242$, rank biserial = 0.101 [$d = 0.107$]). The Mann–Whitney *U* conducted on SMB residualized for basal testosterone further confirmed the lack of difference ($U = 3601.0$, $p = 0.510$, rank biserial = 0.058 [$d = 0.064$]). However, when participants were split by treatment, the Wilcoxon rank test indicated that those receiving placebo did not evince SMB ($M = -0.548$, $SD = 3.028$, $W(83) = 596.0$, $p = 0.145$, rank biserial = −0.226 [$d = -0.181$]), whereas those receiving testosterone did ($M = -0.901$, $SD = 3.537$, $W(90) = 758.5$, $p = 0.017$, rank biserial = −0.334 [$d = -0.244$]). Despite the treatment contingent effect, when these analyses were conducted using SMB partial residuals controlling for basal testosterone, neither group evinced SMB (testosterone: $M = 0.128$, $SD = 3.483$; $W(90) = 2061.0$, $p = 0.901$, rank biserial = −0.015[$d = 0.037$]; placebo: $M = 0.338$, $SD = 3.014$; $W(83) = 2073.0$, $p = 0.200$, rank biserial = 0.161 [$d = 0.112$]).

3.1.1 Perception of sexual interest

With respect to his perceptions of her sexual interest (PSI), a regression analysis revealed that there was no evidence for a main effect of drug-treatment on PSI [$ß = 0.114$, $t(186) = 0.794$, $p = 0.428$]. Her affiliative behavior however was associated with PSI [$ß = 0.132$, $t(186) = 1.996$, $p = 0.047$]. This main effect was qualified by a drug-treatment-by-affiliation interaction [$ß = 0.287$, $t(186) = 2.214$, $p = 0.028$]. Simple slopes analysis indicated that among men receiving placebo, her affiliation behaviors were not correlated with his PSI [$ß = -0.014$, $t(186) = -0.162$, $p = 0.872$]; however, among men receiving testosterone, her affiliative behaviors were positively correlated with his PSI [$ß = 0.280$, $t(186) = 2.848$, $p = 0.005$], indicating

TABLE 1 Zero order correlations, means and standard deviations.

	1	2	3	4	5	6	7	8	9	10
1. Drug-Tx										
2. SPA	0.07									
3. FST	−0.01	0.02								
4. SMB	−0.05	**−0.33*****	**0.70*****							
5. PSI	0.05	**0.46*****	0.07	**−0.59*****						
6. ST	−0.01	0.05	0.05	**−0.27*****	**0.52*****					
7. LT	−0.05	−0.04	−0.04	**−0.21****	**0.44*****	**0.53*****				
8. Aff	−0.08	0.07	**0.14**†	0.05	0.12	−0.05	0.05			
9. Basal Testo	**0.13**†	0.10	−0.07	**−0.15***	**0.12**†	>0.01	0.09	−0.09		
10. Rel. Status	0.07	−0.03	−0.10	−0.07	−0.07	**−0.27*****	**−0.29*****	**0.13**†	**−0.18***	
Mean	0.52	6.33	2.51	−0.73	6.16	5.00	3.98	23.26	59.92	0.51
SD	0.50	1.82	2.45	3.30	4.23	3.24	2.86	3.66	31.53	0.50

Drug-Tx (0 = placebo, 1 = testosterone). SPA, self-perceived attractiveness; FST, female short-term; SMB, sexual misperception bias; PSI, perception of sexual interest; ST, short-term; LT, long-term; Aff, affiliation behavior; Basal Testo (pg/mL) Winsorized to ± 3 SDs; Rel. Status, relationship status (0 = single, 1 = paired). Bold indicates statistical significance at the following levels: †$p < 0.1$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

that exogenous testosterone may sensitize men to affiliation cues when inferring sexual interest. Including basal testosterone in the model, both as a covariate and as an interaction term did not alter the results ($\Delta|\beta|$'s < 0.016, Δp's < 0.020), though when the interaction terms were included, the main effect of basal testosterone was trending [$\beta = 0.121$, $t(182) = 1.684$, $p = 0.094$], whereas when only included as a control variable, the main effect of basal testosterone was not significant ($\beta = 0.117$, $p = 0.125$). None of the basal testosterone interaction terms were significant (p's > 0.228). Likewise, including relationship status in the model as either a covariate or moderator did not alter the results ($\Delta|\beta|$'s < 0.037, Δp's < 0.018) and was unrelated to PSI (p's > 0.151).

3.1.2 Perception of sexual interest and self-perceived attractiveness

In a regression analysis predicting PSI using her affiliative behaviors, drug-treatment, and his self-perceived attractiveness, self-perceived attractiveness was strongly associated with PSI [$\beta = 0.514$, $t(182) = 8.879$, $p < 0.001$]. No other main effects emerged (p's > 0.128). The two-way interaction between self-perceived attractiveness and treatment was not significant [$\beta = 0.152$, $t(182) = 1.368$, $p = 0.173$], while affiliation moderated both the previously observed treatment effect ($p = 0.020$) and self-perceived attractiveness [$\beta = 0.111$, $t(182) = 2.049$, $p = 0.042$]. These effects were qualified by a significant three-way interaction [$\beta = 0.363$, $t(182) = 3.434$, $p < 0.001$]. The simple slopes analysis revealed that among those low in self-perceived attractiveness (−1 SD), her affiliation was not associated with PSI irrespective of drug-treatment [placebo: $\beta = 0.021$, $t(182) = 0.307$, $p = 0.759$; testosterone: $\beta = -0.041$, $t(182) = -0.514$, $p = 0.608$]; however, the effect of her affiliation behaviors among men of average self-perceived attractiveness who had received the placebo was not significant [$\beta = -0.053$, $t(182) = -0.539$, $p = 0.591$], while among those who had received testosterone, her affiliation behavior was positively associated with PSI [$\beta = 0.251$, $t(182) = 2.984$, $p = 0.003$]. Increasing self-perceived attractiveness 1SD further sharpened this effect [Placebo: $\beta = -0.127$, $t(182) = -0.784$, $p = 0.434$; Testosterone: $\beta = 0.544$, $t(182) = 3.94$, $p < 0.001$]. Fisher Z-tests indicated that the testosterone/high self-perceived attractiveness slope was steeper than the testosterone/average self-perceived attractiveness slope ($z = -1.82$, $p < 0.07$; see Figure 1). Once more, including basal testosterone in the model as a covariate and as an interaction term did not alter the results ($\Delta|\beta|$'s < 0.043, Δp's < 0.033), nor did including relationship status ($\Delta|\beta|$'s < 0.050, Δp's < 0.009; p's > 0.108).

3.1.3 Perception of sexual interest, self-perceived attractiveness, and sexual interest

The influence of short-term and long-term interest in the confederate were considered separately. In the first regression analysis in which PSI was regressed onto the participants' short-term interest, the main effect of short-term interest was significant [$\beta = 0.502$, $t(182) = 9.486$, $p < 0.001$] as was the previous main effect of self-perceived attractiveness ($p < 0.001$). These main effects were qualified by a significant two-way interaction between self-perceived attractiveness and short-term interest [$\beta = 0.233$, $t(182) = 4.093$, $p < 0.001$]. Simple slopes analyses characterizing the two-way interaction between self-perceived attractiveness and short-term interest indicated that short-term interest was positively associated with PSI at each level of self-perceived attractiveness and increased monotonically (β's = 0.270, 0.503, 0.736, −1 SD, mean, +1 SD respectively; p's < 0.001). The two-way interactions between drug-treatment and self-perceived attractiveness and drug-treatment and short-term interest were not significant ($|\beta$'s$|$ < 0.149, p's > 0.088), nor was the three-way interaction ($\beta = -0.032$, $p = 0.791$). Including basal testosterone did not alter the results ($\Delta|\beta|$'s < 0.021, Δp's < 0.001), nor did basal testosterone moderate any of the above results (p's > 0.249). While including relationship status as a covariate or moderator did not alter any of the results ($\Delta|\beta|$'s < 0.038, Δp's < 0.001), there was a significant relationship for the main effect relationship status when included as a moderator ($\beta = 0.205$, $p = 0.040$), which was qualified by a significant relationship status by short-term interest by self-perceived attractiveness interaction ($\beta = -0.200$, $p = 0.035$; see Supplementary material).

The results of the regression analysis of long-term interest paralleled those of short-term interest except for a significant three-way interaction between drug-treatment, self-perceived attractiveness, and long-term interest [$\beta = 0.180$, $t(182) = 2.076$, $p = 0.039$]. The same monotonic pattern emerged but only for the testosterone condition; the simple interaction between self-perceived attractiveness and long-term interest was not significant among the placebo condition ($p = 0.063$) but was in the testosterone group ($p < 0.001$). Fisher Z-test indicated that the testosterone/low self-perceived attractiveness slope was significantly flatter than the testosterone/average- ($z = -2.73$, $p = 0.006$) and testosterone/high self-perceived attractiveness ($z = -4.01$, $p < 0.001$); the difference between the latter were trending ($z = -1.92$, $p = 0.055$; see Figure 2). Including basal testosterone did not change any of the above effects with the exception that when included as a covariate, the three-way interaction was no longer significant and only trending ($p = 0.054$), however, the simple effects remained unchanged. When included as a moderator, none of the effects changed in significance ($\Delta|\beta|$'s < 0.048, Δp's < 0.015). The two-way and three-way interactions between basal testosterone, long-term mating interest, and self-perceived attractiveness were not significant (p's > 0.612), with the exception of a trending three-way interaction between basal testosterone, treatment, and long-term interest [$t(175) = -1.805$, $\beta = -0.232$, $p = 0.073$; see Supplementary material]. Relationship status did not alter these results ($\Delta|\beta|$'s < 0.049, Δp's < 0.015), with once again, the exception of the previous three-way interaction no longer reaching significance ($p = 0.054$).

4 Discussion

The present experiment explored the causal influence of a single dose of exogenous testosterone on men's perception of a novel woman's sexual interest, while also considering the role of her affiliation behavior, his self-perceived attractiveness, and short- and long-term interest in her. SMB was observed in the overall sample but was absent in the placebo group while present in the testosterone group; however, contrary to our hypothesis, testosterone did not significantly increase the magnitude of SMB ($d = 0.11$), suggesting that testosterone does not directly influence SMB. Furthermore, after controlling for basal testosterone, which was associated with greater overperception, SMB was not observed in either the overall sample nor in the testosterone group. However, we did find that men's perception of the woman's sexual interest, when considered alongside her affiliative behaviors, was indeed influenced by

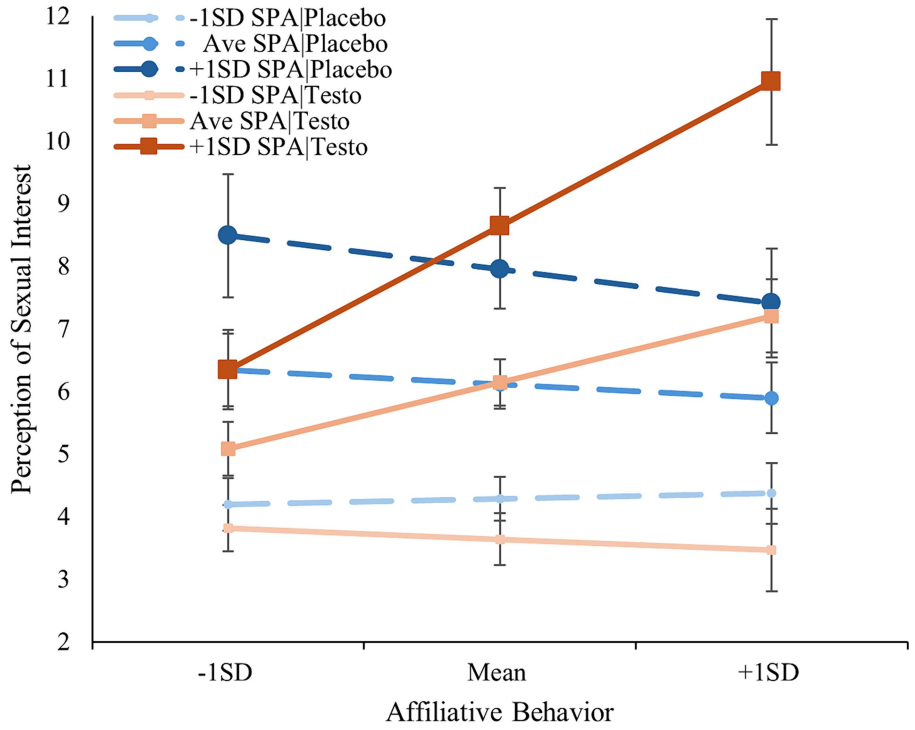


FIGURE 1
Men's perception of sexual interest as a function of drug-treatment, self-perceived attractiveness, and female affiliative behaviors. Error bars represent standard errors. SPA, self-perceived attractiveness.

testosterone. Specifically, testosterone appears to have sensitized men to behavioral cues, with affiliative behavior positively correlated with perception of sexual interest, but only in the testosterone condition. Notably, when her affiliation was low, testosterone decreased men's perception of her sexual interest. Additionally, we observed that the salience of affiliation on perception only occurred above a threshold of his self-perceived attractiveness, beyond which affiliative behaviors were increasingly influential. Thus, it appears that testosterone sensitives men to affiliation cues, but only among men with positive self-perceptions of their own mate-value. Consistent with projectionist accounts (Shotland and Craig, 1988; Henningsen and Henningsen, 2010; Lemay and Wolf, 2016; Lee et al., 2020; Samara et al., 2021), men's short-term and long-term interest were strongly associated with perception of sexual interest, as was their self-perceived attractiveness (Perilloux et al., 2012; Lee et al., 2020; but see Samara et al., 2021). Intriguingly, as men's self-perceived attractiveness increased, they were more likely to project their own short-term sexual interest onto her. A similar effect was described by Lemay and Wolf (2016), only for mate-value rather than self-perceived attractiveness, although the latter was a component of the scale. For long-term interest however, this augmentation was only present among men receiving testosterone suggesting that testosterone boosts the tendency to project one's own sexual interest, particularly among individuals for whom that interest is more likely to be mutual (Lee et al., 2020). Similarly, to the extent that self-perceived attractiveness indexes self-confidence (Bale and Archer, 2013), testosterone may promote courtship by amplifying the tendency to project desire more readily among those high in confidence. While we did observe that basal testosterone was associated with greater overperception, similar to what was found by Perilloux (2011, p. 73), and was higher among unpaired men (e.g., van Anders and Watson, 2006, 2007), neither moderated nor diminished the relationships described above.

Our findings also contribute to an ongoing debate regarding whether projection is itself a sex-specific mechanism selected to promote male overperception. Lee et al. (2020) argued that a more parsimonious evolutionary model assumes that projection leads to mating success regardless of sex, requiring only quantitative changes in the tendency to project one's desire irrespective of sex, rather than a qualitative sex-specific projection mechanism. Empirically, both men and women do project their own interest when making cross-sex inferences about a target's sexual desire (e.g., Lemay and Wolf, 2016). However, as pointed out by Roth and colleagues (2021), this account fails to take into consider men's greater baseline interest in potential partners (Kurzban and Weeden, 2005; Samara et al., 2021) and the greater inherent costs of selecting a suboptimal mate faced by women (Todd et al., 2007). Furthermore, evidence indicates that even when interest is present, the tendency to project is higher among men (Samara et al., 2021). Our finding that testosterone moderates projection further suggests that the tendency might be sex-linked.

Recent research suggests that women tend to signal interest more frequently than men, despite being less interested (Bendixen et al., 2019). While this often may produce misunderstandings, it is interesting to consider whether this dynamic evolved as a means for women to bias the composition of the pool of suitors in favor attractive

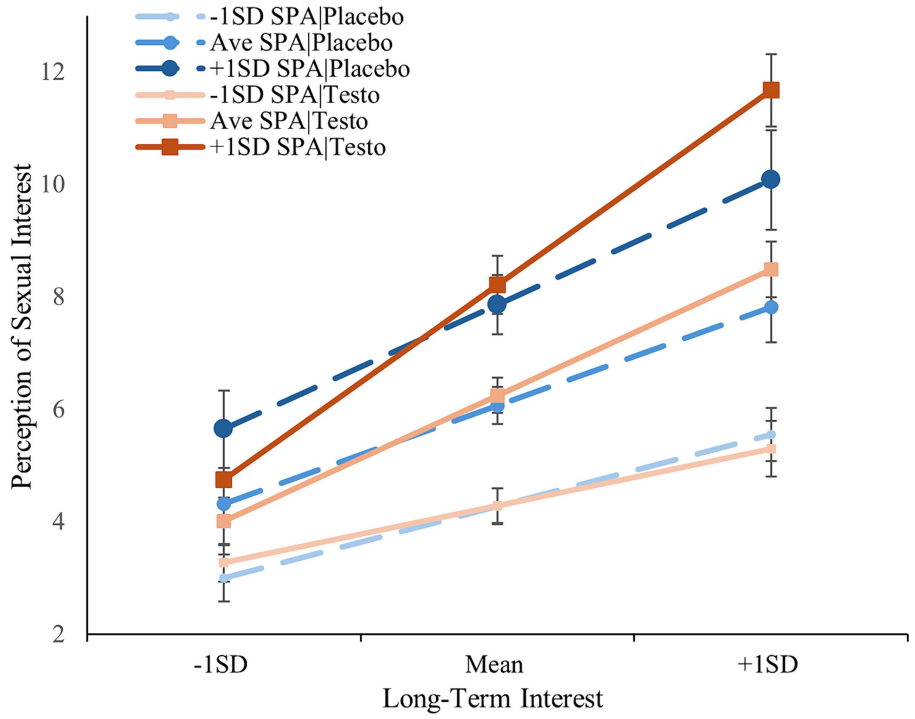


FIGURE 2
Men's perception of sexual interest as a function of drug-treatment, self-perceived attractiveness, and their long-term interest. Error bars represent standard errors. SPA, self-perceived attractiveness.

high testosterone men, given that we found that the effect of testosterone on behavioral cue salience was moderated by self-perceived attractiveness. Although signaling disinterest would also reduce the proportion of these men, disinterest is also more easily detected in general (Hall et al., 2015). Given this dynamic, the proportions of quality suitors under the high signaling scenario is likely to be higher, which could explain why women signal more in the first place. Even though affiliation cues were misperceived, displays of interest by the perceiver do sometimes promote a self-fulfilling prophecy, particularly if the target finds the suitor attractive (Lemay and Wolf, 2016).

Several limitations suggest both a cautious interpretation of the findings and other future directions. First, we only had a single confederate, which may have rendered the SMB measure susceptible to her own idiosyncratic mating criteria. However, we also considered outcome variables that were not subject to her judgments. Second, the study was not originally designed to test the sexual overperception bias and reflects exploratory analyses that should be confirmed by subsequent research (see Goetz, 2020 for published dissertation and the preregistration available on the Open Science Framework: https://osf.io/65btc/).

Another concern arises regarding both the context in which testosterone was administered and the timing of the interaction. The circumstances under which participants experienced the increase in testosterone were artificial, involving a series of economic decision tasks before interacting with the confederate (for a similar critique of oxytocin studies, see Gangestad, 2016). While the interaction itself aligns with a putative functional domain of testosterone, the circumstances leading to its increase fail to model those of any evolutionary relevance. At best we can conclude that we tested the causal role of basal testosterone and at worst that of falling levels. Regarding the latter, our protocol positioned the interaction around 120 min post-administration, a time at which testosterone was likely declining—albeit while remaining above baseline (Geniole et al., 2019)—the effect of which is unknown. Future research should aim to administer testosterone under ecologically relevant circumstances (e.g., in the context of competition or courtship) and coordinate subsequent measurements along the pharmacokinetic curve to isolate the specific effects of interest.

Another limitation is that we could not measure pre-post changes in testosterone due to potential sample contamination via postnasal drip that commonly occurs with nasal testosterone administration, obviating our ability to directly compare our results to those of Perilloux (2011) who found that acute change in testosterone were associated with SMB. Nevertheless, the observed effects of exogenous testosterone provide a close proxy to acute endogenous changes.

This study provides mixed evidence for testosterone's role in sexual perception. Nonetheless, testosterone might still play a role in shaping the development of these mechanisms via organizational effects (e.g., Sisk and Zehr, 2005; Berenbaum and Beltz, 2011; Shirazi et al., 2020), and in supporting their expression. Indeed, research indicates that sociosexuality—a potential mediator of sexual overperception (Howell et al., 2012; Lee et al., 2020)—is related to pubertal timing (Shirazi et al., 2020); crucially, developmental sensitivity to steroid hormones wanes with age, suggesting that organizational effects are involved (Berenbaum and Beltz, 2011). Furthermore, this exact mechanism [SOI] has been shown to be unrelated to circulating testosterone (Stern et al., 2020).

Although speculative, self-perceived attractiveness may provide an index of organization effects given that many of the secondary sexual characteristics, on which men's self-perceived attractiveness are based (Lukaszewski et al., 2014; Sneade and Furnham, 2016; Kanavakis et al., 2021), are developmentally driven by androgens (e.g., muscularity and facial masculinity; Lassek and Gaulin, 2009; Whitehouse et al., 2015; Hodges-Simeon et al., 2016).

These limitations notwithstanding, this study provides the first evidence that exogenous testosterone may influence the sexual overperception bias. Although we did not show a direct effect of exogenous testosterone on SMB, we found basal testosterone was associated with greater overperception and that exogenous testosterone amplified the impact of the woman's affiliation behavior on perceived sexual interest, contingent upon men's self-perceived attractiveness, with the pattern emerging primarily among men of average and above attractiveness. The projection effect was observed for both short-term and long-term interest, with self-perceived attractiveness strengthening the effect for short-term interest and being contingent upon testosterone for long-term interest such that the projection effect was strengthened by self-perceived attractiveness only among those receiving testosterone. These results highlight both testosterone's role as a social hormone influencing person perception and the importance of considering individual differences in moderating its effects.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/65btc/.

Ethics statement

The studies involving humans were approved by Nipissing University Research Ethics Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SG: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing. TL: Formal analysis, Supervision, Writing – review & editing. JC: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to JMC (RPGIN-2020-05120).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1425389/full#supplementary-material

References

Abbey, A. (1982). Sex differences in attributions for friendly behavior: do males misperceive females' friendliness? *J. Pers. Soc. Psychol.* 42, 830–838. doi: 10.1037/0022-3514.42.5.830

Arnold, A. P. (2009). The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Horm. Behav.* 55, 570–578. doi: 10.1016/j.yhbeh.2009.03.011

Bale, C., and Archer, J. (2013). Self-perceived attractiveness, romantic desirability and self-esteem: a mating sociometer perspective. *Evol. Psychol.* 11, 68–84. doi: 10.1177/147470491301100107

Baumeister, R. F., Catanese, K. R., and Vohs, K. D. (2001). Is there a gender difference in strength of sex drive? Theoretical views, conceptual distinctions, and a review of relevant evidence. *Personal. Soc. Psychol. Rev.* 5, 242–273. doi: 10.1207/S15327957PSPR0503_5

Bendixen, M., Kennair, L. E. O., Biegler, R., and Haselton, M. G. (2019). Adjusting signals of sexual interest in the most recent naturally occurring opposite-sex encounter in two different contexts. *Evol. Behav. Sci.* 13, 345–365. doi: 10.1037/ebs0000162

Berenbaum, S. A., and Beltz, A. M. (2011). Sexual differentiation of human behavior: effects of prenatal and pubertal organizational hormones. *Front. Neuroendocrinol.* 32, 183–200. doi: 10.1016/j.yfrne.2011.03.001

Bird, B. M., Welling, L. L., Ortiz, T. L., Moreau, B. J., Hansen, S., Emond, M., et al. (2016). Effects of exogenous testosterone and mating context on men's preferences for female facial femininity. *Horm. Behav.* 85, 76–85. doi: 10.1016/j.yhbeh.2016.08.003

Brandner, J. L., Pohlman, J., and Brase, G. L. (2021). On hits and being hit on: error management theory, signal detection theory, and the male sexual overperception bias. *Evol. Hum. Behav.* 42, 331–342. doi: 10.1016/j.evolhumbehav.2021.01.002

Buss, D. M. (2001). Cognitive biases and emotional wisdom in the evolution of conflict between the sexes. *Curr. Dir. Psychol. Sci.* 10, 219–223. doi: 10.1111/1467-8721.00153

Carré, J. M., Hemsworth, C. T., and Alabi, I. A. (2023). Does testosterone modulate aggression and mating behavior in humans? A narrative review of two decades of single-dose testosterone administration research. *Adapt. Hum. Behav. Physiol.* 9, 400–425. doi: 10.1007/s40750-023-00229-8

Fuxjager, M. J., and Schuppe, E. R. (2018). Androgenic signaling systems and their role in behavioral evolution. *J. Steroid Biochem. Mol. Biol.* 184, 47–56. doi: 10.1016/j.jsbmb.2018.06.004

Gangestad, S. W. (2016). An evolutionary perspective on oxytocin and its behavioral effects. *Curr. Opin. Psychol.* 7, 115–119. doi: 10.1016/j.copsyc.2015.08.016

Geniole, S. N., Bird, B. M., Ruddick, E. L., and Carré, J. M. (2017). Effects of competition outcome on testosterone concentrations in humans: an updated meta-analysis. *Horm. Behav.* 92, 37–50. doi: 10.1016/j.yhbeh.2016.10.002

Geniole, S. N., Proietti, V., Bird, B. M., Ortiz, T. L., Bonin, P. L., Goldfarb, B., et al. (2019). Testosterone reduces the threat premium in competitive resource division. *Proc. R. Soc. B* 286:20190720. doi: 10.1098/rspb.2019.0720

Geniole, S. N., Proietti, V., Robinson, B. A., Bird, B. M., Watson, N. V., Bonin, P. L., et al. (2022). Relatively rapid effects of testosterone on men's ratings of female attractiveness depend on relationship status and the attractiveness of stimulus faces. *Horm. Behav.* 142:105174. doi: 10.1016/j.yhbeh.2022.105174

Goetz, S.M.M. (2020). *Testosterone and interpersonal attraction: A placebo-controlled design.* Doctoral thesis, Wayne State University.

Hall, J. A., Xing, C., and Brooks, S. (2015). Accurately detecting flirting: error management theory, the traditional sexual script, and flirting base rate. *Commun. Res.* 42, 939–958. doi: 10.1177/0093650214534972

Han, C., Zhang, Y., Lei, X., Li, X., Morrison, E. R., and Wu, Y. (2020). Single dose testosterone administration increases men's facial femininity preference in a Chinese population. *Psychoneuroendocrinology* 115:104630. doi: 10.1016/j.psyneuen.2020.104630

Haselton, M. G., and Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *J. Pers. Soc. Psychol.* 78, 81–91. doi: 10.1037/0022-3514.78.1.81

Hau, M. (2007). Regulation of male traits by testosterone: implications for the evolution of vertebrate life histories. *BioEssays* 29, 133–144. doi: 10.1002/bies.20524

Henningsen, D. D., and Henningsen, M. L. M. (2010). Testing error management theory: exploring the commitment skepticism bias and the sexual overperception bias. *Hum. Commun. Res.* 36, 618–634. doi: 10.1111/j.1468-2958.2010.01391.x

Hodges-Simeon, C. R., Hanson Sobraske, K. N., Samore, T., Gurven, M., and Gaulin, S. J. (2016). Facial width-to-height ratio (fWHR) is not associated with adolescent testosterone levels. *PLoS One* 11:e0153083. doi: 10.1371/journal.pone.0153083

Hooven, C. (2021). The story of testosterone, the hormone that dominates and divides us. New York: Holt.

Howell, E. C., Etchells, P. J., and Penton-Voak, I. S. (2012). The sexual overperception bias is associated with sociosexuality. *Personal. Individ. Differ.* 53, 1012–1016. doi: 10.1016/j.paid.2012.07.024

Kanavakis, G., Halazonetis, D., Katsaros, C., and Gkantidis, N. (2021). Facial shape affects self-perceived facial attractiveness. *PLoS One* 16:e0245557. doi: 10.1371/journal.pone.0245557

Koenig, B. L., Kirkpatrick, L. A., and Ketelaar, T. (2007). Misperception of sexual and romantic interests in opposite-sex friendships: four hypotheses. *Pers. Relat.* 14, 411–429. doi: 10.1111/j.1475-6811.2007.00163.x

Kurzban, R., and Weeden, J. (2005). HurryDate: mate preferences in action. *Evol. Hum. Behav.* 26, 227–244. doi: 10.1016/j.evolhumbehav.2004.08.012

La France, B. H., Henningsen, D. D., Oates, A., and Shaw, C. M. (2009). Social–sexual interactions? Meta-analyses of sex differences in perceptions of flirtatiousness, seductiveness, and promiscuousness. *Commun. Monogr.* 76, 263–285. doi: 10.1080/03637750903074701

Lanciotti, L., Cofini, M., Leonardi, A., Penta, L., and Esposito, S. (2018). Up-to-date review about minipuberty and overview on hypothalamic-pituitary-gonadal axis activation in fetal and neonatal life. *Front. Endocrinol.* 9:410. doi: 10.3389/fendo.2018.00410

Lassek, W. D., and Gaulin, S. J. (2009). Costs and benefits of fat-free muscle mass in men: relationship to mating success, dietary requirements, and native immunity. *Evol. Hum. Behav.* 30, 322–328. doi: 10.1016/j.evolhumbehav.2009.04.002

Lee, A. J., Sidari, M. J., Murphy, S. C., Sherlock, J. M., and Zietsch, B. P. (2020). Sex differences in misperceptions of sexual interest can be explained by sociosexual orientation and men projecting their own interest onto women. *Psychol. Sci.* 31, 184–192. doi: 10.1177/0956797619900315

Lemay, E. P., and Wolf, N. R. (2016). Projection of romantic and sexual desire in opposite-sex friendships: how wishful thinking creates a self-fulfilling prophecy. *Personal. Soc. Psychol. Bull.* 42, 864–878. doi: 10.1177/0146167216646077

Lippa, R. A. (2009). Sex differences in sex drive, sociosexuality, and height across 53 nations: testing evolutionary and social structural theories. *Arch. Sex. Behav.* 38, 631–651. doi: 10.1007/s10508-007-9242-8

Luberti, F. R., and Carré, J. M. (2023). Testosterone's role in modulating human behaviors relevant to mating and parenting. *Front. Neuroendocrinol.* 1–21. doi: 10.1016/j.yfrne.2023.101112

Lukaszewski, A. W., Larson, C. M., Gildersleeve, K. A., Roney, J. R., and Haselton, M. G. (2014). Condition-dependent calibration of men's uncommitted mating orientation: evidence from multiple samples. *Evol. Hum. Behav.* 35, 319–326. doi: 10.1016/j.evolhumbehav.2014.03.002

MacKinnon, J. G., and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econ.* 29, 305–325. doi: 10.1016/0304-4076(85)90158-7

Penke, L., and Asendorpf, J. B. (2008). Beyond global sociosexual orientations: a more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *J. Pers. Soc. Psychol.* 95:1113. doi: 10.1037/0022-3514.95.5.1113

Perilloux, C.J. (2011). Sexual misperception: individual differences and context effects. Doctoral Dissertation, The University of Texas at Austin. doi: 10.1177/0956797611424162

Perilloux, C., Easton, J. A., and Buss, D. M. (2012). The misperception of sexual interest. *Psychol. Sci.* 23, 146–151.

Phoenix, C. H., Goy, R. W., Gerall, A. A., and Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology* 65, 369–382.

Polo, P., Fajardo, G., Muñoz-Reyes, J. A., Valenzuela, N. T., Belinchón, M., Figueroa, O., et al. (2024). The role of exogenous testosterone and social environment on the expression of sociosexuality and status-seeking behaviors in young Chilean men. *Horm. Behav.* 161:105522. doi: 10.1016/j.yhbeh.2024.105522

Ronay, R., and Hippel, W. V. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Soc. Psychol. Personal. Sci.* 1, 57–64. doi: 10.1177/1948550609352807

Roney, J. R., Lukaszewski, A. W., and Simmons, Z. L. (2007). Rapid endocrine responses of young men to social interactions with young women. *Horm. Behav.* 52, 326–333. doi: 10.1016/j.yhbeh.2007.05.008

Roth, T. S., Samara, I., and Kret, M. E. (2021). Ultimate and proximate factors underlying sexual overperception bias: a reply to Lee et al. (2020). *Evol. Hum. Behav.* 42, 73–75. doi: 10.1016/j.evolhumbehav.2020.06.002

Samara, I., Roth, T. S., and Kret, M. E. (2021). The role of emotion projection, sexual desire, and self-rated attractiveness in the sexual overperception bias. *Arch. Sex. Behav.* 50, 2507–2516. doi: 10.1007/s10508-021-02017-5

Schmitt, D. P. (2005). Sociosexuality from Argentina to Zimbabwe: a 48-nation study of sex, culture, and strategies of human mating. *Behav. Brain Sci.* 28, 247–275. doi: 10.1017/S0140525X05000051

Sellers, J. G., Mehl, M. R., and Josephs, R. A. (2007). Hormones and personality: testosterone as a marker of individual differences. *J. Res. Pers.* 41, 126–138. doi: 10.1016/j.jrp.2006.02.004

Shirazi, T. N., Self, H., Dawood, K., Cárdenas, R., Welling, L. L., Rosenfield, K. A., et al. (2020). Pubertal timing predicts adult psychosexuality: evidence from typically developing adults and adults with isolated GnRH deficiency. *Psychoneuroendocrinology* 119:104733. doi: 10.1016/j.psyneuen.2020.104733

Shotland, R. L., and Craig, J. M. (1988). Can men and women differentiate between friendly and sexually interested behavior? *Soc. Psychol. Q.* 51, 66–73. doi: 10.2307/2786985

Sisk, C. L., and Zehr, J. L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Front. Neuroendocrinol.* 26, 163–174. doi: 10.1016/j.yfrne.2005.10.003

Sneade, M., and Furnham, A. (2016). Hand grip strength and self-perceptions of physical attractiveness and psychological well-being. *Evol. Psychol. Sci.* 2, 123–128. doi: 10.1007/s40806-016-0042-z

Stern, J., Karastoyanova, K., Kandrik, M., Torrance, J., Hahn, A. C., Holzleitner, I., et al. (2020). Are sexual desire and sociosexual orientation related to men's salivary steroid hormones? *Adapt. Hum. Behav. Physiol.* 6, 447–466. doi: 10.1007/s40750-020-00148-y

Todd, P. M., Penke, L., Fasolo, B., and Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proc. Natl. Acad. Sci.* 104, 15011–15016. doi: 10.1073/pnas.0705290104

van Anders, S. M., and Watson, N. V. (2006). Relationship status and testosterone in north American heterosexual and non-heterosexual men and women: cross-sectional and longitudinal data. *Psychoneuroendocrinology* 31, 715–723. doi: 10.1016/j.psyneuen.2006.01.008

van Anders, S. M., and Watson, N. V. (2007). Testosterone levels in women and men who are single, in long-distance relationships, or same-city relationships. *Horm. Behav.* 51, 286–291. doi: 10.1016/j.yhbeh.2006.11.005

Van der Meij, L., Almela, M., Buunk, A. P., Fawcett, T. W., and Salvador, A. (2012). Men with elevated testosterone levels show more affiliative behaviours during interactions with women. *Proc. R. Soc. B Biol. Sci.* 279, 202–208. doi: 10.1098/rspb.2011.0764

Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., and Maybery, M. T. (2015). Prentatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society: B.* 282:20151351. doi: 10.1098/rspb.2015.1351

Wu, Y., Ou, J., Wang, X., Zilioli, S., Tobler, P. N., and Li, Y. (2022). Exogeneous testosterone increases sexual impulsivity in heterosexual men. *Psychoneuroendocrinology* 145:105914. doi: 10.1016/j.psyneuen.2022.105914

Zilioli, S., and Bird, B. M. (2017). Functional significance of men's testosterone reactivity to social stimuli. *Front. Neuroendocrinol.* 47, 1–18. doi: 10.1016/j.yfrne.2017.06.002

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 18 October 2024
DOI 10.3389/fpsyg.2024.1462891

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Zhiyi Chen,
Army Medical University, China
Kolbjørn Kallesten Brønnick,
University of Stavanger, Norway
Summer Mengelkoch,
Texas Christian University, United States

*CORRESPONDENCE
Elizabeth Hampson
✉ ehampson@uwo.ca

RECEIVED 10 July 2024
ACCEPTED 07 October 2024
PUBLISHED 18 October 2024

CITATION
Hampson E, Abrahamson SN, Breddy TN, Iqbal M and Wolff ER (2024) Current oral contraceptive use affects explicit and implicit measures of depression in women. *Front. Psychol.* 15:1462891. doi: 10.3389/fpsyg.2024.1462891

Current oral contraceptive use affects explicit and implicit measures of depression in women

Elizabeth Hampson[1,2]*, Sara N. Abrahamson[2], Taylor N. Breddy[1], Maisha Iqbal[2] and Elena R. Wolff[1]

[1]Laboratory of Neuroendocrinology, Department of Psychology, University of Western Ontario, London, ON, Canada, [2]Neuroscience Program, Schulich School of Medicine & Dentistry, University of Western Ontario, London, ON, Canada

Some data suggest that increased depressive symptoms may occur in women using combined oral contraceptives (OCs). However, this idea is controversial and the existing evidence is conflicting. The present study compared negative affect in 53 healthy women (M_{age} = 19.9 years) during intervals of active daily OC hormone intake and during the washout week of the contraceptive cycle when no exogenous estrogens or progestins are used. A prospective counterbalanced repeated-measures study design was employed. Depressive affect was evaluated using standard psychometric tests of explicit (self-perceived) and implicit negative affect. Implicit measures are considered less subject to bias related to social expectations, self-awareness, or willingness to disclose. Other than their usual OCs, participants were medication-free and had been using OCs for a median of 12 mo. We found that measures of implicit affect (e.g., Affect Misattribution Procedure, Emotional Stroop Test) displayed a more depressive-like pattern of performance during active hormone intake, particularly among a subgroup of OC users who reported experiencing high levels of depressive affect more generally. In contrast, participants' self-perceptions suggested that they perceived their negative symptoms to be greater during the 'off' phase of the OC cycle, when OC steroids are withdrawn and menses occurs. The present findings reinforce the possibility of depressive mood effects associated with OC usage, and highlight the utility of including implicit measures, but also illustrate the complexity of mood assessment in OC users.

KEYWORDS

oral contraceptive, hormonal contraceptive, mood, affect, depression, depressive, implicit, progestin

Introduction

Oral contraceptives (OCs) are used by over 150 million women worldwide (United Nations Population Division, 2019). Most OCs are 'combined' oral contraceptives consisting of 2 hormonal constituents: a synthetic estrogen and a progestin. Individual brands vary in dose and the exact progestin used, but all OCs inhibit ovulation and thus reduce the possibility of conception (Dickey and Seymour, 2021). While their gynecological actions are well-documented, little is known about OC effects in the central nervous system (CNS). Animal studies have indicated that estrogen and progesterone receptors are widely distributed in the CNS (Brinton et al., 2008; Östlund et al., 2003; Weiser et al., 2008), and hormone-receptor binding can produce a wide spectrum of regional changes in neurotransmission and synaptic

plasticity (Barth et al., 2015; Galea et al., 2017). The synthetic hormones used in OCs can bind to these receptors with affinities that may equal or sometimes even exceed their endogenous ligands (Escande et al., 2006; see Hampson, 2023 for an overview). Despite this fact, scientific knowledge of OC effects in the CNS is still rudimentary.

Although high in contraceptive efficacy, a frequent reason why women discontinue their use of OCs is self-perceived mood changes (Bancroft and Sartorius, 1990; Lundin et al., 2017; Rosenberg and Waugh, 1998). These include increased irritability, mood lability, or depressive affect. Although reported as far back as the 1970s, most evidence for depressive mood changes among OC users has been anecdotal or based on self-reports. Over the past 8–10 years an upsurge in research in the neuroscience and epidemiology communities has sought to document the occurrence, prevalence, scope, and severity of OC-related mood effects, including a number of large-scale population-level studies (e.g., Gawronska et al., 2024; Johansson et al., 2023; Skovlund et al., 2016; Toffol et al., 2011; Wiréhn et al., 2010; Zettermark et al., 2018). However, the findings have been mixed and conflicting. Several recent large-scale studies confirmed an increase in rates of clinical depression among OC users relative to non-users (e.g., Skovlund et al., 2016; see also Skovlund et al., 2018) based on markers such as antidepressant use, suicidal behavior, or documented clinical diagnoses, but other work also using large samples has failed to confirm any differences (McKetta and Keyes, 2019; Lundin et al., 2022) or, less frequently, even suggested a protective effect of hormonal contraception (Keyes et al., 2013; Toffol et al., 2011). Some data suggest adverse mood effects may be limited to adolescents or those who initiate OC use during adolescence (Anderl et al., 2020; de Wit et al., 2020; Johansson et al., 2023; Zettermark et al., 2018). Yet other studies have found that adult women who otherwise experience premenstrual dysphoria when in a non-medicated state may in fact experience greater stabilization in mood under OC use (Bäckström et al., 1992; Lundin et al., 2017). Studies addressing potential CNS mechanisms for depressive changes are even fewer in number (e.g., Larsen et al., 2020; Larsen et al., 2022; Porcu et al., 2019) and are at present, inconclusive. Consequently, the question of negative mood change under OC use remains unanswered.

Even among those studies that report increased rates of depression, the numbers of women who experience clinical depression under OC use is small (an estimated 4–10%; Skovlund et al., 2016; Hampson, 2023; Sundström Poromaa and Segebladh, 2012). This is far below the self-reported rates of mood change among OC users, which range as high as 30–50% (Grant and Pryse-Davies, 1968) and might reflect the magnitude of the changes. We propose that larger numbers of women may experience subclinical increases in dysphoria which are milder and thus not identified by recent clinical studies that have focused on formal clinical diagnoses of major depressive disorder and/or the use of antidepressant therapies. Indeed, a historical review of the question of mood change in OC users (Oinonen and Mazmanian, 2002) advocated for the use of a dimensional approach to evaluate mood as key to future progress. The possibility of subclinical mood disturbance has received little systematic research to date, and is thus a significant gap in developing a complete understanding of the spectrum of OC effects on emotional processing. Factors that determine whether mood changes occur under OC use, and their severity, likely include both person-specific vulnerabilities and drug-related variables such as differences in dose, type of progestin, or duration of OC use, all of which have a potential to contribute to variability in outcomes. Using a dimensional approach, we recently found that negative affect varied by OC progestin subtype in a sample of 193 long-term OC users taking various standard OC drug formulations (Hampson, 2023). In general, the phenomenon and scope of mood changes in response to OC initiation has been inadequately explored. Developing an improved understanding is important because of the significance of mood status for women's quality of life and interpersonal functioning (including their perceptions of, and interactions with, significant others in their social environment).

In the present pilot study, a repeated-measures design was used to evaluate affective status in female users of OCs. We employed a formal self-report measure of depression and other dimensions of mood, but we also used implicit measures, in which depressive affect manifests itself automatically, without overt participant awareness, via mood-driven changes in accuracy or speed of responding to affectively-laden visual stimuli. To our knowledge, the present work is the first to include implicit measures to detect depressive affect in OC users. Implicit measures are sometimes considered a 'truer' window into the affective system, as responses are rapid and driven by automatic associative learning processes that operate outside the direct awareness of participants and are thus less subject to reporting or disclosure biases that can impede explicit, deliberate, self-reports (Suslow et al., 2019; DeCoster et al., 2006). Though not used previously to study OC users, implicit tasks have been used in past studies of major depression in other contexts, where they have revealed selective differences in the emotional processing characteristics of depressed versus non-depressed individuals (Leppänen, 2006; Weightman et al., 2014). Implicit affectivity is also predictive of spontaneous behavioral, autonomic, and HPA (hypothalamic–pituitary–adrenal) responses to everyday emotional stimuli and stress (e.g., Bodenschatz et al., 2018; van der Ploeg et al., 2016; Quirin et al., 2009).

In the present work, our major focus was users' perceptions and reactions to facial expressions of emotion. If hormones in OCs do engender a more depressive mindset, we expected to find higher depression scores on an explicit measure of mood during the 'active' phase of the contraceptive cycle (when synthetic hormones are actively used each day) than during the monthly 'inactive' phase when hormones are not used. We also expected to find a more depressive profile of responses on implicit measures during active intake.

Materials and methods

Participants

Participants were 62 female university students or administrative staff ages 18–26 years, who had used a standard combined OC for at least the past 3 months (median = 12.0 months) and had no chronic health conditions or other prescription medications. One volunteer diagnosed recently with depression was retained in the sample. Demographics are given in Table 1. Participants were recruited through flyers at the university and were reimbursed a total of $25 to cover their travel costs.

General procedure

A repeated-measures design was used. Two counterbalanced test sessions were scheduled, and were performed during the active intake

TABLE 1 Participant Demographics ($N = 62$).

Mean age at testing (years)	19.92 (SD = 1.85)
Mean age at first OC use (years)	17.45 ($SD = 1.74$)
Mean years of university study	2.55 ($SD = 1.55$)
Median time on current OC pill (months)	12.00
Type of progestin used:	
Norethindrone acetate	$n = 4$
Levonorgestrel	$n = 26$
Norgestimate or desogestrel	$n = 21$
Drospirenone or cyproterone	$n = 11$
	Total = 62
Current brand of OC:	
Number of monophasic brands (%)	54 (87%)
Number bi- or triphasic brands (%)	8 (12%)

OC, oral contraceptive.

of OC hormones ("Active") and during no hormone intake ("Inactive"). Although infrequently used to study OC effects, this type of study design can identify short-term effects of OC use that depend on changes in the serum availability of exogenous steroids. Such designs have been used previously to identify cognitive or affective alterations associated with active intake of OCs (e.g., Noachtar et al., 2023; Hampson et al., 2022) and are especially powerful if used in a within-subjects context as recommended by several recent reviews (Beltz, 2022; Hampson et al., 2022).

The active test session took place during the second or third week of the active phase of each woman's contraceptive cycle (i.e., after 14–21 consecutive days of daily ethinyl estradiol and progestin use). Timing of the testing was controlled and coincided with the days of maximum OC steroid dosage based on the particular brand of OC pill used by each individual. Thus the timing was individualized to each participant's OC regimen, based on contraceptive pill and health information shared in advance via a brief online screening. The Inactive session was also tightly controlled. It took place during the 'washout' or 'inactive' days of the cycle, after at least 3 days of no hormone use ($M = 4.81$ days after the last active pill was taken) to allow adequate washout of the OC steroids to occur. Once again, the exact timing was personalized and was dictated by each individual's specific brand of OC.

On the test day, participants reported to a university laboratory for a 40–45 min test session where explicit and implicit measures of mood were administered one-on-one by a trained examiner. Although the timing of each session was targeted prospectively to coincide with specific timepoints in the OC cycle, we also monitored and verified the timing retrospectively. Specifically, participants brought their current package of OC pills to each study visit to allow the researchers to visually inspect and verify the exact prescription details, brand name of OC pill, and the number of pills used (or still remaining) on the date of testing, to confirm that the desired timepoints had been properly targeted (or detect any scheduling errors). In addition to assessment of mood, participants completed several control tasks during each session and provided details about their past and present OC use. The study was approved by the University of Western Ontario Non-Medical Research Ethics Board and was performed in compliance with the Declaration of Helsinki.

