

- Pillsworth, E. G., & Haselton, M. G. (2006). Male sexual attractiveness predicts differential ovulatory shifts in female extra-pair attraction and male mate retention. *Evolution and Human Behavior*, 27, 247–258.
- Roney, J. R., & Simmons, Z. L. (2013). Hormonal predictors of sexual motivation in natural menstrual cycles. *Hormones and Behavior*, 63, 636–645.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General*, 143, 534–547.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Thornhill, R., & Gangestad, S. W. (1999). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior*, 20, 175–201.
- Thornhill, R., Gangestad, S. W., Miller, R., Scheyd, G., Knight, J., & Franklin, M. (2003). MHC, symmetry, and body scent attractiveness in men and women. *Behavioral Ecology*, 14, 668–678.
- Wilcox, A. J., Duncan, D. B., Weinberg, C. R., Trussell, J., & Baird, D. D. (2001). Likelihood of conception with a single act of intercourse: Providing benchmark rates for assessment of post-coital contraceptives. *Contraception*, 63, 211–215.
- Wood, W., Kressel, L., Joshi, P. D., & Louie, B. (2014). Meta-analysis of menstrual cycle effects on mate preferences. *Emotion Review*, 6, 229–249.

Reply to Gangestad's (2016) Comment on Wood, Kressel, Joshi, and Louie (2014)

Wendy Wood

Department of Psychology, University of Southern California, USA

Abstract

Wood, Kressel, Joshi, and Louie's (2014) meta-analysis of menstrual cycle influences on mate preferences identified three artifacts that influenced study findings: imprecise estimates of the fertile phase, decline over time, and publication effects. These artifacts also were evident in another recent meta-analysis by Gildersleeve, Haselton, and Fales (2014a). This consistent evidence of artifacts is not challenged by Gildersleeve et al.'s (2014b) failure to find another artifact—chasing significance levels. In addition, Wood et al. correctly coded the findings of Gangestad and colleagues' research, given the variation in their reporting formats and inclusion criteria, which in some studies included only 54% of the sample. The controversy over menstrual cycle effects could be beneficial in increasing interest in publishing null results as well as in identifying evolutionary models that build on women's capacity to regulate reproduction according to societal roles.

Keywords

mate preferences, menstrual cycle, ovulatory shift hypothesis

I appreciate Stephen Gangestad's (2016) interest in Wood et al.'s (2014) meta-analytic review of menstrual cycle effects on mate preferences. Although Gangestad's comment focused primarily on women's preferences for symmetry, Wood et al. actually evaluated a number of male attributes and generally failed to find evidence for the evolutionary psychology prediction that women in the fertile phase of their menstrual cycle prefer to have short-term affairs with men of supposed genetic quality—as represented by high testosterone, symmetry, masculinity, or

dominance. The only significant effect was unanticipated: That is, women did not prefer symmetric men for short-term relationships but only preferred them in studies that did not specify a type of relationship. Further challenging the ideas that women wanted to have affairs with these symmetric men, they were not rated especially high on scales of sexiness, but instead received high ratings on scales of general attractiveness (see Wood et al.'s supplemental results).

Wood et al. (2014) concluded that the few significant findings in the literature were research artifacts. First, the preference for symmetry effect was disproportionately influenced by a few large-effect-size studies, which primarily assessed scent of symmetry and were conducted by Gangestad and colleagues (Gangestad, Garver-Apgar, Simpson, & Cousins, 2007; Gangestad, Simpson, Cousins, Garver-Apgar, & Christensen, 2004; Gangestad & Thornhill, 1998; Thornhill & Gangestad, 1999; Thornhill et al., 2003). Second, across the literature, three central artifacts influenced the findings: imprecision in estimating the fertile phase, a decline effect over time, and publication bias. Wood et al. adopted a conservative strategy by reporting only artifacts that consistently emerged across multiple literatures. Furthermore, Wood and Carden (2014) replicated these three artifacts in their reanalysis of another meta-analysis of this literature by Gildersleeve, Haselton, and Fales (2014a).

The two meta-analyses reported highly similar findings, with both showing considerable variability in the reviewed results. In the subset of studies in each review that assessed women's

preferences in short-term relationships or did not specify a relationship type, the reviews reported the following (for Gildersleeve et al., 2014a/Wood et al., 2014, respectively): 43/37 findings in which fertile women preferred men of purported high genetic quality, 5/4 findings of zero value indicating no difference,¹ and 20/20 findings in which nonfertile women preferred these quality men.