Explicit and implicit tasks

Profile of Mood States (POMS)

The POMS (McNair et al., 1992) was given at the beginning of each session before any other task. It is a widely used, extensively standardized, well-validated 65-item psychometric mood scale suitable for assessment of both clinical/psychiatric and non-clinical populations. It was considered an explicit measure of mood because respondents must reflect and self-assess their own emotional state. For each of 65 mood descriptors (e.g., 'happy', 'worthless', and 'grouchy'), participants rated how intensely they felt that way over the past week using a Likert scale that ranged from 0 (*Not at All*) to 4 (*Extremely*). Total scores for 6 factor analytically-derived mood dimensions can be computed (Anxiety, Depression, Anger/Irritability, Vigor, Fatigue, and Confusion). Only the Anxiety, Depression, Anger, and Vigor subscales were relevant to the present hypotheses and were analyzed here.

Details of procedures for all tasks are described in the Supplementary material.

Facial Emotion Identification Task (FEIT)

This task was used to evaluate the emotional decoding of faces. Stimuli consisted of 96 images of adult faces expressing six basic emotions (Happy, Sad, Anger, Fear, Disgust, and Neutral), selected from the Pictures of Facial Affect (Ekman and Friesen, 1976) or the Racially Diverse Affective Expression (RADIATE) database (Conley et al., 2018; Tottenham et al., 2009). Stimuli were presented by computer in a random sequence using E-Prime 3.0 software (Psychology Software Tools, Sharpsburg, PA). Each image was a full-frontal view of a face.

Following 4 practice trials, 96 test faces were presented one at a time. Participants were asked to identify each emotional expression by pressing a button as soon as it was recognized. Response time (RT) was measured in milliseconds (ms) from image onset to keypress response. Participants also rated the perceived intensity of each image on a 10-point scale. Past work suggests that intensities are perceived to be more intense in the presence of depression (Weightman et al., 2014). On FEIT tasks, accuracy of identification is typically at or near ceiling in healthy individuals if an open-ended exposure duration is used (e.g., Hampson et al., 2006). Because our intent was to use the FEIT to evaluate implicit not explicit processing, the dependent variable used for analysis was the median time required for each participant to correctly recognize each of the 6 emotions at the active phase versus inactive phase of the OC cycle. Medians were used instead of means to limit the influence of occasional outlier trials where a participant had an unusually long (or short) RT. Previous literature shows that negative stimuli, such as sad or angry faces, capture visual attention more readily in depressed individuals than non-depressed and elicit greater difficulty in attentional disengagement (for reviews see LeMoult and Gotlib, 2019; Leppänen, 2006; Phillips et al., 2010; Ros et al., 2023). This 'negativity bias' reflects the greater attentional salience of negative stimuli in major depression, which is seen for faces but is also seen for negative stimuli or events more broadly (LeMoult and Gotlib, 2019).

Affect Misattribution Procedure (AMP)

The AMP (Payne et al., 2005) is a well-established implicit task. In the classic AMP, a sequence of affectively neutral visual stimuli (unfamiliar Chinese characters) is presented rapidly on a computer screen. Participants are asked to classify each target as aesthetically pleasing or unpleasing by pressing a response key as each character is shown. If a stimulus is preceded by an image of a flower or insect presented at an extremely short or even imperceptible duration, the percentage of the neutral stimuli classified as pleasant or unpleasant is biased by the priming stimulus toward a more pleasant (for flowers) or unpleasant (for insects) classification, instead of the 50% expected by chance. In the present work, the classic AMP was used to test for any difference in AMP scores between the active and inactive sessions, which might indicate a covert change in underlying affective processing. A randomized set of 48 images was presented via an Inquisit software script (Inquisit Lab 5, Millisecond Software LLC, Seattle WA; see Figure 1). The percentage of targets judged as pleasant when preceded by a flower, and when preceded by an insect, was tabulated.

To look more directly at perceptions of facial emotion, we also created an identical AMP task that exactly paralleled the classic AMP but where emotional faces (happy, and sad or angry) were used as primes instead of flowers and insects. Consistent with the hypothesis of greater depressive affect during OC usage, we predicted that positive stimuli would be less effective as primes when women were evaluated during active OC use, and that this would be true for both the classic and face versions of the AMP. Negative primes were expected to capture attention more strongly during active use, but we did not make a directional prediction for the negative stimuli due to conceptual ambiguity as to whether greater capture would be expected to increase or decrease transfer to the Chinese character that followed.

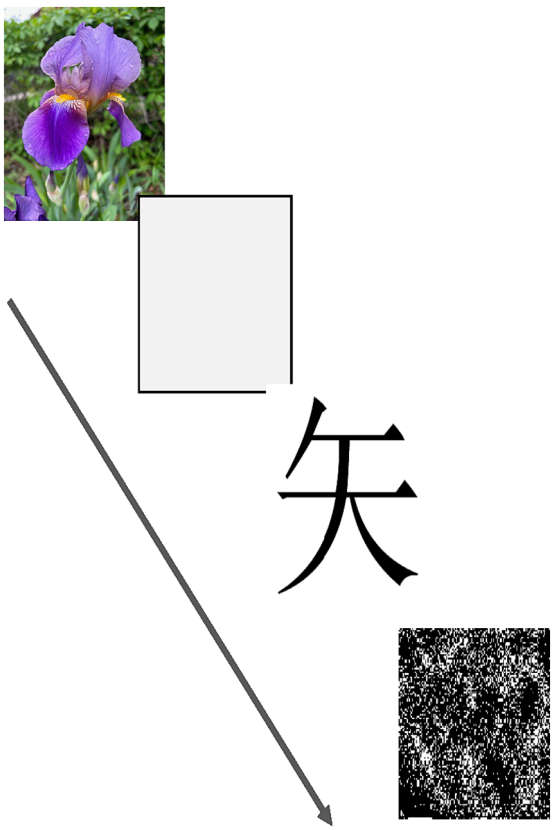


FIGURE 1
The sequence of events seen on each trial of the AMP (Affect Misattribution Procedure). The priming stimulus (a flower in this example) appeared first, followed by a randomly selected Chinese character. A visual mask then appeared, which served to terminate visual processing via backward masking. The arrow indicates the timeline and order of events on each trial.

Emotional Stroop

An Emotional Stroop task was our final measure of implicit processing. We used a task previously described by Başgöze et al. (2015) in a study of major depression. Stimuli were presented and response times recorded (ms) using E-Prime 3.0. Briefly, a randomized series of faces from the RADIATE set were presented on a computer screen at a short duration (max 1,000 ms). Each compound stimulus ($N = 128$ trials) consisted of an adult face in the background (happy or else angry/sad) on which an English word printed in 60-point font (e.g., "harmony") was super-imposed. The word was positioned at about mid-face. Participants were instructed to classify, as rapidly as possible, each word as having a positive or negative meaning, while ignoring the face in the background (which was irrelevant to word classification).

Emotional faces receive implicit processing if present, even if they are to be ignored or are irrelevant to the task at hand (e.g., Strand et al., 2013). In healthy controls, emotional faces typically interfere with the processing of words if the faces and words are of an incongruent valence (Strand et al., 2013; Başgöze et al., 2015). The interference effect is manifested as slower response times and/or increased word classification errors on the incongruent trials. Individuals with clinical depression show an interference effect, but display longer processing latencies (e.g., Epp et al., 2012) and an interference effect that may be amplified (see Epp et al., 2012 for a review; Ros et al., 2023). Some investigators, however, argue that the increased attentional salience of negative stimuli in people with depression can erode the ability of a conflicting positive word to generate a Stroop interference effect (e.g., Başgöze et al., 2015; Hu et al., 2012). If OC hormones are associated with increased dysphoria, we predicted that a larger Stroop effect would be found in the active condition where hormones are actively used, compared with the inactive condition where OC use is absent.

To rule out the possibility that any observed changes in the Emotional Stroop reflected, instead, an alteration in inhibitory control associated with OC use, rather than a change in affective processing, we gave a conventional non-affective Color-Word Stroop task to all participants as a control task. Only colors and words were used as stimuli; no affective processing was involved (Supplementary material).

Statistical analysis

Data were analyzed using IBM SPSS Statistics 29.0. Repeated-measures ANOVA was employed to analyze each task, with OC Intake (Active, Inactive) as a within-subjects factor. For tasks with more than one condition (e.g., subscales of the POMS), Condition served as a second within-subjects factor. As described further below, depressive status was a between-subjects factor in the ANOVAs for the implicit tasks, with *post hoc* simple effects analysis where relevant, to test for

an influence of depressive affect on the implicit scores. Bonferroni correction was used to evaluate significance for the explicit measure (the POMS). Because the present study is the first of its kind, performance on the implicit tests was evaluated at the $\alpha = 0.05$ threshold for significance in order to minimize Type II error. Partial eta squared was used to quantify effect sizes.

The dataset from this study is openly available from the Open Science Framework (Hampson et al., 2024, http://doi.org/10.17605/OSF.IO/WCRN4).

Results

Of the original 62 participants who enrolled in the study, 6 had valid data for one session only, due to scheduling error ($n = 3$) or failure to return for the second assessment ($n = 3$). In addition, 3 used Lolo, an atypical OC that has an inactive interval just 2 days long (limiting metabolic decline in the exogenous hormones that would otherwise occur during the inactive phase, Edelman et al., 2014; Dickey and Seymour, 2021). Because our hypotheses assumed decreased hormone concentrations are present at that time, Lolo did not afford a valid test of our study hypotheses. Thus, the final sample size available for statistical analysis was 53.

Of note, all 3 of the women lost to follow-up prior to their second testing showed elevated POMS depression scores at their initial assessment ($M = 28.67$), relative to population norms for their age and biological sex.

Profile of Mood States

Self-evaluated mood was analyzed via repeated-measures ANOVA with OC Intake (Active, Inactive) and POMS Subscale (Anxiety, Depression, Anger, and Vigor) as factors. The raw data were log-transformed before analysis to correct skewness on 2 of the subscales. However, in Figure 2 raw means are shown to permit easier comparisons with past literature.

Mean scores on all POMS subscales, including Depression, fell within established norms for female college/university students ages 20–21 (McNair and Heuchert, 2010, $N = 516$). However, this masked considerable within-group variability in our sample. ANOVA showed a significant OC Intake × Subscale interaction, $F(2.19, 113.64) = 4.66$, $p = 0.009$; $\eta^2_p = 0.082$, with Huynh-Feldt correction (see Figure 2A). Contrary to *a priori* prediction, negative affect was higher during the inactive not active interval. Differences on the Depression ($p = 0.009$) and Anger ($p = 0.005$) subscales were significant by *post hoc* test (after Bonferroni correction, see Figure 2A).

Importantly, a subset of the women (29% of our sample of 62) showed substantially elevated POMS Depression during Session1 ($M = 27.22$, $SD = 13.22$) and/or Session2 ($M = 26.73$, $SD = 14.57$). In this subgroup, total Depression scores reached a magnitude commonly seen in outpatients diagnosed with anxiety or mild to moderate depression ($M = 28.0$, $SD = 15.9$, $N = 650$ adult female outpatients), according to the clinical test norms of the POMS (McNair and Heuchert, 2010). Below, we refer to this subset as the "HiD-subgroup."

An exploratory ANOVA taking Subgroup into account confirmed greater negative affect in the HiD-subgroup (see Figure 2B), but also revealed that effects of OC intake on mood were accentuated in the HiD women compared with Non-HiD, OC Intake × Subscale × Subgroup interaction: $F(1.75, 89.19) = 6.85$, $p = 0.003$; $\eta^2_p = 0.118$. In the Non-HiD subgroup the same pattern of means was observed, but more weakly. However, in the Non-HiD women the Depression subscale nevertheless exhibited a significant OC Intake effect ($p = 0.038$).

Facial Emotion Identification Test

Accuracy on the FEIT was at or close to ceiling for several of the emotions we tested. These high accuracies indicate careful responding throughout the task. This was achieved via participants self-adjusting their response latencies (RTs) to maintain accuracy. Consequently, response latency was considered a superior indicator of any alterations in emotional processing that may be present.

Only the RTs were examined statistically. Data were analyzed using a three-way ANOVA with OC Intake (Active, Inactive), and Emotion (Disgust, Sad, Fear, Anger, Happy) as within-subjects factors and Subgroup (HiD, Non-HiD) as a between-subjects factor. The data were screened for outliers prior to analysis, and two outliers with RTs >3SD above the mean for their group were removed.

Responses were noticeably slower (longer RT) in the HiD-subgroup, consistent with past work on depressive disorder, main effect of Subgroup: $F(1, 49) = 7.26$, $p = 0.010$; $\eta^2_p = 0.129$; mean for HiD = 1651.33 ms, mean for Non-HiD = 1236.15 ms. Importantly, in the sample of OC users as a whole (see Figure 3A), active OC intake was associated with slower RTs than inactive, particularly for certain key emotions, Emotion × Intake interaction: $F(4, 196) = 4.29$, $p = 0.002$; $\eta^2_p = 0.081$. As seen in Figure 3A, slowing was clearest for sad and angry expressions, consistent with greater attentional capture reported in depressed individuals for these two emotions. This effect was further modified by the presence of overt depressive status, Emotion × OC Intake × Subgroup interaction: $F(4, 196) = 3.17$, $p = 0.015$; $\eta^2_p = 0.061$ (see Figure 3B). Participants with a high level of depressive symptoms (HiD) were slower than Non-HiD women, but also exhibited a larger effect of OC intake, particularly for the sad expressions (Figure 3B).

Perceived intensity of the emotions varied with OC intake. Repeated-measures ANOVA with Intake (Active, Inactive) and Emotion (Happy, Sad, Fear, Anger, Neutral, Disgust) as factors revealed that valenced facial expressions were perceived as more intense during active OC intake, main effect of OC Intake: $F(1, 51) = 4.34$, $p = 0.042$. The effect size was small to medium, Active: $M = 6.42$, $SD = 0.81$; Inactive: $M = 6.27$, $SD = 0.81$; $\eta^2_p = 0.078$. Depressive status was found to moderate the effect of active intake. However, simple effects confirmed a significant OC intake effect even if Non-HiD were considered on their own, $F(1, 36) = 8.51$, $p = 0.006$.

Affect Misattribution Procedure (AMP)

The classic AMP and Faces AMP were each analyzed using repeated-measures ANOVA, with OC Intake (Active, Inactive) and Type of Prime (Flower, Insect or else Happy, Sad or Angry) as factors. To test if depressive affect influenced the AMP, Subgroup was included as a between-subjects factor.

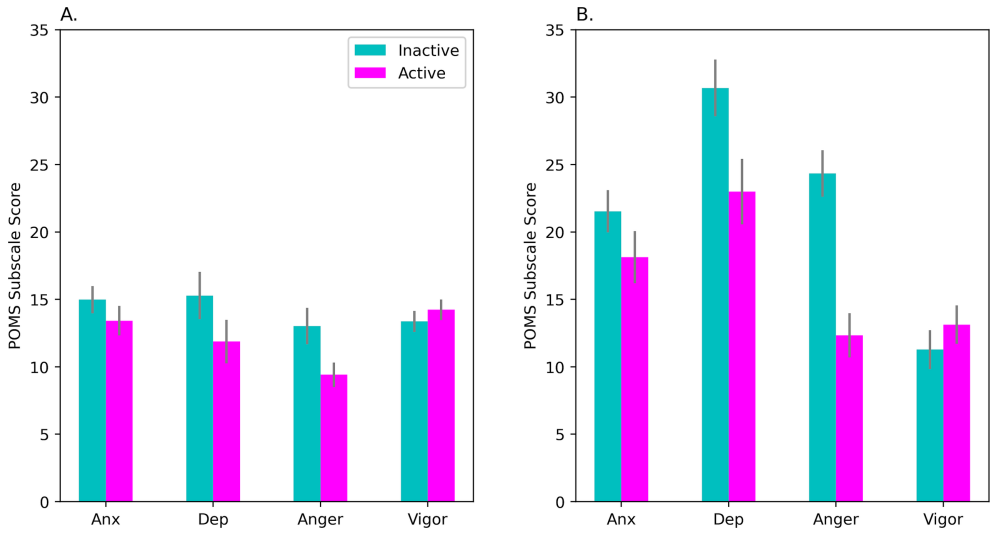


FIGURE 2
Mean scores on the POMS during days of active hormone intake ('Active') and no-intake ('Inactive'). In **(A)**, results are shown for our sample of OC-users as a whole ($N = 53$). Panel **(B)** shows results for the HiD subgroup ($N = 15$) who had elevated self-reported Depression scores that matched levels usually observed in clinical patients with mild to moderate depression. Scores on the Anger and Depression subscales varied with OC intake (Bonferroni correction). The intake effect was seen for the OC-users as a whole, but was significantly larger in the HiD-subgroup, as seen in **(B)**. On the Anxiety subscale, the effect of OC intake was marginally significant ($p = 0.057$; see panel **A**). POMS, Profile of Mood States; Anx, anxiety subscale; Dep, depression subscale; Anger, anger subscale; Vigor, vigor subscale. Bars show the standard error of the mean.

Both AMP tasks showed the expected implicit priming effects (p's < 0.001). Chinese characters immediately following a positive prime were more likely to be classified as "pleasant," whereas targets that followed a negative prime were more likely to be classified as "unpleasant" (see Figure 4A). There was no significant effect of OC Intake on the priming effect for the Faces AMP. However, OC Intake did influence priming on the classic version of the AMP, both in the OC sample as a whole, $F(1, 49) = 4.49$, $p = 0.039$; $\eta^2_p = 0.084$ (Figure 4A) and particularly in the HiD-subgroup, as revealed by simple effects analysis, $F(1, 13) = 9.23$, $p = 0.010$; $\eta^2_p = 0.415$ (Figure 4B). Compared with no-intake, priming was visibly weaker during the active use of OCs.

Although the effect of OC intake was very prominent in the HiD women, no intake effect was seen in the Non-HiD group, if considered on its own, $F(1, 36) = 0.160$, $p = 0.691$; $\eta^2_p = 0.004$. This suggests that the AMP effect was linked to the presence of depressive affect.

Emotional Stroop

On the Emotional Stroop, repeated-measures ANOVA of the RTs showed the expected interference effects ($p < 0.001$). Overall, RTs were slower on incongruent trials where the valence of the face and word were in conflict. RTs were also slower for negative than positive words overall. In the sample as a whole, the presence of a negative target word was associated with slowed responding on trials where positive faces were shown ($M = 627$ ms on congruent trials, $M = 689$ ms on incongruent trials where a negative word was present), while the presence of a positive word exerted little interference effect on RTs when sad/angry faces were shown ($M = 663$ ms on congruent trials, $M = 652$ ms on incongruent), $F(1, 50) = 84.82$, $p < 0.001$; $\eta^2_p = 0.629$. This pattern was moderated by the presence of explicit depressive affect and OC intake, $F(1, 50) = 12.69$, $p < 0.001$; $\eta^2_p = 0.202$. In the HiD-subgroup, shown in Figure 5, the effect of OC intake was statistically significant, $F(1, 14) = 8.41$, $p = 0.012$; $\eta^2_p = 0.375$ (simple effects). Interference by negative words was stronger during active intake. If anything, a conflicting positive word shortened the RTs in the negative face condition during that time (Figure 5). In the Non-HiD subsample (with HiD removed), the OC intake effect was qualitatively similar but weaker, $F(1, 36) = 4.04$, $p = 0.052$; $\eta^2_p = 0.101$ (data not shown in figure).

Accuracy of word classification was close to 100% in all conditions. Thus, individual differences in the present study were expressed primarily in the response times.

Results for the control task, the classic Color-Word Stroop, are shown in Supplementary material. The Color-Word Stroop was procedurally similar to the Emotional Stroop, but lacked any affective element in that only colors and words were presented. A significant interference effect was found on the incongruent trials ($p < 0.001$), but neither OC intake nor depression influenced performance.

Discussion

In a group of healthy young women using conventional OCs, we observed a depression-like pattern of performance on implicit measures of mood during periods of active hormone intake. This result was driven partly, but not completely, by a subset of OC-users who acknowledged having elevated levels of negative affect. Effects were seen during active hormone use relative to the washout week of the contraceptive cycle when absence of hormone intake for as long as 7 days results in diminishing bodily concentrations of OC steroids (Edelman et al., 2014). The fact that a difference was visible over such a short timeframe suggests a neuroendocrine effect that is relatively

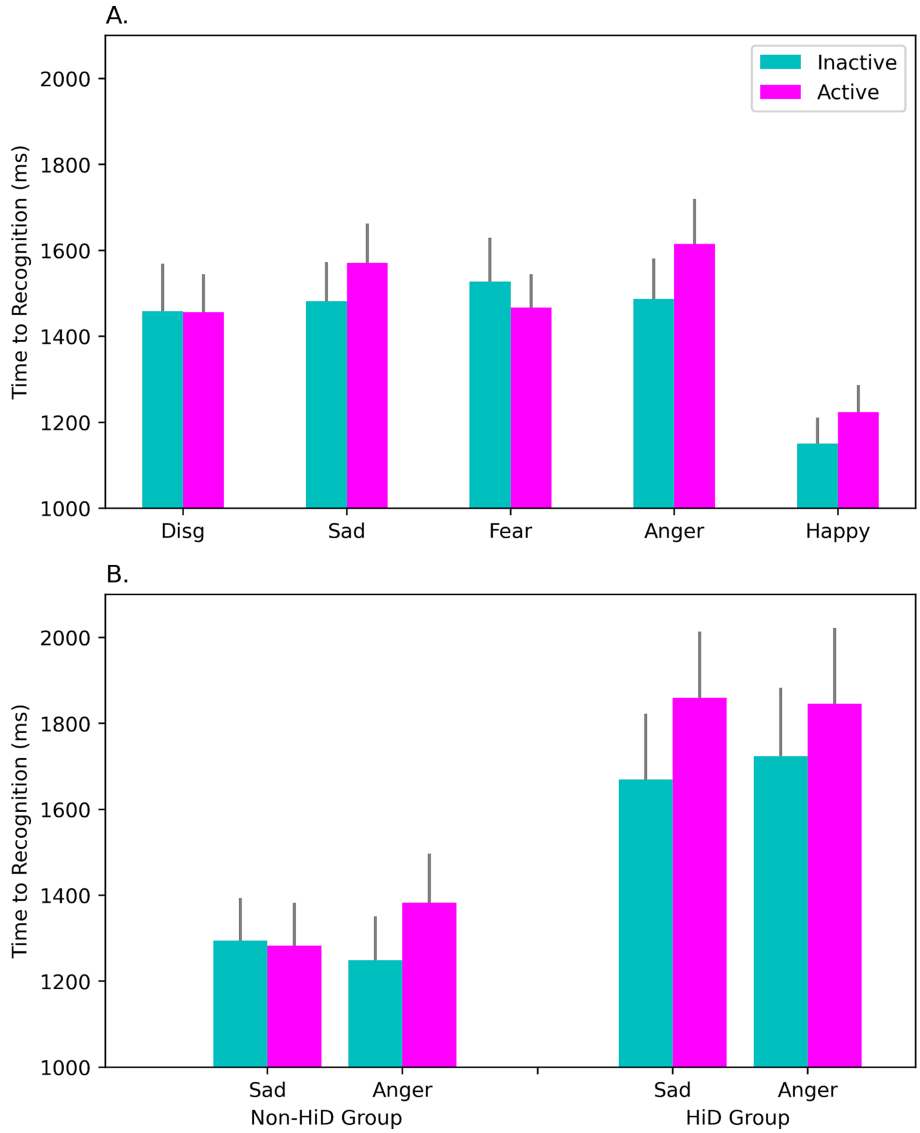


FIGURE 3
Response times (RT) to emotional faces on the FEIT were slower during active OC intake ('Active') than during the inactive interval ('Inactive'). Panel **(A)** shows mean RTs in the sample as a whole (*N* = 51), for each facial emotion we assessed (disgust, sadness, fear, anger, and happiness). RTs were slower during active intake, and this was most evident for the Sad and Angry faces, whereas fewer attentional resources were allocated to the processing of happy expressions (which showed a more rapid RT response). Slowing was greatest among women in the HiD-group (rightmost bars in panel **B**), who reported high levels of negative (depressed) affect. Significant slowing was observed at both phases in the HiD women but was especially marked during the active intake of OC steroids. Slowing may signify the capture of attention by negative emotional stimuli relevant to depression and/or may indicate greater difficulty in recognizing the emotions displayed. FEIT = Facial Emotion Identification Task.

short-term (although our study design cannot address if longer-term effects might also occur; cf. Anderl et al., 2020). The differences we observed were consistent and conformed to our *a priori* predictions. In contrast, the POMS (an explicit measure of 'subjective' mood), showed that participants perceived their own moods to be worse at the inactive not active phase of the contraceptive cycle.

To our knowledge, this is the first study to use implicit tasks to evaluate mood in OC users. Implicit measures have been used to study current or remitted depression in many other contexts (Payne and Lundberg, 2014). Compared with non-depressed individuals, people with major depression display differences on several implicit tests, including larger interference effects on the Emotional Stroop (Epp et al., 2012; Hu et al., 2012; Başgöze et al., 2015), altered decoding of facial signals including decreased attention to positive stimuli and decreased accuracy in identifying positive expressions as happy, greater attentional allocation to negative expressions (especially sad ones), are slower to react to facial expressions if RT is measured, and judge negative but not positive expressions as more intense than controls (Leppänen, 2006; Münkler et al., 2015; Weightman et al., 2014). All of these are reminiscent of the patterns we observed among the HiD in the present study. Our HiD-subgroup showed elevated Depression subscale scores on the POMS, indicating high negative affect by self-report. However, the POMS is a screening instrument only. Although it quantifies negative affect, it lacks items that evaluate the appetitive, cognitive, or motivational dimensions of clinical depression (American Psychiatric Association, 2013). Accordingly,

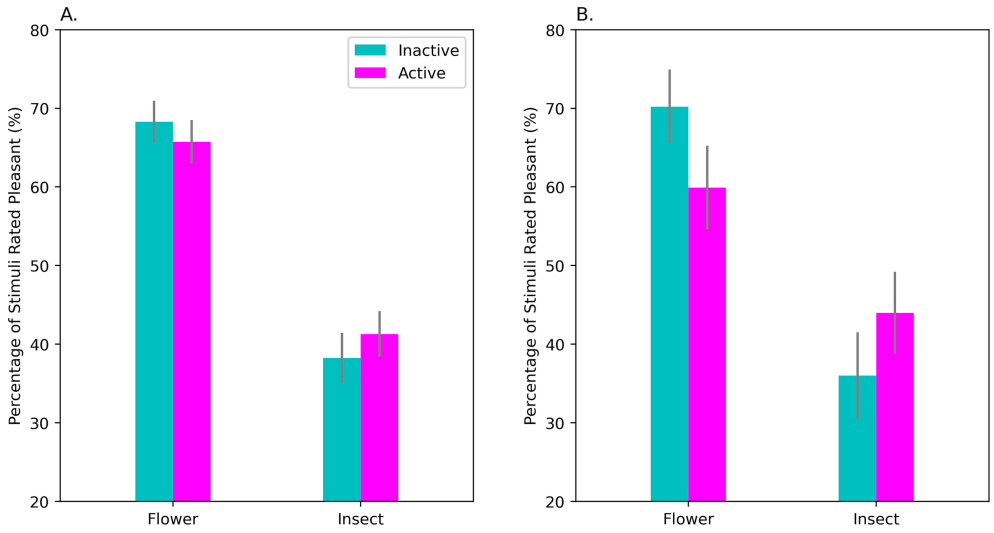


FIGURE 4
Performance on the classic AMP by OC-users in the sample as a whole **(A)** and in the HiD-Subgroup **(B)**. Shown on the *Y*-axis is the mean percentage of neutral characters rated as pleasant for each type of affective prime (flowers and ominous-looking insects). A classic priming effect was found. If flowers were used as primes, a high percentage of the succeeding characters were rated as pleasant, whereas insect primes elicited unpleasant ratings on a majority of the trials. Comparing performance during active ('Active') and inactive ('Inactive') OC intake revealed that affective priming was significantly weaker during the active intake of OC steroids. This effect was evident in the sample as a whole **(A)** but was largest for the HiD women **(B)** where a large effect of OC intake was observed. Note that the effect of an affective prime was weaker during active intake for both positively- and negatively-valenced primes. AMP, Affect Misattribution Procedure. Bars show the standard error of the mean.

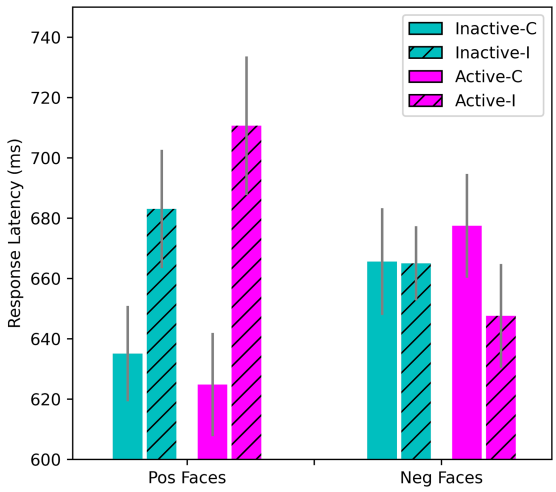


FIGURE 5
Effects of stimulus congruency and OC intake on responses to positively- and negatively-valenced faces on the Emotional Stroop. Hatched bars show the incongruent conditions and open bars the congruent. Only data for the HiD-subgroup are shown here ($N = 15$), but results for the OC sample as a whole were essentially the same. RTs for positive faces combined with positive words were the fastest, whereas negative faces elicited slower processing. For the positive faces, there was a strong Stroop interference effect under incongruent conditions (when a negative word was shown simultaneously). For the negative faces, there was little evidence of any additive slowing caused by the presence of a conflicting positive word.

we cannot conclude that our HiD-subgroup would meet criteria for clinical depression based on gold standard clinician interviews. However, the HiD-subgroup did display high negative affect that was very comparable in severity to norms for outpatients with clinically-ascertained mood conditions (McNair and Heuchert, 2010). They also performed on implicit tests like patients with verified depression. Importantly, while the implicit effects were largest for the HiD-group, a weaker version of the same patterns was evident in the Non-HiD subgroup as well (when the HiD were removed). This might reflect a gradient in mood status or milder negative affect among the Non-HiD women.

Our data support recent epidemiological studies which suggest that depressive changes do occur in a subset of women who opt to use hormonal contraception. For example, a prospective cohort study of over a million women based on the National Prescription Register of Denmark showed that women prescribed hormonal contraceptives subsequently required antidepressants at higher rates than non-users or were more likely to be diagnosed with clinical depression (Skovlund et al., 2016). Hormonal contraceptive use is also associated with an increased risk for attempted or completed suicides (Skovlund et al., 2018). Similarly, Anderl et al. (2020) found that adolescent OC use was associated with an increase in depression risk compared with never-users. The focus of such studies, however, has been depression of a clinical magnitude, which does not speak to lesser mood changes that might be present among OC-users more prevalently. Depressive changes and their time course in OC-users are important to document, whether clinical or subclinical in severity. This knowledge is especially relevant to prospective OC users and their healthcare providers in order to enable fully informed reproductive decision-making and contraceptive counseling, especially given the great importance of negative affect to women's everyday quality of life.

Estimates of the percentage of OC-users who develop dysphoria are widely variable. This reflects differences across studies in how depression is defined, but also the impact of several moderator variables (e.g., personal past history of depression, choice of progestin,

Bengtsdotter et al., 2018; Hampson, 2023; Joffe et al., 2003). In a double-blind, placebo-controlled trial, Lundin et al. (2017) reported that the proportion of women with new-onset subclinical depression during a 3-mo trial of OCs versus placebo was <10%, based on the MADRS scale (Montgomery-Åsberg Depression Rating Scale). Using the POMS, we previously found an incidence of about 7–8% who scored in the clinical range (Hampson, 2023). However, in the study by Lundin et al. (2017), 24% of women allocated to OC treatment had significant mood deterioration by their own self-reports, in agreement with the higher numbers based on self-report in the present study (also see Johansson et al., 2023). Both estimates fall within the ranges reported previously by other labs. Hampson (2023) also identified significantly higher POMS Depression scores among users of OCs that contained third- or fourth-generation progestins, as compared to first- or second-generation OC pills. The same trend could be identified in the present dataset (see Supplementary material). Depressive symptoms associated with the use of hormone-eluting intrauterine devices (IUDs) also have been described (Skovlund et al., 2016; Stenhammar et al., 2023; Zettermark et al., 2018), and may be dose-dependent (Roland et al., 2023). Because IUDs release a progestin but contain no estrogens, the IUD findings combined with our own past and present data implicate the progestin constituent of OCs as a potential driver of mood changes.

To further explore the possibility of depressive symptoms associated with the active use of OCs, we adopted innovative implicit techniques in the present work. Elevated scores on the POMS relative to normative data for age and sex, plus a depressive shift in scores on the implicit tasks (which can be sensitive to mood even when it is not explicitly acknowledged by self-report) suggest that depressive side-effects do occur among a subset of OC-users. In a small number of women, these symptoms may rise to the threshold of a clinical impairment. Ethinyl estradiol is a co-constituent (along with a progestin) of standard combined OC pills and may help to counteract adverse mood changes, given that high estradiol has been linked with positive mood states in other contexts (e.g., Backström et al., 1983; Hampson, 1990; see Hampson, 2002), but the dose of ethinyl estradiol in recently developed OCs is low (Dickey and Seymour, 2021), and perhaps not always sufficient to act as an effective counterweight.

Precisely which hormonally-sensitive brain regions underlie the observed effects on mood is not presently known. Functional brain imaging, notably studies using functional MRI, are beginning to identify differences in brain activity between hormonal contraceptive users and non-users in resting-state activity, functional connectivity, or BOLD activation evoked during imaging tasks designed to elicit affective processing. These include decreased activity in OC users in the inferior frontal cortex, amygdala, insula; reduced amygdala reactivity to emotionally-charged images; and alterations in reward-related processing in response to monetary incentives, food, or erotic stimuli (for review see Brønnick et al., 2020; Pletzer et al., 2023). In a placebo-controlled trial, Hidalgo-Lopez et al. (2023) found connectivity changes during OC use in 2 forebrain networks that were predictive of adverse PMS-like mood effects. Differences in the brain regions that underlie implicit and explicit affect are thought to exist (e.g., DeCoster et al., 2006; Lane, 2008), but it is unknown how these might map onto the differences now being described between OC users and non-users in brain activation during emotional processing. Structural brain changes have also been reported in OC users, but evidence is still sparse and must be viewed as tentative (e.g., Lisofsky et al., 2016; Brouillard et al., 2023; for a critical review see Brønnick et al., 2020). Given that several of the regions implicated in OC studies participate more generally in emotional regulation, social cognition, or the subjective experience of emotion, and are also sites where progesterone receptors have been identified in animal studies (Sundström-Poromaa et al., 2020), they are possible candidate regions where OC steroids might exert neuroregulatory actions to bring about alterations in mood. Further research will be important to help illuminate the pathways concerned.

In their natural form, sex steroids exert a wide spectrum of cellular and molecular effects in the CNS, that operate at various time scales. Modifications in neurogenesis, synaptogenesis, myelination, and neurotransmitter signaling have been reported in animal studies (see Barth et al., 2015; Pletzer et al., 2023 for recent reviews). Estrogen- or progesterone-mediated effects in serotonergic, dopaminergic, noradrenergic, GABA, and several other brain systems (e.g., McEwen and Alves, 1999) are of special interest, as are hormone-mediated effects on regional synapse densities (Woolley and McEwen, 1993), because they occur in the adult CNS, vary dynamically with hormone availability in the bloodstream, and occur rapidly, within minutes, hours, or days. Their time course makes them attractive prospects to explain active versus inactive OC phase differences. It should be noted that contraceptive steroids have received little study in a neurobiological context (Pletzer et al., 2023). However, many of the effects seen for the natural forms of the steroids are likely shared by contraceptive steroids, to the extent that they can occupy the same hormonal receptors in the CNS. The situation in OC users is complex, however, because some synthetic progestins used in OCs possess a wider hormone receptor binding profile than the natural form of progesterone (Dickey and Seymour, 2021; Hampson, 2023; Sitruk-Ware, 2006), may alter the production of neuroactive steroids (Pletzer et al., 2023; Rapkin et al., 2006), and also impact the HPA (hypothalamic–pituitary–adrenal) axis, stress reactivity, and immune system responses (e.g., Kirschbaum et al., 1999), all of which may contribute to mood. The monoamines are especially important for the regulation of mood and depression. One recent study showed a ~10% lower brain serotonin receptor-4 binding potential (5-HT4R) in OC users (Larsen et al., 2020; but see Larsen et al., 2022). An advantage of the active vs. inactive study design used here is its potential to narrow the range of potential mechanisms by pointing to brain effects that are acute, relatively short-acting, and reversible.

A surprising aspect of the present work was that while our implicit tasks suggested negative mood was greater during the active use of exogenous hormones, consistent with our *a priori* theoretical predictions, our *explicit* measure of mood, the POMS, suggested that self-perceived negative affect was greater during the inactive ('washout') phase. During the inactive portion of the contraceptive cycle, a dropoff occurs in serum concentrations of the exogenous steroids, which triggers shedding of the uterine lining in the form of menstrual flow (Edelman et al., 2014). We observed that POMS Depression scores were elevated in our HiD participants (and to some extent the Non-HiD) at both phases of the contraceptive cycle, and therefore the inferences drawn above apply. Nevertheless, a lack of direct parallelism between the implicit and explicit effects was unexpected (even though the psychological literature shows that dissociations can occur in certain situations, consistent with separability in the neural circuitry that underlies implicit and explicit emotional responses; DeCoster et al., 2006; Suslow et al., 2019). Our

study allows rare insight into the explicit mood changes in OC-users, as past studies of the impact of OCs have understandably limited their assessments mostly to periods of active hormone usage. However, our data reinforce sporadic research reports indicating that self-perceptions of negative affect might in fact be higher during the inactive phase of the contraceptive cycle (Noachtar et al., 2023). Future studies will be needed to clarify this finding.

It is unclear why our implicit and explicit measures did not change in a more synchronized fashion, but explicit ratings of negative affect can demonstrate rapid change in other contexts (e.g., under a short-term course of psychotherapy, Suslow et al., 2019) and are subject to top-down control from other cortical inputs that leave implicit responses unaffected, such as conscious verbal representations or inputs from declarative memory systems. In general, a differentiation between implicit and explicit affect is theoretically valuable because they are thought to stem from two interacting but independently-operating emotional processing networks—an automatic/autonomic system and a reflective system that operates within conscious control. Implicit affect can predict individuals' behavioral and psychophysiological reactions to stress or emotional stimuli (e.g., cortisol or cardiovascular reactivity) either exclusively or beyond the predictive value of explicit measures alone (Suslow et al., 2019). Independent of any physiological driver, conscious self-estimations of mood may be more malleable than implicit measures are (DeCoster et al., 2006; Suslow et al., 2019) and thus perhaps more subject to day-to-day variation in subjective perceptions, attentional biases, self-presentation concerns, the demand characteristics inherent in a research setting, or societal expectations based on traditional stereotypes regarding mood and the menstrual cycle or personal past history. It is also possible that other kinds of interacting physiological variables, such as cycle-related variations in sleep quality, exert a larger temporary influence on explicit ratings during the inactive than active phase (even though somatic and appetitive effects tend to be milder and less prominent under OC use; Sundström Poromaa and Segebladh, 2012). If true, then implicit techniques might offer a more stable and perhaps accurate window into changes in underlying neurophysiology related to primitive affective processes.

In non-OC users, self-reported negative affect among a susceptible group of women is most clearly identified with the premenstrual days of the ovarian cycle. At that time, serum concentrations of estradiol and progesterone transition from the highs of the preceding midluteal phase to the lows of the menses that follow (Steiner et al., 2003). Although temporally displaced, certain endocrine parallels exist between the premenstrual phase of the natural menstrual cycle and the no-intake phase of the OC cycle, as intake of exogenous steroids in OC-users is discontinued abruptly at the end of the active phase. Furthermore, evidence implicates progestogens or, more particularly, progestogen withdrawal, in both types of mood phenomena (for a review see Sundström-Poromaa et al., 2020). The timing of the self-perceived negative mood in OC users may also become more comprehensible given that ethinyl estradiol in the bloodstream is metabolized and reaches low steady-state values more rapidly at the beginning of the inactive phase than progestin, leaving unopposed residual progestin activity for 2–3 days at the onset of the inactive interval (Edelman et al., 2014). Mood lability and vulnerability to negative affect occur during other hormonal transition points in the female lifecycle too, such as parturition or the perimenopausal transition (Gordon et al., 2018; for review see Steiner et al., 2003), adding further plausibility to the possibility that mood changes might occur in a subset of users in response to OC transitions. Interestingly, there are also data to suggest that women who are susceptible to major depression may experience larger self-perceived mood effects than non-susceptible women under the influence of ovarian steroids (e.g., Bloch et al., 2000; Sundström-Poromaa et al., 2020), including OCs (Oinonen and Mazmanian, 2002). This might help explain why intake-related mood effects were larger for the HiD-group in the present study.

The present study is the first to demonstrate depressive-like effects on implicit measures of mood in OC users. On implicit tests, depression reveals itself without a respondent's awareness, in the form of objective differences in performance on automatic measures of attentional or perceptual bias, affective reactions, or valenced emotional processing. The potential for desynchronization of implicit and explicit measures highlights the complexity of affective processing in general, and the importance of using several different types of measures to assess affective processing. Inclusion of implicit tasks may prove useful in future studies of OCs and their possible mood effects. These might include the AMP (Payne et al., 2005; Murphy and Zajonc, 1993), which has been widely used in many other contexts. Although the classic version of the AMP was successful in the present study, our use of emotional faces as priming stimuli was not. This was due to our unfortunate decision to use an identical exposure time, which was adequate for simple flower and insect images but turned out to be inadequate for participants to apprehend the greater visual complexity of the emotional expressions we chose as primes (cf. LeMoult et al., 2012, who used a 500 ms prime exposure time). As a result, the face version of our AMP task significantly underperformed. A longer prime exposure is therefore recommended for future studies if faces are to be used as primes.

The present work was intended as a pilot-level study. It will thus be helpful for future larger-scale investigations to recruit a greater sample size. Among other things, this would enable comparisons to be made among the different families of contraceptive progestins. This could help determine if any differences exist in their potential for adverse mood effects. Studying implicit mood in non-OC users would likewise be useful, to allow direct comparisons between OC-users and non-users. In a previous study done by our laboratory (Hampson, 2023), non-users were in fact assessed. Relative to non-users, POMS Depression subscale scores were found to be significantly higher among a group of healthy women taking OCs ($N = 193$). Careful thought will need to be given to recruitment methods and which stages of the menstrual cycle to target if making mood comparisons between OC users and non-users in future studies. While the current study used a prospective design, which is desirable, we were not able to control the length of time post-OC initiation when mood status in our sample was evaluated. If we assume that OC-users who encounter troublesome mood side-effects following the initiation of OC use are more likely to discontinue taking their OCs, the present study might under-estimate the percentage of women who experience depressive changes as a function of OC use due to differential dropout over time. On the other hand, it is a strength of the present work that the median number of months on the OC pill in the current study was 12 months, as some evidence suggests mood effects are seen most frequently during the first 2 years of OC usage (Johansson et al., 2023). Other strengths of the present work are the enrollment of medication-free participants, which avoids confounds introduced by other concurrent

medication use or pre-existing health conditions that might otherwise create ambiguities in attributing depressive symptoms to OC use *per se*. On the other hand, in clinical practice OCs are often prescribed under less pristine conditions to women who have other complex health conditions that may affect OC outcomes such as a personal history of affective disorder (Bengtsdotter et al., 2018; Oinonen and Mazmanian, 2002). This means that effect sizes seen for associations between OC use and implicit measures of depressed mood might potentially be even larger under real-world conditions. Of course, a final but significant strength of the present study is its inclusion of implicit tests instead of the usual exclusive reliance on self-report to detect depressive symptoms.

The present data provide novel insights into affective processing in women who use OCs. Our work converges with other recent investigations which suggest that endocrine states are surprisingly relevant to social cognition. Effects of the menstrual cycle on women's evaluations of men's sexual attractiveness have been reported, as have effects of testosterone on men's aggressive responses but also self-initiated acts of generosity and trust toward others (Boksem et al., 2013; Dreher et al., 2016; Penton-Voak and Perrett, 2000). The ability of sex steroids to modulate brain pathways involved in emotion makes them particularly powerful regulators of human behavior. Because the implicit tasks used in the present study required accurate emotional decoding of faces (e.g., the FEIT, Emotional Stroop), they further illustrate the subtle impact of an ostensibly neutral endocrine variable—oral contraceptive use—on social appraisals, social functioning, and the interpersonal environment, in this case mediated by depressive brain changes linked to the use of OCs. Successful processing of facial signals is vital to the ability of a 'receiver' to generate accurate emotional appraisals and well-calibrated responses to social cues, in turn helping to maintain the quality of personal relationships and close emotional bonds with loved ones (Fischer and Manstead, 2016). Any effects of OC steroids on social cognition are therefore of potential consequence to the millions of women who choose a hormonal form of contraception.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) are: Open Science Framework (OSF), http://doi.org/10.17605/OSF.IO/WCRN4.

Ethics statement

The studies involving humans were approved by The University of Western Ontario Non-Medical Research Ethics Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

EH: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. SA: Investigation, Writing – review & editing. TB: Data curation, Investigation, Visualization, Writing – review & editing. MI: Investigation, Writing – review & editing. EW: Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by a grant to author EH from the Natural Sciences and Engineering Research Council of Canada (#2015–03662 and #2020–04989).

Acknowledgments

We thank P. Istasy for assistance with data coding and scoring.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1462891/full#supplementary-material

References

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders, 5th edition (DSM-5). Washington, DC: American Psychiatric Publishing.

Anderl, C., Li, G., and Chen, F. S. (2020). Oral contraceptive use in adolescence predicts lasting vulnerability to depression in adulthood. *J. Child Psychol. Psychiatry* 61, 148–156. doi: 10.1111/jcpp.13115

Bäckström, T., Hansson-Malmström, Y., Lindhé, B. A., Cavalli-Björkman, B., and Nordenström, S. (1992). Oral contraceptives in premenstrual syndrome: a randomized comparison of triphasic and monophasic preparations. *Contraception* 46, 253–268. doi: 10.1016/0010-7824(92)90006-F

Backström, T., Sanders, D., Leask, R., Davidson, D., Warner, P., and Bancroft, J. (1983). Mood, sexuality, hormones, and the menstrual cycle. II. Hormone levels and their

relationship to the premenstrual syndrome. *Psychosom. Med.* 45, 503–507. doi: 10.1097/00006842-198312000-00004

Bancroft, J., and Sartorius, N. (1990). The effects of oral contraceptives on well-being and sexuality. *Oxf. Rev. Reprod. Biol.* 12, 57–92

Barth, C., Villringer, A., and Sacher, J. (2015). Sex hormones affect neurotransmitters and shape the adult female brain during hormonal transition periods. *Front. Neurosci.* 9:37. doi: 10.3389/fnins.2015.00037

Başgöze, Z., Gönül, A. S., Baskak, B., and Gökçay, D. (2015). Valence-based word-face Stroop task reveals differential emotional interference in patients with major depression. *Psychiatry Res.* 229, 960–967. doi: 10.1016/j.psychres.2015.05.099

Beltz, A. M. (2022). Hormonal contraceptive influences on cognition and psychopathology: Past methods, present inferences, and future directions. *Front. Neuroendocrinol.* 67:101037. doi: 10.1016/j.yfrne.2022.101037

Bengtsdotter, H., Lundin, C., Gemzell Danielsson, K., Bixo, M., Baumgart, J., Marions, L., et al. (2018). Ongoing or previous mental disorders predispose to adverse mood reporting during combined oral contraceptive use. *Eur. J. Contracept. Reprod. Health Care* 23, 45–51. doi: 10.1080/13625187.2017.1422239

Bloch, M., Schmidt, P. J., Danaceau, M., Murphy, J., Nieman, L., and Rubinow, D. R. (2000). Effects of gonadal steroids in women with a history of postpartum depression. *Am. J. Psychiatry* 157, 924–930. doi: 10.1176/appi.ajp.157.6.924

Bodenschatz, C. M., Skopinceva, M., Kersting, A., Quirin, M., and Suslow, T. (2018). Implicit negative affect predicts attention to sad faces beyond self-reported depressive symptoms in healthy individuals: an eye-tracking study. *Psychiatry Res.* 265, 48–54. doi: 10.1016/j.psychres.2018.04.007

Boksem, M. A. S., Mehta, P. H., Van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., et al. (2013). Testosterone inhibits trust but promotes reciprocity. *Psychol. Sci.* 24, 2306–2314. doi: 10.1177/0956797613495063

Brinton, R. D., Thompson, R. F., Foy, M. R., Baudry, M., Wang, J., Finch, C. E., et al. (2008). Progesterone receptors: form and function in brain. *Front. Neuroendocrinol.* 29, 313–339. doi: 10.1016/j.yfrne.2008.02.001

Brønnick, M. K., Økland, I., Graugaard, C., and Brønnick, K. K. (2020). The effects of hormonal contraceptives on the brain: a systematic review of neuroimaging studies. *Front. Psychol.* 11:e556577. doi: 10.3389/fpsyg.2020.556577

Brouillard, A., Davignon, L. M., Turcotte, A. M., and Marin, M. F. (2023). Morphologic alterations of the fear circuitry: the role of sex hormones and oral contraceptives. *Front. Endocrinol.* 14:1228504. doi: 10.3389/fendo.2023.1228504

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., et al. (2018). The racially diverse affective expression (RADIATE) face stimulus set. *Psychiatry Res.* 270, 1059–1067. doi: 10.1016/j.psychres.2018.04.066

de Wit, A. E., Booij, S. H., Giltay, E. J., Joffe, H., Schoevers, R. A., and Oldehinkel, A. J. (2020). Association of use of oral contraceptives with depressive symptoms among adolescents and young women. *JAMA Psychiatry* 77, 52–59. doi: 10.1001/jamapsychiatry.2019.2838

DeCoster, J., Banner, M. J., Smith, E. R., and Semin, G. R. (2006). On the inexplicability of the implicit: differences in the information provided by implicit and explicit tests. *Soc. Cogn.* 24, 5–21. doi: 10.1521/soco.2006.24.1.5

Dickey, R. P., and Seymour, M. L. (2021). Managing contraceptive pill patients and other hormonal contraceptives. *17th* Edn. Fort Collins, CO: EMIS Medical Publishers.

Dreher, J.-C., Dunne, S., Pazderska, A., Frodl, T., Nolan, J. J., and O'Doherty, J. P. (2016). Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. *PNAS* 113, 11633–11638. doi: 10.1073/pnas.1608085113

Edelman, A. B., Cherala, G., Munar, M. Y., McInnis, M., Stanczyk, F. Z., and Jensen, J. T. (2014). Correcting oral contraceptive pharmacokinetic alterations due to obesity. A randomized controlled trial. *Contraception* 90, 550–556. doi: 10.1016/j.contraception.2014.06.033

Ekman, P., and Friesen, W. V. (1976). Pictures of facial affect. Palo Alto, CA: Consulting Psychologists Press.

Epp, A. M., Dobson, K. S., Dozois, D. J. A., and Frewen, P. A. (2012). A systematic meta-analysis of the Stroop task in depression. *Clin. Psychol. Rev.* 32, 316–328. doi: 10.1016/j.cpr.2012.02.005

Escande, A., Pillon, A., Servant, N., Cravedi, J. P., Larrea, F., Muhn, P., et al. (2006). Evaluation of ligand selectivity using reporter cell lines stably expressing estrogen receptor alpha or beta. *Biochem. Pharmacol.* 71, 1459–1469. doi: 10.1016/j.bcp.2006.02.002

Fischer, A. H., and Manstead, A. S. R. (2016). "Social functions of emotion and emotion regulation" in Handbook of emotions. eds. L. Feldman Barrett, M. Lewis and J. M. Haviland-Jones (New York, NY: Guilford Press), 424–439.

Galea, L. A. M., Frick, K. M., Hampson, E., Sohrabji, F., and Choleris, E. (2017). Why estrogens matter for behavior and brain health. *Neurosci. Biobehav. Rev.* 76, 363–379. doi: 10.1016/j.neubiorev.2016.03.024

Gawronska, J., Meads, C., Smith, L., Cao, C., Wang, N., and Walker, S. (2024). Association of oral contraceptive pill use and depression among US women. *J. Affect. Disord.* 344, 132–140. doi: 10.1016/j.jad.2023.10.041

Gordon, J. L., Rubinow, D. R., Eisenlohr-Moul, T. A., Xia, K., Schmidt, P. J., and Girdler, S. S. (2018). Efficacy of transdermal estradiol and micronized progesterone in the prevention of depressive symptoms in the menopause transition: a randomized clinical trial. *JAMA Psychiatry* 75, 149–157. doi: 10.1001/jamapsychiatry.2017.3998

Grant, E. C. G., and Pryse-Davies, J. (1968). Effect of oral contraceptives on depressive mood changes and on endometrial monoamine oxidase and phosphatases. *Br. Med. J.* 3, 777–780. doi: 10.1136/bmj.3.5621.777

Hampson, E. (1990). Estrogen-related variations in human spatial and articulatory-motor skills. *Psychoneuroendocrinology* 15, 97–111. doi: 10.1016/0306-4530(90)90018-5

Hampson, E. (2002). "Sex differences in human brain and cognition: the influence of sex steroids in early and adult life" in Behavioral endocrinology. eds. J. B. Becker, S. M. Breedlove, D. Crews and M. M. McCarthy. *2nd* ed (Cambridge, MA: MIT Press), 578–628.

Hampson, E. (2023). Oral contraceptives in the central nervous system: basic pharmacology, methodological considerations, and current state of the field. *Front. Neuroendocrinol.* 68:101040. doi: 10.1016/j.yfrne.2022.101040

Hampson, E., Abrahamson, S. N., Breddy, T. N., Iqbal, M., and Wolff, E. R. (2024). Data from: current oral contraceptive use affects explicit and implicit measures of depression in women. *Open Sci. Framework*. doi: 10.17605/OSF.IO/WCRN4

Hampson, E., Morley, E. E., Evans, K. L., and Fleury, C. (2022). Effects of oral contraceptives on spatial cognition depend on pharmacological properties and phase of the contraceptive cycle. *Front. Endocrinol.* 13:888510. doi: 10.3389/fendo.2022.888510

Hampson, E., van Anders, S. M., and Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: test of an evolutionary hypothesis. *Evol. Hum. Behav.* 27, 401–416. doi: 10.1016/j.evolhumbehav.2006.05.002

Hidalgo-Lopez, E., Engman, J., Poromaa, I. S., Gingnell, M., and Pletzer, B. (2023). Triple network model of brain connectivity changes related to adverse mood effects in an oral contraceptive placebo-controlled trial. *Transl. Psychiatry* 13, 209–220. doi: 10.1038/s41398-023-02470-x

Hu, Z., Liu, H., Weng, X., and Northoff, G. (2012). Is there a valence-specific pattern in emotional conflict in major depressive disorder? An exploratory psychological study. *PLoS One* 7:e31983. doi: 10.1371/journal.pone.0031983

Joffe, H., Cohen, L. S., and Harlow, B. L. (2003). Impact of oral contraceptive pill use on premenstrual mood: predictors of improvement and deterioration. *Am. J. Obstet. Gynecol.* 189, 1523–1530. doi: 10.1016/S0002-9378(03)00927-X

Johansson, T., Vinther Larsen, S., Bui, M., Ek, W. E., Karlsson, T., and Johansson, Å. (2023). Population-based cohort study of oral contraceptive use and risk of depression. *Epidemiol. Psychiatr. Sci.* 32:e39. doi: 10.1017/S2045796023000525

Keyes, K. M., Cheslack-Postava, K., Westhoff, C., Heim, C. M., Haloossim, M., Walsh, K., et al. (2013). Association of hormonal contraceptive use with reduced levels of depressive symptoms: a national study of sexually active women in the United States. *Am. J. Epidemiol.* 178, 1378–1388. doi: 10.1093/aje/kwt188

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., and Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosom. Med.* 61, 154–162. doi: 10.1097/00006842-199903000-00006

Lane, R. D. (2008). Neural substrates of implicit and explicit emotional processes: a unifying framework for psychosomatic medicine. *Psychosom. Med.* 70, 214–231. doi: 10.1097/PSY.0b013e3181647e44

Larsen, S. V., Köhler-Forsberg, K., Dam, V. H., Poulsen, A. S., Svarer, C., Jensen, P. S., et al. (2020). Oral contraceptives and the serotonin 4 receptor: a molecular brain imaging study in healthy women. *Acta Psychiatr. Scand.* 142, 294–306. doi: 10.1111/acps.13211

Larsen, S. V., Ozenne, B., Köhler-Forsberg, K., Poulsen, A. S., Dam, V. H., Svarer, C., et al. (2022). The impact of hormonal contraceptive use on serotonergic neurotransmission and antidepressant treatment response: results from the NeuroPharm 1 study. *Front. Endocrinol.* 13:799675. doi: 10.3389/fendo.2022.799675

LeMoult, J., and Gotlib, I. H. (2019). Depression: a cognitive perspective. *Clin. Psychol. Rev.* 69, 51–66. doi: 10.1016/j.cpr.2018.06.008

LeMoult, J., Yoon, K. L., and Joormann, J. (2012). Affective priming in major depressive disorder. *Front. Integr. Neurosci.* 6:76. doi: 10.3389/fnint.2012.00076

Leppänen, J. M. (2006). Emotional information processing in mood disorders: a review of behavioral and neuroimaging findings. *Curr. Opin. Psychiatry* 19, 34–39. doi: 10.1097/01.yco.0000191500.46411.00

Lisofsky, N., Riediger, M., Gallinat, J., Lindenberger, U., and Kühn, S. (2016). Hormonal contraceptive use is associated with neural and affective changes in healthy young women. *NeuroImage* 134, 597–606. doi: 10.1016/j.neuroimage.2016.04.042

Lundin, C., Danielsson, K. G., Bixo, M., Moby, L., Bengtsdotter, H., Jawad, I., et al. (2017). Combined oral contraceptive use is associated with both improvement and worsening of mood in the different phases of the treatment cycle—a double-blind, placebo-controlled randomized trial. *Psychoneuroendocrinology* 76, 135–143. doi: 10.1016/j.psyneuen.2016.11.033

Lundin, C., Wikman, A., Lampa, E., Bixo, M., Gemzell-Danielsson, K., Wikman, P., et al. (2022). There is no association between combined oral hormonal contraceptives

and depression: a Swedish register-based cohort study. *BJOG* 129, 917–925. doi: 10.1111/1471-0528.17028

McEwen, B. S., and Alves, S. E. (1999). Estrogen actions in the central nervous system. *Endocr. Rev.* 20, 279–307. doi: 10.1210/er.20.3.279

McKetta, S., and Keyes, K. M. (2019). Oral contraceptive use and depression among adolescents. *Ann. Epidemiol.* 29, 46–51. doi: 10.1016/j.annepidem.2018.10.002

McNair, D. M., and Heuchert, J. W. P. (2010). Profile of mood states technical update. North Tonawanda, NY: Multi-Health Systems Inc.

McNair, D. M., Lorr, M., and Droppleman, L. F. (1992). Profile of mood states (revised). San Diego, CA: EdITS/Educational and Industrial Testing Service.

Münkler, P., Rothkirch, M., Dalati, Y., Schmack, K., and Sterzer, P. (2015). Biased recognition of facial affect in patients with major depressive disorder reflects clinical state. *PLoS One* 10:e0129863. doi: 10.1371/journal.pone.0129863

Murphy, S. T., and Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *J. Pers. Soc. Psychol.* 64, 723–739. doi: 10.1037//0022-3514.64.5.723

Noachtar, I. A., Frokjaer, V. G., and Pletzer, B. (2023). Mental health symptoms in oral contraceptive users during short-term hormone withdrawal. *JAMA Netw. Open* 6:e2335957. doi: 10.1001/jamanetworkopen.2023.35957

Oinonen, K. A., and Mazmanian, D. (2002). To what extent do oral contraceptives influence mood and affect? *J. Affect. Disord.* 70, 229–240. doi: 10.1016/S0165-0327(01)00356-1

Östlund, H., Keller, E., and Hurd, Y. L. (2003). Estrogen receptor gene expression in relation to neuropsychiatric disorders. *Ann. N. Y. Acad. Sci.* 1007, 54–63. doi: 10.1196/annals.1286.006

Payne, B. K., Cheng, C. M., Govorun, O., and Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *J. Pers. Soc. Psychol.* 89, 277–293. doi: 10.1037/0022-3514.89.3.277

Payne, K., and Lundberg, K. (2014). The affect misattribution procedure: ten years of evidence on reliability, validity, and mechanisms. *Soc. Personal. Psychol. Compass* 8, 672–686. doi: 10.1111/spc3.12148

Penton-Voak, I. S., and Perrett, D. I. (2000). Female preference for male faces changes cyclically: further evidence. *Evol. Hum. Behav.* 21, 39–48. doi: 10.1016/S1090-5138(99)00033-1

Phillips, W. J., Hine, D. W., and Thorsteinsson, E. B. (2010). Implicit cognition and depression: a meta-analysis. *Clin. Psychol. Rev.* 30, 691–709. doi: 10.1016/j.cpr.2010.05.002

Pletzer, B., Winkler-Crepaz, K., and Hillerer, K. M. (2023). Progesterone and contraceptive progestin actions on the brain: a systematic review of animal studies and comparison to human neuroimaging studies. *Front. Neuroendocrinol.* 69:101060. doi: 10.1016/j.yfrne.2023.101060

Porcu, P., Serra, M., and Concas, A. (2019). The brain as a target of hormonal contraceptives: evidence from animal studies. *Front. Neuroendocrinol.* 55:100799. doi: 10.1016/j.yfrne.2019.100799

Quirin, M., Kazén, M., Rohrmann, S., and Kuhl, J. (2009). Implicit but not explicit affectivity predicts circadian and reactive cortisol: using the implicit positive and negative affect test. *J. Pers.* 77, 401–426. doi: 10.1111/j.1467-6494.2008.00552.x

Rapkin, A. J., Morgan, M., Sogliano, C., Biggio, G., and Concas, A. (2006). Decreased neuroactive steroids induced by combined oral contraceptive pills are not associated with mood changes. *Fertil. Steril.* 85, 1371–1378. doi: 10.1016/j.fertnstert.2005.10.031

Roland, N., Baricault, B., Weill, A., Bouillon, K., Dray-Spira, R., Duranteau, L., et al. (2023). Association between doses of levonorgestrel intrauterine systems and subsequent use of psychotropic drugs in France. *JAMA* 329, 257–259. doi: 10.1001/jama.2022.21471

Ros, L., Satorres, E., Fernández-Aguilar, L., Delhom, I., López-Torres, J., Latorre, J. M., et al. (2023). Differential effects of faces and words in cognitive control in older adults with and without major depressive disorder: an emotional Stroop task study. *Appl. Neuropsychol. Adult* 30, 239–248. doi: 10.1080/23279095.2021.1927037

Rosenberg, M. J., and Waugh, M. S. (1998). Oral contraceptive discontinuation: a prospective evaluation of frequency and reasons. *Am. J. Obstet. Gynecol.* 179, 577–582. doi: 10.1016/S0002-9378(98)70047-X

Sitruk-Ware, R. (2006). New progestogens for contraceptive use. *Hum. Reprod. Update* 12, 169–178. doi: 10.1093/humupd/dmi046

Skovlund, C. W., Mørch, L. S., Kessing, L. V., Lange, T., and Lidegaard, O. (2018). Association of hormonal contraception with suicide attempts and suicides. *Am. J. Psychiatry* 175, 336–342. doi: 10.1176/appi.ajp.2017.17060616

Skovlund, C. W., Mørch, L. S., Kessing, L. V., and Lidegaard, O. (2016). Association of hormonal contraception with depression. *JAMA Psychiatry* 73, 1154–1162. doi: 10.1001/jamapsychiatry.2016.2387

Steiner, M., Dunn, E., and Born, L. (2003). Hormones and mood: from menarche to menopause and beyond. *J. Affect. Disord.* 74, 67–83. doi: 10.1016/S0165-0327(02)00432-9

Stenhammar, E., Wikman, P., Gemzell Danielsson, K., Kopp-Kallner, H., and Sundström Poromaa, I. (2023). Levonorgestrel intrauterine device and depression: a Swedish register-based cohort study. *Int. J. Psychophysiol.* 193:112230. doi: 10.1016/j.ijpsycho.2023.08.003

Strand, M., Oram, M. W., and Hammar, Å. (2013). Emotional information processing in major depression remission and partial remission: faces come first. *Appl. Neuropsychol. Adult* 20, 110–119. doi: 10.1080/09084282.2012.670159

Sundström-Poromaa, I., Comasco, E., Sumner, R., and Luders, E. (2020). Progesterone—Friend or foe? *Front. Neuroendocrinol.* 59:e100856. doi: 10.1016/j.yfrne.2020.100856

Sundström Poromaa, I., and Segebladh, B. (2012). Adverse mood symptoms with oral contraceptives. *Acta Obstet. Gynecol. Scand.* 91, 420–427. doi: 10.1111/j.1600-0412.2011.01333.x

Suslow, T., Bodenschatz, C. M., Kersting, A., Quirin, M., and Günther, V. (2019). Implicit affectivity in clinically depressed patients during acute illness and recovery. *BMC Psychiatry* 19:376. doi: 10.1186/s12888-019-2365-3

Toffol, E., Heikinheimo, O., Koponen, P., Luoto, R., and Partonen, T. (2011). Hormonal contraception and mental health: results of a population-based study. *Hum. Reprod.* 26, 3085–3093. doi: 10.1093/humrep/der269

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 168, 242–249. doi: 10.1016/j.psychres.2008.05.006

United Nations Population Division (2019). Contraceptive use by method. Available at: https://digitallibrary.un.org/record/3849735 (accessed May 28, 2024)

van der Ploeg, M. M., Brosschot, J. F., Thayer, J. F., and Verkuil, B. (2016). The implicit positive and negative affect test: validity and relationship with cardiovascular stress-responses. *Front. Psychol.* 7:425. doi: 10.3389/fpsyg.2016.00425

Weightman, M. J., Air, T. M., and Baune, B. T. (2014). A review of the role of social cognition in major depressive disorder. *Front. Psych.* 5:179. doi: 10.3389/fpsyt.2014.00179

Weiser, M. J., Foradori, C. D., and Handa, R. J. (2008). Estrogen receptor beta in the brain: from form to function. *Brain Res. Rev.* 57, 309–320. doi: 10.1016/j.brainresrev.2007.05.013

Wiréhn, A. B., Foldemo, A., Josefsson, A., and Lindberg, M. (2010). Use of hormonal contraceptives in relation to antidepressant therapy: a nationwide population-based study. *Eur. J. Contracept. Reprod. Health Care* 15, 41–47. doi: 10.3109/13625181003587004

Woolley, C. S., and McEwen, B. S. (1993). Role of estradiol and progesterone in regulation of hippocampal dendritic spine density during the estrous cycle in the rat. *J. Comp. Neurol.* 336, 293–306. doi: 10.1002/cne.903360210

Zettermark, S., Vicente, R. P., and Merlo, J. (2018). Hormonal contraception increases the risk of psychotropic drug use in adolescent girls but not in adults: a pharmacoepidemiological study on 800,000 Swedish women. *PLoS One* 13:e0194773. doi: 10.1371/journal.pone.0194773

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 December 2024
DOI 10.3389/fpsyg.2024.1401158

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Geoff Kushnick,
Australian National University, Australia
Jan Havlicek,
Charles University, Czechia

*CORRESPONDENCE
Janek S. Lobmaier
✉ j.lobmaier@psychologie.uzh.ch

RECEIVED 14 March 2024
ACCEPTED 18 November 2024
PUBLISHED 13 December 2024

CITATION
Lobmaier JS, Klatt WK and Schweinberger SR (2024) Voice of a woman: influence of interaction partner characteristics on cycle dependent vocal changes in women.
Front. Psychol. 15:1401158.
doi: 10.3389/fpsyg.2024.1401158

Voice of a woman: influence of interaction partner characteristics on cycle dependent vocal changes in women

Janek S. Lobmaier[1]*, Wilhelm K. Klatt[2] and Stefan R. Schweinberger[3,4]

[1]Department of Clinical Psychology and Psychotherapy, University of Zurich, Zurich, Switzerland, [2]Sigma-Zentrum für Akutmedizin, Fachkrankenhaus für Psychiatrie, Psychotherapie, Psychosomatische Medizin, Bad Säckingen, Germany, [3]Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University, Jena, Germany, [4]Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Introduction: Research has shown that women's vocal characteristics change during the menstrual cycle. Further, evidence suggests that individuals alter their voices depending on the context, such as when speaking to a highly attractive person, or a person with a different social status. The present study aimed at investigating the degree to which women's voices change depending on the vocal characteristics of the interaction partner, and how any such changes are modulated by the woman's current menstrual cycle phase.

Methods: Forty-two naturally cycling women were recorded once during the late follicular phase (high fertility) and once during the luteal phase (low fertility) while reproducing utterances of men and women who were previously assessed to have either attractive or unattractive voices.

Results: Phonetic analyses revealed that women's voices in response to speakers changed depending on their menstrual cycle phase (F0 variation, maximum F0, Centre of gravity) and depending on the stimulus speaker's vocal attractiveness (HNR, Formants 1–3, Centre of gravity), and sex (Formant 2). Also, the vocal characteristics differed when reproducing spoken sentences of the stimulus speakers compared to when they read out written sentences (minimum F0, Formants 2–4).

Discussion: These results provide further evidence that women alter their voice depending on the vocal characteristics of the interaction partner and that these changes are modulated by the menstrual cycle phase. Specifically, the present findings suggest that cyclic shifts on women's voices may occur only in social contexts (i.e., when a putative interaction partner is involved).

KEYWORDS

menstrual cycle, social mimicry, accommodation, attractiveness, phonetics

Introduction

Social interactions are an integral part of our daily lives. In the vast tapestry of human social interactions, the voice serves as a powerful medium to convey not only factual information, but also rich dynamic information about the speaker's identity, emotions, and intentions (Schweinberger et al., 2014). From an evolutionary perspective, the voice may be understood as an "honest signal," a concept rooted in the theory that certain traits and behaviors have evolved to convey truthful information to others about the individual's fitness, health, or reproductive status. The human voice, particularly in women, may carry indicators of reproductive viability or general health. The pitch and tonal quality of a woman's voice can

signal her age and hormonal status, which are closely linked to fertility. In fact, research suggests that women's voices sound most attractive during the high-fertility phase of the menstrual cycle (Pipitone and Gallup, 2008). Meanwhile, there is also evidence that people change their voices depending on the context, for example when speaking to a highly attractive person or a person with a different social status (e.g., Fraccaro et al., 2011; Hughes et al., 2010; Zraick et al., 2006). The aim of the present study was to investigate the extent to which women's voices change depending on the vocal characteristics of the interaction partner, and whether such changes are modulated by the current phase of the woman's menstrual cycle.

Hearing a person's voice allows a listener to form impressions about the speaker's personality. For example, men and women with a lower voice pitch are perceived as more competent and trustworthy than those with higher-pitched voices. Conversely, women with a higher voice pitch are judged as having a warmer personality than women with lower-pitched voices (Oleszkiewicz et al., 2017). In return, a speaker's voice can be adapted, according to the circumstances in which a conversation takes place. For instance, women have been shown to speak with higher voice pitch, larger pitch range, expanded intonation contours, and slower speech rate when speaking to infants (i.e., "motherese"), compared to when speaking to adults (Fernald and Simon, 1984; Grieser and Kuhl, 1988), and this phenomenon has been shown to be universal in western and traditional cultures (Broesch and Bryant, 2015). Voice pitch is also affected by changes in tension (Titze, 1989), intonation, stress, and loudness of speech (Raphael et al., 2011), and women have been reported to speak in a lower or higher voice pitch when speaking to a superior or a subordinate, respectively (Zraick et al., 2006).

Voice pitch is typically measured by the mean fundamental frequency (mean F0), depicting the rate of vocal fold vibration. The standard deviation of the fundamental frequency (F0 SD) represents the variability in perceived voice pitch (i.e., intonation). Low values in F0 SD are perceived as a monotonous voice, higher values are found in melodious voices. Other measures often used to objectively measure voice characteristics include minimum fundamental frequency ($F0_{min}$) and maximum fundamental frequency ($F0_{max}$), Centre of gravity, Formants 1 to 4 (F1–F4), harmonics-to-noise ratio (HNR), Jitter, Shimmer, and variation in intensity (Intensity SD). $F0_{min}$ and $F0_{max}$ represent the upper and lower limits of the pitch range. Centre of gravity is the frequency which divides the voice spectrum into two halves, so a higher Centre of gravity means that a voice has more high-frequency energy in its spectrum. F1 to F4 are frequency ranges that are intensified within the spectrum. HNR is the ratio of harmonic to non-harmonic components within the voice spectrum and reflects breathiness of the voice. Jitter is a local variation in frequency, Shimmer is a local variation in amplitude; both Jitter and Shimmer are caused by irregular vocal fold vibration and are perceived as roughness in a speaker's voice. Finally, Intensity SD indicates the variability in perceived loudness.

People seem to alter their voice when speaking to interaction partners they find attractive. Fraccaro et al. (2011) found that women speak at a higher pitch (higher mean F0) to men to whom they are attracted, whereas Hughes et al. (2010) reported that both women and men lower their voice pitch (mean F0) when speaking to an attractive opposite-sex target. When speaking to attractive women, both men and women appear to show a greater voice pitch variability (higher F0 SD; Leongómez et al., 2014). Farley et al. (2013) provided evidence suggesting that naïve listeners were able to identify whether a person was speaking to a romantic partner or a friend. Women used a deeper voice (lower mean F0) when talking to their romantic partners than to their friends. In contrast, men used a higher voice pitch when addressing their romantic partners. As a limitation, this study compared voice samples directed to opposite-sex partners and voice samples directed to same-sex friends, conflating sex of the interaction partner with intimacy. Nevertheless, these studies together suggest that women change their voices depending on the sex and attractiveness of the interaction partner. However, none of these studies controlled for a potential influence of the menstrual cycle.

Regarding attractiveness of voices, there is evidence that attractiveness increases with "averageness" of a voice – an effect that is potentially enhanced by larger harmonics-to-noise ratio in attractive voices (Bruckert et al., 2010; Zaske et al., 2020a). Several studies suggest that women's voices, and voice attractiveness in particular, change during the menstrual cycle. Pipitone and Gallup (2008) found that the attractiveness of naturally cycling women's voices increases with their conception probability. In a follow-up study, Shoup-Knox and Pipitone (2015) demonstrated that voices during high fertility period are not only rated as more attractive than low fertility voices but also lead to a higher galvanic skin response in the listeners, suggesting an increased arousal at a physiological level. Shoup-Knox et al. (2019) analysed Pipitone and Gallup's (2008) recordings phonetically, finding a significantly lower Shimmer in high fertility compared to low fertility recordings of naturally cycling women. Bryant and Haselton (2009) found that women's voice pitch (mean F0) was increased during high compared to low fertility when speaking the sentence "Hi, I'm a student at UCLA." At first sight, these findings suggest that a higher pitch in women's voices might convey a cue to fertility. Karthikeyan and Locke (2015) replicated the finding that highly fertile women's voices are rated as more attractive, but they found that the women in their sample actually spoke in a *lower* voice pitch compared to when not fertile. A further study recorded women's voices on a daily basis throughout their cycle (Fischer et al., 2011), finding that mean voice pitch and variation in voice pitch increase prior to ovulation and show a distinct drop on the day of ovulation.

Consistently, men rated the voices to be more attractive during the pre-ovulatory period than on the day of ovulation itself (Fischer et al., 2011). Banai (2017) compared the voices of naturally cycling women and women using hormonal contraceptives during menstruation, the late follicular and the luteal phase. She found naturally cycling women to have a higher *minimum* voice pitch ($F0_{min}$) in the late follicular phase and a lower voice intensity in the luteal phase, each compared to the other phases. In hormonal contraceptive users, no voice changes across the cycle were detected. In contrast to Fischer et al. (2011), Banai (2017) did not find a significant effect in *mean* voice pitch (mean F0). This indicates that mean F0 alone does not seem to be a reliable cue to fertility. Similar to Banai (2017), La and Polo (2020) compared the voices of naturally cycling women with those taking hormonal contraceptives at three different times of the cycle, but in a double-blind design. In naturally cycling women, they found a significantly lower mean F0, F0 SD, and maximum F0 during menstruation, compared to the follicular and luteal phase. In women using hormonal contraceptives there was no cycle-dependent difference in these measures.

However, overall, and despite considerable variability in study designs and findings reported above, a sizeable number of published findings indicate the existence of voice changes depending on the menstrual cycle.

A common explanation for cycle-dependent voice changes is based on the fact that levels of women's reproductive hormones vary during the menstrual cycle, with a surge of estradiol in the late-follicular phase and high levels of progesterone in the luteal phase. Regarding the vocal apparatus, estradiol promotes cell differentiation and mucosa secretion, while progesterone enhances the acidity and viscosity of the mucus for water retention, resulting in increased mass of the vocal folds, which in turn promotes lower-frequency vibration (Abitbol et al., 1999; Karthikeyan and Locke, 2015). In addition, there is evidence for specific sex hormone receptors within the vocal fold mucosa (Schneider et al., 2007; but see Nacci et al., 2011). Together, these studies suggest that a woman's vocal apparatus may change under the influence of cycle dependent hormone concentration, leading to perceivable changes in her voice.

It should be noted that not all studies found an effect of the menstrual cycle on women's voices. For example, Barnes and Latman (2011) found no significant voice changes across the cycle and no differences between naturally cycling women and those taking hormonal contraceptives in mean F0, jitter, shimmer, relative average perturbation, peak-to-peak amplitude variation, HNR, degree of voice breaks, and number of voice breaks. Likewise, Meurer et al. (2009), Raj et al. (2010), Celik et al. (2013), and Plexico et al. (2020) found no evidence for cycle-dependent voice changes.

An alternative explanation for cycle effects on women's voices – which might also explain conflicting results – refers to psychological changes which may lead to increased mating motivation when currently fertile (Haselton and Gildersleeve, 2016). Karthikeyan and Locke (2015) suggest that only when provided with a mating context, women may be motivated to speak and behave more attractively, showing subtle vocal behaviors that are phase-specific (see also Klatt et al., 2020). However, even within a mating context, Pavela Banai et al. (2022) did not detect cycle-related changes in mean F0 and F0 SD in naturally cycling women who were leaving voice messages directed to masculinized and feminized pictures of men and women. Given these inconsistent results, Pavela Banai et al. (2022) name several methodological issues in earlier studies on cycle-dependent voice changes. First, different cycle-tracking methods have been used (counting method, body temperature, assessment of hormone levels). Second, many studies focused on mean F0 only, disregarding other voice parameters. Third, a variety of vocal stimuli have been analysed (vowels, numbers, sentences, free speech) which differ in experimental control and ecological validity. For example, women's mean F0 has been shown to differ significantly depending on whether they were counting, producing vowels, reading out, or speaking spontaneously (Zraick et al., 2000).

In addition to these issues, one could argue that when a mating context is created, it makes a difference whether or not you are leaving a voice message for a target with an attractive face (Hughes et al., 2010; Pavela Banai et al., 2022) or an attractive voice (e. g., Karthikeyan and Locke, 2015). Furthermore, we want to point out that for phonetic analyses, previous studies used different software applications and different frequency ranges when analysing voice recordings for mean F0, F0 SD, $F0_{min}$, and $F0_{max}$. It is reasonable that phonetic analyses will produce different results when the predefined search area for F0 parameters is different (see Table 1 for an overview of different studies and the respective F0 definitions). Also, despite the fact that mean F0 may be the easiest voice parameter to quantify, researchers interested in social mimicry of voices are well advised to also quantify other relevant vocal parameters.

TABLE 1 Overview of frequency ranges used in phonetic analyses in previous studies.

Study	Software	Frequency range analysed for F0
Banai (2017)	Praat	100–500 Hz
Pavela Banai et al. (2022)	Praat	100–500 Hz
Barnes and Latman (2011)	N/A	N/A
Bryant and Haselton (2009)	Praat	100–600 Hz
Celik et al. (2013)	Praat	N/A
Farley et al. (2013)	Praat	N/A*
Feinberg et al. (2006)	Praat	100–600 Hz
Fischer et al. (2011)	Kurt Hammerschmidt's custom acoustic analysis tool	Automatic**
Fraccaro et al., 2011, referring to Feinberg et al. (2008), referring to Feinberg et al. (2005)	Praat	100–600 Hz
Hughes et al. (2010)	Praat	75–600 Hz***
Karthikeyan and Locke (2015)	Multi-speech software	75–300 Hz
La and Polo (2020)	Electrolaryngographic recording, sopran software by svante granqvist	N/A
Leongómez et al. (2014)	Praat	100–500 Hz
Meurer et al. (2009)	Motor speech profile program	N/A
Plexico et al. (2020)	Computerized speech laboratory; analysis of dysphonia in speech and voice program	E-Mail 08.10.23, no answer
Raj et al. (2010)	Vaughmi software	N/A
Shoup-Knox et al. (2019)	Praat	E-Mail 08.10.23, no answer

*Personal communication via e-mail, 16.09.2023. **Personal communication via e-mail, 15.09.2023. ***Personal communication via e-mail, 15.09.2023.

Given that women's voices can change depending on their current menstrual cycle phase, characteristics of the interaction partner and speech context, we aimed at further scrutinizing potential internal and external factors that may have an effect on women's voices. Specifically, as an internal factor of the participants, we tested whether the menstrual cycle phase affects women's voices. In addition, as an external factor, we manipulated the interaction partner of the women (operationalized here as the stimulus speaker). This interaction partner could either be a man or woman, with an attractive or unattractive voice. In contrast to studies in which conception probability was calculated using counting methods (Pipitone and Gallup, 2008; Puts et al., 2013), we verified the menstrual cycle phases by hormone tests based on urine (luteinising hormone LH) and saliva (estradiol, progesterone). We hypothesized that women alter their voices depending on sex and attractiveness of the stimulus speaker's voice, and that these changes are modulated by the current menstrual cycle phase of the women. An effect of menstrual cycle was thought to occur more pronounced in response to men's voices than to women's voices, and more pronounced in response to attractive voices than to unattractive voices. Voice recordings were analysed phonetically for a large range of parameters that included mean F0, $F0_{min}$, $F0_{max}$, F0 SD, HNR, Jitter, Shimmer, Formants 1 to 4, Centre of gravity, and Intensity SD.

Materials and methods

Participants

Eighty-three women were initially recruited for the present study. Participants were recruited from the subject pool of the university, via flyers and leaflets, the institutional website, free internet advertisements, and word-of-mouth recommendation. All of them provided written informed consent to participate and the study was approved by the local ethics committee (approval number: 2012-8-167070) and participants were treated in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). They were compensated either with course credits or 50 CHF (approximately 60 US$). The final sample consisted of 42 women (see below).

Stimuli

Recordings of stimulus speakers (spoken sentences) were taken from the Jena Speaker Set (JESS; Zaske et al., 2020b) which at the time consisted of 64 speakers (half of them women) who were recorded while speaking a series of neutral sentences. All sentences consisted of exactly five words and had the same syntactic structure (e.g., "Der Fahrer lenkt den Wagen"/"The driver steers the car," see Supplementary Table S1 for the complete list of sentences used).

In order to identify the most attractive and most unattractive voices of the 64 stimulus speakers, an attractiveness rating was performed with an independent sample of participants using PsychoPy software (Peirce, 2007). Twenty-four listeners (half of them women, ranging in age between 20 and 39 years, $M = 25.0$, $SD = 4.6$) who were recruited via flyers and word-of-mouth recommendation were received individually in a quiet room. After giving informed consent, they were asked to take place in front of a laptop and to attach Sennheiser HD 439 headphones. Listeners were told that they would be hearing spoken sentences via headphones and that they were asked to respond to the question "How attractive do you find this voice?" on a 7-point likert scale that was showed on the screen (labelled "very attractive" and "very unattractive" at each end). Instructions were given verbally and onscreen. Listeners first completed three practice trials with sentences and stimulus speakers which were not used afterwards to get used to the task and to adjust the volume of the headphones. Voice recordings of the 64 stimulus speakers uttering the same two neutral sentences ("The train passes the town," "The lemonade quenches the thirst") were presented and rated for attractiveness separately in randomised order, resulting in 128 trials. The experiment took about 10 min to complete. Rank analyses identified the eight most attractive and eight most unattractive female and male speakers, respectively, resulting in 32 different stimulus speakers which were then used in the present study.

Procedure

Initially, all interested women were asked to complete an online survey with questions regarding age, smoking habits, mother tongue, reading disabilities, hearing problems, sexual orientation, relationship status, use of hormonal contraception, pregnancy, onset of last menstruation, regularity and length of menstrual cycle. To take part in this study, participants had to meet following inclusion criteria: No use of hormonal contraceptives (contraceptive pill, morning-after pill, hormonal contraceptive coil, contraceptive implants); no pregnancy, and no breastfeeding within the last three months; regular menstrual cycle (23–35 days in length); heterosexual orientation; German mother tongue; no dyslexia; no hearing problems; no chronic smoking (more than 20 cigarettes a week). All reported to be healthy (no mental and/or physical diseases) and not to have a hoarse voice, cough, or nasal congestion on the days of recording. Women who met the above-mentioned inclusion criteria were contacted by phone by a female research assistant who gave them detailed information about the procedure.

To determine time of highest fertility, participants completed a series of urine tests measuring a metabolite of luteinising hormone (LH) using one-step urine LH tests with a reported sensitivity of 10 mIU/ml (David One Step Ovulation Tests, Runbio Biotech, China, http://www.runbio-bio.com). Women were instructed to perform urine tests twice a day (morning and evening) starting three days before the date of predicted peak fertility (based on the average cycle length of each individual woman using forward and backward counting method). After a positive test result, participants continued performing LH tests until the results became negative for two consecutive days. Participants photographed each test using their smartphones and sent the picture to the research assistant, who verified whether the test was positive or not.

The women were either scheduled to be tested approximately two days before the calculated day of peak fertility and again seven days after a positive LH test result (late follicular–luteal menstrual cycle condition) or they were scheduled seven days after the LH surge (luteal–late follicular menstrual cycle condition). Participants of the luteal–late follicular menstrual cycle condition performed LH tests again in the following cycle and were scheduled to be tested two days before the calculated day of peak fertility. Thus, LH tests were used to

determine peak fertility and to verify that the cycle was ovulatory. Late follicular recording sessions took place between 4 days before and 24 h after the LH surge, luteal phase recording sessions took place 6 to 13 days after the LH surge (see Table 2 for an overview of the time of recording relative to the LH surge). Order of recording sessions was counterbalanced across participants: half of the women completed the first session at high fertility and their second session in the luteal phase (late follicular–luteal menstrual cycle condition), and the other half were tested first during the luteal phase and then during high fertility (luteal–late follicular menstrual cycle condition).

In order to assess phase-specific hormone levels, participants provided saliva samples from which estradiol, progesterone, testosterone, and cortisol levels were determined. Participants were instructed to refrain from eating and drinking anything but water for at least 30 min prior to saliva collection. Samples were collected by passive drool using a commercially available sampling device (SaliCaps, IBL International, Hamburg, Germany). The saliva samples were stored at −28°C and were later analysed by an independent laboratory (Dresden Lab Service GmbH, Dresden, Germany) using liquid chromatography with coupled tandem mass spectrometry (LC–MS/MS). LC–MS/MS has become the method of choice for steroid analysis because of its high sensitivity, better reproducibility, greater specificity, and ability to analyse multiple steroids simultaneously (see Gao et al., 2015 for methodological details on LC–MS/MS). Both recording sessions took place between 8 and 11 AM in order to control for circadian variability of hormone levels (Dabbs and Delarue, 1991; Wust et al., 2000).

Women's voices were recorded in a soundproof recording booth under standardised conditions. A Beyerdynamic MC 930 condenser microphone with a popkiller and 48 V phantom power was placed about 20 centimetres away from the participant's mouth and connected to a Zoom H4n digital audio recorder (uncompressed WAV, 48 kHz, 16-bit sampling rate). Before starting the experiment proper, participants were asked to read out a short newspaper article about voice research presented on a computer screen to warm up their voices and to adjust the recording level to −12 to −6 dB. Each woman absolved two recording sessions (one in the late follicular and one in the luteal phase), and each recording session consisted of two blocks. In the first block the woman read aloud presented written sentences; in the second block, the task was to reproduce spoken sentences verbally which were presented via headphones. Written and spoken sentences were identical in content. At the beginning of the first block, the instruction to read out presented written sentences in a natural manner was given verbally and onscreen. Twenty-four sentences of affectively neutral content (see Supplementary Table S1) were presented. Sentences were displayed on a laptop screen in 25 pt. Arial using PsychoPy software (Peirce, 2007). The screen was placed in front of the participant at a distance of approximately 50 centimetres. Sentences were presented consecutively in randomised order for five seconds each. After three practice trials with sentences which were not used subsequently, the experimenter left the booth and the participant started the first block of the recording session (written sentences). The participant indicated by knocking on the booth's door when the block was finished. The experimenter returned, asked the participant to attach Sennheiser HD 439 headphones and prepared the second block of the recording session (spoken sentences). Participants were told that they would be hearing spoken sentences via headphones and that they were to reproduce these sentences in a natural manner. Instructions were given verbally and onscreen. Sentences were presented separately in randomised order. Participants first completed three practice trials with sentences and stimulus speakers which were not used in the experiment proper. Subsequently, they were given the opportunity to adjust the volume of the headphones. After that, the experimenter left the booth and the participant started the second block (spoken sentences). Again, the participant indicated by knocking on the booth's door when the block was finished. Both recording sessions (late follicular and luteal phase) consisted of these two blocks, following the same procedure except that participants were fully debriefed after the second recording session. Each recording session took about 45 min to complete.

Of the 83 women who initially entered the study, some had to be excluded because their recordings were unusable due to misspeaking ($N = 12$), because they had anovulatory cycles during the recording period (i.e., no LH surge, $N = 8$), did not conduct the required LH tests during the peri-ovulatory period ($N = 3$), were tested in the wrong cycle phase (too early/too late as revealed by LH tests, $N = 8$), or dropped out due to personal reasons ($N = 10$). Thus, the final sample consisted of 42 women between 19 and 35 years of age ($M = 22.6$, $SD = 3.4$). From these, a total of 2,016 uncompressed WAV voice samples (42 participants × 2 menstrual cycle phases × 24 sentences) were cut from raw recordings using Audacity® software (Audacity Team, 2023).

TABLE 2 Time of recording sessions relative to peak fertility as indicated by LH surge.

Days relative to LH surge	Participants in late follicular session	Days relative to LH surge	Participants in luteal session
− 4	2	+ 6	5
− 3	3	+ 7	8
− 2	4	+ 8	12
− 1	9	+ 9	10
0	13	+ 10	2
+ 1	11	+ 11	3
		+ 12	1
		+ 13	1

Variables and statistical analyses

Praat software version 6.3.16 (Boersma and Weenink, 2023) was used to analyse the voice samples for the following phonetic parameters: Mean F0, F0 SD, $F0_{min}$, $F0_{max}$, Centre of gravity, Formants 1 to 4, HNR, Jitter, Shimmer, and Intensity SD. For analyses of mean F0, F0 SD, $F0_{min}$, and $F0_{max}$, the frequency range was set to 100–500 Hz and for formant analyses, the maximum frequency was set to 5,500 Hz, as suggested by Boersma and Weenink (2023) for female voices. Apart from that, default settings were used.

Control condition: reading written sentences aloud

In the control condition (where participants read the sentences out loud from the computer screen), for each woman, values of

individual vocal parameters were averaged over the 24 sentences, separately for both menstrual cycle phases for further analyses. We then calculated one-factor (participant's menstrual cycle phase: Late follicular vs. luteal phase) repeated measures ANOVA's, separately for each of the vocal parameters using Bonferroni correction.

We also ran Pearson correlations between the hormone levels and the vocal parameters, separately for the late follicular and luteal phase. We present these correlations in the Supplementary material.

Experimental condition: reproduction of spoken sentences

In the Experimental condition (where participants reproduced the sentences after hearing them being spoken), for each woman, values of individual vocal parameters were averaged over the individual sentences, separately for menstrual cycle phase of the woman, stimulus speaker's vocal attractiveness, and stimulus speaker's sex, for further analyses. We then calculated 2 (participant's menstrual cycle phase: Late follicular vs. luteal phase) × 2 (stimulus speaker's vocal attractiveness: Attractive vs. unattractive) × 2 (stimulus speaker's sex: Female vs. male) repeated measures ANOVA separately for each of the vocal parameters. The Huynh-Feldt epsilon correction for heterogeneity of covariances (Huynh and Feldt, 1976) was used when sphericity could not be assumed. For post-hoc pairwise comparisons we used the Bonferroni correction. To correct for multiple testing, we used the Holm-Bonferroni method. We report the corrected alongside the uncorrected *p*-values.

We also ran Pearson correlations between the hormone levels and the vocal parameters, separately for the late follicular and luteal phase. We present these correlations in the Supplementary material.

Phonetic analysis of the stimulus speakers' voices

To examine whether the effects in women's voices in reaction to the stimulus speakers' voice characteristics may be driven by social mimicry or accommodation, also the stimulus speakers' voices were analysed phonetically for the same parameters as the participants using Praat software version 6.3.16 (Boersma and Weenink, 2023). Female and male stimulus speakers' voices were analysed separately because of sex-specific pre-adjustments in the phonetic software (Leongómez et al., 2014). For analyses of mean F0, F0 SD, $F0_{min}$, and $F0_{max}$ of the female stimulus speakers, frequency range was set to 100–500 Hz; for formant analyses, the maximum frequency was set to 5,500 Hz (same as for analyses performed with the participants' voices). For analyses of mean F0, F0 SD, $F0_{min}$, and $F0_{max}$ of the male stimulus speakers, frequency range was set to 75–300 Hz;for formant analyses, the maximum frequency was set to 5,000 Hz, as suggested by Boersma and Weenink (2023). Apart from that, default settings were used.