Gangestad (2016) claimed that Wood et al. (2014) provided no direct evidence of the influence of research artifacts in this literature. However, meta-analytic techniques link features of research methods to study outcomes, and this approach represents the standard, accepted format to test for methodological artifacts (see Cooper, Hedges, & Valentine, 2009). This well-established approach has been used widely in medical and psychological research to track the influence of research paradigms and publication decisions. In contrast, Gildersleeve, Haselton, and Fales (2014b) used the null effects obtained in analyses modeling the frequency distributions of statistically significant *p* values (aka *p*-hacking analyses) to argue against artifacts in this literature. However, these analyses addressed only one source of bias stemming from the original researchers chasing statistical significance, and they did not directly evaluate the artifacts identified by Wood et al. In short, Gildersleeve et al.'s (2014b) failure to find one particular artifact does not nullify evidence for the other artifacts apparent in this literature (see also Publication Bias section next).

Accurate Fertile Windows: 6 Days is the Best Estimate

In both Wood et al.'s (2014) and Gildersleeve et al.'s (2014a) meta-analytic reviews, menstrual cycle shifts were evident primarily in studies that used less precise measures of the fertile phase. That is, women's preferences did not shift across the cycle in studies that used hormonal assessments to identify time of ovulation. They only shifted across the cycle in studies that relied on women's self-reports of cycle day, and only in studies that used longer, less accurate fertile phases spanning 9, 13, or 15 days (see Wood & Carden, 2014, Figure 1; Wood et al., 2014, Figures 2a and 2b).

Gangestad (2016) speculated that larger fertile phases more accurately captured the 6-day fertile window, given that some women have short cycles and others long cycles, so that a larger window might capture the varying timing of fertility across women (see also Gildersleeve et al., 2014b). Wood et al. (2014) judged this unlikely given the lack of shifts in studies using hormonal assessments—the gold standard for estimating fertility. Also, direct evidence contradicting Gangestad's speculation that 9 days are needed to capture the fertile phase given variability in the days that women ovulate comes from Wilcox, Dunson, and Baird (2000; see also Wilcox, Dunson, Weinberg, Trussell, & Baird, 2001). Using hormonal tests to identify women's ovulatory day, Wilcox et al. (2000) estimated for each day of the month the likelihood of women being in the 6-day fertile window. Their findings are represented in Figure 1. Importantly, the curve averages across the naturally occurring variation in cycle

length and date of ovulation. Because each cycle includes many days in which women are not fertile and only a few fertile days, increasing the length of the fertile window increases the number of women incorrectly classified as fertile. For example, extending the fertile phase to include Day 7 designates additional women as fertile—fully 82% of these would be incorrectly classified because they are not actually fertile, and only 18% would be correct. Thus, Wilcox et al.'s (2000) results demonstrate that, because fertility rises and falls steeply across the cycle, smaller fertile windows centered around the day of ovulation yield the most accurate measures of the fertile window. These estimates reflect the most reliable values after taking into account variability in the day of ovulation.

Wood et al. (2014) puzzled over why the preference shifts emerged only in studies with less precise measures of the fertile phase, noting that, “we can only speculate why studies with larger fertile windows produced larger preference shifts” (p. 242). This larger window/stronger effect pattern emerged from correlational analyses that were conducted at the meta-analytic level, and thus are open to multiple explanations. However, the pattern was not due to any particular research methods, given that it emerged across multiple paradigms involving preferences for masculinity as well as symmetry. Ultimately, Wood et al. postulated confirmatory hypothesis-testing procedures in which the original study authors began with a standard 6-day fertile window and successively widened this in exploratory analyses until they identified an effect—a strategy that would have capitalized on chance. Adding further plausibility to this artifactual account of mate preference findings, Gildersleeve et al.'s (2014a) review that claimed to find evidence of menstrual cycle shifts used a broader fertile window (8.61 days) in calculating effect sizes than Wood et al.'s review (6.85 days) that concluded cycle effects are unreliable.