Results

Control condition: reading written sentences aloud

The one-factor (participant's menstrual cycle phase: Late follicular vs. luteal phase) repeated measures ANOVAs indicated no effect of menstrual cycle phase on any of the phonetic measures (all *p*s > 0.30). ANOVA results and descriptive statistics are provided in Table 3.

Experimental condition: reproduction of spoken sentences

The 2 (participant's menstrual cycle phase: Late follicular vs. luteal phase) × 2 (stimulus speaker's vocal attractiveness: Attractive vs. unattractive) × 2 (stimulus speaker's sex: Female vs. male) repeated measures ANOVA was calculated on each of the vocal parameters. The results of the 2 (late follicular vs. luteal phase) × 2 (attractive vs. unattractive) × 2 (stimulus speaker's sex) repeated measures ANOVAs are shown in Table 4, descriptive statistics are given in Table 5.

TABLE 3 Control condition (reading aloud written sentences): phonetic parameters depending on participants' menstrual cycle phase.

	Menstrual cycle phase				ANOVA		
	Late follicular		Luteal		(follicular vs. luteal)		
	M	*SD*	*M*	*SD*	*F*	*p*	η_p^2
F0	213.25 Hz	16.75	213.39 Hz	17.46	0.011	0.92	0.000
F0 SD	44.35 Hz	8.85	44.67 Hz	9.76	0.067	0.80	0.002
$F0_{min}$	147.99 Hz	19.31	148.32 Hz	17.94	0.021	0.89	0.001
$F0_{max}$	386.88 Hz	40.25	389.12 Hz	39.89	0.127	0.72	0.003
Centre of gravity	684.99 Hz	254.10	668.83 Hz	193.61	0.936	0.34	0.022
F1	674.71 Hz	42.28	678.43 Hz	36.78	0.377	0.54	0.009
F2	1842.96 Hz	53.84	1842.06 Hz	56.57	0.041	0.84	0.001
F3	2863.71 Hz	59.82	2860.69 Hz	63.29	0.288	0.60	0.007
F4	3925.51 Hz	63.81	3923.96 Hz	65.48	0.052	0.82	0.001
HNR	13.31 dB	2.21	13.17 dB	2.16	1.027	0.32	0.024
Jitter	0.03%	0.01	0.03%	0.01	0.511	0.48	0.012
Shimmer	0.09%	0.01	0.09%	0.01	0.000	1.00	0.000
Intensity SD	11.14 dB	1.37	11.25 dB	1.20	0.473	0.50	0.011

TABLE 4 Experimental condition (reproduction of spoken sentences), inferential statistics: phonetic parameters depending on menstrual cycle phase of the participants, voice attractiveness of the stimulus speaker, and sex of the stimulus speaker.

	Participant's menstrual cycle phase (follicular vs. luteal)				Stimulus speaker's voice attractiveness (attractive vs. unattractive)				Stimulus speaker's sex (female vs. male)			
	F	*p*	*p* (corr)	η_p^2	F	*p*	*p* (corr)	η_p^2	F	*p*	*p* (corr)	η_p^2
F0	0.510	0.48	>0.999	0.012	0.472	0.50	>0.999	0.011	0.689	0.41	>0.999	0.017
F0 SD	9.008	**0.005**	0.06	0.180	2.136	0.15	>0.999	0.050	0.468	0.50	>0.999	'011
$F0_{min}$	3.46	0.07	0.7	0.078	0.467	0.50	>0.999	0.011	2.584	0.116	>0.999	0.059
$F0_{max}$	15.379	**0.001**	**0.013**	0.273	1.570	0.22	>0.999	0.037	0.911	0.35	>0.999	0.022
CoG	5.128	**0.029**	0.319	0.111	13.601	**0.001**	**0.013**	0.249	0.079	0.78	>0.999	0.002
F1	0.215	0.65	>0.999	0.005	17.907	**<0.001**	**0.013**	0.304	2.611	0.11	>0.999	0.060
F2	0.218	0.353	>0.999	0.009	8.922	**0.005**	0.055	0.179	11.436	**0.002**	**0.026**	0.218
F3	0.548	0.46	>0.999	0.013	7.364	**0.01**	0.09	0.152	0.171	0.68	>0.999	0.004
F4	0.546	0.50	>0.999	0.011	0.332	0.57	>0.999	0.008	1.703	0.20	>0.999	0.040
HNR	1.822	0.18	>0.999	0.043	8.885	**0.005**	0.055	0.178	0.260	0.61	>0.999	0.006
Jitter	0.093	0.76	>0.999	0.002	0.882	0.35	>0.999	0.021	1.277	0.27	>0.999	0.030
Shimmer	0.177	0.77	>0.999	0.004	0.669	0.42	>0.999	0.016	2.196	0.15	>0.999	0.051
Intens SD	0.412	0.53	>0.999	0.010	0.115	0.74	>0.999	0.003	0.592	0.47	>0.999	0.013

p (corr): *p*-value corrected for multiple testing using the Holm-Bonferroni Method. Bold values indicate significant effects.

Menstrual cycle phase of the participants had a significant main effect on F0 SD, $F0_{max}$, and Centre of gravity. When recorded in the late follicular phase, women spoke with lower F0 SD, lower $F0_{max}$, and higher Centre of gravity compared to the luteal phase. When correcting for multiple testing, menstrual cycle phase had an effect only on $F0_{max}$. Stimulus speaker's voice attractiveness showed a significant effect on five phonetic measures. HNR was lower in response to stimulus speakers with attractive voices than in response to speakers with less attractive voices, implying that participants used a breathier and hoarser voice when reacting to attractive voices than when reacting to unattractive voices. Formants 1 to 3, reflecting frequencies that are intensified relative to the rest of the vocal spectrum, were significantly higher in frequency when responding to attractive stimulus speakers compared to unattractive stimulus speakers. Additionally, women showed a higher frequency in Centre of gravity in response to attractive stimulus speakers compared to unattractive stimulus speakers. When correcting for multiple testing, the effect remained significant only for Center of Gravity and F1. The sex of the stimulus speaker had a significant effect only on Formant 2, which was higher in frequency when women responded to a female stimulus speaker, compared to when responding to a male stimulus speaker. This effect was also significant after correcting for multiple testing. No significant main effects were observed in mean F0, $F0_{min}$, Jitter, Shimmer, Formant 4 and Intensity SD. No interaction reached statistical significance (all $ps \geq 0.05$; see Supplementary Table S2).

Phonetic analysis of the stimulus speakers' voices

Phonetic analyses of the stimulus speakers revealed that stimulus speakers with attractive voices of both sexes had higher formants and a higher Centre of gravity than stimulus speakers with less attractive voices (for more detailed results, see Supplementary Tables S3, S4). The differences in these parameters were reflected in women's voices when responding to attractive versus unattractive stimulus speakers (Formants 1–3, centre of gravity).

Regarding stimulus speaker's sex, not surprisingly, analyses showed that female stimulus speakers had higher frequencies in mean F0, F0 SD, $F0_{min}$, $F0_{max}$, and Formants 2–4 than male stimulus speakers, as well as higher voice quality according to higher HNR, lower Jitter, and lower Shimmer (see Supplementary Table S5). The difference in Formant 2 was reflected in women's voices when responding to male versus female stimulus speakers.

Taken together, phonetic analyses of the stimulus speakers' voices suggest that the shifts observed in women's voices – depending on stimulus speaker attractiveness and sex – are at least in part explainable by social mimicry or accommodation (Chartrand and Lakin, 2013; Gregory and Webster, 1996).

General effect of the predefined frequency range in phonetic analysis

Previous studies (e.g., Hughes et al., 2010; Karthikeyan and Locke, 2015) used different frequency ranges in phonetic analysis of their voice recordings compared to other studies (e.g., Leongómez et al., 2014). To test whether phonetic analyses produce different results depending on the search area in frequency range, we repeated the phonetic analysis with Praat software's default frequency range (75–600 Hz instead of 100–500 Hz as suggested for female voices by Boersma and Weenink, 2023). As expected, the results of mean F0, F0 SD, $F0_{min}$, and $F0_{max}$ differed significantly between predefined frequency range of 75–600 Hz and 100–500 Hz. Also, HNR and Shimmer were significantly different when Praat's default frequency range was used (see Supplementary Table S6, S7).

TABLE 5 Experimental condition (reproduction of spoken sentences), descriptive statistics: phonetic parameters depending on menstrual cycle phase of the participants, attractiveness of the stimulus speaker, and sex of the stimulus speaker.

	Participant's menstrual cycle phase		Stimulus speaker's voice attractiveness		Stimulus speaker's sex	
		M (SD)		*M (SD)*		*M (SD)*
F0 [Hz]	Fol	211.94 (16.98)	Attr	212.18 (17.10)	Female	212.41 (17.09)
	Lut	212.63 (18.01)	Unattr	212.39 (17.41)	Male	212.17 (17.41)
F0 SD [Hz]	Fol	44.48 (6.93)	Attr	45.84 (7.84)	Female	45.65 (7.56)
	Lut	46.55 (8.53)	Unattr	45.19 (7.32)	Male	45.37 (7.55)
$F0_{min}$ [Hz]	Fol	139.14 (18.24)	Attr	140.62 (18.01)	Female	140.21 (17.62)
	Lut	142.61 (16.69)	Unattr	141.13 (18.30)	Male	141.54 (18.74)
$F0_{max}$ [Hz]	Fol	385.81 (33.24)	Attr	395.37 (36.57)	Female	394.83 (36.17)
	Lut	401.59 (40.51)	Unattr	392.04 (34.88)	Male	392.58 (34.85)
Centre of gravity [Hz]	Fol	710.03 (286.57)	Attr	692.90 (249.08)	Female	687.96 (242.36)
	Lut	664.89 (214.70)	Unattr	682.03 (240.85)	Male	686.96 (247.79)
F1 [Hz]	Fol	681.71 (36.56)	Attr	682.08 (32.00)	Female	681.08 (31.96)
	Lut	679.30 (36.12)	Unattr	678.94 (32.60)	Male	679.93 (32.61)
F2 [Hz]	Fol	1855.07 (51.74)	Attr	1855.29 (47.79)	Female	1855.56 (47.79)
	Lut	1853.03 (46.38)	Unattr	1852.81 (48.09)	Male	1852.55 (48.11)
F3 [Hz]	Fol	2872.12 (62.45)	Attr	2871.26 (58.27)	Female	2870.24 (57.69)
	Lut	2867.96 (58.02)	Unattr	2868.81 (56.78)	Male	2869.84 (57.39)
F4 [Hz]	Fol	3941.56 (52.81)	Attr	3939.70 (48.96)	Female	3940.12 (50.67)
	Lut	3937.21 (55.45)	Unattr	3939.06 (50.89)	Male	3938.64 (49.21)
HNR [dB]	Fol	13.00 (1.72)	Attr	13.05 (1.84)	Female	13.11 (1.77)
	Lut	13.20 (2.02)	Unattr	13.15 (1.81)	Male	13.09 (1.87)
Jitter [%]	Fol	0.03 (0.01)	Attr	0.03 (0.01)	Female	0.03 (0.01)
	Lut	0.03 (0.01)	Unattr	0.03 (0.01)	Male	0.03 (0.01)
Shimmer [%]	Fol	0.09 (0.01)	Attr	0.09 (0.01)	Female	0.09 (0.01)
	Lut	0.09 (0.01)	Unattr	0.09 (0.01)	Male	0.09 (0.01)
Intensity SD [dB]	Fol	11.04 (1.28)	Attr	10.99 (1.08)	Female	10.98 (1.09)
	Lut	10.94 (1.11)	Unattr	10.98 (1.10)	Male	10.99 (1.08)

"fol" = participant in the late follicular phase, "lut" = participant in the luteal phase, "attr" = stimulus speaker with an attractive voice, "unattr" = stimulus speaker with an unattractive voice.

Hormone assays

Hormone levels of the participants during the late follicular and luteal phase are shown in Table 6. A Kolmogorov–Smirnov test indicated that hormonal data were not normally distributed. Hence, nonparametric Wilcoxon signed-rank tests were used to compare the hormone levels between both cycle phases. These analyses revealed that, as expected, progesterone levels were significantly higher in the luteal phase than in the late follicular phase ($Z = -3.748$, $p < 0.001$). Levels of estradiol ($Z = -1.389$, $p = 0.17$), testosterone ($Z = -0.312$, $p = 0.76$), and cortisol ($Z = -0.772$, $p = 0.44$), however, did not differ between the two phases.

We found no significant correlation between hormones and vocal parameters during the late follicular phase, neither in the baseline condition (reading sentences out loud) nor in the treatment condition (repeating spoken sentences). During the luteal phase, however, estradiol levels were negatively correlated with mean F0 values and HNR, and positively correlated with jitter and shimmer. In the treatment condition, testosterone levels were negatively correlated with F1, F2, F3 and F4. Cortisol levels were negatively correlated with $F0_{min}$ and F1. Estradiol levels were positively correlated with jitter (see correlation matrices in the Supplementary material).

Discussion

The aim of the present study was to investigate whether and how women's voices change during the menstrual cycle when responding to female or male speakers with attractive or unattractive voices. For this purpose, the voice of naturally cycling women was recorded during the late follicular phase and during the luteal phase while speaking sentences in response to a stimulus speaker (experimental condition) and when reading aloud the sentences from the computer screen (control condition). Based on earlier studies, we expected that the menstrual cycle of the participants would have an influence on

TABLE 6 Hormone levels in the two cycle phases of the participants.

	Estradiol (pg/ml)	Progesterone (pg/ml)	Testosterone (pg/ml)	Cortisol (nmol/l)
Late follicular (*M*, *SD*)	3.47 (1.49)	37.44 (70.94)	10.97 (4.52)	8.21 (5.51)
Luteal phase (*M*, *SD*)	4.23 (1.98)	110.36 (100.55)	11.69 (6.22)	7.02 (4.88)
Wilcoxon signed-rank	$p = 0.17$	$\mathbf{p < 0.001}$	$p = 0.76$	$p = 0.44$

Bold values indicate significant effects.

women's voices. We also hypothesized that women would alter their voices depending on whether they were responding to male (vs. female) stimulus speakers with attractive (vs. unattractive) voices. Phonetic analyses confirmed these predictions in part. In the experimental condition, some vocal parameters of women's voices were indeed affected by their current menstrual cycle phase (but only $F0_{max}$ when correcting for multiple testing), and by the vocal attractiveness (only Center of Gravity and F1 when correcting for multiple testing) and sex of the stimulus speaker (only F2, when correcting for multiple testing). We observed no interaction between cycle phase and stimulus attractiveness, suggesting that women in the fertile phase did not react specifically to attractive voices. By contrast, in the control condition, in which women merely read out sentences off a computer screen, we observed no effects of menstrual cycle on women's voices.

In the experimental condition, where women "responded" to recordings of other speakers by repeating spoken sentences, we found some evidence for an effect of the current menstrual cycle phase on women's vocal characteristics. In the control condition however, where the sentences had to be read aloud without any external vocal input, phonetic analyses showed no effect of participants' current menstrual cycle phase. While this is inconsistent with some studies which found a cycle effect on women's voices (Banai, 2017; Bryant and Haselton, 2009; Fischer et al., 2011; Tatar et al., 2015), it is in line with other studies that failed to find a menstrual cycle effect (Barnes and Latman, 2011; Celik et al., 2013; Meurer et al., 2009; Raj et al., 2010). The fact that we only found menstrual cycle effects when women responded to stimulus speakers but not when reading sentences out loud could contribute in part to understanding such inconsistencies across previous published studies and supports the notion that cycle-dependent voice changes need a social trigger to unfold.

When responding to stimulus speakers with attractive voices, women spoke with a breathier and hoarser voice, characterized by lower HNR (note that this effect just failed to reach statistical significance when correcting for multiple testing). Breathiness has been argued to be a feminine trait and is related to desirability in women (Henton and Bladon, 1985). We also observed heightened formant frequencies (in Formants 1–3, only in Formant 1 when correcting for multiple testing), and a higher Centre of gravity when the women responded to more attractive stimulus speakers compared to relatively unattractive stimulus speakers. Centre of gravity is the frequency which divides the voice spectrum into two halves, so a higher Centre of gravity when responding to attractive stimulus speakers means that women's voices had more high-frequency energy compared to when responding to unattractive stimulus speakers. Of relevance, higher-frequency female voices have been reported to be more attractive than female voices with lower frequencies (Collins and Missing, 2003; Jones et al., 2010). Taken together, these findings suggest that women tried to make their voices sound more desirable and more attractive when speaking to stimulus speakers with attractive voices. In case of male targets, this could be a sign of increased mating motivation, whereas in case of female targets it could serve competitive needs in order to sound more attractive than a rival.

Notably, we did not observe a systematic variation in mean F0, suggesting that women did not generally speak in a higher (Fraccaro et al., 2011) or lower voice pitch (Hughes et al., 2010) when speaking to stimulus speakers with attractive voices. In contrast to Fraccaro et al. (2011) and Hughes et al. (2010), who presented the speakers with photographs of people they were allegedly speaking to, we asked our participants to respond to more or less attractive male and female voice recordings. Our finding might either relate to the inconsistency of the published findings discussed above, or alternatively suggests that the vocal channel alone may not be sufficient to evoke an effect.

With regard to sex of the stimulus speakers, women responded with a higher-frequency Formant 2 to female speakers than to male speakers. Phonetic analysis revealed that this effect might be the result of social mimicry or accommodation, as the speakers were mirroring characteristics of the stimulus speakers (Chartrand and Lakin, 2013; Gregory and Webster, 1996).

The observation that a cycle effect occurred in the experimental condition (responding to stimulus speakers, social context) but not in the control condition (reading sentences aloud, no social context) does not support the assumption that hormone-driven changes in laryngeal mucus and vocal folds are responsible for cycle dependent vocal changes (Abitbol et al., 1999). Furthermore, according to Abitbol et al. (1999), we would expect higher vocal frequencies during the follicular phase. Instead, we observed a higher $F0_{max}$ in the luteal phase than in the follicular phase. Overall, our findings suggest that an effect of the menstrual cycle on a woman's voice does not occur by default and as a result of hormone-driven biological inevitabilities, but instead needs a social trigger to unfold. This interpretation corresponds to the findings of Bryant and Haselton (2009) who found an effect of menstrual cycle only when women spoke a social sentence, not vowels, further suggesting that cycle-dependent voice changes may occur during social communication only. Likewise, Karthikeyan and Locke (2015) supposed that only when women are motivated to speak and behave attractively, subtle cycle-dependent voice changes may occur. Accordingly, Klatt et al. (2020) found an effect of menstrual cycle phase only if the women uttered sentences with a social content, but not when they spoke neutral sentences.

In previous studies, different software applications and different predefined frequency ranges have been used in phonetic analyses of human voices, which is problematic in terms of comparability of the results (see Table 1). By using different presettings in frequency range, we demonstrate that the chosen frequency range has an effect on phonetic raw data. We repeated phonetic analyses with the default presettings of Praat software (75–600 Hz instead of 100–500 Hz), as did for example Hughes et al. (2010). The Praat output was

significantly different in mean F0, F0 SD, $F0_{min}$, $F0_{max}$, HNR, and Shimmer, depending on recording condition. Given these results, it is possible that the inconsistent results of previous studies on female voices partially trace back to the variety of phonetic analysis software and different frequency ranges which have been used. Voice changes during the menstrual cycle are subtle and therefore difficult to detect. In order to increase the comparability of future studies, we recommend that researchers develop unified standards for phonetic analysis. Specifically, we recommend using Praat software with the respective settings for male and female voices suggested by the Praat developers.

This study has some limitations. First, mostly Swiss-German speaking individuals with different regional dialects were asked to speak standard German during recording. Potentially this may have resulted in moderately elevated stress for the speakers, making them feeling slightly uncomfortable, and this could have affected their voices over and above any effect of their current menstrual cycle phase or the stimulus speakers' voices. Second, reproducing the same sentences multiple times may lead to spontaneously occurring variation during speech since a speaker does not pronounce a sentence in exactly the same way every time (Fitch, 1990). This spontaneous variation might have interfered with voice changes evoked by the stimulus speaker. Thirdly, and most obviously, the participants were asked to repeat sentences spoken by stimulus speakers without any real interaction taking place. Although we opted for the present experimental design because allowing natural oral interaction between participants would have significantly confined experimental control, it is plausible that stronger social mimicry effects can be observed in a real social interaction situation. Finally, this was a complex study that demanded enormous commitment from the study participants. It is therefore understandable that some women did not take part until the end. Unfortunately, we no longer have access to the full set of data from the excluded participants, as in some cases, the ethical protocol demanded us to delete them. However, as the women left the study at different stages and for different reasons, we assume that there is no systematic reason for their quitting the study.

In the present study, we took great care in scheduling the cycle-dependent recording sessions and in standardisation of the recording procedure which has sometimes been neglected in previous studies. In the present study, the fertile window of the speakers was determined using LH tests and confirmed with hormone assays of saliva. We minimized potential confounding variables such as irregularity of menstrual cycle, mother tongue, smoking, respiratory diseases, time of the day, and background noise.

Taken together, the present study offers additional evidence that women's voices subtly change depending on their current menstrual cycle phase and depending on the attractiveness and sex of the stimulus speaker. Importantly, an effect of menstrual cycle was only found when responding to stimulus speakers and not when reading the sentences aloud, suggesting that cycle-dependent voice changes need a social trigger to unfold.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethics commission of the Faculty of Human Sciences, University of Bern. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JL: Data curation, Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. WK: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. SS: Writing – review & editing, Writing – original draft, Resources, Methodology, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by SNF Project no. PP00P1_139072.

Acknowledgments

The authors wish to thank Sabrina Beeler, Vera Bergamaschi, Gina Camenzind, Emilia da Costa, Fion Emmenegger, Rahel Gfeller, Nik Hunziker, Claudia Ramseier, and Arjeta Velii for their assistance at various stages of this project. We also thank Pavel Šebesta for his help with the phonetic analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1401158/full#supplementary-material

References

Abitbol, J., Abitbol, P., and Abitbol, B. (1999). Sex hormones and the female voice. *J. Voice* 13, 424–446. doi: 10.1016/s0892-1997(99)80048-4

Audacity Team. (2023). Audacity (R): Free audio editor and recorder [Computer program]. Version 3.3. Available at: https://www.audacityteam.org/ (Accessed May 06, 2023).

Banai, I. P. (2017). Voice in different phases of menstrual cycle among naturally cycling women and users of hormonal contraceptives. *PLoS One* 12:e0183462. doi: 10.1371/journal.pone.0183462

Barnes, L., and Latman, N. (2011). Acoustic measure of hormone affect on female voice during menstruation. *Int. J. Humanit. Soc. Sci.* 1, 5–10.

Boersma, P., and Weenink, D. (2023). Available at: https://www.fon.hum.uva.nl/praat/ (last accessed on 03.09.2023).

Broesch, T. L., and Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *J. Cogn. Dev.* 16, 31–43. doi: 10.1080/15248372.2013.833923

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., et al. (2010). Vocal attractiveness increases by averaging. *Curr. Biol.* 20, 116–120. doi: 10.1016/j.cub.2009.11.034

Bryant, G. A., and Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biol. Lett.* 5, 12–15. doi: 10.1098/rsbl.2008.0507

Celik, Ö., Celik, A., Atespare, A., Boyaci, Z., Celebi, S., Gündüz, T., et al. (2013). Voice and speech changes in various phases of menstrual cycle. *J. Voice* 27, 622–626. doi: 10.1016/j.jvoice.2013.02.006

Chartrand, T. L., and Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annu. Rev. Psychol.* 64, 285–308. doi: 10.1146/annurev-psych-113011-143754

Collins, S. A., and Missing, C. (2003). Vocal and visual attractiveness are related in women. *Anim. Behav.* 65, 997–1004. doi: 10.1006/anbe.2003.2123

Dabbs, J. M., and Delarue, D. (1991). Salivary testosterone measurements among women—relative magnitude of circadian and menstrual cycles. *Horm. Res.* 35, 182–184. doi: 10.1159/000181899

Farley, S. D., Hughes, S. M., and LaFayette, J. N. (2013). People will know we are in love: evidence of differences between vocal samples directed toward lovers and friends. *J. Nonverbal Behav.* 37, 123–138. doi: 10.1007/s10919-013-0151-3

Feinberg, D. R., DeBruine, L. M., Jones, B. C., and Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Percept.* 37, 615–623. doi: 10.1068/p5514

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Smith, M. J. L., Cornwell, R. E., et al. (2005). The voice and face of woman: One ornament that signals quality?. *Evol. Hum. Behav.* 26, 398–408. doi: 10.1016/j.evolhumbehav.2005.04.001

Feinberg, D. R., Jones, B. C., Smith, M. L., Moore, F. R., DeBruine, L. M., Cornwell, R. E., et al. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Horm Behav.* 49, 215–222. doi: 10.1016/j.yhbeh.2005.07.004

Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113. doi: 10.1037/0012-1649.20.1.104

Fischer, J., Semple, S., Fickenscher, G., Juergens, R., Kruse, E., Heistermann, M., et al. (2011). Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime. *PLoS One* 6:e24490. doi: 10.1371/journal.pone.0024490

Fitch, J. L. (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *J. Speech Hear. Disord.* 55, 360–363. doi: 10.1044/jshd.5502.360

Fraccaro, P. J., Jones, B. C., Vukovic, J., Smith, F. G., Watkins, C. D., Feinberg, D. R., et al. (2011). Experimental evidence that women speak in a higher voice pitch to men they find attractive. *J. Evol. Psychol.* 9, 57–67. doi: 10.1556/JEP.9.2011.33.1

Gao, W., Stalder, T., and 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. doi: 10.1016/j.talanta.2015.05.004

Gregory, S. W., and Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *J. Pers. Soc. Psychol.* 70, 1231–1240. doi: 10.1037/0022-3514.70.6.1231

Grieser, D. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol.* 24, 14–20. doi: 10.1037/0012-1649.24.1.14

Haselton, M. G., and Gildersleeve, K. (2016). Human ovulation cues. *Curr. Opin. Psychol.* 7, 120–125. doi: 10.1016/j.copsyc.2015.08.020

Henton, C. G., and Bladon, R. A. W. (1985). Breathiness in Normal female speech—inefficiency versus desirability. *Lang. Commun.* 5, 221–227. doi: 10.1016/0271-5309(85)90012-6

Hughes, S. M., Farley, S. D., and Rhodes, B. C. (2010). Vocal and physiological changes in response to the physical attractiveness of conversational partners. *J. Nonverbal Behav.* 38, 107–127. doi: 10.1007/s10919-013-0163-z

Huynh, H., and Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *J. Educ. Behav. Stat.* 1, 69–82. doi: 10.3102/10769986001001069

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., and Vukovic, J. (2010). A domain-specific opposite-sex biasin human preferences for manipulated voice pitch. *Anim. Behav.* 79, 57–62. doi: 10.1016/j.anbehav.2009.10.003

Karthikeyan, S., and Locke, J. L. (2015). Men's evaluation of women's speech in a simulated dating context: effects of female fertility on vocal pitch and attractiveness. *Evol. Behav. Sci.* 9, 55–67. doi: 10.1037/ebs0000014

Klatt, W. K., Mayer, B., and Lobmaier, J. S. (2020). Content matters: cyclic effects on women's voices depend on social context. *Horm. Behav.* 122:104762. doi: 10.1016/j.yhbeh.2020.104762

La, F. M. B., and Polo, N. (2020). Fundamental frequency variations across the menstrual cycle and the use of an Oral contraceptive pill. *J. Speech Lang. Hear. Res.* 63, 1033–1043. doi: 10.1044/2020_JSLHR-19-00277

Leongómez, J. D., Binter, J., Kubicová, L., Stolarová, P., Klapilová, K., Havlicek, J., et al. (2014). Vocal modulation during courtship increases proceptivity even in naive listeners. *Evol. Hum. Behav.* 35, 489–496. doi: 10.1016/j.evolhumbehav.2014.06.008

Meurer, E. M., Garcez, V., von Eye Corleta, H., and Capp, E. (2009). Menstrual cycle influences on voice and speech in adolescent females. *J. Voice* 23, 109–113. doi: 10.1016/j.jvoice.2007.03.001

Nacci, A., Fattori, B., Basolo, F., Filice, M. E., De Jeso, K., Giovanni, L., et al. (2011). Sex hormone receptors in vocal fold tissue: a theory about the influence of sex hormones in the larynx. *Folia Phoniatr. Logop.* 63, 77–82. doi: 10.1159/000316136

Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., and Sorokowska, A. (2017). Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychon. Bull. Rev.* 24, 856–862. doi: 10.3758/s13423-016-1146-y

Pavela Banai, I., Burriss, R. P., and Šimić, N. (2022). Voice changes across the menstrual cycle in response to masculinized and feminized man and woman. *Adapt. Hum. Behav. Physiol.* 8, 238–262. doi: 10.1007/s40750-022-00190-y

Peirce, J. W. (2007). PsychoPy - psychophysics software in python. *J. Neurosci. Methods.* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Pipitone, R. N., and Gallup, G. G. Jr. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evol. Hum. Behav.* 29, 268–274. doi: 10.1016/j.evolhumbehav.2008.02.001

Plexico, L. W., Sandage, M. J., Kluess, H. A., Franco-Watkins, A. M., and Neidert, L. E. (2020). Blood plasma hormone-level influence on vocal function. *J. Speech Lang. Hear. Res.* 63, 1376–1386. doi: 10.1044/2020_JSLHR-19-00224

Puts, D. A., Bailey, D. H., Cardenas, R. A., Burriss, R. P., Welling, L. L. M., Wheatley, J. R., et al. (2013). Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle. *Horm. Behav.* 63, 13–19. doi: 10.1016/j.yhbeh.2012.11.007

Raj, A., Gupta, B., Chowdhury, A., and Chadha, S. (2010). A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *J. Voice* 24, 363–368. doi: 10.1016/j.jvoice.2008.10.005

Raphael, L. J., Borden, G. J., and Harris, K. S. (2011). Speech science primer. Philadelphia, PA: Lippincott Williams and Wilkins.

Schneider, B., Cohen, E., Stani, J., Kolbus, A., and Rudas, M. (2007). Towards the expression of sex hormone receptors in the human vocal fold. *J. Voice* 21, 502–507. doi: 10.1016/j.jvoice.2006.01.002

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., and Zäske, R. (2014). Speaker perception. *WIREs Cogn. Sci.* 5, 15–25. doi: 10.1002/wcs.1261

Shoup-Knox, M. L., Ostrander, G. M., Reimann, G. E., and Pipitone, R. N. (2019). Fertility-dependent acoustic variation in Women's voices previously shown to affect listener physiology and perception. *Evol. Psychol.* 17:1474704919843103. doi: 10.1177/1474704919843103

Shoup-Knox, M. L., and Pipitone, R. N. (2015). Physiological changes in response to hearing female voices recorded at high fertility. *Physiol. Behav.* 139, 386–392. doi: 10.1016/j.physbeh.2014.11.028

Tatar, E. C., Sahin, M., Demiral, D., Bayir, O., Saylam, G., Ozdek, A., et al. (2015). Normative values of voice analysis parameters with respect to menstrual cycle in healthy adult Turkish women. *J. Voice* 30, 322–328. doi: 10.1016/j.jvoice.2015.04.014

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *J. Acoust. Soc. Am.* 85, 901–906. doi: 10.1121/1.397562

Wust, S., Wolf, J., Hellhammer, D. H., Federenko, I., Schommer, N., and Kirschbaum, C. (2000). The cortisol awakening response—normal values and confounds. *Noise Health* 2, 79–88.

Zaske, R., Skuk, V. G., Golle, J., and Schweinberger, S. R. (2020a). The Jena speaker set (JESS)-a database of voice stimuli from unfamiliar young and old adult speakers. *Behav. Res. Methods* 52, 990–1007. doi: 10.3758/s13428-019-01296-0

Zaske, R., Skuk, V. G., and Schweinberger, S. R. (2020b). Attractiveness and distinctiveness between speakers' voices in naturalistic speech and their faces are uncorrelated. *R. Soc. Open Sci.* 7:201244. doi: 10.1098/rsos.201244

Zraick, R. I., Gentry, M. A., Smith-Olinde, L., and Gregg, B. A. (2006). The effect of speaking context on elicitation of habitual pitch. *J. Voice* 20, 545–554. doi: 10.1016/j.jvoice.2005.08.008

Zraick, R. I., Skaggs, S. D., and Montague, J. C. (2000). The effect of task on determination of habitual pitch. *J. Voice* 14, 484–489. doi: 10.1016/s0892-1997(00)80005-3

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 02 April 2025
DOI 10.3389/fpsyg.2025.1531021

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Ray Garza,
Texas A&M International University, United States
Cristina O. Mosso,
University of Turin, Italy
David Puts,
The Pennsylvania State University (PSU), United States

*CORRESPONDENCE
Philippa Hüpen
✉ rhuepen@ukaachen.de

RECEIVED 19 November 2024
ACCEPTED 19 March 2025
PUBLISHED 02 April 2025

CITATION
Friedrich S, Brodkin ES, Derntl B, Habel U and Hüpen P (2025) Assessing the association between menstrual cycle phase and voice-gender categorization: no robust evidence for an association.
Front. Psychol. 16:1531021.
doi: 10.3389/fpsyg.2025.1531021

Assessing the association between menstrual cycle phase and voice-gender categorization: no robust evidence for an association

Sarah Friedrich[1], Edward S. Brodkin[2], Birgit Derntl[3,4], Ute Habel[1,5] and Philippa Hüpen[1,5]*

[1]Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University Hospital, Aachen, Germany, [2]Department of Psychiatry, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, United States, [3]Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health (TüCMH), University of Tübingen, Tübingen, Germany, [4]LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany, [5]Institute of Neuroscience and Medicine, JARA-Institute Brain Structure Function Relationship (INM 10), Research Center Jülich, Jülich, Germany

Introduction: Hormone fluctuations during the menstrual cycle are known to influence a wide variety of cognitive-emotional processes and behavior. Mate choice and changes in attractiveness ratings for faces and voices are often investigated in this context, but research on changes in voice-gender perception independent of attractiveness ratings is rare even though the voice is an essential element in social interactions. For this reason, we investigated the influence of cycle phase and levels of estrogen and progesterone on performance in a voice-gender categorization task. Our expectation was to find a more pronounced other-sex effect, so faster and more accurate reactions for masculine voices, in the follicular (fertile) phase than in the luteal phase.

Methods: We measured 65 healthy, naturally-cycling women, half of them in the follicular phase and the other half in the luteal phase. For the analyses, we used signal detection theory (SDT) measures in addition to reaction times and percent of correct reactions. The study was preregistered after measuring the first 33 participants and prior to any data analyses (https://osf.io/dteyn).

Results: Cycle phase and hormone levels showed no significant effect on reaction time or SDT measures. This was the case both using frequentist analyses and Bayesian statistics. Reaction time was influenced by voice-gender, with faster reactions for feminine voices compared to masculine voices in both cycle phases.

Discussion: Taken together, our results add to the increasing number of studies that do not find an interaction of menstrual cycle phase and reaction to gendered stimuli.

KEYWORDS

menstrual cycle, voice-gender categorization, estradiol, progesterone, signal detection theory, mating cues, follicular phase, luteal phase

1 Introduction

The menstrual cycle, and associated changes in hormonal levels, have been shown to influence a variety of human functions, including cognition, emotion, physiology, and brain activity (Albert et al., 2015; Derntl et al., 2008; Derntl et al., 2013; Haraguchi et al., 2021; Hidalgo-Lopez et al., 2020; Pletzer et al., 2019). The primary hormonal fluctuations throughout the menstrual cycle involve changes in estradiol and progesterone levels. Both hormone levels are low during menses and the early follicular phase. In the later follicular phase, estradiol rises, peaking before ovulation which characterizes high fertility. After ovulation, the luteal phase begins, characterized by a rise in progesterone, which peaks mid-luteal phase, accompanied by a second estradiol increase. Finally, concentrations of both hormones begin to decline to their lowest levels during menses (Farage et al., 2008).

Associated with these hormonal fluctuations, changes in social behavior (Maner and Miller, 2014; Anderl et al., 2015) and mate selection (Puts et al., 2012) can be observed. For example, during the fertile phase of the cycle, women tend to prefer more masculine partners (Penton-Voak et al., 1999; Penton-Voak and Perrett, 2000) whereas during the nonfertile phase of the cycle, masculinity seems to be less important. Especially for faces, studies suggest a menstrual cycle related change in women's preference with a higher preference for more masculine faces during the late follicular phase and during days prior to and directly after ovulation (Penton-Voak and Perrett, 2000; Johnston et al., 2001; Little et al., 2008). Evolutionary explanations suggest that higher levels of testosterone, which are associated with greater virility (Penton-Voak and Chen, 2004), may be associated with healthier offspring (Jones et al., 2008). This idea is supported by studies finding robust cycle shift effects for evaluation of potential short-term partners as opposed to only small or nul effects on choice of ponential long-term partners (Little et al., 2002; for an extensive review on cycle-shifts for attraction ratings see e.g., Gildersleeve et al., 2014 and Jones et al., 2008). Overall, the gender-categorization of human faces seems to be influenced by a variety of biological factors such as the viewers own gender role and sexual orientation (Luther et al., 2021), their age and experience with human faces (Hillairet de Boisferon et al., 2019) as well as previously activated categorial knowledge (Macrae et al., 2002).

Challenging the idea of a menstrual cycle-shift, a growing body of literature does not find a clear association between cycle phase and preferences for masculinized faces (Jones et al., 2018; Marcinkowska et al., 2016; Peters et al., 2009). Inconsistent findings may be related to methodological shortcomings such as inconsistent methods even within the same lab (Harris et al., 2013), low statistical power, and a lack of objective measures of cycle phase (Jones et al., 2019; Lewis, 2020). Another explanation for inconsistent findings is changes in participants' visual processing (Garza and Byrd-Craven, 2019) and their visual discrimination abilities over the cycle. These changes have been demonstrated in increased visual sensitivity (Lewis, 2020; Parlee, 1983) and an increased ability to identify facial symmetry during fertile cycle phases (Lewis, 2017). Facial symmetry is generally interpreted as a sign of advantageous genetic traits and health (Fink and Penton-Voak, 2002; Foo et al., 2017), associated with a strong preference for symmetrical faces, independent of conscious detection (Little and Jones, 2006) thus substantiating the idea of evolutionary mating strategies exerting a strong influence on face perception and preference. Subconscious changes in the ability to identify facial symmetry or asymmetry may contribute to variations in face preference over the cycle. These hormonal variations not only affect preference and perception of faces, but also the appearance and attractiveness of women's own faces. Through subtle changes in shape and skin structure, female faces are perceived as more attractive during the fertile phase (Bobst and Lobmaier, 2012; Puts et al., 2013).

While much of the literature focuses on changes in face preference during the menstrual cycle, less attention has been paid to other mating cues. One such cue is the human voice, which, like facial features, plays a key role for both mating (Hughes and Puts, 2021; Pisanski et al., 2018) and other social interactions (Guldner et al., 2020; Hellbernd and Sammler, 2016; McGettigan, 2015). The voice conveys pertinent characteristics which allow us not only to identify known voices, but also to characterize a stranger's age, gender (Gallup and Frederick, 2010) or even health status (Arnocky et al., 2018). To do so, people rely mainly on two properties of the voice: the fundamental frequency (F0) and formant frequencies (Hillenbrand and Clark, 2009). The F0 is the average rate of vibration of the vocal folds per second and is closely related to the perceived overall pitch of a voice, while the formant frequencies are the result of the movement of the vocal apparatus during formatting vowels and consonants (Goldstein, 2014).

For women's own voices, a robust association between menstrual cycle phase and voice quality can be seen. During phases of lower estradiol, the voice quality decreases in naturally cycling women, showing higher tension, roughness and instability (Arruda et al., 2019; Raj et al., 2010). Additionally, changes in F0 have been reported across the cycle, though with inconsistent direction (Bryant and Haselton, 2009; Fischer et al., 2011; Karthikeyan and Locke, 2015; Lã and Polo, 2020). However, research findings on changes in voice preference for male voices associated with the menstrual cycle are mixed. Whereas some studies find a clear inclination of women to more masculine voices during fertile cycle phases (Feinberg et al., 2006; Puts, 2005), other studies fail to find an effect of hormonal fluctuations on voice preferences (Jünger et al., 2018). These mixed results on the influence of the menstrual cycle on voice preference parallel the mixed results on face preference and perception.

However, studies on changes in voice perception over the cycle are still scarce. Nonetheless, this is an important factor in understanding the underlying mechanisms of potential preference changes, as illustrated by the previously described influences on face perception. Therefore, the goal of this study is to shed light on potential differences between cycle phases in voice-gender categorization as an important part of social interactions, where not only visual facial cues but also vocal information has to be integrated into a multisensory perception that guides behavior.

According to previous studies, similar to the perceived gender of faces, voice-gender categorization is influenced by both listeners' gender (Junger et al., 2013; Smith et al., 2018) and sexual orientation (Smith et al., 2019). One way to study voice-gender perception is via a voice-gender categorization paradigm which uses words spoken by both natural male and female speakers alongside with voices morphed toward the opposite sex to investigate a person's reaction to increasingly ambiguous stimuli. This kind of paradigm allows for the analysis of both the accuracy of responses and the response bias, meaning the inclination to a certain response in ambiguous situations. Overall, the aforementioned studies found an opposite sex effect for

response accuracy in highly ambiguous trials together with a tendency for a response bias toward the opposite sex in those trials both heterosexual men and women (Smith et al., 2018; Junger et al., 2013; Junger et al., 2014). For both homosexual men and women however, response bias in ambiguous trials show a pattern more similar to heterosexual men, underlining an association of sexual orientation (Smith et al., 2019). Strikingly, the effect of listeners' own sex on voice-gender categorization seems to be less robust in women than in men, as response patterns across studies show greater variance for women than they do for men. This may be related to hormonal fluctuations associated with the menstrual cycle. In fact, previous studies on voice-gender categorization have not taken the menstrual cycle phase into account (Smith et al., 2018).

Due to the high relevance of voices as cues for social interaction and mating and because of the influence of female sex hormones on mating cues, we first expected to find a behavioral difference between women in different cycle phases. Because the late follicular phase is associated with higher fertility and a greater preference for masculinity in heterosexual women, we expected a stronger other-sex effect in that cycle phase for response accuracy and reaction time. Secondly, we expected a difference in response bias, meaning the inclination to categorize a voice rather as masculine or feminine when the categorization is unclear. Specifically, we expected the response bias to be influenced by the estimated cost of a wrong decision. According to error management theory (Haselton and Buss, 2000), the response bias can be expected to be influenced by the estimated costs of incorrect decisions. The favored decision should be the one that results in the less costly error, so the smallest loss of resources, if the decision was wrong. Applied to female mate choice, it is unclear if the cost is higher for mistaking a male for a female and thereby missing a potential mate or if it is higher for mistaking a female for a male and thereby investing in a non-reproductive mate (Johnston et al., 2008). Thus, this is a more exploratory question, and we do not have prior assumptions for the direction of the difference. The influence of choice costs is expected to be more pronounced in the follicular phase since mating is more likely to result in offspring. Hence, we expect a stronger response bias in the follicular compared to the luteal phase.

2 Methods

All procedures were in accordance with the Declaration of Helsinki and were approved by the Independent Ethics Committee of the RWTH Aachen Faculty of Medicine. All participants gave written informed consent and received financial compensation of 10 €.

2.1 Participants

A total of 78 naturally cycling cisgender heterosexual women between 18 and 35 years ($M = 25.48$, $SD = 4.11$) participated in the study. Thirty-three of these datasets were collected in the context of an earlier study (unpublished data) but have not been analyzed before. The required sample size was 62 as calculated *a priori* using G*Power 3.1 (Faul et al., 2007). Based on the mixed results on associations of voice perception and cycle phase, as described in the introduction, we expected a small effect (Cohen's $f = 0.15$) with a power of $\beta = 0.80$ and an α error probability of 0.05. In the context of mixed results and varying degrees of uncertainty, sample sizes in studies on the influence of cycle phase on perception and reaction to cues with mate value vary quite substantially, ranging from 50 or less (Oinonen and Mazmanian, 2007; Sanders and Wenmoth, 1998; Rosenberg and Park, 2002) to 200 or more (Jones et al., 2018; Jünger et al., 2018; Stern et al., 2021) and yield mixed results independent of sample size. Therefore, we decided to base our sample on the basic power calculation described above.

Participants were recruited using public flyers and online postings. All participants reported a regular menstrual cycle and did not take any contraceptives. Participants were recruited to be either in the follicular ($N = 31$) or the luteal ($N = 34$) phase (as determined by self-reports and hormonal profiles; see below) at the time of measurement. Only women whose reported menstrual cycle phase matched the cycle phase measured by blood samples were included. Consequently, 13 participants were be excluded from all analyses, because their self-reported cycle phase differed from the cycle phase determined, making a clear classification impossible. Both progesterone ($t(63) = -8.25$, $p < 0.001$) and estradiol ($t(63) = -6.91$, $p = < 0.001$) levels were significantly different between the two groups. There was no significant difference in age ($t(63) = -1.45$ $p = 0.152$) or years of education ($t(62) = -1.29$, $p = 0.202$) between groups. Demographic information and hormonal levels *(M ± SD)* are presented in Table 1.

Prior to enrolment, each woman took part in a telephone interview to assess eligibility for the study and to assess the current day of the cycle. Exclusion criteria were hearing or speech impairment, use of oral contraceptives or other hormones, diseases or medications known to affect the endocrine system, pregnancy, or breastfeeding, and neurological or mental disorders. Physical illness, medication and pregnancy were assessed by self-report. The absence of mental disorders was assessed using the clinical version of the structured clinical interview for DSM-5 (SCID-5 CV; First et al., 2016). One participant had to be excluded due to an assumed presence of a mental illness based on the SCID interview. To control for a potential influence of sexual orientation, only heterosexual women were included in this study. Sexual orientation was assessed via self-report.

2.2 Cycle phase determination

To determine the cycle phase, participants were asked for the first day of their last menses during the telephone interview. To schedule the measurement date, they had to inform the study team via email as soon as their next menses started. If the time frame between both menses fell into a regular cycle (23–35 days), participants were randomly assigned to either the follicular phase group (7–11 days after

TABLE 1 Final sample characteristics for both cycle phases.