Accuracy in Representing Gangestad and Colleague's Research

Gangestad (2016) argued that the fertility calculations in his research were misrepresented by Wood et al. (2014) as varying across study, given that he used a standard 9-day fertile window as well as a continuous measure of conception risk associated with each cycle day. However, Gangestad and colleagues' research actually varied widely in reporting format. Of the five studies Wood et al. located by this research group, two provided sufficient information for Wood et al. to calculate an effect size comparing the fertile versus nonfertile phase (Gangestad & Thornhill, 1998; Thornhill et al., 2003), one provided sufficient data to calculate an effect only for continuous estimates of conception risk (Thornhill & Gangestad, 1999), and two did not provide data sufficient to estimate an effect size (Gangestad et al., 2007; Gangestad et al., 2004). It might also seem surprising that, despite Gangestad's endorsement of the 9-day fertile phase, two of these studies did not even report results comparing a fertile and nonfertile phase (Gangestad et al., 2007; Gangestad et al., 2004). Additionally suggesting flexibility in reporting, all five studies excluded substantial numbers of participants based on shifting sets of criteria, with

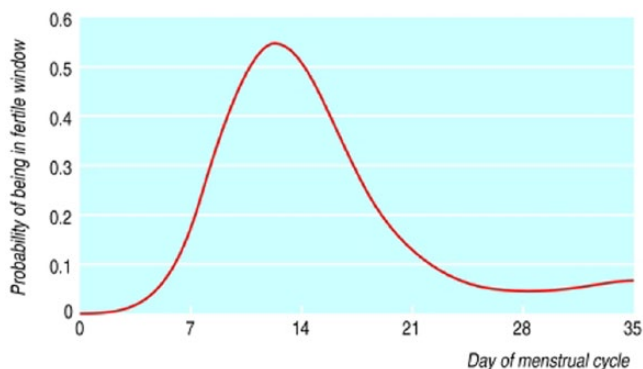


Figure 1. Wilcox, Dunson, and Baird's (2000) probability of being in the 6-day fertile phase at each menstrual cycle day.

exclusion rates ranging from 14% in Gangestad et al. (2004) to 46% in Gangestad and Thornhill (1998).

We do not know why Gangestad and colleagues varied their decision rules in these five studies with respect to sample selection, analytic strategy, and reporting procedures. However, it seems certain that the standard definition of p-hacking is not relevant. Gangestad and colleagues' effect sizes are extremely large outliers in this literature and clearly were not tailored to a .05 reporting criterion. Instead, the results they reported were of sufficient magnitude to skew the overall findings of both reviews to show significant shifts in mate preferences across the cycle. Contrary to Gangestad's (2016) claims, however, his research using a particular paradigm is not responsible for the finding that only studies with large fertile phases found that fertile women preferred symmetric men more than nonfertile ones. That is, this effect for length of the fertile phase holds whether or not Gangestad and colleagues' research is included in the regression models (compare Figure 2 with Wood et al.'s [2014] Figure 2b).

In general, Gangestad (2016) attempted to justify variability in the length of the fertile window in this research literature, but he failed to address other forms of flexibility. Although citing Harris, Pashler, and Mickes (2014), Gangestad failed to acknowledge one of their main points: Variations in the literature include not just the number of days counted as fertile or infertile, but also the placement of these windows, the selection of moderators, the exclusion criteria of female participants, and the reported analyses (e.g., see Figure 1 of Harris, Chabot, & Mickes, 2013). Along with other researchers, Gangestad has studied the same variable (e.g., symmetry) using different criteria for categorization and analysis across different articles (as noted before). This flexibility in reporting format potentially enabled a variety of patterns to be interpreted as supporting evolutionary psychology predictions about mate preferences.

Decline Effect Over Time

Research findings that capitalize on chance tend to decline over time as subsequent research fails to document the initial pattern (Ioannidis, 2005; Schooler, 2011). In both Wood et al.'s (2014)

and Gildersleeve et al.'s (2014a) data, cycle shifts approximated zero in more recent studies. Gangestad (2016) argued that this decline effect in studies of preference for symmetry was due to changes in the research paradigms used over time. However, this account is not plausible given that the decline effect also emerged in studies of masculinity in Wood et al.'s review and was found broadly across all studies in Gildersleeve et al.'s (2014a) review.

A more plausible account for the decline effect is that recent studies have improved methods for assessing fertility. Studies published more recently were in fact found to use more rigorous methods and to more precisely estimate the fertile phase (see Wood & Carden, 2014). Alternatively, Schooler (2011) tied decline effects to publication bias and regression toward the mean, speculating that "early published studies benefit from being at one statistical end of a larger body of (unpublished) findings" (p. 437). Whether the increasing precision in estimating the fertile phase or the increasing availability of (formerly) unpublished results is responsible, recent studies appear to be less susceptible to false-positive conclusions about the effects of cycle phase on women's judgments.

Publication Bias

Publication bias proved to be a strong artifact influencing the size of effect in this literature. Only published studies revealed shifts in women's preferences across the cycle in Wood et al.'s (2014) review of symmetry, masculinity, and dominance, and this same publication effect was evident in Gildersleeve et al.'s (2014a) review. Specifically, unpublished studies in both reviews revealed null effects.