	Follicular phase ($n = 31$) $M \pm SD$	Luteal phase ($n = 34$) $M \pm SD$	p
Age (years)	24.71 ± 3.04	26.18 ± 4.83	0.202
Education (years)	15.50 ± 2.68	16.33 ± 2.48	0.152
β-estradiol (pg/ml)	55.20 ± 89.0	130.9 ± 55.0	<0.001
Progesterone (ng/ml)	0.142 ± 0.08	9.565 ± 6.35	<0.001

onset of the current menses) or the luteal phase group (17–34 days after onset of the current menses). Since the cycle phase is often inaccurately self-reported (Farrar et al., 2015), we confirmed the estimated cycle phase by assessing levels of progesterone (P) and estradiol (E) via blood serum samples, assessed using ElektroChemiLumineszenz-ImmunoAssays (ECLIAs). The reference range for estradiol was 20.5–233 pg./mL for the follicular and 30.2–305 pg./mL for the luteal phase. For progesterone, the reference range for the follicular phase was <0.05–0.323 ng/mL and for the luteal phase 0.537–20.9 ng/mL. Progesterone levels below the detection limit (0.05 ng/mL) were entered as half the detection limit (0.025 ng/mL). In our sample, this was the case for 4 women, all of them in the follicular phase. For higher reliability of cycle phase determination, people not involved in data collection or analyses rated the cycle phase for each woman based on hormonal levels according to reference ranges. Additional information about the procedure can be found in the Supplementary information 1. This combined approach of forward counting of cycle days and assessing the level of reproductive hormones allows us to substantially reduce the uncertainty of true cycle phase for each participant. This procedure enables detecting effects with a much smaller sample size than usually required in studies using counting methods alone (Gangestad et al., 2016; Jonge et al., 2019; Maki et al., 2002).

2.3 Procedure and paradigm

At the beginning of the session, each woman completed a short interview including questions on demographic data, current cycle phase and exclusion of hormone intake. This interview was followed by a screening version of the SCID-5 to exclude mental disorders, and a blood withdrawal (ca. 7 mL) to assess blood serum hormone levels. Due to practical reasons, we could not control for time of day for the blood withdrawal. Subsequently, participants completed a voice-gender categorization paradigm. An extensive description of the paradigm can be found in Junger et al. (2013). The stimuli consisted of 6 trisyllabic, neutral nouns, each spoken by 5 male and 5 female speakers. The resulting 60 words were each morphed 2, 4 and 6 semitones (st) toward the speaker's other sex by adjusting the pitch contour and the formant structure as a reflection of vocal tract length accordingly using the "change gender" function implemented in the software Praat Version 5.2.03 (Boersma and Weenink, 2010). These final 240 words were presented pseudorandomized in a way that no speaker and no word was presented consecutively. The presentation was divided into 80 blocks, each consisting of 3 words spoken by the same sex and morphed to the same degree. The stimuli were delivered via headphones using the software Presentation Version 21.1 (Neurobehavioral Systems, Inc., 2019). Participants were instructed to categorize the speaker's sex for each stimulus as male or female as fast as possible by pressing the number key "7" for male speakers and number key "8" for female speakers on a laptop keyboard.

2.4 Data analyses

Behavioral data were analyzed using Matlab2019a (MathWorks, Inc., 2019). Sociodemographics, group differences and correlations were analyzed using the software R version 4.1.2 (RStudio Team, 2021). The study has been preregistered after measuring the first 33 participants and prior to any data analyses.[1]

2.4.1 Between- and within-group behavioral differences

Reaction time differences were calculated using three mixed-model ANOVAs, with overall reaction time, reaction time for correct trials and reaction time for incorrect trials as dependent variables, respectively. The models contained the between-subjects factor cycle phase (follicular vs. luteal phase) and the two within-subject factors voice-gender (masculine vs. feminine) and morphing level (0, 2, 4, or 6 st morphing) as independent variables. For significant effects, post-hoc pairwise comparisons were calculated. All post-hoc comparisons were Bonferroni-corrected to account for multiple testing. Effect sizes were calculated using generalized eta squared.

The frequentist approach to statistics has been increasingly criticized (Jarosz and Wiley, 2014). Major points of criticism have been the arbitrariness of a *p*-value of 0.05 as a cut-off as well the influence of sampling and sample size on the p-value, which can lead to significant results, that are only valid within the given sample (for an extensive overview, refer to Wagenmakers, 2007). Therefore, we decided to validate effects of cycle phase using Bayesian ANOVAs to assess the likelihood for H_0 (cycle phase does not influence task performance) over H_1 (there is a significant difference between cycle phases) using the Bayes factor (BF_{01}). Conventions for interpreting the resulting BF are provided by Raftery (1995) and define a BF between 1 and 3 as weak, between 3 and 20 as positive, between 20 and 150 as strong and larger than 150 as very strong. Bayesian analyses were conducted using the BayesFactor package for R. For all models 10,000 iterations were run, participants were included as a random factor.

2.4.2 Signal detection theory

To allow for a more detailed investigation of the underlying mechanisms of potential performance differences, we employed signal detection theory (SDT) measures. The SDT is a well-established model for decision making processes (Lynn and Barrett, 2014; Stanislaw and Todorov, 1999). It was originally developed for signal vs. noise psychophysical perception tasks (Green and Swets, 1974) and differentiates between discriminability (i.e., ability to detect a target stimulus from background events) and a response bias (i.e., a tendency toward a certain response independent of the stimulus) (Stanislaw and Todorov, 1999). SDT measures have been employed in previous studies using the same voice-gender categorization paradigm and proved to be suitable for detecting differences in gender categorization ability (Junger et al., 2013; Smith et al., 2018; Smith et al., 2019; Junger et al., 2014). To calculate SDT measures for our paradigm, male voices were defined as target and female voices were defined as noise. This definition is arbitrary and chosen to match previous studies ability (Junger et al., 2013; Smith et al., 2018; Smith et al., 2019; Junger et al., 2014). This results in a definition of correct reactions to male voices as hits and incorrect reactions to male voices as misses. For female voices, correct reactions are defined as correct rejections, while incorrect reactions are defined as false alarms.

1 https://osf.io/dteyn

Since the Shapiro–Wilk test revealed deviances from normality of the data, we used non-parametrical measures. These measures were A′ for discriminability (Equation 1) and B''_D for response bias (Equation 2) and can be calculated using the following formulas (Pallier, 2002):

If hits > false alarms:
$$A' = \frac{1}{2} + \frac{(hit - fa) * (1 + hit - fa)}{4 * hit * (1 - fa)} \quad (1)$$

If false alarms > hits:
$$A' = \frac{1}{2} + \frac{(fa - hit) * (1 + fa - hit)}{4 * fa * (1 - hit)}$$

$$B''_D = \frac{(1 - hit) * (1 - fa) - hit * fa}{(1 - hit) * (1 - fa) + hit * fa} \quad (2)$$

A′ ranges from 0 to 1 with higher values indicating better discriminability (i.e., a high rate of correct reactions to both male and female speakers) and values near 0.5 indicating performance on chance level (Pallier, 2002). B''_D ranges from −1 to 1 with a value of 0 indicating no response bias, positive values indicating in this case a tendency to categorize a voice as feminine and negative values indicating a tendency to categorize a voice as masculine. As each trial must be categorized as either a hit/correct rejection or a miss/false alarm for SDT analysis, non-response trials were not considered.

To test for differences between groups and conditions, three mixed-model ANOVAs were performed with cycle phase (follicular vs. luteal phase) as between-subjects factor and morphing level (0, 2, 4, 6) as within-subject factor. A′ was the dependent variable for the first ANOVA, whereas B''_D was the dependent variable for the second. Additional to the SDT measures, a mixed-model ANOVA with percent of correct responses as dependent variable was calculated. While the normality assumption was not fulfilled, homoscedasticity was given, therefore the results of a mixed-model ANOVA can still be assumed to be robust despite the non-parametric distribution (Harwell et al., 1992). For all variables where Mauchly test indicated a lack of sphericity, degrees of freedom were corrected using Greenhouse–Geisser adjustment (Greenhouse and Geisser, 1959).

2.4.3 Signal detection theory and reaction times

One of the major drawbacks of the SDT is that it does not take RT into account. Therefore, we assessed potential associations between SDT values and RT in both cycle phases using correlations. The Shapiro–Wilk-Test indicated a lack of normal distribution for SDT values ($p > 0.05$), so Kendall's τ was chosen as a measure of correlation. SDT measures and mean reaction times for male and female speakers were correlated over all morphing levels. To control for multiple testing, Holm correction was used (Holm, 1979).

2.4.4 Influences of morphing, speaker, and hormone levels on performance

To allow us to look at hormonal influences and their interactions with morphing level and the voice-gender regardless of cycle phase, we ran multiple regressions with response bias B''_D and with RT as dependent variables and hormone levels (P and E) as independent variables. This approach gave us the possibility to include hormone levels as a continuous variable as regression analyses have a higher statistical power than ANOVAs. To identify influential predictor variables, we compared increasingly complex models by stepwise adding predictors and interactions. For both B''_D and RT we started with simple models including only variables as determined by the design of the paradigm: we started with morphing level, followed by voice-gender as a predictor for RT. Finally, hormone levels were added as further predictors. Additive models as well as interactions were tested. Since our previous analyses did not yield a significant difference between the two cycle groups, we did not include cycle phase as a factor. Model fits were compared based on Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. AIC and BIC are both calculated using a model's maximum likelihood estimate and correcting for number of model parameters (Vrieze, 2012). A major difference between both criteria is that BIC is growing more restrictive with an increasing number of parameters. Thus, BIC is more consistent, as long as the true model has a finite number of parameters and is one of the models, that are tested. In cases, where the model is more complex, the AIC is preferred (Vrieze, 2012). Due to the different calculations, we included both measures in our model selection. Additionally, we tested for significant differences between the models using Chi-Square test.

3 Results

3.1 Reaction time

For overall RT, we found significant main effects of voice-gender ($F(1, 63) = 44.18$, $p < 0.0001$, $\eta^2 = 0.023$) and morphing level ($F(3, 189) = 206.38$, $p < 0.0001$, $\eta^2 = 0.193$) as well as an interaction for both variables ($F(3, 189) = 39.96$, $p < 0.0001$, $\eta^2 = 0.022$), but no main effect for cycle phase ($F(1, 64) = 2.15$, $p = 0.148$). Post-hoc comparisons showed lower reaction times for feminine voices compared to male voices ($t(259) = -8.20$, $p < 0.001$). For morphing level, we found significant RT differences for each morphing level compared to another (all $p < 0.001$). Post-hoc tests for interaction effects of voice-gender and morphing level showed a significant influence of voice-gender on RT for each morphing level with faster RTs for feminine voices for 4 st ($F(1, 64) = 52.32$, $p < 0.0001$, $\eta^2 = 0.065$) and 6 st morphing ($F(1, 64) = 100.94$, $p < 0.0001$, $\eta^2 = 0.085$).

Comparable to overall RT, RT for correct trials also showed significant effects of voice-gender ($F(1, 63) = 38.67$, $p < 0.0001$, $\eta^2 = 0.027$) and morphing level ($F(3, 189) = 231.88$, $p < 0.0001$, $\eta^2 = 0.204$) as well as an interaction for both variables ($F(3, 189) = 27.1$, $p < 0.0001$, $\eta^2 = 0.018$). Again, post-hoc pairwise comparisons showed faster RT for feminine voices over all morphing levels ($t(259) = -8.21$, $p < 0.001$), significant differences between each morphing level (all $p < 0.001$) and a significant interaction of voice-gender and morphing level for 4 st ($F(1, 64) = 69.33$, $p < 0.001$, $\eta^2 = 0.082$) and 6 st ($F(1, 64) = 33.28$, $p < 0.001$, $\eta^2 = 0.061$) morphing (see Figure 1) with faster RT for feminine voices. We did not find any significant RT effects of incorrect reaction trials (all $p > 0.05$).

To substantiate the null effect for menstrual cycle phase, we conducted a Bayesian ANOVA. Results for overall RT ($BF_{01} = 1.975$), RT for correct responses ($BF_{01} = 2.038$) and RT for incorrect responses ($BF_{01} = 2.922$) supported the H_0, though only to a weak extent.

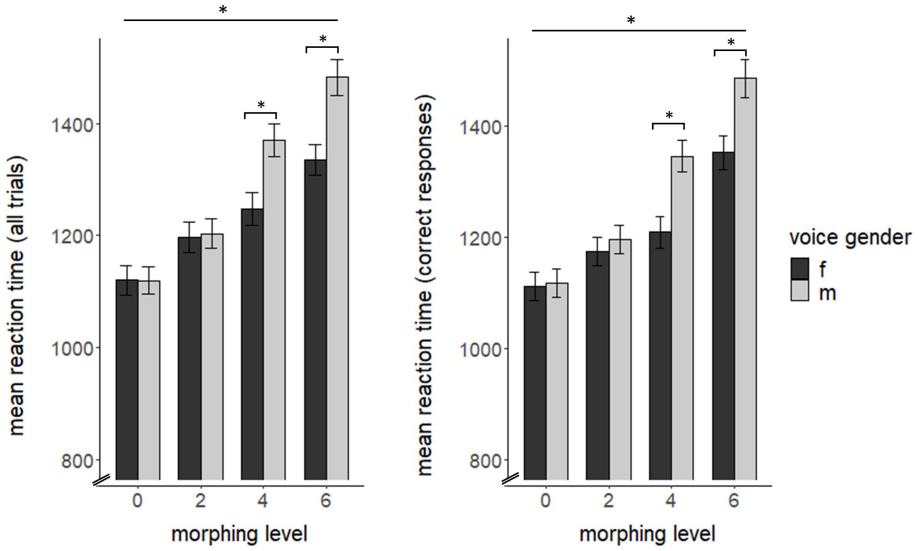


FIGURE 1
Reaction time as a function of morphing level and voice gender for all trials (left) and for correct reactions (right) across cycle phase. f = female speaker, m = male speaker. Reaction time is depicted in ms. The asterisk on the upper line indicates significant differences between all morphing levels. Results of the pairwise comparisons are listed in the Supplementary material.

3.2 Signal detection theory—discriminability and response bias

Results of the mixed-model ANOVA revealed a significant main effect of morphing level for both discriminability A′ (F(1.17, 73.65) = 460.9, $p < 0.001$, $\eta^2 = 0.829$) and response bias B''_D (F(2.27, 143.17) = 9.43, $p < 0.001$, $\eta^2 = 0.046$). Pairwise comparisons showed significant differences between all morphing levels (see Figure 2) with decreased values for A′ for higher morphing levels (all $p < 0.001$). Pairwise comparisons for B''_D showed differences for 0 compared to 2st morph (t(64) = 2.88, $p = 0.033$), 2st to 6st morph (t(64) = −5.03, $p < 0.001$) and 4 st to 6 st morph (t(64) = −5.35, $p < 0.001$). Results of the pairwise comparisons are listed in the Supplementary material.

There were no significant main effects of cycle phase ($F_{A'}$(1, 63) = 2.55, $p_{A'} = 0.115$, $\eta^2 = 0.013$; $F_{B''D}$ (1, 63) = 0.24, $p_{B''D} = 0.625$, $\eta^2 = 0.003$) and no interaction effects of cycle phase and morphing level ($F_{A'}$(1.17, 73.65) = 2.02, $p_{A'} = 0.157$, $\eta^2 = 0.021$; $F_{B''D}$ (2.27, 143.17) = 1.04, $p_{B''D} = 0.364$, $\eta^2 = 0.005$).

In line with that finding, Bayesian ANOVA results were in favor of the H_0 for both A′ ($BF_{01} = 5.263$) and $B_{''D}$ ($BF_{01} = 3.124$).

In line with SDT measures, for percent of correct answers we found a main effect for morphing level (F(1.87, 112.14) = 998.60, $p < 0.001$, $\eta^2 = 0.679$), but no effect for voice-gender or cycle phase and no interaction effects (all $p > 0.05$). Bayesian ANOVAs also favored H_0 regarding effects of menstrual cycle phase ($BF_{01} = 7.136$). Post-hoc tests showed significant differences in percent of correct answers for each morphing level compared to another (all $p < 0.001$).

3.3 Signal detection theory and reaction times

Within the follicular phase group, we found negative correlations between A′ and mean RT for both feminine ($r_\tau = -0.21$, $p = 0.003$) and masculine ($r_\tau = -0.35$, $p < 0.001$) voices. In the luteal phase we also found negative correlations between A' and mean RT for feminine ($r_\tau = -0.26$, $p < 0.001$) and masculine ($r_\tau = -0.35$, $p < 0.001$) voices. Additionally, we found a positive correlation between B''_D and mean RT for masculine voices ($r_\tau = 0.20$, $p = 0.003$) that did not become apparent in the follicular phase ($r_\tau = 0.06$, $p = 0.545$).

3.4 Influence of morphing, speaker, and hormone levels on performance

For B''_D model comparisons indicated the best model fit when only morphing level was included as a predictor (see Table 2). This model showed a significant influence only for 2 st Morphing (t(192) = −3.67, $p < 0.001$).

For RT, the model including the interaction of speaker and morphing level showed the best model fit according to AIC and BIC (see Table 3). All morphing levels influenced RT significantly as well as the interaction of speaker and morphing level for 4 st and 6 st morphing (see Table 4).

As a comparison, we also took a closer look at the second-best model, which includes estrogen and progesterone. Again, the model showed a significant influence of voice-gender and morphing level, but no influence of either hormone level (see Table 4).

Model estimates for all models can be found in the Supplementary information.

4 Discussion

Voice-gender categorization is an important part of everyday social interaction, further influencing mate choice and preferences (Weston et al., 2015). Despite the growing body of evidence, studies investigating the influencing factors on the ability to categorize a voice

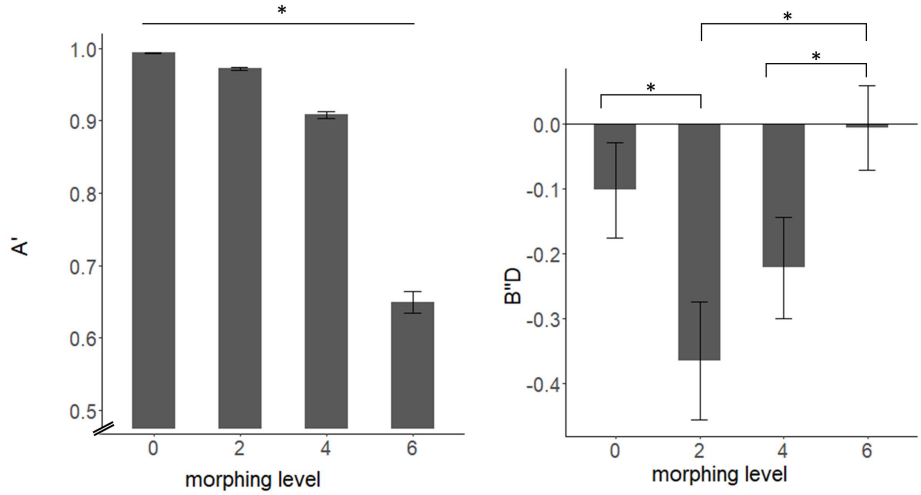


FIGURE 2
Differences in discriminability A' (left) and response bias B''D (right) for different morphing levels. The asterisk on the upper line indicates significant differences between all morphing levels.

TABLE 2 Model comparison for mixed models for response bias B''_D.

Model	npar	AIC	BIC	X^2	Df	$Pr(>X^2)$
B''_D ~ Morph + (1|Proband)	**6**	**401.456**	**422.820**			
B''_D ~ Morph + E + P+ (1|Proband)	8	405.088	433.574	0.367	2	0.832
B''_D ~ Morph * E * P+ (1|Proband)	18	407.120	471.213	17.968	10	0.056

Morph = morphing level, E = Estradiol, P = Progesterone, npar = number of parameters. The best fitting model is highlighted in bold.

TABLE 3 Mixed models for RT.

	npar	AIC	BIC	X^2	Df	$Pr(>X^2)$
RT ~ Morph	6	6541.675	6567.198			
RT ~ Morph + Speaker	7	6482.133	6511.91	61.542	1	<0.001
RT ~ Morph + Speaker + E + P	9	6483.31	6521.595	2.823	2	0.244
RT ~ Morph * Speaker	**10**	**6421.501**	**6464.039**	**63.809**	**1**	**<0.001**
RT ~ Morph * Speaker + E + P	12	6422.678	6473.724	2.823	2	0.244
RT ~ Morph * Speaker + E * P	13	6423.59	6478.89	1.088	1	0.297
RT ~ Morph * Speaker * E * P	34	6448.943	6593.573	16.647	21	0.732

Only random effects are depicted in the table, for all models fixed effects were + (1|Proband). Morph = morphing level, Speaker = voice-gender, E = Estradiol, P = Progesterone, npar = number of parameters. The best fitting model is highlighted in bold.

as masculine or feminine remain scarce. Important influences could be the menstrual cycle phase and female sex hormone profile. Hence, we investigated the associations of menstrual cycle phase with the performance in a voice-gender categorization task using masculine and feminine voices which were morphed toward the other sex, resulting in increasingly ambiguous stimuli.

In contrast to our hypotheses, we did not find any significant differences between cycle phases (follicular vs. luteal) for reaction times (RT), discriminability or response bias. These potential null effects were assessed using different statistical approaches, increasing the likelihood of the null hypothesis being true. In accordance with earlier studies, we observed an increase in RT as well as a decrease in discriminability and correct responses with increasing morphing levels (Smith et al., 2018; Smith et al., 2019). In other words, with increasing ambiguity of the voices' sex, participants needed more time to decide and were less accurate in doing so. This effect most likely reflects a higher cognitive workload with increasing stimulus ambiguity. Thus, while accuracy was generally high, the paradigm also elicited consistent effects of increasing task difficulty on response times and choices. This may also increase behavioral differences between menstrual cycle phases if present.

Congruent with the findings for reaction times, we did not observe a significant difference between menstrual cycle phases for voice-gender categorization accuracy within our sample. In addition, E and P do not seem to show associations with voice-gender categorization, as investigated by regression models that treated these two variables as continuous predictors. In other studies, estradiol is associated with increased preferences for higher masculinity, both for

TABLE 4 Model estimates for the model with the best model fit (upper model) and the simplest model including hormones estrogen and progesterone (lower model).

Coefficient	*B*	Df	*t*	*p*
Intercept	1120.734	86.568	40.01	**<0.001**
Morphing 2	76.554	448	4.80	**<0.001**
Morphing 4	126.454	448	7.93	**<0.001**
Morphing 6	214.403	448	13.44	**<0.001**
Speaker	−0.715	448	−0.05	0.964
Morphing 2: Speaker	6.7153	448	0.30	0.766
Morphing 4: Speaker	124.566	448	5.52	**<0.001**
Morphing 6: Speaker	148.031	448	6.56	**<0.001**
Intercept	1085.82	76.250	39.96	**<0.001**
Morphing 2	79.911	451	6.61	**<0.001**
Morphing 4	188.737	451	15.61	**<0.001**
Morphing 6	288.419	451	23.85	**<0.001**
Speaker	69.113	451	8.08	**<0.001**
Estradiol	42.741	62	1.11	0.270
Progesteron	−63.376	62	−1.65	0.104

The number behind Morphing indicates semitones morphed, Speaker = voice-gender. Significant results ($p < 0.05$) are highlighted in bold.

voices (Feinberg et al., 2006) and faces (Hromatko et al., 2008). Our results did not show such straightforward connection between hormone levels and reaction to male voices. However, it is important to note that fluctuations in progesterone and estradiol are not the only possible explanation for behavioral changes across the cycle. Thus, potential effects could also be driven by an interplay of different sex hormones, which were not investigated in this study such as testosterone, which also influences changes in face preferences (Welling et al., 2007; Niu and Zheng, 2020). Another possible explanation is the influence of other variables affected by the menstrual cycle, such as mood (Cohen et al., 1987; Collins et al., 1985; Pierson et al., 2021) or attention (Pletzer et al., 2017; Thimm et al., 2014). These changes can in turn influence the response behavior.

In addition to the overall effect of morphing level, we found faster RTs for feminine voices compared to masculine voices with higher morphing levels in both cycle phases. This finding also replicates earlier findings on voice-gender categorization, which identified faster reaction times for feminine compared to masculine voices regardless of sex (Junger et al., 2014) or in female participants for morphed voices (Smith et al., 2018; Smith et al., 2019). A possible explanation for faster RTs for female voices could be a higher sensitivity for higher frequencies and specifically female voices in general (Lattner et al., 2005) resulting in a faster reaction to higher pitched voices. As the signal detection theory does not take reaction time into account, this slightly higher sensitivity could accelerate the responses to feminine voices while SDT measures can still be expected to show an other-sex effect for discriminability and response bias.

Regarding response bias, an overall tendency to categorize a voice as male – especially with higher uncertainty - became apparent, thus supporting the idea of an other-sex bias in heterosexual women. The resulting distribution resembles an U shape with smaller biases for original voices and the highest morphing level (6 st) and higher biases for ambiguous voices (2 st and 4 st) with the most pronounced effects for 2 st morphing. While this stronger tendency for 2 st morphing has been found before (Junger et al., 2013, 2014), the underlying mechanism is not yet clear. A possible explanation is the slight ambiguity of the stimuli which increases the effect of the reaction bias. While gender-categorization for unmorphed voices is in most cases an easy task, increased ambiguity heightens the cognitive processing load (Junger et al., 2013). The highest response bias for 2 st morphed voices could reflect the automatic use of heuristics to lighten the processing load (Gigerenzer and Gaissmaier, 2011). As the morphing level increases, overall performance decreases and categorization develops into a more conscious decision process, making the influence of reaction bias less pronounced. Since our sample showed a high variance across all morphing levels, our results must be interpreted with caution.

When looking at the menstrual cycle phases separately, we found a correlation between response bias and reaction time for masculine voices present in the luteal phase, even though there was no significant difference in response bias between phases. A possible explanation could be the influence of cycle phase on hearing sensitivity. Studies on hearing sensitivity could show that sensitivity is higher in the follicular phase than in the luteal phase (Emami et al., 2018; Williamson et al., 2020). A proposed explanation is an enhancing effect of estradiol on hearing which is influenced by the interaction of estradiol with progesterone levels (Williamson et al., 2020). Since there was no interaction of hormone levels in our data, this seems an unlikely explanation for our results. A more likely explanation could be a blunting effect of progesterone on the enhancing effect of estradiol (Al-Mana et al., 2010). Therefore, our results could hint toward a weaker influence of the response bias on reaction times in phases with higher acoustic sensitivity. As we did not assess general hearing sensitivity in this study, this explanation remains speculative and needs to be explored in future studies. Moreover, since interindividual differences in hormone levels are higher than intraindividual fluctuations (Gann et al., 2001), the finding above could be driven by individual hormone profiles that are not strictly linked to cycle phase.

In an exploratory analysis, we looked at regression models in which we examined the potential effect of female sex hormones (P and E) regardless of cycle phase to check for an overall effect of hormone levels. Model comparisons showed that there was no effect of hormones on RT or response bias. Significant predictors of those models were only morphing level and voice-gender. Thus, for the chosen sample, even dimensional models (which have more power compared to ANOVAs) suggest that female sex hormones do not influence sex-voice categorization as measured by RT, and response bias. In line with our previous results, we did not find a clear influence of either cycle phase or sex hormone levels on both accuracy and speed of voice-gender categorization.

Due to the manifold influences on hormonal effects, some limitations of the current study should be considered. For the current sample, we did not differentiate between early and late follicular phase. This may have impacted our results, as estradiol levels change from early to late follicular phase (Sacher et al., 2013). Similarly, we did not account for early, mid, or late luteal phase. However, previous studies show, that differences across cycle phases can be detected using this broader differentiation (Penton-Voak and Perrett, 2000; Johnston et al., 2001; Thimm et al., 2014; Nielsen et al., 2013; Senior et al., 2007). Furthermore, using hormone blood serum levels determined by ECLIAS entails additional uncertainties. Despite their high specificity and common use, a known limitation for all

Immunoassays is the potential cross-reactivity with compounds similar to the target hormone, which could potentially influence the calculated blood serum level (Krasowski et al., 2014). Additionally, due to the wider reference range for cycle-phases, misclassification cannot be ruled out, especially in cases where blood serum levels are close to boundaries between cycle phases. Since we only included women in our study, who were rated as being unambiguously within one cycle phase by two independent raters and those ratings had to correspond with the women's self-reports, we estimate the likelihood of misclassification within our sample as rather small though.

For feasibility reasons, our study used a between-subject design. Within-subject designs are more suitable at detecting changes over the menstrual cycle (Jones et al., 2019; Schmalenberger et al., 2021), so the current study design could have missed subtle changes over the cycle as well as the effect of intraindividual hormone fluctuations. The use of between-subject designs for investigating menstrual-cycle effects is suspected to have a substantial effect on the validity of cycle-phase determination and thus on the statistical power attainable. Referring to estimations by Gonzales and Ferrer (2016), the use of a between-subject design inevitably leads to a higher required sample size, independent of accuracy of cycle-phase determination accuracy. Future studies on voice-gender categorization are therefore advised to use a within-subject design to achieve more robust results, whenever possible. Further methodological precision could be achieved by directly measuring ovulation, using for instance LH tests in a standardized way as recommended by Blake et al. (2016).

Besides, additional hormones may also be considered in their interaction possibly influencing voice-gender categorization. For example, testosterone and thyroid hormones both influence cochlear development and hearing and may be investigated in future studies on voice-gender categorization tasks (Frisina et al., 2021). This is especially important considering that for feasibility reasons we could not measure all participants at the same time of day. Since all sex hormones present a specific circadian rhythm (Dabbs and de La Rue, 1991; Rahman et al., 2019), potential effects of time of day on the measured blood-serum level as well as potential interactions of progesterone and estradiol with the aforementioned additional hormones cannot be included in the analyses.

Looking at the manifold influences involved in analyses of hormonal effects, the sample size used in our study was likely not sufficient after all. Each factor brings a certain degree of variance into the analyses that cannot be accounted for in a regular power-analysis for between-subject comparisons. Therefore, sample size calculations should be based on methods considering the specifics of menstrual-cycle research (e.g., Gangestad et al., 2016; Gonzales and Ferrer, 2016; Schmalenberger et al., 2021). The shortcoming of not taking those specifics into account likely lead to a decrease in statistical power and therefore to a higher probability of accepting the null hypothesis while it was not true. Thus, we strongly recommend replicating the study design using a higher number of participants as well as a within-subjects design to increase statistical power.

As the originally spoken words were manipulated for the paradigm, especially the higher morphing levels sounded less natural, which could lead to diminished ecological validity and thus to a weakened influence of mechanisms important for mate choice. Additionally, stimuli were controlled for and changed in F0, but not in degree of breathiness, which could further influence voice-gender perception (Whitling et al., 2023). Nevertheless, considering that the exact same paradigm was used multiple times before and robustly showed effects of both gender and sexual orientation (Junger et al., 2013; Junger et al., 2014; Smith et al., 2018; Smith et al., 2019), we do assume this effect does not exert major influence on our results, especially in view of various studies successfully using mechanically morphed voices to examine even more complex social cues such as perceived dominance (e.g., Brown et al., 1973; Feinberg et al., 2005; Wang et al., 2018).

As mentioned earlier, there is a growing number of studies that do not find clear influences of menstrual cycle phase on female behavior such as mating behavior (Stern et al., 2021; Stern et al., 2022; Holzleitner et al., 2022). In two different studies, Harris (2011, 2013) tested different methods commonly used in studies on menstrual cycle effects on attractiveness ratings for male faces but did not find any influence of cycle phase on the participants' ratings. Likewise, eye-tracking studies testing the influence of hormonal changes on attractiveness ratings for male faces and bodies did not find associations with cycle phase either (Garza and Byrd-Craven, 2019; Garza and Byrd-Craven, 2023). Furthermore, meta-analyses on preference shifts across sensory modalities (i.e., faces, voices and scent) showed only a few effects, which were likely due to an imprecise definition of the fertile phase (Wood, 2016; Wood et al., 2014). Nevertheless, there is still an ongoing debate about the presence or absence as well as the magnitude of the influence of menstrual cycle phase on behavior, as other studies using similar tasks still find large effects on similar questions (Jones et al., 2008). Additional disagreement arises through divergent interpretations of the increasing number of null effects (Wood, 2016; Gangestad, 2016).

Suspected reasons for an overestimation of a cycle dependent shift in attractiveness ratings are a high publication bias as well as a high degree of freedom when it comes to researchers' decisions in sampling, study design and methods. Thus, confirmatory hypothesis testing can lead to arbitrary exclusion of participants, a broader definition of the fertile window and an inconsistent choice of moderators across analyses to achieve significant results (Wood et al., 2014; Harris, 2013). However, some of the studies presenting null effects show similar methodological shortcomings and are faced with criticism concerning sample sizes and statistical power (e.g., DeBruine et al., 2010) and thus regarding the interpretation of null effects (Gildersleeve et al., 2013). Moreover, studies seem to show only sparse evidence for the evolutionary perspective of female mate selection being driven by increasing chances for optimal offspring (Harris, 2013). One possible explanation is the negligence of socio-economic and sociosexual influences on mate choice which can be expected to play a stronger role than potential hormonal influences (Wood et al., 2014; Albert et al., 2018). Taken together, our results fall in line with the accumulating findings, that the associations of menstrual cycle phase with female (mating) behavior found in earlier studies might be a less robust effect than originally assumed, thus contributing to the ongoing debate about factors that might influence the complex interplay of sex hormones and behavior.

In conclusion, in this first study on associations of menstrual cycle phase with performance in voice-gender categorization, we did not find any significant differences between cycle phases for discriminability, response bias or reaction time in a between-subject design. Investigating effects regardless of cycle phase also did not show any significant associations of hormone levels and performance. Therefore, there might be no straightforward association between menstrual cycle phase or sex hormone level and voice-gender perception, supporting a growing body of literature reporting no or only subtle effects of menstrual cycle phase on female mating behavior. However, interpretation of results is impeded by multiple factors

relevant not only in analyses, but also in study design for menstrual cycle research, such as sampling method and cycle phase confirmation. Keeping those drawbacks in mind, our study seems to support the idea, that earlier studies on the matter might have overestimated the influence of sex hormone fluctuations on women's behavior. Further research in this line of research is needed to shed light on the interplay of hormones, socioeconomic factors, and behavior. The diverging results thus far highlight the importance of standardized best practices guidelines for sampling, sample size and interpretation of results for menstrual cycle research to minimize confounding factors and allow for a higher comparability of results across studies.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by The Independent Ethics Committee of the RWTH Aachen Faculty of Medicine. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SF: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. EB: Writing – review & editing, Writing – original draft. BD: Conceptualization, Writing – review & editing, Writing – original draft. UH: Conceptualization, Funding acquisition, Supervision, Writing – review & editing, Writing – original draft. PH: Formal analysis, Project administration, Supervision, Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work has been funded by the DFG (HA 3202/7-3 and HA 3202/7-4) and supported by the International Research Training Group the Neuroscience of modulating aggression and impulsivity in psychopathology (IRTG-2150) of the German Research Foundation (DFG).

Acknowledgments

This work was supported by the Brain Imaging Facility of the Interdisciplinary Centre for Clinical Research (IZKF) Aachen within the Faculty of Medicine at the RWTH Aachen University. Furthermore, we want to thank Mirjam Leitner and Franziska Feldmeyer for their help with data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1531021/full#supplementary-material

References

Albert, G., Pearson, M., Arnocky, S., Wachowiak, M., Nicol, J., and Murphy, D. R. (2018). Effects of masculinized and feminized male voices on men and Women's distractibility and implicit memory. *J. Individ. Differ.* 39, 151–165. doi: 10.1027/1614-0001/a000259

Albert, K., Pruessner, J., and Newhouse, P. (2015). Estradiol levels modulate brain activity and negative responses to psychosocial stress across the menstrual cycle. *Psychoneuroendocrinology* 59, 14–24. doi: 10.1016/j.psyneuen.2015.04.022

Al-Mana, D., Ceranic, B., Djahanbakhch, O., and Luxon, L. M. (2010). Alteration in auditory function during the ovarian cycle. *Hear. Res.* 268, 114–122. doi: 10.1016/j.heares.2010.05.007

Anderl, C., Hahn, T., Notebaert, K., Klotz, C., Rutter, B., and Windmann, S. (2015). Cooperative preferences fluctuate across the menstrual cycle. *Judgm. Decis. Mak.* 10, 400–406. doi: 10.1017/S1930297500005544

Arnocky, S., Hodges-Simeon, C. R., Ouellette, D., and Albert, G. (2018). Do men with more masculine voices have better immunocompetence? *Evol. Hum. Behav.* 39, 602–610. doi: 10.1016/j.evolhumbehav.2018.06.003

Arruda, P., Da Diniz Rosa, M. R., Almeida, L. N. A., Pernambuco, A., De, L., and Almeida, A. A. (2019). Vocal acoustic and auditory-perceptual characteristics during fluctuations in estradiol levels during the menstrual cycle: a longitudinal study. *J. Voice* 33, 536–544. doi: 10.1016/j.jvoice.2018.01.024

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

Bobst, C., and Lobmaier, J. S. (2012). Men's preference for the ovulating female is triggered by subtle face shape differences. *Horm. Behav.* 62, 413–417. doi: 10.1016/j.yhbeh.2012.07.008

Boersma, P., and Weenink, D. (2010). Praat: Doing phonetics by computer. Amsterdam.

Brown, B. L., Strong, W. J., and Rencher, A. C. (1973). Perceptions of personality from speech: effects of manipulations of acoustical parameters. *J. Acoust. Soc. Am.* 54, 29–35. doi: 10.1121/1.1913571

Bryant, G. A., and Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biol. Lett.* 5, 12–15. doi: 10.1098/rsbl.2008.0507

Cohen, I. T., Sherwin, B. B., and Fleming, A. S. (1987). Food cravings, mood, and the menstrual cycle. *Horm. Behav.* 21, 457–470. doi: 10.1016/0018-506X(87)90004-3

Collins, A., Eneroth, P., and Landgren, B. M. (1985). Psychoneuroendocrine stress responses and mood as related to the menstrual cycle. *Psychosom. Med.* 47, 512–527. doi: 10.1097/00006842-198511000-00002

Dabbs, J. J. M., and de La Rue, D. (1991). Salivary testosterone measurements among women: relative magnitude of circadian and menstrual cycles. *Horm. Res.* 35, 182–184. doi: 10.1159/000181899

DeBruine, L., Jones, B. C., Frederick, D. A., Haselton, M. G., Penton-Voak, I. S., and Perrett, D. I. (2010). Evidence for menstrual cycle shifts in women's preferences for masculinity: a response to Harris (in press) "menstrual cycle and facial preferences reconsidered". *Evol. Psychol.* 8, 768–775. doi: 10.1177/147470491000800416

Derntl, B., Hack, R. L., Kryspin-Exner, I., and Habel, U. (2013). Association of menstrual cycle phase with the core components of empathy. *Horm. Behav.* 63, 97–104. doi: 10.1016/j.yhbeh.2012.10.009

Derntl, B., Kryspin-Exner, I., Fernbach, E., Moser, E., and Habel, U. (2008). Emotion recognition accuracy in healthy young females is associated with cycle phase. *Horm. Behav.* 53, 90–95. doi: 10.1016/j.yhbeh.2007.09.006

Emami, S. F., Gohari, N., Ramezani, H., and Borzouei, M. (2018). Hearing performance in the follicular-luteal phase of the menstrual cycle. *Int. J. Otolaryngol.* 2018, 7276359–7276356. doi: 10.1155/2018/7276359

Farage, M. A., Osborn, T. W., and MacLean, A. B. (2008). Cognitive, sensory, and emotional changes associated with the menstrual cycle: a review. *Arch. Gynecol. Obstet.* 278, 299–307. doi: 10.1007/s00404-008-0708-2

Farrar, D., Neill, J., Scally, A., Tuffnell, D., and Marshall, K. (2015). Is objective and accurate cognitive assessment across the menstrual cycle possible? A feasibility study. *SAGE Open Med.* 3:205031211456519. doi: 10.1177/2050312114565198

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146

Feinberg, D. R., Jones, B. C., Law Smith, M. L., Moore, F. R., DeBruine, L. M., Cornwell, R. E., et al. (2006). Menstrual cycle, trait estrogen level, and masculinity preferences in the human voice. *Horm. Behav.* 49, 215–222. doi: 10.1016/j.yhbeh.2005.07.004

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., and Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Anim. Behav.* 69, 561–568. doi: 10.1016/j.anbehav.2004.06.012

Fink, B., and Penton-Voak, I. S. (2002). Evolutionary psychology of facial attractiveness. *Curr. Dir. Psychol. Sci.* 11, 154–158. doi: 10.1111/1467-8721.00190

First, M. B., Williams, J. B., Karg, R. S., and Spitzer, R. L. (2016). Structured clinical interview for DSM-5 disorders, Clinician Version (SCID-5-CV). 0141–0768.

Fischer, J., Semple, S., Fickenscher, G., Jürgens, R., Kruse, E., Heistermann, M., et al. (2011). Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime. *PLoS One* 6:e24490. doi: 10.1371/journal.pone.0024490

Foo, Y. Z., Simmons, L. W., and Rhodes, G. (2017). Predictors of facial attractiveness and health in humans. *Sci. Rep.* 7:39731. doi: 10.1038/srep39731

Frisina, R. D., Bazard, P., Bauer, M., Pineros, J., Zhu, X., and Ding, B. (2021). Translational implications of the interactions between hormones and age-related hearing loss. *Hear. Res.* 402:108093. doi: 10.1016/j.heares.2020.108093

Gallup, G. G., and Frederick, D. A. (2010). The science of sex appeal: an evolutionary perspective. *Rev. Gen. Psychol.* 14, 240–250. doi: 10.1037/a0020451

Gangestad, S. W. (2016). Comment: Wood et al.'s (2014) speculations of inappropriate research practices in ovulatory cycle studies. *Emot. Rev.* 8, 87–90. doi: 10.1177/1754073915580400

Gangestad, S. W., Haselton, M. G., Welling, L. L., Gildersleeve, K., Pillsworth, E. G., Burriss, R. P., et al. (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evol. Hum. Behav.* 37, 85–96. doi: 10.1016/j.evolhumbehav.2015.09.001

Gann, P. H., Giovanazzi, S., van Horn, L., Branning, A., and Chatterton, R. T. (2001). Saliva as a medium for investigating intra- and Interindividual differences in sex hormone levels in premenopausal women. *Cancer Epidemiol. Biomarkers Prev.* 10, 59–64

Garza, R., and Byrd-Craven, J. (2019). Fertility status in visual processing of men's attractiveness. *Evol. Psychol. Sci.* 5, 328–342. doi: 10.1007/s40806-019-00190-4

Garza, R., and Byrd-Craven, J. (2023). The role of hormones in attraction and visual attention to facial masculinity. *Front. Psychol.* 14:1067487. doi: 10.3389/fpsyg.2023.1067487

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Annu. Rev. Psychol.* 62, 451–482. doi: 10.1146/annurev-psych-120709-145346

Gildersleeve, K., DeBruine, L., Haselton, M. G., Frederick, D. A., Penton-Voak, I. S., Jones, B. C., et al. (2013). Shifts in Women's mate preferences across the ovulatory cycle: a critique of Harris (2011) and Harris (2012). *Sex Roles* 69, 516–524. doi: 10.1007/s11199-013-0273-4

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

Goldstein, E. B. (2014). Sensation and perception. *Internat. ed., 9th* Edn. Belmont: Calif. Wadsworth Cengage Learning.

Gonzales, J. E., and Ferrer, E. (2016). Efficacy of methods for ovulation estimation and their effect on the statistical detection of ovulation-linked behavioral fluctuations. *Behav. Res. Methods* 48, 1125–1144. doi: 10.3758/s13428-015-0638-4

Green, D. M., and Swets, J. A. (1974). Signal detection theory and psychophysics. *rev.* Edn. Huntington, NY: RF Krieger.

Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24, 95–112. doi: 10.1007/BF02289823

Guldner, S., Nees, F., and McGettigan, C. (2020). Vocomotor and social brain networks work together to express social traits in voices. *Cereb. Cortex* 30, 6004–6020. doi: 10.1093/cercor/bhaa175

Haraguchi, R., Hoshi, H., Ichikawa, S., Hanyu, M., Nakamura, K., Fukasawa, K., et al. (2021). The menstrual cycle alters resting-state cortical activity: a magnetoencephalography study. *Front. Hum. Neurosci.* 15:652789. doi: 10.3389/fnhum.2021.652789

Harris, C. R. (2011). Menstrual cycle and facial preferences reconsidered. *Sex Roles* 64, 669–681. doi: 10.1007/s11199-010-9772-8

Harris, C. R. (2013). Shifts in masculinity preferences across the menstrual cycle: still not there. *Sex Roles* 69, 507–515. doi: 10.1007/s11199-012-0229-0

Harris, C. R., Chabot, A., and Mickes, L. (2013). Shifts in methodology and theory in menstrual cycle research on attraction. *Sex Roles* 69, 525–535. doi: 10.1007/s11199-013-0302-3

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., and Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. *J. Educ. Stat.* 17, 315–339. doi: 10.3102/10769986017004315

Haselton, M. G., and Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *J. Pers. Soc. Psychol.* 78, 81–91. doi: 10.1037/0022-3514.78.1.81

Hellbernd, N., and Sammler, D. (2016). Prosody conveys speaker's intentions: acoustic cues for speech act perception. *J. Mem. Lang.* 88, 70–86. doi: 10.1016/j.jml.2016.01.001

Hidalgo-Lopez, E., Mueller, K., Harris, T., Aichhorn, M., Sacher, J., and Pletzer, B. (2020). Human menstrual cycle variation in subcortical functional brain connectivity: a multimodal analysis approach. *Brain Struct. Funct.* 225, 591–605. doi: 10.1007/s00429-019-02019-z

Hillairet de Boisferon, A., Dupierrix, E., Uttley, L., DeBruine, L. M., Jones, B. C., and Pascalis, O. (2019). Sex categorization of faces: the effects of age and experience. *Iperception* 10:2041669519830414. doi: 10.1177/2041669519830414

Hillenbrand, J. M., and Clark, M. J. (2009). The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Atten. Percept. Psychophys.* 71, 1150–1166. doi: 10.3758/APP.71.5.1150

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Holzleitner, I. J., Driebe, J. C., Arslan, R. C., Hahn, A. C., Lee, A. J., O'Shea, K. J., et al. (2022). No increased inbreeding avoidance during the ovulatory phase of the menstrual cycle. *Evol. Hum. Sci.* 4, 1–48. doi: 10.1017/ehs.2022.41

Hromatko, I., Tadinac, M., and Vranic, A. (2008). Femininity and masculinity across the menstrual cycle: a relation to mate value. *Coll. Antropol.* 32, 81–86

Hughes, S. M., and Puts, D. A. (2021). Vocal modulation in human mating and competition. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 376:20200388. doi: 10.1098/rstb.2020.0388

Jarosz, A. F., and Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *J. Probl. Solving* 7:2. doi: 10.7771/1932-6246.1167

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., and Grammer, K. (2001). Male facial attractiveness: evidence for hormone-mediated adaptive design. *Evol. Hum. Behav.* 22, 251–267. doi: 10.1016/S1090-5138(01)00066-6

Johnston, L., Miles, L., and Macrae, C. N. (2008). Was that a man? Sex identification as a function of menstrual cycle and masculinity. *Appl. Cognit. Psychol.* 22, 1185–1194. doi: 10.1002/acp.1436

Jones, B. C., DeBruine, L. M., Perrett, D. I., Little, A. C., Feinberg, D. R., and Law Smith, M. J. (2008). Effects of menstrual cycle phase on face preferences. *Arch. Sex. Behav.* 37, 78–84. doi: 10.1007/s10508-007-9268-y

Jones, B. C., Hahn, A. C., and DeBruine, L. M. (2019). Ovulation, sex hormones, and Women's mating psychology. *Trends Cogn. Sci.* 23, 51–62. doi: 10.1016/j.tics.2018.10.008

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., Han, C., et al. (2018). No compelling evidence that preferences for facial masculinity track changes in Women's hormonal status. *Psychol. Sci.* 29, 996–1005. doi: 10.1177/0956797618760197

Jonge, X. J. D. E., Thompson, B., and Han, A. (2019). Methodological recommendations for menstrual cycle research in sports and exercise. *Med. Sci. Sports Exerc.* 51, 2610–2617. doi: 10.1249/MSS.0000000000002073

Junger, J., Habel, U., Brohr, S., Neulen, J., Neuschaefer-Rube, C., Birkholz, P., et al. (2014). More than just two sexes: the neural correlates of voice gender perception in gender dysphoria. *PLoS One* 9:e111672. doi: 10.1371/journal.pone.0111672

Jünger, J., Motta-Mena, N. V., Cardenas, R., Bailey, D., Rosenfield, K. A., Schild, C., et al. (2018). Do women's preferences for masculine voices shift across the ovulatory cycle? *Horm. Behav.* 106, 122–134. doi: 10.1016/j.yhbeh.2018.10.008

Junger, J., Pauly, K., Brohr, S., Birkholz, P., Neuschaefer-Rube, C., Kohler, C., et al. (2013). Sex matters: neural correlates of voice gender perception. *NeuroImage* 79, 275–287. doi: 10.1016/j.neuroimage.2013.04.105

Karthikeyan, S., and Locke, J. L. (2015). Men's evaluation of women's speech in a simulated dating context: effects of female fertility on vocal pitch and attractiveness. *Evol. Behav. Sci.* 9, 55–67. doi: 10.1037/ebs0000014

Krasowski, M. D., Drees, D., Morris, C. S., Maakestad, J., Blau, J. L., and Ekins, S. (2014). Cross-reactivity of steroid hormone immunoassays: clinical significance and two-dimensional molecular similarity prediction. *BMC Clin. Pathol.* 14:33. doi: 10.1186/1472-6890-14-33

Lã, F. M. B., and Polo, N. (2020). Fundamental frequency variations across the menstrual cycle and the use of an Oral contraceptive pill. *J. Speech Lang. Hear. Res.* 63, 1033–1043. doi: 10.1044/2020_JSLHR-19-00277

Lattner, S., Meyer, M. E., and Friederici, A. D. (2005). Voice perception: sex, pitch, and the right hemisphere. *Hum. Brain Mapp.* 24, 11–20. doi: 10.1002/hbm.20065

Lewis, M. B. (2017). Fertility affects asymmetry detection not symmetry preference in assessments of 3D facial attractiveness. *Cognition* 166, 130–138. doi: 10.1016/j.cognition.2017.05.034

Lewis, M. B. (2020). Challenges to both reliability and validity of masculinity-preference measures in menstrual-cycle-effects research. *Cognition* 197:104201. doi: 10.1016/j.cognition.2020.104201

Little, A. C., and Jones, B. C. (2006). Attraction independent of detection suggests special mechanisms for symmetry preferences in human face perception. *Proc. Biol. Sci.* 273, 3093–3099. doi: 10.1098/rspb.2006.3679

Little, A. C., Jones, B. C., and DeBruine, L. M. (2008). Preferences for variation in masculinity in real male faces change across the menstrual cycle: women prefer more masculine faces when they are more fertile. *Personal. Individ. Differ.* 45, 478–482. doi: 10.1016/j.paid.2008.05.024

Little, A. C., Jones, B. C., Penton-Voak, I. S., Burt, D. M., and Perrett, D. I. (2002). Partnership status and the temporal context of relationships influence human female preferences for sexual dimorphism in male face shape. *Proc. Biol. Sci.* 269, 1095–1100. doi: 10.1098/rspb.2002.1984

Luther, T., Lewis, C. A., Grahlow, M., Hüpen, P., Habel, U., Foster, C., et al. (2021). Male or female? - influence of gender role and sexual attraction on sex categorization of faces. *Front. Psychol.* 12:718004. doi: 10.3389/fpsyg.2021.718004

Lynn, S. K., and Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychol. Sci.* 25, 1663–1673. doi: 10.1177/0956797614541991

Macrae, C. N., Hood, B. M., Milne, A. B., Rowe, A. C., and Mason, M. F. (2002). Are you looking at me? Eye gaze and person perception. *Psychol. Sci.* 13, 460–464. doi: 10.1111/1467-9280.00481

Maki, P. M., Rich, J. B., and Rosenbaum, R. S. (2002). Implicit memory varies across the menstrual cycle: estrogen effects in young women. *Neuropsychologia* 40, 518–529. doi: 10.1016/s0028-3932(01)00126-9

Maner, J. K., and Miller, S. L. (2014). Hormones and social monitoring: menstrual cycle shifts in progesterone underlie women's sensitivity to social information. *Evol. Hum. Behav.* 35, 9–16. doi: 10.1016/j.evolhumbehav.2013.09.001

Marcinkowska, U. M., Ellison, P. T., Galbarczyk, A., Milkowska, K., Pawlowski, B., Thune, I., et al. (2016). Lack of support for relation between woman's masculinity preference, estradiol level and mating context. *Horm. Behav.* 78, 1–7. doi: 10.1016/j.yhbeh.2015.10.012

MathWorks, Inc. (2019). MATLAB: 2019a. Natick, Massachusetts: The MathWorks, Inc.

McGettigan, C. (2015). The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. *Front. Hum. Neurosci.* 9:129. doi: 10.3389/fnhum.2015.00129

Neurobehavioral Systems, Inc. (2019). Presentation. Berkeley, CA.

Nielsen, S. E., Ahmed, I., and Cahill, L. (2013). Sex and menstrual cycle phase at encoding influence emotional memory for gist and detail. *Neurobiol. Learn. Mem.* 106, 56–65. doi: 10.1016/j.nlm.2013.07.015

Niu, C., and Zheng, L. (2020). Androstadienone resulted in a rightward shift of women's preference for sexually dimorphic male faces across the continuum of femininity-masculinity. *Horm. Behav.* 118:104635. doi: 10.1016/j.yhbeh.2019.104635

Oinonen, K. A., and Mazmanian, D. (2007). Facial symmetry detection ability changes across the menstrual cycle. *Biol. Psychol.* 75, 136–145. doi: 10.1016/j.biopsycho.2007.01.003

Pallier, C. (2002). Computing discriminability and bias with the R software. Available online at: https://www.pallier.org/pdfs/aprime.pdf, 1–6.

Parlee, M. B. (1983). Menstrual rhythm in sensory processes: a review of fluctuations in vision, olfaction, audition, taste, and touch. *Psychol. Bull.* 93, 539–548. doi: 10.1037/0033-2909.93.3.539

Penton-Voak, I. S., and Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evol. Hum. Behav.* 25, 229–241. doi: 10.1016/j.evolhumbehav.2004.04.003

Penton-Voak, I. S., and Perrett, D. I. (2000). Female preference for male faces changes cyclically: further evidence. *Evol. Hum. Behav.* 21, 39–48. doi: 10.1016/S1090-5138(99)00033-1

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

Peters, M., Simmons, L. W., and Rhodes, G. (2009). Preferences across the menstrual cycle for masculinity and symmetry in photographs of male faces and bodies. *PLoS One* 4:e4138. doi: 10.1371/journal.pone.0004138

Pierson, E., Althoff, T., Thomas, D., Hillard, P., and Leskovec, J. (2021). Daily, weekly, seasonal and menstrual cycles in women's mood, behaviour and vital signs. *Nat. Hum. Behav.* 5, 716–725. doi: 10.1038/s41562-020-01046-9

Pisanski, K., Oleszkiewicz, A., Plachetka, J., Gmiterek, M., and Reby, D. (2018). Voice pitch modulation in human mate choice. *Proc. R. Soc. B Biol. Sci.* 285:20181634. doi: 10.1098/rspb.2018.1634

Pletzer, B., Harris, T.-A., and Ortner, T. (2017). Sex and menstrual cycle influences on three aspects of attention. *Physiol. Behav.* 179, 384–390. doi: 10.1016/j.physbeh.2017.07.012

Pletzer, B., Harris, T.-A., Scheuringer, A., and Hidalgo-Lopez, E. (2019). The cycling brain: menstrual cycle related fluctuations in hippocampal and fronto-striatal activation and connectivity during cognitive tasks. *Neuropsychopharmacology* 44, 1867–1875. doi: 10.1038/s41386-019-0435-3

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

Puts, D. A., Bailey, D. H., Cárdenas, R. A., Burriss, R. P., Welling, L. L. M., Wheatley, J. R., et al. (2013). Women's attractiveness changes with estradiol and progesterone across the ovulatory cycle. *Horm. Behav.* 63, 13–19. doi: 10.1016/j.yhbeh.2012.11.007

Puts, D. A., Jones, B. C., and DeBruine, L. M. (2012). Sexual selection on human faces and voices. *J. Sex Res.* 49, 227–243. doi: 10.1080/00224499.2012.658924

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociol. Methodol.* 25:111. doi: 10.2307/271063

Rahman, S. A., Grant, L. K., Gooley, J. J., Rajaratnam, S. M. W., Czeisler, C. A., and Lockley, S. W. (2019). Endogenous circadian regulation of female reproductive hormones. *J. Clin. Endocrinol. Metabol.* 104, 6049–6059. doi: 10.1210/jc.2019-00803

Raj, A., Gupta, B., Chowdhury, A., and Chadha, S. (2010). A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *J. Voice* 24, 363–368. doi: 10.1016/j.jvoice.2008.10.005

Rosenberg, L., and Park, S. (2002). Verbal and spatial functions across the menstrual cycle in healthy young women. *Psychoneuroendocrinology* 27, 835–841. doi: 10.1016/S0306-4530(01)00083-X

RStudio Team. (2021). RStudio. Boston, MA.

Sacher, J., Okon-Singer, H., and Villringer, A. (2013). Evidence from neuroimaging for the role of the menstrual cycle in the interplay of emotion and cognition. *Front. Hum. Neurosci.* 7, 1–7. doi: 10.3389/fnhum.2013.00374

Sanders, G., and Wenmoth, D. (1998). Verbal and music dichotic listening tasks reveal variations in functional cerebral asymmetry across the menstrual cycle that are phase and task dependent. *Neuropsychologia* 36, 869–874. doi: 10.1016/S0028-3932(98)00022-0

Schmalenberger, K. M., Tauseef, H. A., Barone, J. C., Owens, S. A., Lieberman, L., Jarczok, M. N., et al. (2021). How to study the menstrual cycle: practical tools and recommendations. *Psychoneuroendocrinology* 123:104895. doi: 10.1016/j.psyneuen.2020.104895

Senior, C., Lau, A., and Butler, M. J. (2007). The effects of the menstrual cycle on social decision making. *Int. J. Psychophysiol.* 63, 186–191. doi: 10.1016/j.ijpsycho.2006.03.009

Smith, E., Junger, J., Pauly, K., Kellermann, T., Dernt, B., and Habel, U. (2019). Cerebral and behavioural response to human voices is mediated by sex and sexual orientation. *Behav. Brain Res.* 356, 89–97. doi: 10.1016/j.bbr.2018.07.029

Smith, E., Junger, J., Pauly, K., Kellermann, T., Neulen, J., Neuschaefer-Rube, C., et al. (2018). Gender incongruence and the brain–behavioral and neural correlates of voice gender perception in transgender people. *Horm. Behav.* 105, 11–21. doi: 10.1016/j.yhbeh.2018.07.001

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/bf03207704

Stern, J., Hildebrand, T., and Casto, K. (2022). Women's Intrasexual competitiveness and jealousy across the ovulatory cycle: a hormone-based study. *Soc. Psychol. Personal. Sci.* 14, 647–661. doi: 10.1177/19485506221117712

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

Thimm, M., Weis, S., Hausmann, M., and Sturm, W. (2014). Menstrual cycle effects on selective attention and its underlying cortical networks. *Neuroscience* 258, 307–317. doi: 10.1016/j.neuroscience.2013.11.010

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol. Methods* 17, 228–243. doi: 10.1037/a0027127

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/bf03194105

Wang, T. Y., Kawaguchi, I., Kuzuoka, H., and Otsuki, M. (2018). Effect of manipulated amplitude and frequency of human voice on dominance and persuasiveness in audio conferences. Proceedings of the ACM on human-computer interaction, 2, 1–18.

Welling, L. L., Jones, B. C., DeBruine, L. M., Conway, C. A., Law Smith, M. J., Little, A. C., et al. (2007). Raised salivary testosterone in women is associated with increased attraction to masculine faces. *Horm. Behav.* 52, 156–161. doi: 10.1016/j.yhbeh.2007.01.010

Weston, P. S., Hunter, M. D., Sokhi, D. S., Wilkinson, I. D., and Woodruff, P. W. (2015). Discrimination of voice gender in the human auditory cortex. *NeuroImage* 105, 208–214. doi: 10.1016/j.neuroimage.2014.10.056

Whitling, S., Botzum, H. M., and van Mersbergen, M. R. (2023). Degree of breathiness in a synthesized voice signal as it differentiates masculine versus feminine voices. *J. Voice*. [in press]. doi: 10.1016/j.jvoice.2023.04.022

Williamson, T. T., Zhu, X., Pineros, J., Ding, B., and Frisina, R. D. (2020). Understanding hormone and hormone therapies' impact on the auditory system. *J. Neurosci. Res.* 98, 1721–1730. doi: 10.1002/jnr.24588

Wood, W. (2016). Reply to Gangestad's (2016) comment on Wood, Kressel, Joshi, and Louie (2014). *Emot. Rev.* 8, 90–94. doi: 10.1177/1754073915580401

Wood, W., Kressel, L., Joshi, P. D., and Louie, B. (2014). Meta-analysis of menstrual cycle effects on Women's mate preferences. *Emot. Rev.* 6, 229–249. doi: 10.1177/1754073914523073

frontiers | Frontiers in Psychology

TYPE Review
PUBLISHED 25 April 2025
DOI 10.3389/fpsyg.2025.1559158

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Hellen Vivianni Veloso Corrêa,
Federal University of Pará, Brazil
John Thomas Manning,
Swansea University, United Kingdom

*CORRESPONDENCE
Ashlyn Swift-Gallant
✉ aswiftgallant@mun.ca
David Puts
✉ dap27@psu.edu

RECEIVED 11 January 2025
ACCEPTED 19 March 2025
PUBLISHED 25 April 2025

CITATION
Swift-Gallant A, Aung T, Salia S, Breedlove SM and Puts D (2025) Sexual orientation is associated with 2D:4D finger length ratios in both sexes: an updated and expanded meta-analysis.
Front. Psychol. 16:1559158.
doi: 10.3389/fpsyg.2025.1559158

Sexual orientation is associated with 2D:4D finger length ratios in both sexes: an updated and expanded meta-analysis

Ashlyn Swift-Gallant[1]*, Toe Aung[2], Stephanie Salia[1], S. Marc Breedlove[3] and David Puts[4]*

[1]Department of Psychology, Memorial University of Newfoundland and Labrador, St. John's, NL, Canada, [2]Department of Psychology and Counseling, Immaculata University, Immaculata, PA, United States, [3]Neuroscience Program, Michigan State University, East Lansing, MI, United States, [4]Departments of Anthropology and Psychology, Pennsylvania State University, University Park, PA, United States

The ratio of the lengths of the 2nd and 4th fingers (2D:4D) is a putative marker for prenatal gonadal hormone signaling and has been linked to human sexual orientation. Although 2D:4D is consistently found to be lower in males than females, the association with sexual orientation is variable across studies, with one meta-analysis finding lower (more masculine) digit ratios in lesbians than heterosexual females, but no overall association in males. However, this previous meta-analysis considered neither unpublished datasets nor bisexual individuals separately from homosexual and heterosexual individuals. Moreover, 17 datasets examining relationships between 2D:4D and sexual orientation have been published since that time, and we located an additional 11 unpublished datasets. We therefore conducted an updated and expanded meta-analysis comprising 51 studies, including 44 male and 34 female datasets, totaling 227,648 participants. This meta-analysis also explored whether 2D:4D differed between heterosexual and bisexual and/or non-exclusive individuals in both sexes. Results indicate lower (more male-typical) digit ratios in homosexual women (right hand $g = 0.26$, left hand $g = 0.16$; both adjusted following trim-and-fill), and higher (more female-typical) ratios in homosexual men (right hand $g = -0.17$, left hand $g = -0.20$; both adjusted) compared to heterosexual same-sex counterparts. Moderator analyses do not support publication bias for females. For males, positive findings were more likely to be published, but robustness tests, including trim-and-fill and leave-one-out, support the findings' robustness. No significant differences were observed in 2D:4D between male or female bisexual and heterosexual individuals. These findings are consistent with evidence that prenatal androgens increase attraction to females and/or that prenatal estrogens increase attraction to males.

KEYWORDS

sexual orientation, bisexuality, digit ratios, 2D:4D, prenatal androgens, prenatal estrogens

Introduction

Perinatal androgens play a central role in shaping sex differences in the brain and behavior across mammalian species by regulating patterns of gene expression in the developing brain (e.g., Xu et al., 2012). Comparisons of people with and without various endocrine conditions suggest that androgens play a similar role in the development of human sex differences in brain and behavior (Shirazi et al., 2022; Swift-Gallant et al., 2022, 2023). However, disentangling

direct effects of androgens on brain development from other biological and/or environmental factors, such as differential treatment by parents or physicians, remains challenging. This problem, coupled with the ethical infeasibility of experimental studies in humans, has led to considerable interest in exploring retrospective biomarkers of early androgen action (Swift-Gallant et al., 2020, 2023).

One such putative biomarker is 2D:4D, the ratio between the lengths of the second (2D) and fourth (4D) manual digits. 2D:4D is consistently lower in males (shorter index finger relative to ring finger; for meta-analysis, see Hönekopp and Watson, 2010), and converging evidence links prenatal androgens to its development (reviewed in Puts et al., 2008; Swift-Gallant et al., 2020, 2023). For example, digit ratios are lower (more male-typical) among women with congenital adrenal hyperplasia, in which prenatal androgens are elevated (Richards et al., 2020a), whereas digit ratios are higher (more female-typical) among chromosomal males with insensitivity to androgens (e.g., androgen insensitivity syndrome; Berenbaum et al., 2009; Van Hemmen et al., 2017).

Some evidence supports a link between 2D:4D and sexual orientation, one of the most strongly sexually differentiated human psychological traits (Balthazart, 2011; Hines, 2011; Kostic and Scofield, 2022). However, reported relationships between 2D:4D and sexual orientation have been mixed, with studies reporting lower ratios, higher ratios, or no significant correlation (reviewed in Swift-Gallant et al., 2020). These varied results prompted a previous meta-analysis (Grimbos et al., 2010), which found lower (i.e., more male-typical) 2D:4D among women with same-sex orientations than their heterosexual counterparts but no association of 2D:4D with male sexual orientation.

Despite its contribution, the Grimbos et al. (2010) meta-analysis did not include unpublished datasets, leaving it vulnerable to the "file drawer problem" that negative results may be less likely to be published, a concern often cited in critiques of digit ratio research (e.g., McCormick and Carré, 2020). Grimbos et al. also treated sexual orientation dichotomously, collapsing bisexual individuals with those exclusively oriented toward same-sex partners, and thus could not test whether bisexual individuals are intermediate between heterosexual and homosexual orientations or more similar to either. Additionally, since the last meta-analysis by Grimbos et al. (2010), there has been a significant increase in 2D:4D research, with the number of publications more than doubling between 2010 and 2020 (Figure 1). We have identified 44 datasets (10 unpublished) for male sexual orientation and 34 datasets (5 unpublished) for female sexual orientation, an increase of 26 male and 21 female datasets from those included in Grimbos et al. (2010). We therefore conducted a new meta-analysis on this larger sample that includes unpublished data to mitigate the effects of publication bias, consider intermediate sexual orientations, increase the precision of effect size estimates, and better assess the robustness of any associations.

Methods

Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021), we conducted a systematic literature search and extracted pertinent data from 60 published studies using two prominent electronic databases, PubMed and Google Scholar. In addition to consulting published works, we reached out to researchers who have published on 2D:4D, regardless of whether they evaluated this marker in relation to sexual orientation. Using keywords such as "2D:4D" and "digit ratio," we identified 296 unique corresponding authors with publications using this marker. From these authors, we requested data on sexual orientation for their published datasets where this information was originally omitted. Simultaneously, we inquired about any unpublished datasets containing the requisite information.

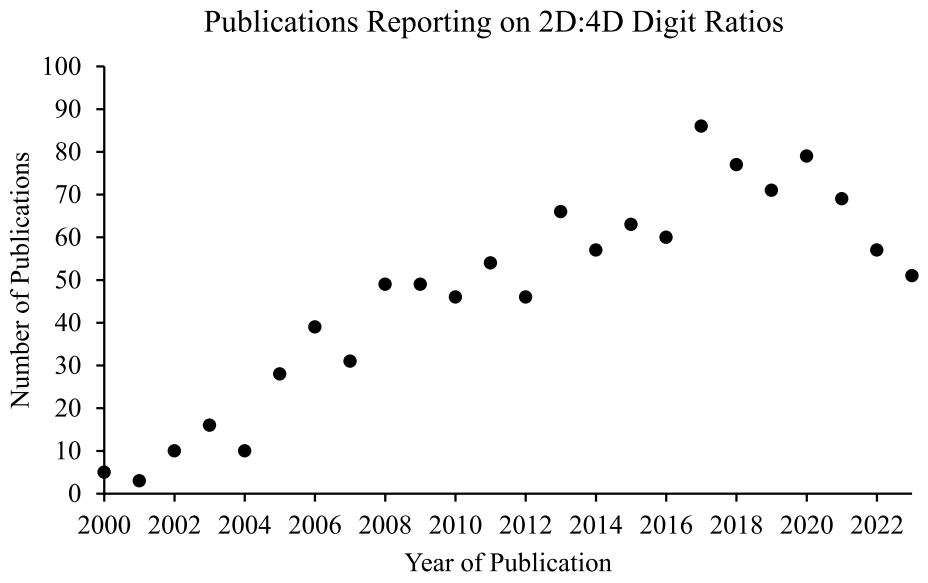


FIGURE 1
Publications reporting on digit ratios. The number of publications reporting on 2D:4D digit ratios has grown substantially since the last comprehensive meta-analysis assessing this measure in relation to sexual orientation in 2010 (Grimbos et al., 2010). A search on PubMed using the keywords "2D:4D" or "2D4D" or "digit ratio" for years 2000–2023 yielded the above number of publications by year (search date: November 26, 2024).

Articles and unpublished data were eligible for inclusion if they reported data on 2D:4D by sex and sexual orientation. Studies were excluded if they did not report an effect size or mean 2D:4D and standard deviation (SD) or standard error (SE), broken down by sex and sexual orientation. Studies were also excluded if they focused exclusively on heterosexual or homosexual individuals, or used or reanalyzed previously published data that had already been selected for inclusion. Out of 60 published papers and 13 unpublished datasets assessed for eligibility, 40 published studies and 11 unpublished datasets (comprising 10 male and 5 female unpublished data sets) were determined to be suitable for inclusion in the meta-analysis (i.e., 22 were excluded; see Supplementary Table S1), comprising information from a total of 227,648 participants (Figure 2). In addition to recording effect sizes for right and left hand 2D:4D from these studies, we recorded methodologically relevant variables and study characteristics for planned moderation analysis, including publication status (published or unpublished) geographic location (North American, UK/Europe, Asia, or other) and digit measurement method (direct, self-report, photocopy/scan, mixed, or unknown). The collated data including, effect sizes and moderator variables, are provided as a supplementary file.

Search strategy and study selection

We conducted an article search from 2000 to 2024 using PubMed and Google Scholar. Two search strategies were employed on PubMed using the terms: (1) (((((2D:4D and Sexual Orientation)); and (2)

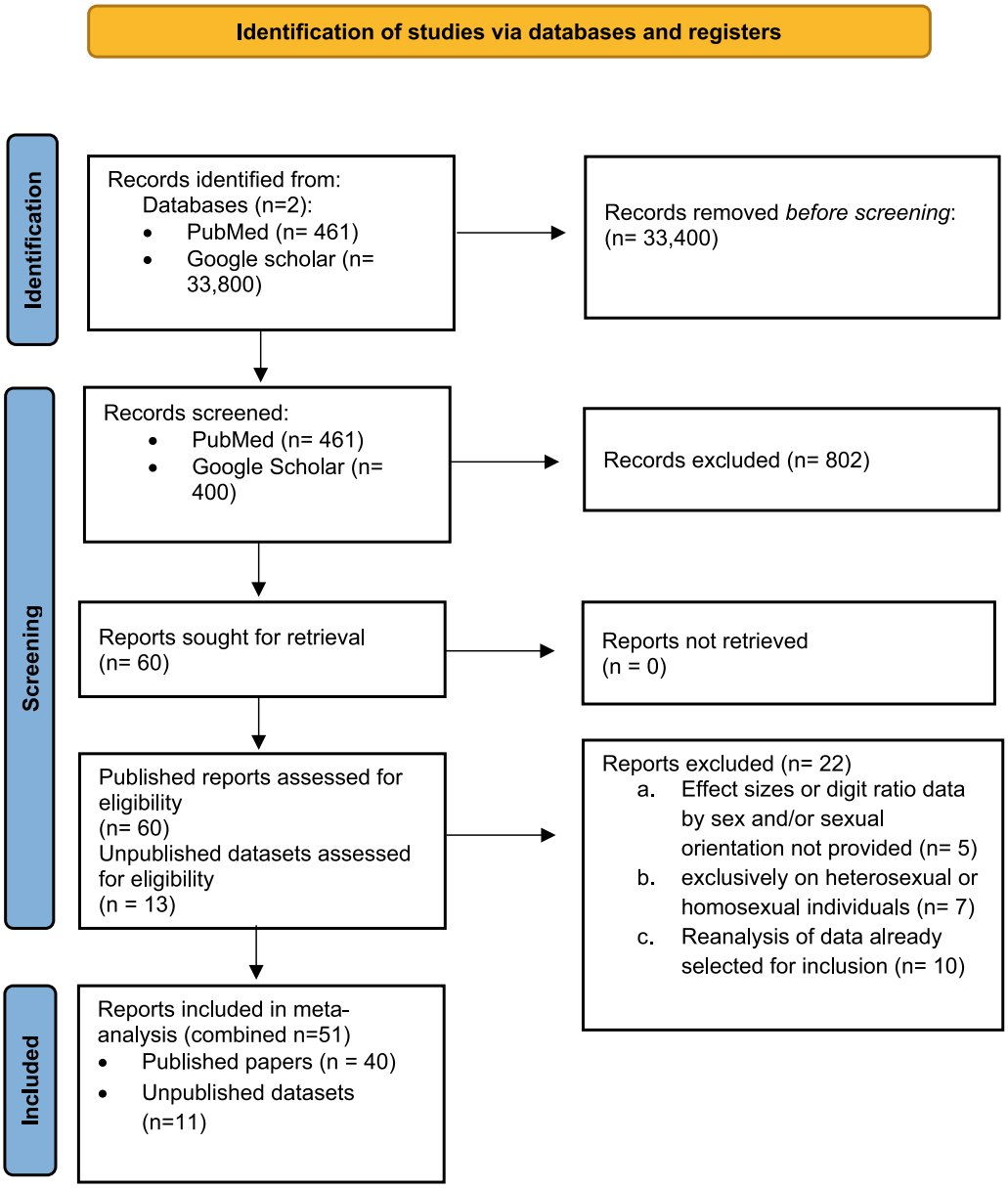


FIGURE 2
Prisma Flow Chart summarizing the records retrieval and workflow.

((((((Digit ratio) OR (2D:4D)) OR (2D4D)) OR (finger length)) OR (digit length)) OR (finger) OR (digit)) AND ((((sexual orientation) OR (lesbian)) OR (bisexual)) OR (heterosexual)). Similarly, on Google Scholar, the search included the terms: (1) "2D:4D and Sexual Orientation" and (2) "Digit ratio"|"2D:4D"|"2D4D"|"finger length"|"digit length"|"finger"|"digit" and "lesbian"|"bisexual"|"sexual orientation"|"heterosexual." PubMed retrieved a combined total of 461 reports, while Google Scholar yielded 33,800 reports; all Pubmed studies and the top 400 articles returned by Google Scholar were reviewed for inclusion. Additionally, we contacted the corresponding authors ($n = 296$) of 2D:4D studies to request information on the sexual orientation of participants, as well as inquire about unpublished data on both 2D:4D and sexual orientation of participants.

Statistical analyses

Meta-analyses were conducted using a random effects model implemented via the "metafor" package (version 4.2–0; Viechtbauer, 2010) in *R* (version 4.3.0). Standardized effect sizes were calculated using Hedges's *g* via the "escalc" function, and the random effects models were tested with the "rma.mv" function. Additionally, leave-one-out analyses were conducted utilizing the "leave1out" function to investigate the robustness of results and their dependence on any individual study. To address publication bias, Duval and Tweedie's trim-and-fill tests were applied using the "trimfill" function. The data and analysis scripts for all tested models are accessible in the supplementary file.

Following Grimbos et al. (2010), we excluded Manning et al. (2007) from primary analyses due to its potential to exert undue influence on meta-analytic results because of its size (>200,000 participants). Results of analyses including Manning et al. (2007) are reported in section Supplementary Results.

Results

Sex differences in digit ratios

Digit ratios exhibited expected sex differences: Heterosexual men had lower 2D:4D than heterosexual women for both the right hand ($g = -0.49$, $p < 0.001$; Figure 3) and left hand ($g = -0.43$, $p < 0.001$).

Male sexual orientation and digit ratios

Exclusive heterosexual vs. exclusive homosexual men

Right 2D:4D

Exclusively heterosexual and homosexual men did not significantly differ in right hand 2D:4D ($g = -0.15$, $p = 0.051$; Figure 4A). Leave-one-out analysis produced Hedge's *g* values ranging from −0.12 to −0.20 (Supplementary Figures S1–S4). No significant moderators were found for the right hand heterosexual and homosexual comparisons (Table 1). However, trim-and-fill analysis imputed seven studies and excluded one (Rahman, 2005), leading to an adjusted point estimate where 2D:4D is lower in heterosexual men than in homosexual men ($g = -0.17$, $p < 0.001$; Supplementary Figures S5–S10).

Left 2D:4D

Exclusively heterosexual men had a lower left hand 2D:4D than exclusively homosexual men ($g = -0.18$, $p < 0.001$; Figure 4B). Leave-one-out analyses produced Hedge's *g* values ranging from −0.16 to −0.20, suggesting that the difference between exclusive heterosexual and homosexual men does not depend on the inclusion of any particular study (Supplementary Figure S2). This relationship was moderated by publication status and present only in published studies, suggesting a tendency for statistically significant effects to be published (Table 1). Two missing studies were imputed during trim-and-fill analysis (Supplementary Figure S8), leading to a point estimate of $g = -0.20$. Geographical location, measurement type, and publication status were significant moderators of left hand digit ratios, but no pairwise comparisons were significant (Table 1).

Heterosexual vs. non-heterosexual men

Right 2D:4D

Exclusively heterosexual men had a lower right 2D:4D than non-heterosexual (bisexual plus homosexual) men ($g = -0.10$, $p = 0.018$; Figure 5A). This relationship was present in published studies, whereas the point estimate in unpublished studies was near zero, suggesting a tendency for statistically significant effects to be published (Table 2). However, following trim-and-fill analyses, which imputed seven studies and excluded one (Rahman, 2005), the relationship remained significant, and the effect size increased (adjusted $g = -0.17$, $p < 0.001$; Supplementary Figure S9). Leave-one-out analyses produced Hedge's g values ranging from −0.08 to −0.11, suggesting that difference between heterosexual and non-heterosexual men is robust to the exclusion of individual studies (Supplementary Figure S3). Measurement type was a significant moderator, with a significant difference between photocopy/scan and mixed or unknown measures, suggesting that mixed methods or studies that did not report how they measured ratios were more likely to find higher right 2D:4D among heterosexual men than non-heterosexual men (Table 2).

Left 2D:4D

Exclusively heterosexual men also had a lower left hand 2D:4D than non-heterosexual (bisexual plus homosexual) men ($g = -0.13$, $p = 0.006$; Figure 5B). Publication status was a significant moderator; however, the point estimate was nearly identical for published ($g = -0.13$, $p = 0.013$) and unpublished ($g = -0.12$, $p = 0.267$) studies, and no studies were imputed in trim-and-fill analysis. Leave-one-out analyses for the left hand produced Hedge's *g* values ranging from −0.11 to −0.15, suggesting these findings are robust to the exclusion of any particular study (Supplementary Figure S4). Measurement type moderated left hand comparisons, with significant differences observed between direct and mixed or unknown measures, between self-reported measures and mixed or unknown measures, and between photocopy/scan and mixed or unknown measures (Table 2). The effect size was in the opposite direction for studies reporting

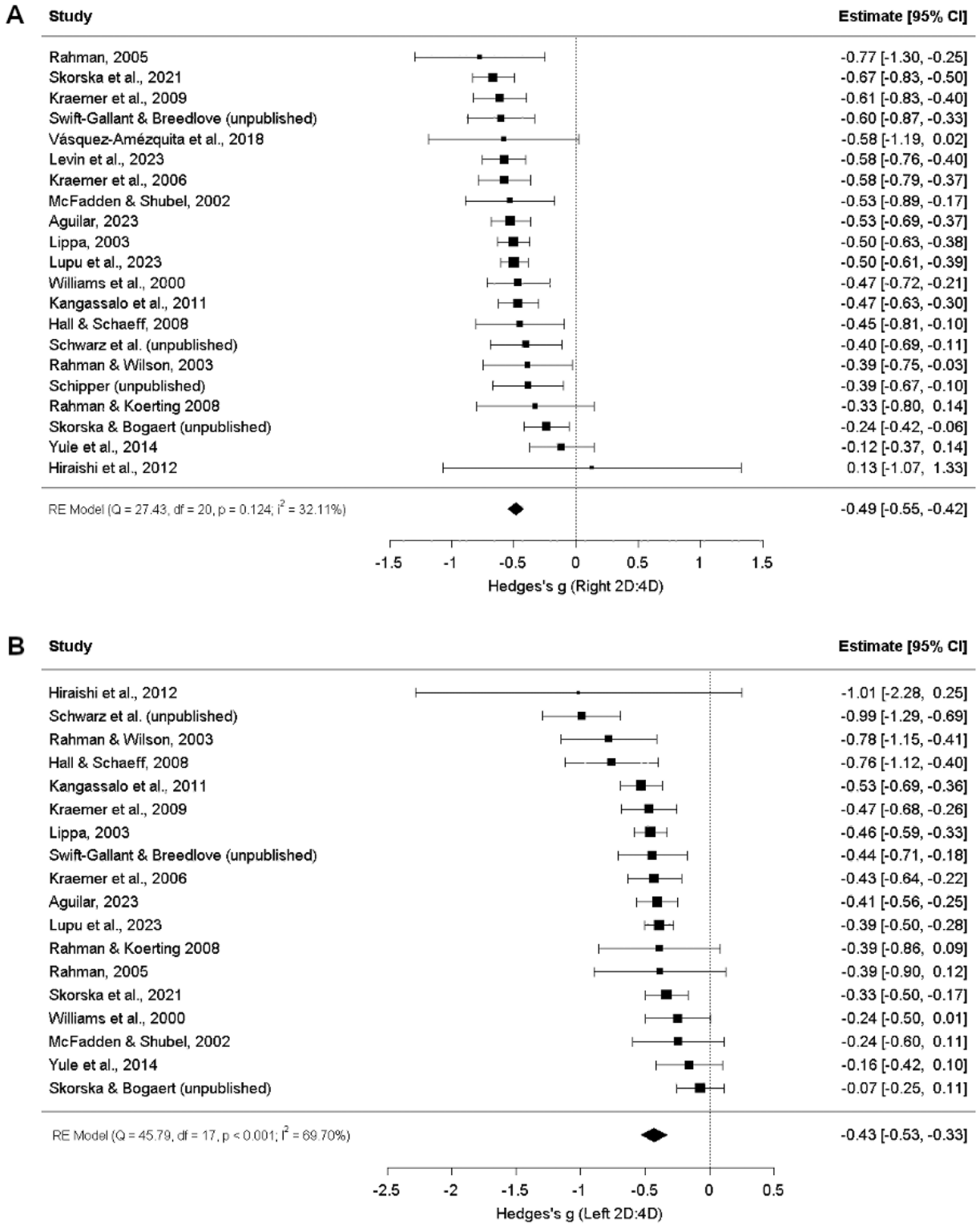


FIGURE 3
Right **(A)** and Left **(B)** 2D:4D comparison between exclusively heterosexual men and exclusively heterosexual women (Rahman, 2005; Skorska et al., 2021; Kraemer et al., 2009; Vásquez-Amézquita et al., 2018; Levin et al., 2023; Kraemer et al., 2006; McFadden and Shubel, 2002; Aguilar, 2023; Lippa, 2003; Lupu et al., 2023; Williams et al., 2000; Kangassalo et al., 2011; Hall and Schaeff, 2008; Rahman and Wilson, 2003; Rahman and Koerting, 2008; Yule et al., 2014; Hiraishi et al., 2012).

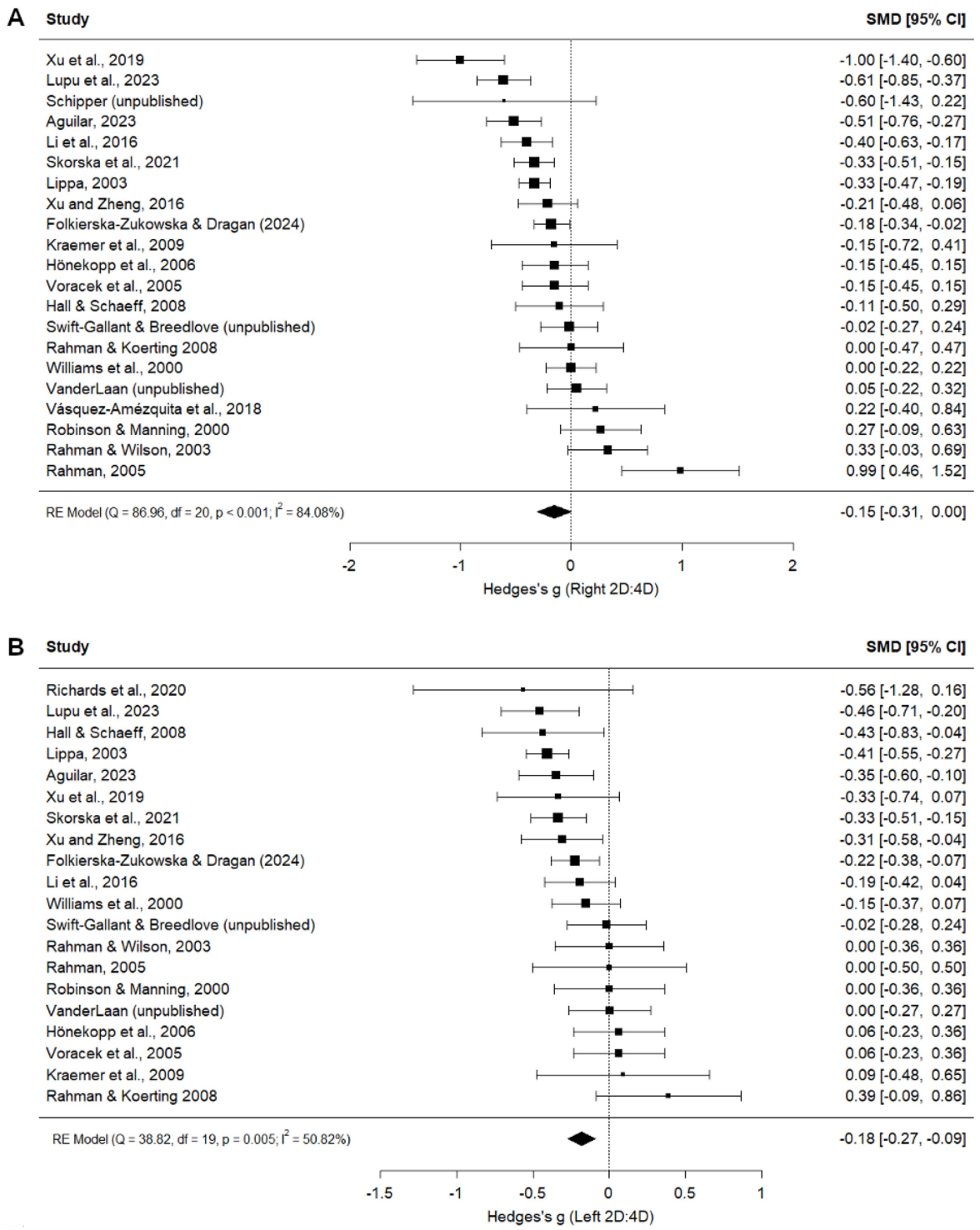


FIGURE 4
Right **(A)** and Left **(B)** 2D:4D comparison between exclusively heterosexual men and exclusively homosexual men (Xu et al., 2019; Lupu et al., 2023; Aguilar, 2023; Li et al., 2016; Skorska et al., 2021; Lippa, 2003; Xu and Zheng, 2016; Folkierska-Żukowska and Dragan, 2024; Kraemer et al., 2009; Hönekopp et al., 2006; Voracek et al., 2005; Hall and Schaeff, 2008; Rahman and Koerting, 2008; Williams et al., 2000; Vásquez-Amézquita et al., 2018; Robinson and Manning, 2000; Rahman and Wilson, 2003; Rahman, 2005; Richards et al., 2020b).

mixed or unknown measurement methods compared to direct, self-report and photocopy/scan methods. Geographical location moderated relationships, but no pairwise comparisons reached significance.

Comparisons of heterosexual, bisexual, and homosexual men

We also tested whether relationships between sexual orientation and 2D:4D differed across comparisons between heterosexual and

TABLE 1 Results from moderator analyses for exclusive heterosexual men vs. exclusive homosexual men.

Right 2D:4D model	Q_m	*k*	*g*	se	*z*	*p*	Lower CI	Upper CI
Geographic location	4.22					0.238		
North America		8	−0.10	0.13	−0.78	0.434	−0.37	0.16
UK/Europe		10	−0.13	0.12	−1.09	0.275	−0.37	0.11
Other[+]								
Asia		3	−0.32	0.20	−1.55	0.120	−0.72	0.08
Measurement type	6.78					0.079		
Direct		1	−0.02	0.13	−0.16	0.871	−0.28	0.24
Self-report[+]								
Photocopy/scan		12	−0.27	0.10	−2.59	0.010	−0.47	−0.07
Mixed or unknown		2	0.05	0.25	0.20	0.838	−0.44	0.54
Publication status	3.66					0.161		
Published		18	−0.16	0.09	−1.87	0.062	−0.33	0.01
Unpublished		3	−0.09	0.23	−0.42	0.674	−0.54	0.35

Left 2D:4D model	Q_m	*k*	*g*	se	*z*	*p*	Lower CI	Upper CI
Geographic location[1]	17.59					0.001		
North America		7	−0.22	0.07	−2.96	0.003	−0.37	−0.07
UK/Europe		10	−0.10	0.07	−1.34	0.179	−0.24	0.05
Other[+]								
Asia		3	−0.28	0.11	−2.65	0.008	−0.49	−0.07
Measurement type[1]	20.02					0.001		
Direct		7	−0.19	0.08	−2.46	0.014	−0.33	−0.04
Self-report		1	−0.56	0.39	−1.44	0.150	−1.33	0.20
Photocopy/scan		10	−0.21	0.06	−3.44	0.001	−0.33	−0.09
Mixed or unknown		2	0.04	0.15	0.24	0.811	−0.26	0.33
Publication status[1]	21.28					<0.001		
Published		18	−0.21	0.05	−4.61	<0.0001	−0.30	−0.12
Unpublished		2	−0.01	0.13	−0.06	0.949	−0.26	0.24

Pairwise comparisons were tested for the model with a significant moderator. [1]No significant difference was observed for all pairwise comparisons ($p > 0.05$). [+]Not identified in data.

bisexual men, bisexual and homosexual men, and heterosexual and homosexual men across the 8 samples for which these comparisons were possible (Supplementary Table S4). In left 2D:4D, homosexual men had a higher (more female-typical) 2D:4D than heterosexual men, whereas bisexual men differed from neither heterosexual nor homosexual men. A similar non-significant trend was evident for right 2D:4D.

Female sexual orientation and digit ratios

Exclusive heterosexual vs. exclusive homosexual women

Right 2D:4D

Exclusively heterosexual women had a higher right 2D:4D than exclusively homosexual women ($g = 0.26$, $p = 0.016$; Figure 6A). No moderators, including publication status, were significant (Table 3), and no studies were imputed or removed in the trim-and fill analysis (Supplementary Figure S8). Leave-one-out analysis produced Hedge's *g* values ranging from 0.14 to 0.30 (Supplementary Figures S13).