Gildersleeve et al. (2014a) failed to report the null effects in their unpublished literature, but instead concluded no evidence for publication bias based on a funnel plot of the effect sizes and associated tests. Yet these tests provide only an indirect assessment of publication bias, by evaluating whether less precise, smaller studies report different results than more precise, larger studies. Because a variety of factors could produce or mask differences between small-sample and large-sample findings, Lau, Ioannidis, Terrin, Schmid, and Olkin (2006) concluded that, "asymmetry of the funnel plot, either visually interpreted or statistically tested, does not accurately predict publication bias" (p. 599). Especially in literatures with considerable heterogeneity, as is the case with studies on mate preferences, publication bias statistics are likely to be misleading (Ioannidis & Trikalinos, 2007). Additionally, under a variety of conditions, these tests have been shown to be insensitive to even large amounts of bias (Kicinski, 2014).

As Gangestad (2016) noted, Gildersleeve et al. (2014b) concluded that menstrual cycle shifts had evidential basis, given the distribution of statistically significant *p* values in this literature. However, their calculations were based on small numbers of studies (e.g., $k = 14$), and the results would likely have revealed little evidential value without the very large effects reported by Gangestad and colleagues (e.g., Gangestad et al., 2007; Gangestad et al., 2004; Gangestad & Thornhill, 1998; Thornhill & Gangestad, 1999; Thornhill et al., 2003).

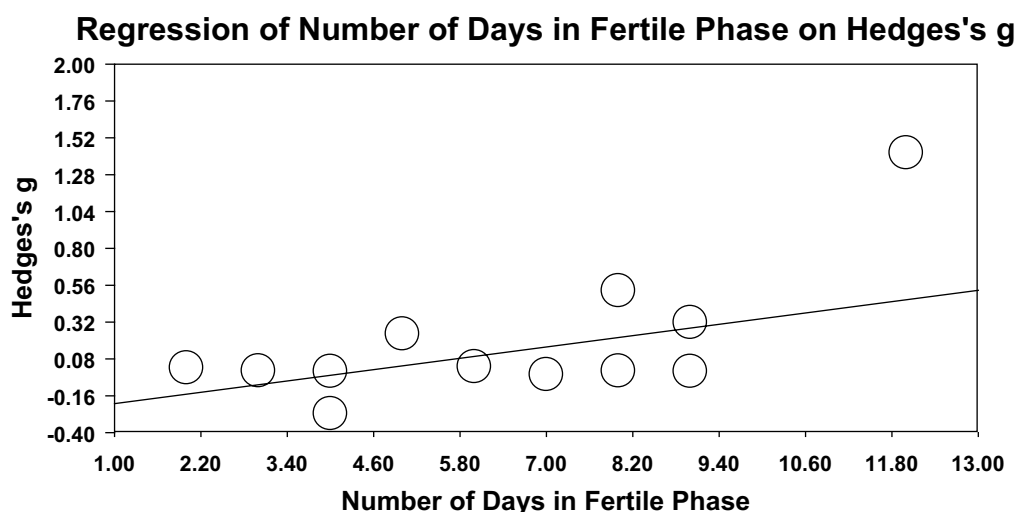


Figure 2. Effects of width of fertile phase on women's preferences for symmetric men, after excluding Gangestad and colleagues' (Gangestad & Thornhill, 1998; Thornhill & Gangestad, 1999; Thornhill et al., 2003) research (Wood et al., 2014).

Given the inaccuracy of existing tests for publication bias, Ioannidis and Trikalinos (2007) recommended that, "whenever both unpublished and published information is available, the results of these 2 types of evidence should be compared" (p. 1095). The results of this comparison in Gildersleeve et al.'s (2014a) and Wood et al.'s (2014) reviews were clear: Published studies overestimated the effects of menstrual cycles on mate preferences.

Future Directions

The controversy over whether menstrual cycles affect women's preferences has focused the field's attention on this issue, potentially increasing interest in reporting and publishing null findings, and in this way reducing the influence of publication bias. In fact, several studies that failed to find any evidence of menstrual cycle effects on women's social preferences were recently accepted for publication in top journals (Hawkins, Fitzgerald, & Nosek, 2015; Scott & Pound, in press), including a particularly impressive investigation of women's preferences for masculinity in 12 societies differing in economic development (Scott et al., 2014). These publications are likely beneficial consequences of the controversy over menstrual cycle effects, and they suggest a basis for optimism about the self-correcting nature of science.

The controversy over whether women's preferences shift across