Left 2D:4D

Exclusively heterosexual women also had a higher left 2D:4D than exclusively homosexual women ($g = 0.17$, $p = 0.006$; Figure 6B). Publication status moderated this effect: the effect was larger in unpublished ($g = 0.31$, $p = 0.040$) than published ($g = 0.14$, $p = 0.033$) studies (Table 3), although both effects were significant in the same direction. Following trim-and-fill analyses, no study was imputed and one study was removed (Kraemer et al., 2006), resulting in a significant adjusted main effect ($g = 0.16$, $p = 0.010$; Supplementary Figure S8). Leave-one-out analyses produced Hedge's *g* values ranging from 0.12 to 0.20, also suggesting that the findings are robust to the exclusion of any particular study (Supplementary Figure S14). Geographical

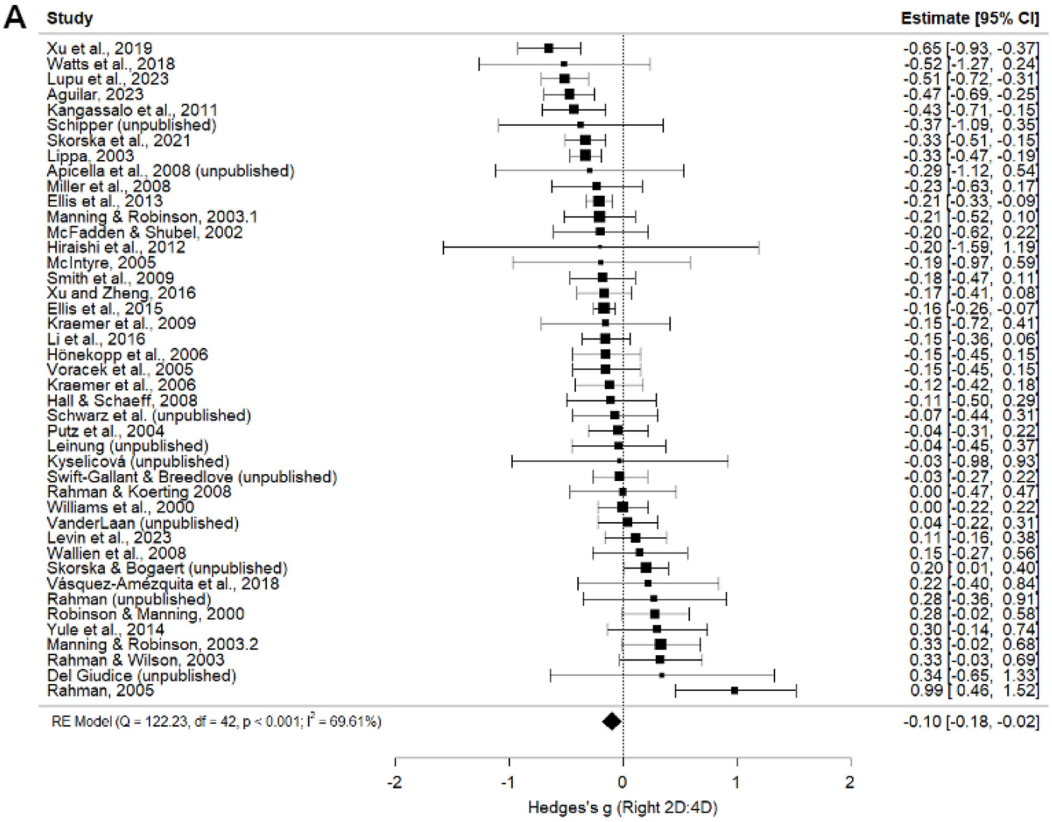


FIGURE 5
Right **(A)** and Left **(B)** 2D:4D comparison between heterosexual men and non-heterosexual men. 1 = UK participants; 2 = Multi-ethnic participants. In the random-effects model where Manning and Robinson (2003) studies were treated as two samples rather than one study, right 2D:4D ratios

(Continued)

FIGURE 5 (Continued)
(g = −0.10 [−0.18, −0.01], p = 0.025) differed significantly between heterosexual men and non-heterosexual men, while left 2D:4D ratios (g = −0.12 [−0.21, −0.03], p = 0.010) do not differ significantly between heterosexual men and non-heterosexual men (Xu et al., 2019; Watts et al., 2018; Lupu et al., 2023; Aguilar, 2023; Kangassalo et al., 2011; Skorska et al., 2021; Lippa, 2003; Miller et al., 2008; Ellis et al., 2013; Manning and Robinson, 2003; McFadden and Shubel, 2002, Hiraishi et al., 2012; McIntyre, 2005; Smith et al., 2010; Xu and Zheng, 2016; Ellis et al., 2015; Kraemer et al., 2009; Li et al., 2016; Hönekopp et al., 2006; Voracek et al., 2005; Kraemer et al., 2006; Hall and Schaeff, 2008; Putz et al., 2004; Rahman and Koerting, 2008; Williams et al., 2000; Levin et al., 2023; Wallien et al., 2008; Vásquez-Amézquita et al., 2018; Robinson and Manning, 2000; Yule et al., 2014; Rahman and Wilson, 2003; Rahman, 2005; Richards et al., 2020b).

TABLE 2 Results from moderator analyses for heterosexual men vs. non-heterosexual men.

Right 2D:4D model	Q_m	k	g	se	z	p	Lower CI	Upper CI
Geographic location	6.49					0.166		
North America		18	−0.06	0.06	−0.99	0.320	−0.19	0.06
UK/Europe		19	−0.10	0.07	−1.54	0.124	−0.24	0.03
Other		2	−0.18	0.26	−0.71	0.477	−0.69	0.32
Asia		4	−0.22	0.14	−1.62	0.105	−0.49	0.05
Measurement type[1]	13.05					0.011		
Direct		6	0.06	0.10	0.55	0.580	−0.15	0.26
Self-report		4	−0.11	0.11	−0.99	0.320	−0.33	0.11
Photocopy/scan		28	−0.17	0.05	−3.25	0.001	−0.27	−0.07
Mixed or unknown		4	0.17	0.15	1.10	0.273	−0.13	0.47
Publication status[2]	8.40					0.015		
Published		33	−0.13	0.05	−2.89	0.004	−0.22	−0.04
Unpublished		10	0.02	0.10	0.25	0.800	−0.16	0.21
Left 2D:4D model		k	g	se	z	p	Lower CI	Upper CI
Geographic location[2]	9.79					0.044		
North America		13	−0.18	0.08	−2.36	0.019	−0.33	−0.03
UK/Europe		7	−0.06	0.07	−0.81	0.416	−0.19	0.08
Other		2	−0.07	0.26	−0.28	0.780	−0.58	0.43
Asia		4	−0.25	0.14	−1.87	0.062	−0.52	0.01
Measurement type[3]	22.38					<0.001		
Direct		6	−0.04	0.09	−0.43	0.667	−0.23	0.14
Self-report		2	−0.13	0.17	−0.75	0.454	−0.45	0.20
Photocopy/scan		23	−0.21	0.05	−4.09	<0.001	−0.30	−0.11
Mixed or unknown		4	0.31	0.14	2.21	0.027	0.04	0.59
Publication status[2]	7.43					0.024		
Published		28	−0.13	0.05	−2.49	0.013	−0.23	−0.03
Unpublished		8	−0.12	0.11	−1.11	0.267	−0.33	0.09

Pairwise comparisons were tested for the model with a significant moderator. [1]Significant difference was observed between photocopy/scan and mixed or unknown measures (p = 0.038). [2]No significant difference was observed for all pairwise other comparisons (p > 0.05). [3]Significant difference was observed between direct and mixed or unknown measures (p = 0.038), between self-reported measures and mixed or unknown measures (p = 0.046), and between photocopy/scan and mixed or unknown measures (p < 0.001).

location moderated left-hand digit ratios. Pairwise comparisons identified that UK/Europe differed significantly from North America, with UK/Europe having a larger positive effect size (lower left 2D:4D in exclusive homosexual women compared to heterosexual women) than North America (Table 2). Measurement type was a significant moderator, but no pairwise comparisons were significant.

Heterosexual vs. non-heterosexual women

Heterosexual women had a higher digit ratio than non-heterosexual women in both right (g = 0.17, p = 0.012; Figure 7A) and left (g = 0.27, p = 0.005; Figure 7B) hands. While publication status was a significant moderator for both hands (Table 4), effect sizes were similar for published and unpublished studies (right hand:

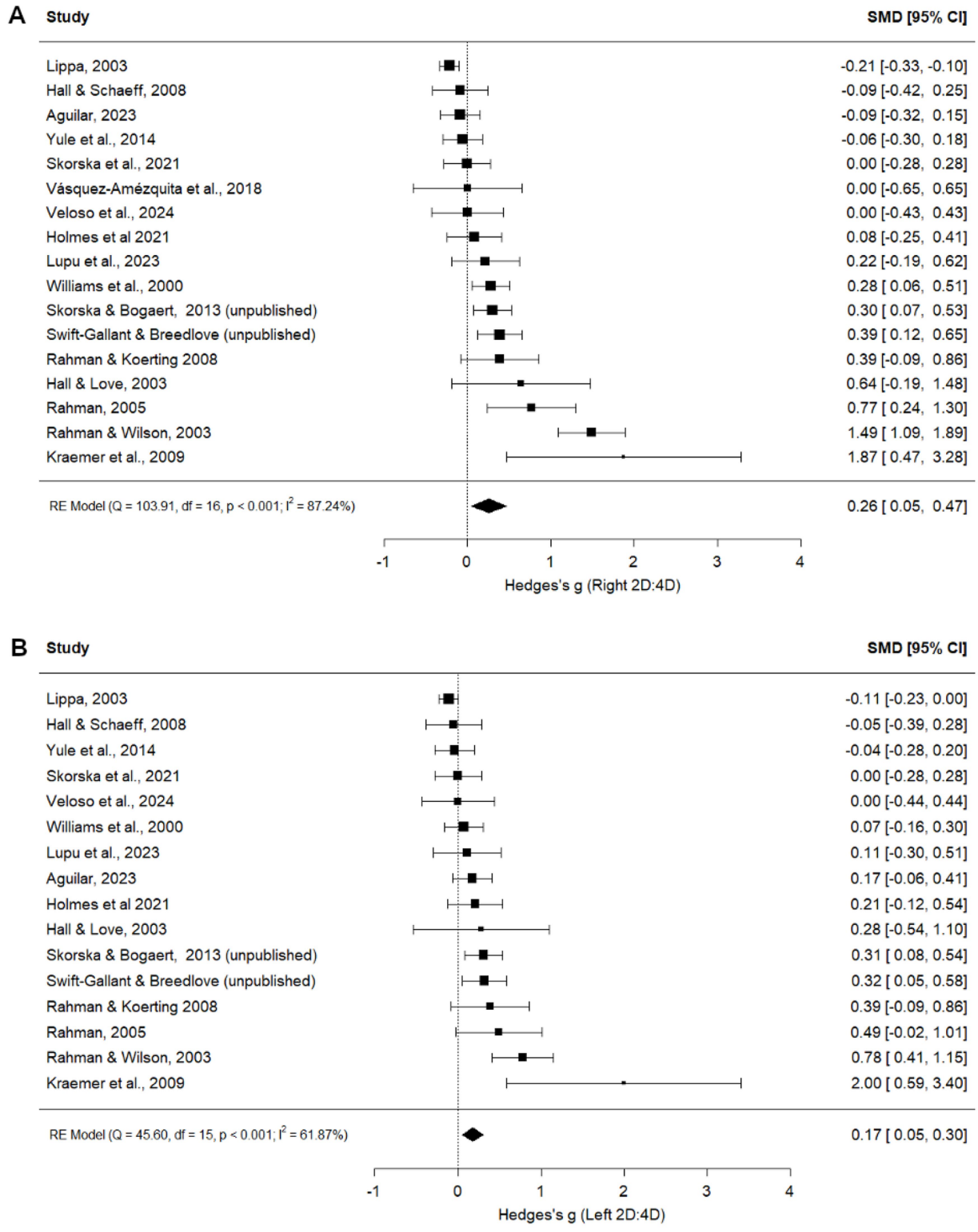


FIGURE 6
Right **(A)** and Left **(B)** 2D:4D comparison between exclusively heterosexual women and exclusively homosexual women (Lippa, 2003; Hall and Schaeff, 2008; Aguilar, 2023; Yule et al., 2014; Skorska et al., 2021; Vásquez-Amézquita et al., 2018; Veloso et al., 2024; Holmes et al., 2021; Lupu et al., 2023; Williams et al., 2000; Rahman and Koerting, 2008; Hall and Love, 2003; Rahman, 2005; Rahman and Wilson, 2003; Kraemer et al., 2009).

10.3389/fpsyg.2025.1559158

TABLE 3 Results from moderator analyses for exclusive heterosexual women vs. exclusive homosexual women.

Right 2D:4D model	Q_m	*k*	*g*	se	*z*	*p*	Lower CI	Upper CI
Geographic location	9.21					0.056		
North America		8	0.13	0.16	0.84	0.401	−0.17	0.44
UK/Europe		7	0.52	0.18	2.92	0.004	0.17	0.87
Other		1	0.00	0.45	0.00	1.000	−0.89	0.89
Asia		1	0.00	0.42	0.00	1.000	−0.83	0.83
Measurement type	6.20					0.111		
Direct		7	0.25	0.18	1.39	0.163	−0.10	0.60
Self-report		1	−0.06	0.44	−0.13	0.898	−0.91	0.80
Photocopy/scan		9	0.32	0.16	2.01	0.044	0.01	0.63
Mixed or unknown[+]								
Publication status	5.65					0.059		
Published		15	0.25	0.12	2.08	0.037	0.01	0.48
Unpublished		2	0.34	0.30	1.15	0.252	−0.24	0.93

Left 2D:4D model	Q_m	*k*	*g*	se	*z*	*p*	Lower CI	Upper CI
Geographic location[2]	13.67					0.008		
North America		7	0.08	0.08	0.99	0.320	−0.08	0.24
UK/Europe		7	0.37	0.10	3.56	<0.001	0.17	0.57
Other		1	0.00	0.28	0.00	1.000	−0.55	0.55
Asia		1	0.00	0.22	0.00	1.000	−0.43	0.43
Measurement type[1]	8.01					0.046		
Direct		7	0.17	0.11	1.59	0.111	−0.04	0.39
Self-report		1	−0.04	0.24	−0.17	0.867	−0.51	0.43
Photocopy/scan		8	0.22	0.09	2.34	0.020	0.03	0.40
Mixed or unknown[+]								
Publication status[1]	8.75					0.013		
Published		14	0.14	0.07	2.13	0.033	0.01	0.28
Unpublished		2	0.31	0.15	2.05	0.040	0.01	0.61

Pairwise comparisons were tested for the model with a significant moderator. [1]No significant difference was observed for all pairwise other comparisons ($p > 0.05$). [2]Significant difference was observed between UK/Europe and North America ($p = 0.031$). All other comparisons were non-significant. [+]Not identified in data.

published $g = 0.17$, $p = 0.023$; unpublished $g = 0.18$, $p = 0.324$; left hand: published $g = 0.28$, $p = 0.010$; unpublished $g = 0.26$, $p = 0.333$). Following trim-and-fill analyses, no studies were imputed, and one study was removed for the right hand, resulting in an adjusted estimate of $g = 0.15$, $p = 0.019$ (Supplementary Figure S9). For the left hand, six studies were imputed and one study was removed, resulting in a non-significant adjusted estimate $g = 0.06$, $p = 0.472$ (Supplementary Figure S9). Leave-one-out analyses for right (Supplementary Figure S15) and left (Supplementary Figure S16) 2D:4D produced Hedge's *g* values of 0.09 to 0.18 and 0.18 to 0.29, respectively, indicating that differences between heterosexual and non-heterosexual women are robust to the exclusion of any particular study. Geographical location and measurement type were significant moderators for left, but not right, hand 2D:4D; however, no pairwise comparisons reached significance (Table 4).

Comparisons of heterosexual, bisexual, and homosexual women

We also tested whether relationships between sexual orientation and 2D:4D differed across comparisons between heterosexual and bisexual women, bisexual and homosexual women, and heterosexual and homosexual women across the 6 samples for which these comparisons were possible (Supplementary Table S5). In both hands, heterosexual and bisexual women had higher (more female-typical) 2D:4D than homosexual women but did not differ from each other.

Discussion

Although relatively few studies were available for comparing heterosexual, bisexual, and homosexual individuals separately, a trend

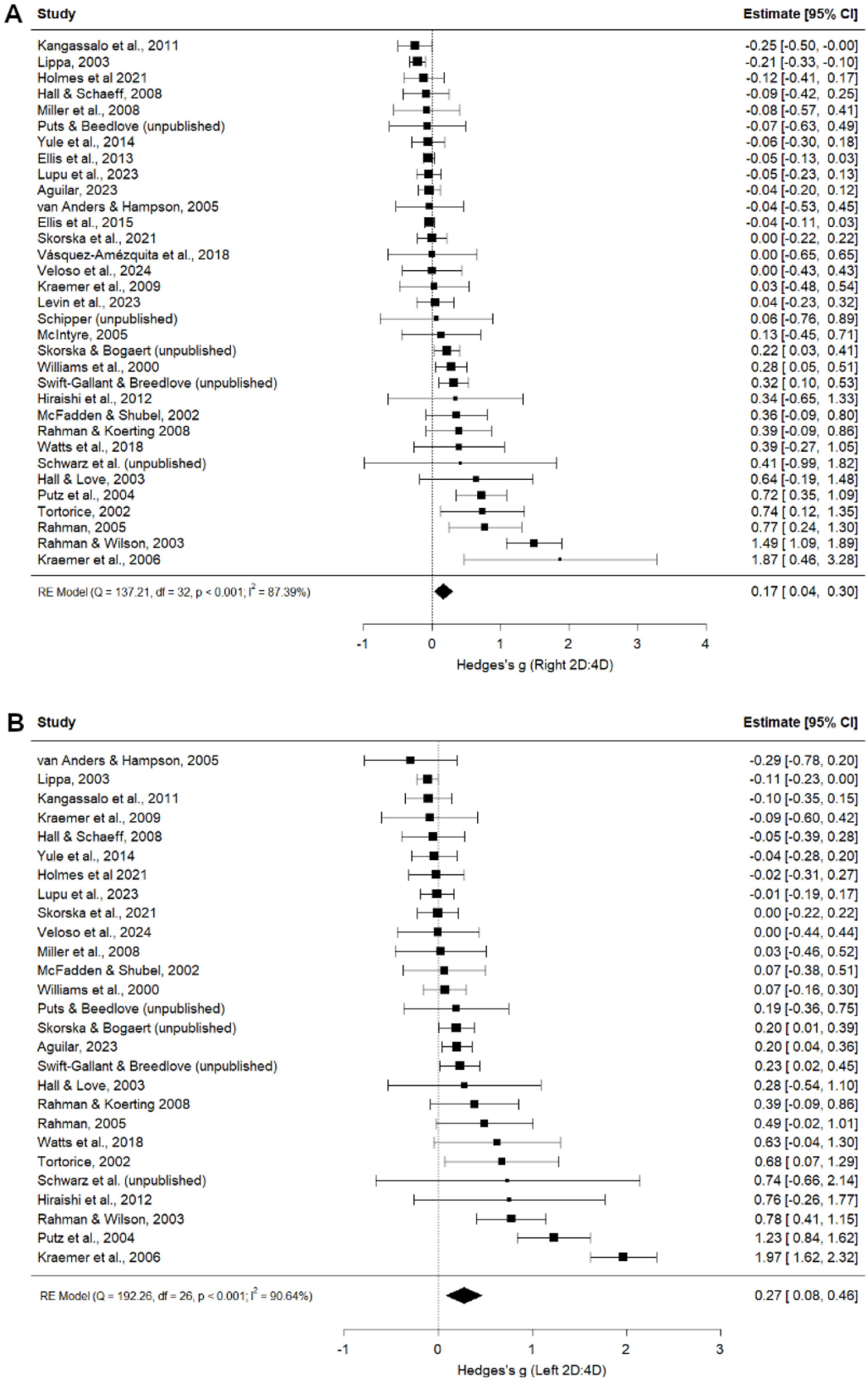


FIGURE 7
Right **(A)** and Left **(B)** 2D:4D comparison between heterosexual women and non-heterosexual women (Kangassalo et al., 2011; Lippa, 2003; Holmes et al., 2021; Hall and Schaeff, 2008; Miller et al., 2008; Yule et al., 2014; Ellis et al., 2013; Lupu et al., 2023; Aguilar, 2023; Van Anders and Hampson,

(Continued)

FIGURE 7 (Continued)
2005; Ellis et al., 2015; Skorska et al., 2021; Vásquez-Amézquita et al., 2018; Veloso et al., 2024; Kraemer et al., 2009; Levin et al., 2023; McIntyre, 2005; Williams et al., 2000; Hiraishi et al., 2012; McFadden and Shubel, 2002; Rahman and Koerting, 2008; Watts et al., 2018; Hall and Love, 2003; Putz et al., 2004; Tortorice, 2002; Rahman, 2005; Rahman and Wilson, 2003; Kraemer et al., 2006).

TABLE 4 Results from moderator analyses for heterosexual women vs. non-heterosexual women.

Right 2D:4D model	Q_m	*K*	g	se	*z*	*P*	Lower CI	Upper CI
Geographic location	7.82					0.099		
North America		18	0.16	0.09	1.75	0.080	−0.02	0.33
UK/Europe		10	0.27	0.13	2.10	0.036	0.02	0.53
Other		4	−0.03	0.19	−0.14	0.890	−0.41	0.35
Asia		1	0.34	0.60	0.57	0.570	−0.84	1.52
Measurement type[1]	7.71					0.053		
Direct		7	0.22	0.14	1.51	0.131	−0.06	0.50
Self-report		3	−0.05	0.19	−0.26	0.798	−0.42	0.32
Photocopy/scan		23	0.19	0.08	2.31	0.021	0.03	0.36
Mixed or unknown[+]								
Publication status[1]	6.18					0.046		
Published		28	0.17	0.07	2.28	0.023	0.02	0.31
Unpublished		5	0.18	0.19	0.99	0.324	−0.18	0.55

Left 2D:4D model	Q_m	*K*	*g*	se	*z*	*P*	Lower CI	Upper CI
Geographic location[1]	11.15					0.025		
North America		12	0.24	0.14	1.70	0.090	−0.04	0.52
UK/Europe		10	0.43	0.16	2.64	0.008	0.11	0.74
Other		4	−0.06	0.25	−0.24	0.809	−0.55	0.43
Asia		1	0.76	0.69	1.10	0.270	−0.59	2.10
Measurement type[1]	8.50					0.037		
Direct		7	0.17	0.20	0.88	0.380	−0.21	0.55
Self-report		1	−0.04	0.48	−0.08	0.934	−0.99	0.91
Photocopy/scan		19	0.33	0.12	2.78	0.006	0.10	0.57
Mixed or unknown[+]								
Publication status[1]	7.62					0.022		
Published		23	0.28	0.11	2.59	0.010	0.07	0.49
Unpublished		4	0.26	0.27	0.97	0.333	−0.27	0.79

Pairwise comparisons were tested for the model with a significant moderator. [1]No significant difference was observed for all pairwise comparisons ($p > 0.05$). [+]Not identified in data.

emerged in both sexes: 2D:4D ratios tended to be more similar between bisexual and heterosexual individuals than between either group and homosexual individuals. There was also a tendency for exclusively heterosexual and homosexual individuals to exhibit the greatest differences. These results indicate that the approach used in prior studies comparing heterosexual to non-heterosexual individuals may be less informative than analyses comparing exclusively heterosexual and homosexual individuals. We therefore focus our discussion on comparisons between exclusively heterosexual and homosexual individuals and consider heterosexual/non-heterosexual comparisons in this light.

Our results replicate the main finding from a previous meta-analysis (Grimbos et al., 2010) demonstrating an association between 2D:4D and women's sexual orientation: Homosexual women tend to have lower (more male-typical) digit ratios in both hands than heterosexual women. However, the inclusion of unpublished data and additional published studies in the present meta-analysis, as well as comparisons of more homogenous groups (exclusive heterosexual vs. homosexual), appears to contribute in two important ways.

First, the present data appear to be less influenced by publication bias than those in Grimbos et al. In the previous meta-analysis, adjusted effect sizes following trim-and-fill (right: 0.13, left: 0.07) were less than half of

unadjusted values (right: 0.29, left: 0.23). In the present meta-analysis, the effect size for right 2D:4D (0.26) was unchanged following trim-and-fill, and the difference between adjusted (0.16) and unadjusted (0.17) effects for the left hand was minimal. Second, adjusted effect sizes were approximately twice as large in the present meta-analysis as in Grimbos et al. These results increase confidence that associations between sexual orientation and 2D:4D are real and meaningful (see below). Results from comparisons of heterosexual to non-heterosexual women were also positive but showed greater evidence of publication bias.

In contrast to Grimbos et al. (2010), our findings showed that exclusively homosexual men tend to have higher (more female-typical) 2D:4D ratios than exclusively heterosexual men. This association was statistically significant in the left hand prior to correction for publication bias, and in both hands following trim-and-fill analysis, which also slightly increased effect size estimates from −0.15 to −0.17 (right hand) and from −0.18 to −0.20 (left hand). Somewhat smaller, but statistically significant, relationships were observed in comparisons between heterosexual and non-heterosexual men.

Similar to Grimbos et al. (2010), we excluded Manning et al. (2007) from primary analyses due to its potential to exert undue influence on meta-analytic results because of its size (>200,000 participants). However, this study is included in the Supplementary Results. Notably, the effect sizes remain nearly identical whether or not this study is included. The only exception is the unadjusted right hand comparison of heterosexual and homosexual men: with Manning et al. included, the effect is significant ($p = 0.045$), whereas it is not when excluded ($p = 0.051$); in both cases, the effect size is $g = -0.15$. All other effect sizes differ by $g = 0.02$ or less, with no changes in significance.

It is noteworthy that Manning et al. (2024) conducted a follow-up to their earlier work, Manning et al. (2007). In the original analysis, Manning et al. (2007) compared discrete sexual orientation categories (i.e., homosexual, bisexual, and heterosexual), while the 2024 study assessed sexual attraction scores on a 7-point Likert scale. The 2007 findings revealed significant differences in men, with homosexual and bisexual men exhibiting higher 2D:4D ratios compared to heterosexual men, aligning with the results of the present meta-analyses. However, no significant relationships were identified for women's sexual orientation categories. Conversely, the 2024 analysis using sexual attraction scores uncovered associations for both men and women, consistent with the findings of the present study. Thus, the association with digit ratio may be obscured when moderately and mostly bisexual individuals are combined. Our present analyses suggest that bisexual women are more similar to heterosexual women in digit ratios, but there may be further nuance, where those falling in the middle of the scale or between heterosexual and bisexual on the scale are more like heterosexual women, while those falling between bisexual and homosexual are more similar to lesbians in digit ratios.

Overall, these results conform to the hypothesis that common endocrine factors influence the development of digit ratios and sexual orientation. Specifically, relatively higher levels of prenatal androgen signaling may simultaneously masculize digit ratios (e.g., Richards et al., 2020a; Swift-Gallant et al., 2020, 2023; Zheng and Cohn, 2011) and increase the probability of gynephilia in females (Puts and Motta-Mena, 2018; Swift-Gallant et al., 2020, 2023). Conversely, relatively lower levels of androgen signaling and/or higher levels of estrogen signaling may feminize digit ratios (e.g., Manning et al., 1998; Zheng and Cohn, 2011) and increase androphilia (Shirazi et al., 2021; Swift-Gallant et al., 2023) in males.

Addressing concerns of publication bias

To address possible publication bias in the digit ratio literature, we contacted nearly 300 researchers, including those who have published on sexual orientation and those who used digit ratio data in relation to other traits and/or behaviors. We were able to include 10 male and 5 female unpublished datasets in the present meta-analysis. Thus, the primary meta-analyses included both published and unpublished datasets, and to assess potential publication bias we assessed publication status (published vs. unpublished) as a moderator. We also conducted trim-and-fill analyses, and conducted leave-one-out analyses to explore the robustness of results following the exclusion of each individual study.

For female sexual orientation comparisons, our analyses revealed no evidence of publication bias. Publication status was not a significant moderator for the right hand comparisons between heterosexual and homosexual women, and while publication status moderated the left hand comparison, the effect size was larger and in the same direction for unpublished datasets than for published ones ($g = 0.31$ vs. 0.14), which is the opposite of what would be expected if the overall association across published studies were due to publication bias. Similarly, while publication status moderated both the right and left heterosexual and non-heterosexual women comparisons, the effect sizes were nearly identical between published and unpublished studies. Looking to confidence intervals, there was more variability in unpublished datasets, likely due to lower statistical power/smaller sample sizes, which may have contributed to authors' decisions to not publish. Leave-one-out and trim-and-fill analyses also supported the robustness of these findings.

For male sexual orientation comparisons, publication status moderated many effects, though leave-one-out and trim-and-fill did not render any significant findings non-significant. Indeed, the corrected effect sizes were slightly larger for both right and left hand comparisons between heterosexual versus homosexual men (right $g = -0.15$ vs. adjusted $g = -0.17$; left hand $g = -0.18$ vs. adjusted $g = -0.20$). While unpublished studies tended to have point estimates near zero, suggesting a bias to publish positive results for male sexual orientation measures, additional tests of robustness and publication bias indicated that the combined datasets are an unbiased representation of the studies conducted on this relationship.

Do sexual orientation effect sizes measure up to sex differences?

Contrary to previous meta-analysis, we found that digit ratios are more female-typical among homosexual and non-heterosexual men compared to heterosexual men, although the effect sizes are small ($g = -0.10$ to -0.17). Effect sizes for female sexual orientation comparisons were slightly larger, ranging from $g = 0.17$ to 0.28 (similar to Grimbos et al., $g = 0.23$–0.29). Because homosexual individuals do not exhibit the pronounced physiological and reproductive differences observed between the sexes (i.e., traits driven by prenatal androgen exposure), effect sizes for sexual orientation comparisons within sexes would be expected to be smaller than the sex difference. Hence, the strengths of associations between digit ratio and sexual orientation within sexes are consistent with the overall medium-sized sex difference in digit ratio ($g = 0.44$–0.5, i.e., see present meta-analysis on

heterosexual sex differences and a meta-analysis by Hönekopp and Watson, 2010).

The modest effect sizes among sexual orientation groups may also relate to three non-mutually exclusive factors. First, smaller effect sizes could result from heterogeneity in the biological pathways underlying sexual orientation. Same-sex orientation likely involves multiple factors, including but not limited to prenatal androgen exposure (e.g., Swift-Gallant et al., 2019; VanderLaan et al., 2022). As a result, aggregating individuals with same-sex orientation into a single group may obscure or dilute associations between digit ratios and sexual orientation. Supporting this view, prior research has found digit ratio differences within subgroups of gay men based on receptive and insertive sex roles (Swift-Gallant et al., 2021). Prior work also supports potential subgroups for female sexual orientation, such that more masculine and/or butch-identifying lesbians present with lower (more male-typical) digit ratios than female-typical or femme-identifying lesbians (reviewed in Swift-Gallant et al., 2020, 2023). Thus, effect sizes may be larger for a subgroup of homosexual males and females. Future research may consider subgroups and/or measuring markers of other biological contributors (e.g., genetics, immune activation) in addition to digit ratios to understand the development of human sexual orientation. In any case, future work should consider the effect sizes reported in the present meta-analysis when designing their studies, to ensure they are sufficiently powered.

Second, digit ratios are an imperfect proxy for prenatal androgen exposure (Swift-Gallant et al., 2020). They are likely influenced not only by prenatal androgens but also by prenatal estrogens and genetic and other factors, which must limit their sensitivity to subtle androgen variations (Swift-Gallant et al., 2020, 2023). This limitation is particularly relevant when studying men and raises the possibility of a "ceiling effect," where once prenatal androgen levels reach the male-typical range sufficient to masculinize digit ratios and/or sexual orientation, additional androgen exposure may not further influence these traits (Swift-Gallant et al., 2023). Consequently, digit ratios and sexual orientation may be more sensitive to variation in prenatal androgens among females than males. Despite these constraints, with the more accurate and precise effect sizes reported here, researchers can now conduct appropriate power analyses in future work.

Finally, while the present study extended a previous meta-analysis to assess whether bisexual individuals and/or those with intermediate Kinsey scores differ from heterosexual individuals in digit ratios, there is evidence that androgens and estrogens contribute to attraction to males and attraction to females separately (Shirazi et al., 2022; reviewed in Swift-Gallant et al., 2023). Because digit ratios may be influenced by both androgenic and estrogenic signaling (Manning et al., 1998; Zheng and Cohn, 2011), it is possible that combining androphilic and gynephilic attraction in one scale may obscure differences between sexual orientation groups. Thus, future work should consider measuring androphilia and gynephilia separately instead of as ends of a continuum.

Moderator analyses

Like Grimbos et al., we conducted moderator analyses for geographical location and measurement type. In contrast to Grimbos et al., we did not find consistent effects of geographical location for male digit ratio associations. Specifically, Grimbos et al. found that North American samples had effect sizes in the negative direction, indicating non-heterosexual men have higher digit ratios than heterosexual men, while Europe had effect sizes in the positive direction, indicating non-heterosexual men have lower digit ratios than heterosexual men. In the current analyses, geographical location emerged as a significant moderator, but none of the pairwise comparisons were significant. This may be due to a difference in the number of ethnically diverse samples and/or that these differences may not emerge with the greater number of studies. However, these results do not negate prior work suggesting geographical location/ethnicity differences in digit ratios (e.g., McFadden et al., 2005; Manning and Robinson, 2003), as the majority of these samples are still predominantly White. Geographical location did emerge as a moderator for both the right and left hand heterosexual vs. non-heterosexual women comparison in the present analyses. Left-hand pairwise comparisons indicated that the positive effects were larger for UK/Europe samples compared to North American and others. As Grimbos et al. found that geographical location did not explain more variation than did ethnicity alone, it is likely ethnicity is also driving these effects. As such, it appears critical for future work to consider ethnicity in digit ratio and sexual orientation research (Savolainen et al., 2024).

While measurement type emerged as a moderator for both sexes, pairwise comparisons were significant only for heterosexual vs. non-heterosexual male comparisons. Specifically, these results indicate that larger positive effect sizes (indicating more female-typical digit ratios among non-heterosexual than heterosexual men) were found with the photocopy/scan method than any other method. This may be due to several factors, including that with direct measures the experimenter is likely not completely blinded to the condition and/or could be distracted by their surroundings or the participant in taking the measures. Prior research has also indicated that self-report, compared to photocopy methods, yields smaller effect sizes (Manning et al., 2005; Manning et al., 2007). For these reasons, along with the benefit of including multiple blind raters, it may be advantageous for future research to consider using photocopies/scans when collecting digit ratio data, or expect to increase sample sizes to be adequately powered.

Conclusion

By examining both published and unpublished datasets, we provide a comprehensive meta-analysis between digit ratios and human sexual orientation in both males and females. These results confirm associations between digit ratios and female sexual orientation, such that same-sex-oriented women tend to have more male-typical ratios than heterosexual women, indicative of higher prenatal androgen exposure among lesbians. In contrast to prior meta-analysis, we also found that both right and left hand digit ratios differed by male sexual orientation, such that homosexual men have more female-typical ratios than heterosexual men. While sexual orientation differences in digit ratios are expected to be smaller than sex differences, we offer several possible limitations to current work that can be addressed in future research. These include the potential existence of subgroups among non-heterosexual individuals that differ in the biological factors

contributing to their sexual orientation, as well as the importance of distinguishing between androphilic and gynephilic orientations when investigating the relationship between digit ratios and male sexual orientation.

Author contributions

AS-G: Data curation, Writing – original draft, Writing – review & editing. TA: Formal analysis, Visualization, Writing – review & editing. SS: Data curation, Writing – review & editing. SB: Writing – review & editing. DP: Formal analysis, Methodology, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor LW declared a past co-authorship with the author DP.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1559158/full#supplementary-material

References

Aguilar, R. (2023). Finger-length ratios and sexual orientation in Southern Spain: a large sample study on sex-linked personality and cognitive traits. *Arch. Sex. Behav.* 52, 3301–3312. doi: 10.1007/s10508-023-02670-y

Balthazart, J. (2011). Minireview: hormones and human sexual orientation. *Endocrinology* 152, 2937–2947. doi: 10.1210/en.2011-0277

Berenbaum, S. A., Bryk, K. K., Nowak, N., Quigley, C. A., and Moffat, S. (2009). Fingers as a marker of prenatal androgen exposure. *Endocrinology* 150, 5119–5124. doi: 10.1210/en.2009-0774

Ellis, L., Lykins, A., Hoskin, A., and Ratnasingam, M. (2015). Putative androgen exposure and sexual orientation: cross-cultural evidence suggesting a modified neurohormonal theory. *J. Sex. Med.* 12, 2364–2377. doi: 10.1111/jsm.13070

Ellis, L., Lykins, A., and Ratnasingan, M. (2013). 2D: 4D, sexual orientation, east versus west: maybe It's true the twain shall never meet (on average). *Mank. Q.* 54, 115–150. doi: 10.46469/mq.2013.54.2.1

Folkierska-Żukowska, M., and Dragan, W. (2024). Biodevelopmental correlates of sexual orientation in men: evidence from a Polish sample. *Arch. Sex. Behav.* 1–14. doi: 10.1007/s10508-024-03018-w

Grimbos, T., Dawood, K., Burriss, R. P., Zucker, K. J., and Puts, D. A. (2010). Sexual orientation and the second to fourth finger length ratio: a meta-analysis in men and women. *Behav. Neurosci.* 124, 278–287. doi: 10.1037/a0018764

Hall, L. S., and Love, C. T. (2003). Finger-length ratios in female monozygotic twins discordant for sexual orientation. *Arch. Sex. Behav.* 32, 23–28. doi: 10.1023/a:1021837211630

Hall, P. A., and Schaeff, C. M. (2008). Sexual orientation and fluctuating asymmetry in men and women. *Arch. Sex. Behav.* 37, 158–165. doi: 10.1007/s10508-007-9282-0

Hines, M. (2011). Prenatal endocrine influences on sexual orientation and on sexually differentiated childhood behavior. *Front. Neuroendocrinol.* 32, 170–182. doi: 10.1016/j.yfrne.2011.02.006

Hiraishi, K., Sasaki, S., Shikishima, C., and Ando, J. (2012). The second to fourth digit ratio (2D: 4D) in a Japanese twin sample: heritability, prenatal hormone transfer, and association with sexual orientation. *Arch. Sex. Behav.* 41, 711–724. doi: 10.1007/s10508-011-9889-z

Holmes, L., Watts-Overall, T. M., Slettevold, E., Gruia, D. C., Raines, J., and Rieger, G. (2021). Sexual orientation, sexual arousal, and finger length ratios in women. *Arch. Sex. Behav.* 50, 3419–3432. doi: 10.1007/s10508-021-02095-5

Hönekopp, J., Voracek, M., and Manning, J. T. (2006). 2nd to 4th digit ratio (2D: 4D) and number of sex partners: evidence for effects of prenatal testosterone in men. *Psychoneuroendocrinology* 31, 30–37. doi: 10.1016/j.psyneuen.2005.05.009

Hönekopp, J., and Watson, S. (2010). Meta-analysis of digit ratio 2D: 4D shows greater sex difference in the right hand. *Am. J. Hum. Biol.* 22, 619–630. doi: 10.1002/ajhb.21054

Kangassalo, K., Pölkki, M., and Rantala, M. J. (2011). Prenatal influences on sexual orientation: digit ratio (2D: 4D) and number of older siblings. *Evol. Psychol.* 9, 496–508. doi: 10.1177/147470491100900402

Kostic, B., and Scofield, J. E. (2022). Sex and sexual orientation differences in sexuality and mate choice criteria. *Arch. Sex. Behav.* 51, 2855–2865. doi: 10.1007/s10508-021-02280-6

Kraemer, B., Noll, T., Delsignore, A., Milos, G., Schnyder, U., and Hepp, U. (2006). Finger length ratio (2D: 4D) and dimensions of sexual orientation. *Neuropsychobiology* 53, 210–214. doi: 10.1159/000094730

Kraemer, B., Noll, T., Delsignore, A., Milos, G., Schnyder, U., and Hepp, U. (2009). Finger length ratio (2D: 4D) in adults with gender identity disorder. *Arch. Sex. Behav.* 38, 359–363. doi: 10.1007/s10508-007-9262-4

Levin, R. N., Schudson, Z. C., Raphel, K., Takahashi, A. N., Franks, J., Fiastro, A., et al. (2023). Gender identity, sexual orientation, and the prenatal androgen theory: Reevaluating definitions, cognitive tests, and somatic markers. *Psychol. Sex. Orientat. Gend. Divers.* doi: 10.1037/sgd0000671

Li, C., Jia, M., Ma, Y., Luo, H., Li, Q., Wang, Y., et al. (2016). The relationship between digit ratio and sexual orientation in a Chinese Yunnan Han population. *Personal. Individ. Differ.* 101, 26–29. doi: 10.1016/j.paid.2016.05.043

Lippa, R. A. (2003). Are 2D: 4D finger-length ratios related to sexual orientation? Yes for men, no for women. *J. Pers. Soc. Psychol.* 85, 179–188. doi: 10.1037/0022-3514.85.1.179

Lupu, D. C., Monedero, I., Rodriguez-Ruiz, C., Pita, M., and Turiegano, E. (2023). In support of 2D: 4D: more data exploring its conflicting results on handedness, sexual orientation and sex differences. *PLoS One* 18:e0280514. doi: 10.1371/journal.pone.0280514

Manning, J. T., Churchill, A. J., and Peters, M. (2007). The effects of sex, ethnicity, and sexual orientation on self-measured digit ratio (2D: 4D). *Arch. Sex. Behav.* 36, 223–233. doi: 10.1007/s40806-017-0098-4

Manning, J. T., Fink, B., Neave, N., and Caswell, N. (2005). Photocopies yield lower digit ratios (2D: 4D) than direct finger measurements. *Arch. Sex. Behav.* 34, 329–333. doi: 10.1007/s10508-005-3121-y

Manning, J. T., Parpa, K., Kobus, M., Mason, L., and Michaelides, M. (2024). Is digit ratio (2D: 4D) a biomarker for lactate? Evidence from a cardiopulmonary test on

professional male footballers. *Early Hum. Dev.* 192:105994. doi: 10.1016/j.earlhumdev.2024.105994

Manning, J. T., and Robinson, S. J. (2003). 2nd to 4th digit ratio and a universal mean for prenatal testosterone in homosexual men. *Med. Hypotheses* 61, 303–306. doi: 10.1016/S0306-9877(03)00184-1

Manning, J. T., Scutt, D., Wilson, J., and Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum. Reprod.* 13, 3000–3004. doi: 10.1093/humrep/13.11.3000

McCormick, C. M., and Carré, J. M. (2020). Facing off with the phalangeal phenomenon and editorial policies: a commentary on Swift-Gallant, Johnson, Di Rita and Breedlove (2020). *Horm. Behav.* 120:104710. doi: 10.1016/j.yhbeh.2020.104710

McFadden, D., Loehlin, J. C., Breedlove, S. M., Lippa, R. A., Manning, J. T., and Rahman, Q. (2005). A reanalysis of five studies on sexual orientation and the relative length of the 2nd and 4th fingers (the 2D: 4D ratio). *Arch. Sex. Behav.* 34, 341–356. doi: 10.1007/s10508-005-3123-9

McFadden, D., and Shubel, E. (2002). Relative lengths of fingers and toes in human males and females. *Horm. Behav.* 42, 492–500. doi: 10.1006/hbeh.2002.1833

McIntyre, M. H. (2005). The validity of digit ratios in approximating perinatal masculinization and applications to the study of human sexuality. Cambridge, Massachusetts: Harvard University.

Miller, S. S., Hoffmann, H. L., and Mustanski, B. S. (2008). Fluctuating asymmetry and sexual orientation in men and women. *Arch. Sex. Behav.* 37, 150–157. doi: 10.1007/s10508-007-9256-2

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. *J. Clin. Epidemiol.* 134, 103–112. doi: 10.1016/j.jclinepi.2021.02.003

Puts, D. A., McDaniel, M. A., Jordan, C. L., and Breedlove, S. M. (2008). Spatial ability and prenatal androgens: meta-analyses of congenital adrenal hyperplasia and digit ratio (2D: 4D) studies. *Arch. Sex. Behav.* 37, 100–111. doi: 10.1007/s10508-007-9271-3

Puts, D., and Motta-Mena, N. V. (2018). Is human brain masculinization estrogen receptor-mediated? Reply to Luoto and Rantala. *Horm. Behav.* 97, 3–4. doi: 10.1016/j.yhbeh.2017.07.018

Putz, D. A., Gaulin, S. J., Sporter, R. J., and McBurney, D. H. (2004). Sex hormones and finger length: what does 2D: 4D indicate? *Evol. Hum. Behav.* 25, 182–199. doi: 10.1016/j.evolhumbehav.2004.03.005

Rahman, Q. (2005). Fluctuating asymmetry, second to fourth finger length ratios and human sexual orientation. *Psychoneuroendocrinology* 30, 382–391. doi: 10.1016/j.psyneuen.2004.10.006

Rahman, Q., and Koerting, J. (2008). Sexual orientation-related differences in allocentric spatial memory tasks. *Hippocampus* 18, 55–63. doi: 10.1002/hipo.20375

Rahman, Q., and Wilson, G. D. (2003). Sexual orientation and the 2nd to 4th finger length ratio: evidence for organising effects of sex hormones or developmental instability? *Psychoneuroendocrinology* 28, 288–303. doi: 10.1016/S0306-4530(02)00022-7

Richards, G., Browne, W. V., Aydin, E., Constantinescu, M., Nave, G., Kim, M. S., et al. (2020a). Digit ratio (2D: 4D) and congenital adrenal hyperplasia (CAH): systematic literature review and meta-analysis. *Horm. Behav.* 126:104867. doi: 10.1016/j.yhbeh.2020.104867

Richards, G., Wei, Y., and Hendriks, O. (2020b). Self-measure digit ratio (2D: 4D) and gender variance. *Endocr. Pract.* 26, 250–251. doi: 10.4158/1934-2403-26.2.250

Robinson, S. J., and Manning, J. T. (2000). The ratio of 2nd to 4th digit length and male homosexuality. *Evol. Hum. Behav.* 21, 333–345. doi: 10.1016/S1090-5138(00)00052-0

Savolainen, V., Bailey, N. W., Diamond, L., Swift-Gallant, A., Gavrilets, S., Raymond, M., et al. (2024). A broader cultural view is necessary to study the evolution of sexual orientation. *Nat. Ecol. Evol.* 8, 181–183. doi: 10.1038/s41559-023-02273-9

Shirazi, T. N., Self, H., Dawood, K., Welling, L. L., Cárdenas, R., Rosenfield, K. A., et al. (2021). Evidence that perinatal ovarian hormones promote women's sexual attraction to men. *Psychoneuroendocrinology* 134:105431. doi: 10.1016/j.psyneuen.2021.105431

Shirazi, T. N., Self, H., Rosenfield, K. A., Dawood, K., Welling, L. L., Cárdenas, R., et al. (2022). Low perinatal androgens predict recalled childhood gender nonconformity in men. *Psychol. Sci.* 33, 343–353. doi: 10.1177/09567976211036075

Skorska, M. N., Coome, L. A., Peragine, D. E., Aitken, M., and VanderLaan, D. P. (2021). An anthropometric study of sexual orientation and gender identity in Thailand. *Sci. Rep.* 11:18432. doi: 10.1038/s41598-021-97845-9

Smith, A. R., Hawkeswood, S. E., and Joiner, T. E. (2010). The measure of a man: associations between digit ratio and disordered eating in males. *Int. J. Eat. Disord.* 43, 543–548. doi: 10.1002/eat.20736

Swift-Gallant, A., Aung, T., Rosenfield, K., Dawood, K., and Puts, D. (2023). Organizational effects of gonadal hormones on human sexual orientation. *Adapt. Hum. Behav. Physiol.* 9, 344–370. doi: 10.1007/s40750-023-00226-x

Swift-Gallant, A., and Breedlove, S. M. (under review). Finger pointing: Why Bias against 2D:4D ratios Doesn't hold up. *Hormon. Behav.*

Swift-Gallant, A., Coome, L. A., Aitken, M., Monks, D. A., and VanderLaan, D. P. (2019). Evidence for distinct biodevelopmental influences on male sexual orientation. *Proc. Natl. Acad. Sci.* 116, 12787–12792. doi: 10.1073/pnas.1809920116

Swift-Gallant, A., Di Rita, V., Major, C. A., Breedlove, C. J., Jordan, C. L., and Breedlove, S. M. (2021). Differences in digit ratios between gay men who prefer receptive versus insertive sex roles indicate a role for prenatal androgen. *Sci. Rep.* 11:8102. doi: 10.1038/s41598-021-87338-0

Swift-Gallant, A., Johnson, B. A., Di Rita, V., and Breedlove, S. M. (2020). Through a glass, darkly: human digit ratios reflect prenatal androgens, imperfectly. *Horm. Behav.* 120:104686. doi: 10.1016/j.yhbeh.2020.104686

Swift-Gallant, A., Shirazi, T., Puts, D. A., and Breedlove, S. M. (2022). Evidence for perinatal steroid influence on human sexual orientation and gendered behavior. *Cold Spring Harb. Perspect. Biol.* 14:a039123. doi: 10.1101/cshperspect.a039123

Tortorice, J. L. (2002). Written on the body: Butch/femme lesbian gender identity and biological correlates. Rutgers, The State University of New Jersey, School of Graduate Studies.

van Anders, S. M., and Hampson, E. (2005). Testing the prenatal androgen hypothesis: measuring digit ratios, sexual orientation, and spatial abilities in adults. *Horm. Behav.* 47, 92–98. doi: 10.1016/j.yhbeh.2004.09.003

Van Hemmen, J., Cohen-Kettenis, P. T., Steensma, T. D., Veltman, D. J., and Bakker, J. (2017). Do sex differences in CEOAEs and 2D: 4D ratios reflect androgen exposure? A study in women with complete androgen insensitivity syndrome. *Biol. Sex Differ.* 8, 1–10. doi: 10.1186/s13293-017-0132-z

VanderLaan, D. P., Skorska, M. N., Peragine, D. E., Coome, L. A., Moskowitz, D. A., Swift-Gallant, A., et al. (2022). "Carving the biodevelopment of same-sex sexual orientation at its joints" in Gender and sexuality development: contemporary theory and research (Cham: Springer International Publishing), 491–537.

Vásquez-Amézquita, M., Leongómez, J. D., Seto, M. C., Bonilla, F. M., Rodríguez-Padilla, A., and Salvador, A. (2018). No relation between digit ratio (2D: 4D) and visual attention patterns to sexually preferred and non-preferred stimuli. *Personal. Individ. Differ.* 120, 151–158. doi: 10.1016/j.paid.2017.08.022

Veloso, V., Miranda, A. C., Rodrigues, C. N. C., Medrado, N. C., Nunes, M. C. S., Silva Júnior, M. D., et al (2024). Testosterone Concentrations and 2D: 4D Digit Ratio in Heterosexual and Masculine and Feminine Lesbian Women. *Adapt. Human Behav. Physiol.* 10, 284–302. doi: 10.1007/s40750-024-00248-z

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48. doi: 10.18637/jss.v036.i03

Voracek, M., Manning, J. T., and Ponocny, I. (2005). Digit ratio (2D: 4D) in homosexual and heterosexual men from Austria. *Arch. Sex. Behav.* 34, 335–340. doi: 10.1007/s10508-005-3122-x

Wallien, M. S., Zucker, K. J., Steensma, T. D., and Cohen-Kettenis, P. T. (2008). 2D: 4D finger-length ratios in children and adults with gender identity disorder. *Horm. Behav.* 54, 450–454. doi: 10.1016/j.yhbeh.2008.05.002

Watts, T. M., Holmes, L., Raines, J., Orbell, S., and Rieger, G. (2018). Finger length ratios of identical twins with discordant sexual orientations. *Arch. Sex. Behav.* 47, 2435–2444. doi: 10.1007/s10508-018-1262-z

Williams, T. J., Pepitone, M. E., Christensen, S. E., Cooke, B. M., Huberman, A. D., Breedlove, N. J., et al. (2000). Finger-length ratios and sexual orientation. *Nature* 404, 455–456. doi: 10.1038/35006555

Xu, X., Coats, J. K., Yang, C. F., Wang, A., Ahmed, O. M., Alvarado, M., et al. (2012). Modular genetic control of sexually dimorphic behaviors. *Cell* 148, 596–607. doi: 10.1016/j.cell.2011.12.018

Xu, Y., Norton, S., and Rahman, Q. (2019). Early life conditions and adolescent sexual orientation: a prospective birth cohort study. *Dev. Psychol.* 55, 1226–1243. doi: 10.1037/dev0000704

Xu, Y., and Zheng, Y. (2016). The relationship between digit ratio (2D: 4D) and sexual orientation in men from China. *Arch. Sex. Behav.* 45, 735–741. doi: 10.1007/s10508-015-0535-z

Yule, M. A., Brotto, L. A., and Gorzalka, B. B. (2014). Biological markers of asexuality: handedness, birth order, and finger length ratios in self-identified asexual men and women. *Arch. Sex. Behav.* 43, 299–310. doi: 10.1007/s10508-013-0175-0

Zheng, Z., and Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci.* 108, 16289–16294. doi: 10.1073/pnas.1108312108

frontiers | Frontiers in Psychology

TYPE Brief Research Report
PUBLISHED 22 May 2025
DOI 10.3389/fpsyg.2025.1447640

OPEN ACCESS

EDITED BY
Lisa L. M. Welling,
Oakland University, United States

REVIEWED BY
Stefan Mattias Maria Goetz,
Wayne State University, United States
Ella Brown,
University of Michigan, United States

*CORRESPONDENCE
Randy Corpuz
✉ randy.corpuz@pepperdine.edu

RECEIVED 11 June 2024
ACCEPTED 06 May 2025
PUBLISHED 22 May 2025

CITATION
Donovan RL and Corpuz R (2025) The dual-hormone hypothesis and first-time fathers' relationship satisfaction at 3 months postpartum.
Front. Psychol. 16:1447640.
doi: 10.3389/fpsyg.2025.1447640

The dual-hormone hypothesis and first-time fathers' relationship satisfaction at 3 months postpartum

Rylei L. Donovan[1] and Randy Corpuz[2]*

[1]Department of Psychology, University of Massachusetts Boston, Boston, MA, United States, [2]Department of Psychology, Pepperdine University, Malibu, CA, United States

Human males face tradeoffs in how they invest resources toward mating and parenting. Research on male's transition to fatherhood has revealed shifts in hormones tied to these tradeoffs. While work has focused on the influence of hormones on parenting during this stage, less is known about how these hormones influence mating (i.e., relationship functioning with partner) in the postnatal period. A father's relationship satisfaction is expected to be related to endocrine activity across the transition to parenthood. We predicted that first-time fathers with high testosterone (T) would report lower relationship satisfaction. We expected this effect to be amplified (moderation) for those males with lower cortisol (CORT) levels (i.e., dual hormone hypothesis). At 3 months postpartum we measured salivary T and CORT (n = 220) and recorded relationship satisfaction using the Investment Model Scale (IMS). We found that fathers with high T and low CORT had the highest relationship satisfaction. While the effect was small, these findings ran counter to our predictions. We speculate that higher T and lower CORT males may report increased satisfaction as they support, retain, and secure additional opportunities from a mate who recently demonstrated her ability (and willingness) to produce offspring. Discussion focuses on numerous limitations of the study, small effect size, and the need for replication with less homogenous samples.

KEYWORDS

fathers, testosterone, cortisol, dual hormone, relationships

Overview

Psychobiological research has progressed from studying the functional role of single hormones to a more nuanced understanding of how hormones interact to facilitate behavior. Recent work using the dual hormone hypothesis (DHH) finds that testosterone (T) and cortisol (CORT) interact to regulate behaviors in specific domains (Mehta and Josephs, 2010). According to the DHH, positive associations between T and behaviors conventionally associated with dominance, competition, and status-seeking increase when CORT levels are low (Mehta and Prasad, 2015; see Dekkers et al., 2019 for meta-analysis). While research on the association between individual hormones and romantic relationships has grown in recent years (see Edelstein, 2022 for review), predictions using the DHH remain untested in the domain of romantic relationships. Extant DHH work has also yet to fully leverage known windows of neuroendocrine fluctuations in baseline levels of CORT and/or T—e.g., first-time fatherhood (Gettler et al., 2011; Saxbe et al., 2017a; Saxbe et al., 2017b; Kuo et al., 2018)—and, instead, focus (mostly) on hormone reactivity in response to laboratory tasks/stimuli (Dekkers et al., 2019). In this study, we focus on the transition to first time fatherhood and the role of T, CORT, and the DHH in a new father's relationship satisfaction with his partner.

Literature review

T and relationships

T is the major output of the hypothalamic–pituitary-gonadal (HPG) axis and fluctuates in response to developmental milestones such as forming new romantic relationships or the transition to parenthood (Grebe et al., 2019). The role of T in forming and maintaining relationships is a common focus of psychobiological work (Saxbe et al., 2017a; Saxbe et al., 2017b; Cárdenas et al., 2023; see Edelstein, 2022). Research has also focused on T as facilitating male dominance and aggression as well as being antagonistic to forming/maintaining long term pair bonds (Gray et al., 2020). For example, T is associated with more lifetime sexual partners and decreased relationship satisfaction (Dhillon et al., 2020; van der Meij et al., 2019; Tackett et al., 2014; Welker et al., 2014). Edelstein et al. (2014) measured relationship satisfaction and commitment in heterosexual couples and salivary T. Findings suggested that a woman's relationship satisfaction is negatively associated with their partner's T. The idea that T is antagonistic to nurturing behavior is commonly presented as an explanation for why higher T has been correlated with lower relationship satisfaction. Dhillon et al. (2020) examined associations between T, perceived partner accommodation, and conversation satisfaction during a relationship stressor task. The findings of this study found that high T was negatively related to perceived partner accommodation during the task. Overall, evidence suggests that high T is negatively related to relationship functioning.

CORT and relationships

CORT is produced by the hypothalamic–pituitary–adrenal (HPA) axis and is often examined in association with depression, anxiety, aggression, and stress (Gordis et al., 2006; Galbally et al., 2019). More recently, CORT has also been found to be related to relationship support and nurturing behaviors. For example, paternal CORT is significantly lower in fathers than non-father male controls (Berg and Wynne-Edwards, 2001). Lower CORT is also related to higher quality caregiving for fathers (Beijers et al., 2022). Other literature suggests that higher paternal CORT on the day of their infant's birth, the day following, and after first holding their infant is correlated with more paternal care (Kuo et al., 2018). Braren et al. (2020) found that, during pregnancy, low paternal CORT acts as a buffer for maternal CORT in mothers that report high stress. Taken together, high CORT is negatively correlated with relationship functioning. While CORT—relative to T—is not as commonly studied for its relationship with supportive and nurturing behaviors, research on CORT moderating T in close relationships is (to our knowledge) wholly absent from the literature to date.

T × CORT interaction

Aside from CORT, the HPA axis is also regulated by androgens—most notably T (Bingham et al., 2011; Chen et al., 2014). For example, androgens have both activational and inhibitory actions on the HPA axis (Zuloaga et al., 2024). The integrated and reciprocal interactions between the HPA and HPG axes has been documented through endocrine manipulations in animal research (Viau, 2002). For example, increases in CORT have inhibitory effects on the release of sex steroids specific to mating and reproduction (Tilbrook et al., 2000). Conversely, male T inhibits the HPA axis' response to relational stressors (Viau and Meaney, 1996). Altogether, these interactions suggest the possibility of a T × CORT interaction being related to romantic relationships where the HPA and HPG axes have been implicated (separately) in humans.

The DHH suggests that the influence of T on behavior is amplified when levels of CORT are low (Mehta and Josephs, 2010). The idea that CORT modulates the androgenicity of T has generated novel findings of varying effect sizes (see Dekkers et al., 2019). However, discoveries are constrained to categories of behavior (e.g., dominance) putatively antagonistic to maintaining close relationships in socially monogamous, paternally investing species like humans (Donovan et al., 2023) and have yet to generate systematic research in domains related to close relationships. As noted above, elevated levels of T are generally associated with poor relationship satisfaction. It is an open question as to whether T's negative association with relationship satisfaction may be more pronounced among human males who also have lower levels of CORT.

Hypotheses

In integrating literature on T, CORT, and relationship functioning with extant DHH research, our primary prediction is that relationship satisfaction will be lowest in individuals with high T and low CORT (i.e., CORT will moderate the relationship between T and relationship satisfaction). Aligned with the DHH literature, we also test for main effects of T and CORT (individually) on relationship satisfaction. We predict that males with high T will report lower satisfaction in their romantic relationships and that males with high CORT will also have lower satisfaction in their romantic relationships.

Methods

Overview and study design

The secondary data used in this study is from a previously completed longitudinal study on paternal postpartum health outcomes for first-time fathers (see Corpuz et al., 2021). All data used for the current report was collected approximately 3 months following childbirth (M = 96.07 days, SD = 16.48 days). All materials and procedures were reviewed and approved by the University's Institutional Review Board (IRB). Participants were provided information on the risks and benefits of participating in this research, signed consent forms prior to data collection, and were compensated for their contributions. The data for this study was collected between 2013 and 2015 and all saliva assays (T and CORT) were conducted between 2014 and 2015.

Participants

First-time fathers (n = 220) completed self-report measures and submitted saliva samples. Fathers were recruited from multiple

sources: hospital birthing or community lactation classes (62.7%), midwife referrals (15.7%), social media ads (13.6%), or community Baby Basics class (2.2%). The remaining 6% of the sample did not report a recruitment source. All participants were residing in Southern California (U.S.A.) at the time of data collection.

As noted in Corpuz et al. (2021), the average age of fathers in this study was $M = 32.9$, $SD = 5.4$, 84.1% of this sample was married to their child's mother (all but one couple reported cohabitating at time of study) and 77.4% of these fathers held at least a college degree. The median income of this sample was $50,000 to $75,000. Fathers self-reported their race/ethnicity as White (70.6%), Latino/Hispanic (12%), Asian American (5.2%), Black/African American (1.7%), Native American (1.3%), multiracial (2.6%), and other (3.9%).

Materials and procedure

Participants completed self-report questionnaires during pre-planned home visits. Following the completion of self-report measures, home visitors trained fathers on how to expectorate saliva through a simulated collection procedure using the exact materials they would use on the day of sampling. Parents were provided with pre-labeled saliva kits (sterile cotton swabs, polypropylene tubes, written instructions, and Ziploc bags) and a video demonstrating the process in detail.

Saliva collection

Fathers were instructed to expectorate saliva "within 30 min of waking up" during their next day off from work where applicable (i.e., a weekend day for most parents; see Corpuz et al., 2021). The specific day of sample selection (and subsequent sample retrieval) was agreed upon between the home visitor and the participant. Mean sampling times across participants was 6:47 am (SD = 1:09).

Fathers were told to abstain from alcohol (12 h prior), all food (1 h prior), and any beverages containing sugar, acid, or caffeine (5 min prior) leading up to their morning sample as per Granger et al. (2012). During sampling, fathers placed a sterilized absorbent cotton swab underneath their tongue for a minimum of 120 s. They then directed the swab into a polypropylene tube (using their tongue) and placed the tube into a freezer safe bag and into the freezer until samples were retrieved by a home visitor. Home visitors retrieved saliva samples from parents within 7 days of each visit[1]. All samples were retrieved from participants, inventoried, and frozen at−50°C for up to 90 days and were then shipped on dry ice.

Saliva assays

Samples were assayed in duplicate at the Institute for Interdisciplinary Salivary Bioscience Research (IISBR; Arizona State University) using a highly sensitive competitive enzyme immunoassay (EIA) without modifications to the recommended protocols from Salimetrics (Carlsbad, CA).

Testosterone (T)

The test volume for T assay was 25 µL, and range of sensitivity was from 1.0 to 600 pg./mL. On average the inter- and intra-assay coefficients of variation were less than 15 and 10%, respectively. Reagents were stored at 2–8 degrees (C); reagents and samples were completed without interruption across a 96-well microtiter plate coated with polyclonal anti-T antibodies. The full assay protocol can be downloaded directly from the manufacturer: https://salimetrics.com/wp-content/uploads/2018/03/testosterone-saliva-elisa-kit.pdf.

Cortisol (CORT)

For CORT, the assay range of sensitivity was 0.004–3.0 µg/dL. The detection limit was 0.018 µg /dl (after accounting for extraction dilution). On average the inter- and intra-assay coefficients of variation were less than 10 and 5%, respectively. Reagents were stored at 2–8 degrees (C); reagents and samples were completed without interruption across a 96-well microtiter plate coated with monoclonal anti-CORT antibodies. The full assay protocol can be downloaded directly from the manufacturer: https://salimetrics.com/wp-content/uploads/2018/03/salivary-CORT-elisa-kit.pdf.

Missing data: T and CORT

Eleven fathers were missing data for both CORT and T (six fathers provided insufficient quantity; four fathers had a concentration below lower limit of sensitivity; one father dropped out of the study prior to saliva retrieval). These cases are retained in analyses following maximum likelihood estimation to address missingness for CORT and T. There were two fathers with outlying values (> 3SDs) for CORT and one father with an outlying value for T. These three cells were replaced as missing and retained in analyses following maximum likelihood estimation to address missingness for these variables (see Corpuz et al., 2021).

Relationship satisfaction

To measure relationship satisfaction participants completed the satisfaction subscale of the Investment Model Scale (IMS; Rusbult et al., 1998). This relationship satisfaction subscale is a widely used self-report questionnaire with high internal consistency and acceptable validity in the literature (Edelstein et al., 2014; Saxbe et al., 2017a; Saxbe et al., 2017b). The subscale is composed of 10 items scored on a Likert scale. The first five items have scores from 0: "Do not Agree At All" to 4: "Agree Completely" with items such as "My partner fulfills my needs for companionship (doing things together, enjoying each other's company, etc.)" and "My partner fulfills my needs for security (feeling trusting, comfortable in a stable relationship, etc.)." The last five items have scores from 0: "Do Not Agree At All" to 8: "Agree Completely" with items such as "My relationship is much better than others' relationships" and "My relationship is close to ideal." In the current sample of fathers, the scale was highly reliable (Cronbach's[2] $\alpha = 0.93$).

1 78.4% of all samples retrieved were collected within 7 days of participant expectorating saliva. There were no differences in assay values from samples picked up within 7 days compared to samples retrieved after 8 + days (all *ps*>0.41).

2 The average covariance divided by the average variance across all 10 items of the scale.

Results

Primary analyses for this report were executed using a structural equation modeling (SEM) framework which includes a robust maximum likelihood estimation (MLE) missing data module in AMOS v.27 (IBM Chicago; Arbuckle, 2019).

Covariates

Demographic covariates

No differences were observed in study variables due to marital status (*p*s>0.79), household income (*p*s>0.61), or self-reported ethnicity (*p*s>0.68).

Endocrine covariates

A series of bivariate correlations were tested to evaluate covariates[3] for inclusion in models: BMI, father's age, and time of morning saliva sample was collected. Fathers' BMI was not associated with their morning T ($r = -0.01$, $p = 0.93$) nor morning CORT ($r = -0.02$, $p = 0.74$). While paternal age was not related to T ($r = -0.07$, $p = 0.32$), it was significantly correlated with morning CORT ($r = 0.20$, $p = 0.004$) and, as a result, age was included as a covariate in subsequent models that included morning CORT. Exact time of morning sample was not related to father's morning T ($r = 0.04$, $p = 0.54$) but was correlated with paternal morning CORT ($r = -0.15$, $p = 0.04$). Subsequent CORT models include time of morning sample as a covariate.

There were 27 fathers that self-reported smoking tobacco which can influence salivary assay values (see Granger et al., 2012). However, smoking status was not related to T ($p = 0.18$) nor CORT ($p = 0.63$) in this sample of fathers. Three fathers reported taking medications with documented effects on either CORT or T production (e.g., aromatase inhibitor). We elected for a conservative approach to handling these medications in our analyses; values for T and CORT for all three fathers were removed and replaced as missing[4].

Missing data

Overall, missingness for the variables tested in this study were moderate (0–6.1%) (Little and Rubin, 2019). To adjust for biases due to missing data, we fitted all models using the maximum likelihood estimation (MLE) missing data module in AMOS v.22. Data were missing completely at random (MCAR): Little's MCAR test ($p = 0.68$).

Hypothesis testing

In this sample, we did not find evidence that paternal T predicted a father's self-reported relationship satisfaction (β = 0.04, p = 0.58).

TABLE 1 Regression model of testosterone T x CORT interaction predicting self-reported relationship satisfaction (n = 220 fathers).

	B	*β*	SE	CR	*p*
Relationship satisfaction					
T	0.01	0.04	0.02	0.55	0.582
CORT	−2.49	−0.05	3.93	−0.63	0.527
T x CORT	−1.15	−0.13	0.62	−1.87	0.062

B indicates unstandardized regression coefficient. β indicates standardized regression coefficients. SE-standard error. CR-critical ratio.

We moved on to test a model whereby paternal CORT predicted a father's self-reported relationship satisfaction (covarying for age and time of morning sample). In this CORT model, we also did not find evidence of a relationship between CORT and self-reported relationship satisfaction ($\beta = -0.05$, $p = 0.53$).

DHH

To explore the relationship between the DHH and relationship satisfaction, we created a T × CORT interaction term (i.e., multiplied standardized values of each) and tested a model where this new variable predicted self-reported relationship satisfaction while covarying for the following variables: T, CORT, age, and time of day.

We found a small, negative effect for dual hormone influence on relationship satisfaction (β = −0.13, p = 0.06) (See Table 1 for regression estimates). We created a simple-slope graph[5] (Figure 1) to aid in interpretation of this small (non-significant) effect. Males with high T and low CORT were more satisfied in their relationships compared to males with lower T[6].

Discussion

We expected to find that first-time fathers with high T and low CORT would report lower relationship satisfaction. Instead, our results suggest that this group of fathers have higher relationship satisfaction. Although these findings provide evidence for the DHH being applicable to studies of relationship satisfaction, they are exactly opposite of our prediction. While we speculate on the nature of this small effect below, we caution readers that this result did not reach conventional levels of statistical significance.

Aside from our findings specific to the DHH, we also did not find that T (main effect) nor CORT (main effect) predicted paternal relationship satisfaction. It is not uncommon to find null relationships between single hormones and a given outcome variable while finding

3 Salivary endocrine data collection is vulnerable to contextual influences (Granger et al., 2012).

4 Most other medications listed by fathers in this sample were for allergies (e.g., OTC nasal spray, decongestants), pain (e.g., ibuprofen), or digestive issues (e.g., acid reflux).

5 As per Dekkers et al. (2019), a negative TxCORT interaction effect is conventionally evidence for dual hormone effects on the outcome variable and should be probed further for interpretation.

6 Johnson-Neyman intervals were explored using the PROCESS (v4.2) macro (Hayes, 2022) on SPSS (V.30). This technique estimates the values of the moderator (CORT) for which the slope of the predictor (T) on the outcome (relationship satisfaction) will be statistically significant. The corresponding Johnson-Neyman procedure estimated that at values of CORT below 0.48, relationship satisfaction was associated with higher T. No significant associations were observed for CORT levels above 0.48.

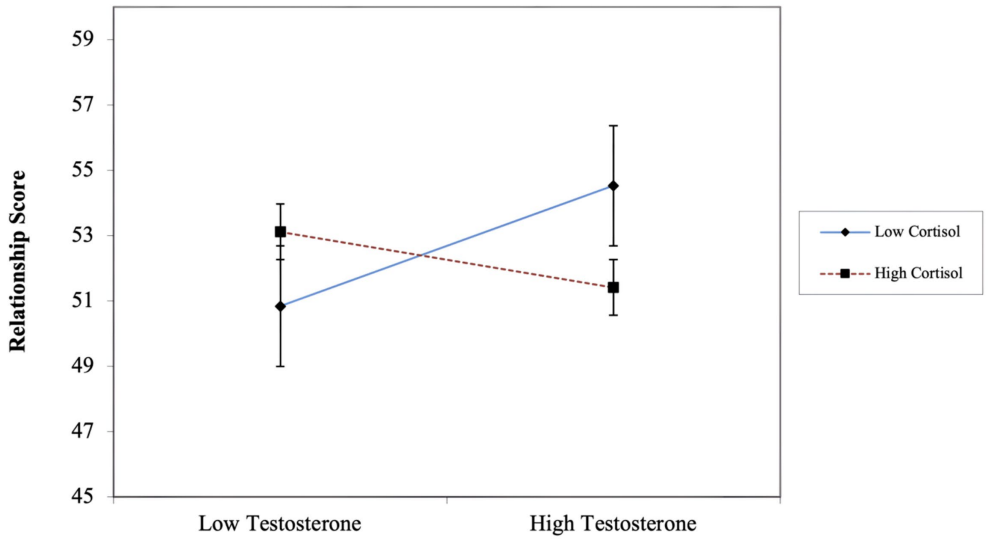


FIGURE 1
DHH interaction: relationship satisfaction as a function of the interactive effects of T and CORT. Low = 1 SD below mean; high = 1 SD above mean. Intercept and slopes were used to plot relationship satisfaction scores one SD above and below means for T and CORT. Error bars = ± 1 SEM.

effects for the DHH—one of the strengths of a DHH approach. For example, work on T and CORT modulating risk-taking found no significant relationship between each individual hormone and risk-taking but the dual-hormone interaction was found to be significantly related to risk-taking (Mehta et al., 2015). Similarly, dual-hormone research on psychopathy identified a significant relationship between psychopathy and the dual-hormone interaction variable but not with T nor CORT alone (Roy et al., 2019). This pattern is suggestive of how important research on the interaction between the HPA and HPG axis is to fully understand the connectedness of hormones and behavior in interpersonal relationships.

Mating effort

The effect we report here, while small, requires some speculation to interpret. Traits conventionally associated with high T—dominance, status-seeking, aggression—have been the focus of DHH research and are the same behaviors historically categorized as traits that collectively index one's mating effort (e.g., Muller and Pilbeam, 2017; Grebe et al., 2019). In this U.S. sample of (mostly) married males, we speculate that mating effort may go beyond behaviors that increase a male's competitiveness to acquire access to mates but might also capture a psychology aimed at retaining one's current mate (Barbaro et al., 2016; Salkicevic et al., 2014). While our original interpretation of the literature's "high T, increased mating effort" relationship led us to predict that high T/low CORT fathers would be less satisfied with their current partner, it is possible that— in humans, a serially monogamous, biparental species (see Donovan et al., 2023 for review)—higher relative androgenicity (specifically, in the postnatal period) may facilitate increased effort directed at one's current mate. Higher T males may self-report increased satisfaction as they support, retain, secure, elicit additional opportunities from a mate who recently demonstrated her ability (and willingness) to produce offspring. We reiterate the speculative nature of this idea as current theories cannot easily accommodate this possibility. For example, Roney and Gettler's (2015) model of T modulation within romantic relationships expects that increased baseline T helps males target partners for committed relationships but subsequent elevations in T within a committed relationship may facilitate additional mate seeking (e.g., van Anders et al., 2007). Past research suggests that male T declines after relationship formation and declines further following the arrival of offspring (Gettler et al., 2013). Future research with improved operationalization of mating effort—particularly acquisition vs. retention—is needed.

Masculinity

The transition to fatherhood presents men with a new identity as fathers. Across cultures the attainment of fatherhood status entails meeting different expectations (Enderstein and Boonzaier, 2015). Involved fathers that take responsibility for paternity and the duties of paternal care are presented with an alternate kind of masculinity (Dunn and Maharaj, 2023; Enderstein and Boonzaier, 2015; Plantin et al., 2003). It is possible that the perception of fatherhood as an enhanced masculine status may help explain the unexpected finding.

Sexual satisfaction

Sexual satisfaction may further explain this contradictory finding. Sexual activity has been shown to be correlated with higher T, having bidirectional effects in men (van Anders et al., 2007). Fathers who experience milder T declines in response to childbirth have been shown to maintain higher frequencies of sexual activity postpartum (Gettler et al., 2013). Postpartum sexual expectations differ between mothers and fathers (Santtila et al., 2007). This difference between desired amount and actual frequency of sexual behaviors within the relationship dyad can be distressing, indicating that fathers reporting greater sexual satisfaction (i.e., alignment between sexual desire and frequency) may experience lower stress (indexed by CORT) in their

relationships (Ahlborg et al., 2005). Future work should include sexual satisfaction to further understand this contradictory finding.

Limitations

Measurement of relationship satisfaction

Satisfaction is a commonly studied variable in the study of close relationships due to the significant impact that it has on downstream behavior (Dhillon et al., 2020). Relationship satisfaction is often collected in the form of a self-report measure where individuals in romantic relationships respond to a questionnaire regarding how they feel about their relationship. Results from these measures have been used as a therapeutic aid (Callaci et al., 2021), to predict postpartum relationship investment (Saxbe et al., 2017a; Saxbe et al., 2017b) or even to assess dyadic endocrine interactions within couples (Edelstein et al., 2014). In this study, we only collected data specific to the relationship satisfaction subscale from Investment Model Scale (Rusbult et al., 1998). The full scale additionally measures commitment level, quality of alternatives, and investment size. While our interests in the current study were exclusively on satisfaction, a more complete understanding of how T, CORT, and their interaction may influence general relationship functioning is an important further consideration. It is possible that the relationships that we uncovered (or did not find) in this study would be different based on our selection of different relationship functioning measures or if we used measures that capture dimensions of relationship functioning beyond mere satisfaction (e.g., close relationship behavior; see Jaremka and Collins, 2017).

Hormone collection

Overall, the precision of our results in this study is notably constrained by the number of collection periods and the timing of the morning sample. The salivary analyses in this study only include one saliva sample per participant on a single day in the postnatal period. Despite reliability in T across consecutive days (Dabbs, 1990), our measure of T would have been more precise if we also had assays across consecutive days. Likewise, the same would apply to CORT; despite demonstrated stability across multiple days (Wang et al., 2014), a single collection day is inadequate to securely propose that our findings might replicate in future studies. In addition, the samples were collected within 30 min of awakening but not at the exact moment of waking up which prevents modeling of the awakening response of T and CORT (Kuzawa et al., 2016). Researchers interested in the neuroendocrine functioning of the family unit are thus highly encouraged to plan to collect significantly more samples at highly precise intervals both within and across days (see Kuzawa et al., 2016).

Sample characteristics

Another limitation to this study is that we did not collect data on the length of the couple's relationship. It is likely that reports of relationship satisfaction during the novel stress of transitioning to parenthood will be related to how long ago the relationship began (Farrelly et al., 2015). Additionally, postpartum anxiety and depression were not considered which would have potential to impact paternal hormones. This sample consisted solely of heterosexual couples, predominantly White, well-educated, and middle- to high-income. Although this limits the generalizability of the results, more recent work suggests that there are correlations between T and relationship satisfaction self-reports in parents during the postpartum period in more diverse samples (Cárdenas et al., 2023). Lastly, the inclusion of maternal data (hormones and/or self-reported relationship satisfaction) in future work should add considerable nuance in our understanding how hormones and relationship functioning are related within a dyad (see Edelstein et al., 2014).

Biological meaningfulness and statistical significance

This paper uses secondary data from previously published research that focused on T and fatherhood (Corpuz et al., 2021). While the larger project was not originally designed to test for the DHH, the small effect that ran counter to predictions among this community sample may provide valuable insight for scientists interested in applying the DHH to romantic relationships—unexplored terrain in literature that has largely focused on male dominance and risk-taking (Dekkers et al., 2019). A more pressing concern than conventional statistical significance is whether the observed effect size is biologically meaningful. Future experimental work (i.e., exogenous T administration) may reveal thresholds for when CORT meaningfully modulates T in the domain of romantic relationships in a manner predicted by the DHH or other extant neuroendocrine theories. In reporting on and speculating about the direction of the small effect size in this correlational design, we aim to promote further study (experimental, cross-sectional, longitudinal) on the role of hormones in human relationship functioning.

Conclusion

The novel findings of this report require replication. To our knowledge this is the first study to apply the DHH to relationship satisfaction. Future research would greatly benefit from conducting this study on a longitudinal timeline to increase the number of collection periods for more in-depth analyses of romantic relationship dynamics for parents in the postpartum period. It is important to understand how interactions within the endocrine system are contributing to the postpartum experience. The role of endocrine interactions for T and CORT between as well as within individuals is an important future direction for this work. Future studies have the potential to uncover what makes some romantic relationships more resilient to the challenges of the transition to parenthood than others.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the data that support the findings of this study are available from the corresponding author RC, upon reasonable request. Requests to access these datasets should be directed to randy.corpuz@pepperdine.edu.

Ethics statement

The studies involving humans were approved by University of California Santa Barbara IRB. The studies were conducted in accordance with the local legislation and institutional requirements.

The participants provided their written informed consent to participate in this study.

Author contributions

RD: Conceptualization, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. RC: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by an NSF graduate fellowship (DGE-1144085) which funded the second author's graduate education. Original data collection was funded by a research award (BCS-1147671).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahlborg, T., Dahlöf, L. G., and Hallberg, L. R. M. (2005). Quality of the intimate and sexual relationship in first-time parents six months after delivery. *J. Sex Res.* 42, 167–174. doi: 10.1080/00224490509552270

Arbuckle, J. L. (2019). *IBM SPSS Amos 27 user's guide*. [Computer Program]. Amos development corporation, SPSS Inc.

Barbaro, N., Shackelford, T. K., and Weekes-Shackelford, V. A. (2016). Mothers and fathers perform more mate retention behaviors than individuals without children. *Hum. Nat.* 27, 316–333. doi: 10.1007/s12110-016-9261-z

Beijers, R., Breugelmans, S., Brett, B., Willemsen, Y., Bos, P., and de Weerth, C. (2022). CORT and testosterone concentrations during the prenatal and postpartum period forecast later caregiving quality in mothers and fathers. *Horm. Behav.* 142:105177. doi: 10.1016/j.yhbeh.2022.105177

Berg, S. J., and Wynne-Edwards, K. E. (2001) Changes in testosterone, cortisol, and estradiol levels in men becoming fathers. *Mayo Clin Proc.* 76:582–592. doi: 10.4065/76.6.582

Bingham, B., Gray, M., Sun, T., and Viau, V. (2011). Postnatal blockade of androgen receptors or aromatase impair the expression of stress hypothalamic-pituitary-adrenal axis habituation in adult male rats. *Psychoneuroendocrinology* 36, 249–257. doi: 10.1016/j.psyneuen.2010.07.015

Braren, S. H., Brandes-Aitken, A., Ribner, A., Perry, R. E., and Blair, C. (2020). Maternal psychological stress moderates diurnal CORT linkage in expectant fathers and mothers during late pregnancy. *Psychoneuroendocrinology* 111:104474. doi: 10.1016/j.psyneuen.2019.104474

Callaci, M., Vaillancourt-Morel, M. P., Labonté, T., Brassard, A., Tremblay, N., and Péloquin, K. (2021). Attachment insecurities predicting romantic disengagement over the course of couple therapy in a naturalistic setting. *Couple Family Psychol.* 10, 54–69. doi: 10.1037/cfp0000158

Cárdenas, S. I., Tse, W., León, G., Kim, A., Tureson, K., Lai, M., et al. (2023). Prenatal testosterone synchrony in first-time parents predicts fathers' postpartum relationship quality. *Horm. Behav.* 156:105440. doi: 10.1016/j.yhbeh.2023.105440

Chen, C. V., Brummet, J. L., Lonstein, J. S., Jordan, C. L., and Breedlove, S. M. (2014). New knockout model confirms a role for androgen receptors in regulating anxiety-like behaviors and HPA response in mice. *Horm. Behav.* 65, 211–218. doi: 10.1016/j.yhbeh.2014.01.001

Corpuz, R., D'Alessandro, S., and Collom, G. K. (2021). The postnatal testosterone rebound in first-time fathers and the quality and quantity of paternal care. *Dev. Psychobiol.* 63, 1415–1427. doi: 10.1002/dev.22064

Dabbs, J. M. (1990). Salivary testosterone measurements: reliability across hours, days, and weeks. *Physiol. Behav.* 48, 83–86. doi: 10.1016/0031-9384(90)90265-6

Dekkers, T. J., van Rentergem, J. A. A., Meijer, B., Popma, A., Wagemaker, E., and Huizenga, H. M. (2019). A meta-analytical evaluation of the dual-hormone hypothesis: does CORT moderate the relationship between testosterone and status, dominance, risk taking, aggression, and psychopathy? *Neurosci. Biobehav. Rev.* 96, 250–271. doi: 10.1016/j.neubiorev.2018.12.004

Dhillon, A., Denes, A., Crowley, J. P., Ponivas, A., Winkler, K. L., and Bennett, M. (2020). Does testosterone influence young adult romantic Partners' accommodation during conversations about stressors? *Hum. Commun. Res.* 46, 444–469. doi: 10.1093/hcr/hqaa008

Donovan, R. L., Kotov, D. A., and Corpuz, R. (2023). "Pair bonding and paternal care" in Encyclopedia of sexual psychology and behavior (Cham: Springer International Publishing), 1–5.

Dunn, S., and Maharaj, P. (2023). More dedicated, more loyal: young men redefining fathering roles and expectations. *J. Fam. Stud.* 29, 2133–2152. doi: 10.1080/13229400.2022.2142151

Edelstein, R. S. (2022). Testosterone tradeoffs in close relationships. *Advance. Experiment. Soc. Psychol.* 65, 235–280. doi: 10.1016/bs.aesp.2021.11.004

Edelstein, R. S., van Anders, S. M., Chopik, W. J., Goldey, K. L., and Wardecker, B. M. (2014). Dyadic associations between testosterone and relationship quality in couples. *Hormones Behav* 65, 401–407. doi: 10.1016/j.yhbeh.2014.03.003

Enderstein, A. M., and Boonzaier, F. (2015). Narratives of young south African fathers: redefining masculinity through fatherhood. *J. Gend. Stud.* 24, 512–527. doi: 10.1080/09589236.2013.856751

Farrelly, D., Owens, R., Elliott, H. R., Walden, H. R., and Wetherell, M. A. (2015). The effects of being in a "new relationship" on levels of testosterone in men. *Evol. Psychol.* 13, 250–261. doi: 10.1177/147470491501300116

Galbally, M., van Rossum, E. F., Watson, S. J., de Kloet, E. R., and Lewis, A. J. (2019). Transgenerational stress regulation: mother-infant CORT and maternal mental health across the perinatal period. *Psychoneuroendocrinology* 109:104374. doi: 10.1016/j.psyneuen.2019.104374

Gettler, L. T., McDade, T. W., Feranil, A. B., and Kuzawa, C. W. (2011). Longitudinal evidence that fatherhood decreases testosterone in human males. *Proceedings of the National Academy of Sciences*, 108, 16194–16199.

Gettler, L. T., McDade, T. W., Agustin, S. S., Feranil, A. B., and Kuzawa, C. W. (2013). Do testosterone declines during the transition to marriage and fatherhood relate to men's sexual behavior? Evidence from the Philippines. *Horm. Behav.* 64, 755–763. doi: 10.1016/j.yhbeh.2013.08.019

Gordis, E. B., Granger, D. A., Susman, E. J., and Trickett, P. K. (2006). Asymmetry between salivary CORT and α-amylase reactivity to stress: relation to aggressive behavior in adolescents. *Psychoneuroendocrinology* 31, 976–987. doi: 10.1016/j.psyneuen.2006.05.010

Granger, D. A., Johnson, S. B., Szanton, S. L., Out, D., and Schumann, L. L. (2012). Incorporating salivary biomarkers into nursing research: an overview and review of best practices. *Biol. Res. Nurs.* 14, 347–356. doi: 10.1177/1099800412443892

Gray, P. B., Straftis, A. A., Bird, B. M., McHale, T. S., and Zilioli, S. (2020). Human reproductive behavior, life history, and the challenge hypothesis: a 30-year review, retrospective and future directions. *Horm. Behav.* 123:104530. doi: 10.1016/j.yhbeh.2019.04.017

Grebe, N. M., Del Giudice, M., Thompson, M. E., Nickels, N., Ponzi, D., Zilioli, S., et al. (2019). Testosterone, cortisol, and status-striving personality features: a review and empirical evaluation of the dual hormone hypothesis. *Horm. Behav.* 109, 25–37. doi: 10.1016/j.yhbeh.2019.01.006

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed). Guilford publications.

Jaremka, L. M., and Collins, N. L. (2017). Cortisol increases in response to brief social exchanges with opposite sex partners. *Biol. Psychol.* 124, 39–46. doi: 10.1016/j.biopsycho.2017.01.001

Kuo, P. X., Braungart-Rieker, J. M., Lefever, J. E. B., Sarma, M. S., O'Neill, M., and Gettler, L. T. (2018). Fathers' CORT and testosterone in the days around infants' births predict later paternal involvement. *Horm. Behav.* 106, 28–34. doi: 10.1016/j.yhbeh.2018.08.011

Kuzawa, C. W., Georgiev, A. V., McDade, T. W., Bechayda, S. A., and Gettler, L. T. (2016). Is there a testosterone awakening response in humans? *Adapt. Human Behav. Physiol.* 2, 166–183. doi: 10.1007/s40750-015-0038-0

Little, R. J., and Rubin, D. B. (2019). Statistical analysis with missing data. New York, NY, USA: John Wiley & Sons.

Mehta, P. H., and Josephs, R. A. (2010). Testosterone and CORT jointly regulate dominance: evidence for a dual-hormone hypothesis. *Horm. Behav.* 58, 898–906. doi: 10.1016/j.yhbeh.2010.08.020

Mehta, P. H., and Prasad, S. (2015). The dual-hormone hypothesis: a brief review and future research agenda. *Curr. Opin. Behav. Sci.* 3, 163–168. doi: 10.1016/j.cobeha.2015.04.008

Mehta, P. H., Welker, K. M., Zilioli, S., and Carré, J. M. (2015). Testosterone and CORT jointly modulate risk-taking. *Psychoneuroendocrinology* 56, 88–99. doi: 10.1016/j.psyneuen.2015.02.023

Muller, M. N., and Pilbeam, D. R. (2017). Mating system. *Chimpanzees Human Evol.* 383:426. doi: 10.4159/9780674982642-011

Plantin, L., Mansson, S., and Kearney, J. (2003). Talking and doing fatherhood: on fatherhood and masculinity in Sweden and England. *Fathering Harriman* 1, 3–26. doi: 10.3149/fth.0101.3

Roney, J. R., and Gettler, L. T. (2015). The role of testosterone in human romantic relationships. *Curr. Opin. Psychol.* 1, 81–86. doi: 10.1016/j.copsyc.2014.11.003

Roy, A. R., Cook, T., Carré, J. M., and Welker, K. M. (2019). Dual-hormone regulation of psychopathy: evidence from mass spectrometry. *Psychoneuroendocrinology* 99, 243–250. doi: 10.1016/j.psyneuen.2018.09.006

Rusbult, C. E., Martz, J. M., and Agnew, C. R. (1998). The investment model scale: measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Pers. Relat.* 5, 357–387. doi: 10.1111/j.1475-6811.1998.tb00177.x

Salkicevic, S., Stanic, A. L., and Grabovac, M. T. (2014). Good mates retain us right: investigating the relationship between mate retention strategies, mate value, and relationship satisfaction. *Evol. Psychol.* 12, 1038–1052. doi: 10.1177/147470491401200512

Santtila, P., Wager, I., Witting, K., Harlaar, N., Jern, P., Johansson, A. D. A., et al. (2007). Discrepancies between sexual desire and sexual activity: gender differences and associations with relationship satisfaction. *J. Sex Marital Ther.* 34, 31–44. doi: 10.1080/00926230701620548

Saxbe, D. E., Edelstein, R. S., Lyden, H. M., Wardecker, B. M., Chopik, W. J., and Moors, A. C. (2017a). Fathers' decline in testosterone and synchrony with partner testosterone during pregnancy predicts greater postpartum relationship investment. *Horm. Behav.* 90, 39–47. doi: 10.1016/j.yhbeh.2016.07.005

Saxbe, D. E., Schetter, C. D., Simon, C. D., Adam, E. K., and Shalowitz, M. U. (2017b). High paternal testosterone may protect against postpartum depressive symptoms in fathers, but confer risk to mothers and children. *Horm. Behav.* 95, 103–112. doi: 10.1016/j.yhbeh.2017.07.014

Tackett, J. L., Herzhoff, K., Harden, K. P., Page-Gould, E., and Josephs, R. A. (2014). Personality× hormone interactions in adolescent externalizing psychopathology. *Personal. Disord.* 5, 235–246. doi: 10.1037/per0000075

Tilbrook, A. J., Turner, A. I., and Clarke, I. J. (2000). Effects of stress on reproduction in non-rodent mammals: the role of glucocorticoids and sex differences. *Rev. Reprod.* 5, 105–113. doi: 10.1530/ror.0.0050105

Van Anders, S. M., Hamilton, L. D., and Watson, N. V. (2007). Multiple partners are associated with higher testosterone in north American men and women. *Horm. Behav.* 51, 454–459. doi: 10.1016/j.yhbeh.2007.01.002

van der Meij, L., Demetriou, A., Tulin, M., Méndez, I., Dekker, P., and Pronk, T. (2019). Hormones in speed-dating: the role of testosterone and CORT in attraction. *Horm. Behav.* 116:104555. doi: 10.1016/j.yhbeh.2019.07.003

Viau, V. (2002). Functional cross-talk between the hypothalamic-pituitary-gonadal and -adrenal axes. *J. Neuroendocrinol.* 14, 506–513. doi: 10.1046/j.1365-2826.2002.00798.x

Viau, V., and Meaney, M. J. (1996). The inhibitory effect of testosterone on hypothalamic-pituitary-adrenal responses to stress is mediated by the medial preoptic area. *J. Neurosci.* 16, 1866–1876. doi: 10.1523/JNEUROSCI.16-05-01866.1996

Wang, X., Sánchez, B. N., Golden, S. H., Shrager, S., Kirschbaum, C., Karlamangla, A. S., et al. (2014). Stability and predictors of change in salivary CORT measures over six years: MESA. *Psychoneuroendocrinology* 49, 310–320. doi: 10.1016/j.psyneuen.2014.07.024

Welker, K. M., Lozoya, E., Campbell, J. A., Neumann, C. S., and Carré, J. M. (2014). Testosterone, CORT, and psychopathic traits in men and women. *Physiol. Behav.* 129, 230–236. doi: 10.1016/j.physbeh.2014.02.057

Zuloaga, D. G., Lafrican, J. J., and Zuloaga, K. L. (2024). Androgen regulation of behavioral stress responses and the hypothalamic-pituitary-adrenal axis. *Horm. Behav.* 162:105528. doi: 10.1016/j.yhbeh.2024.105528

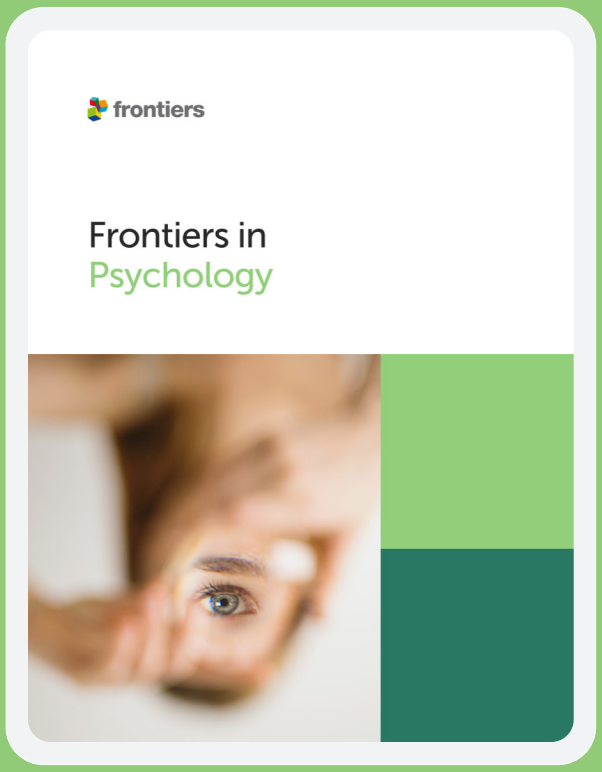
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact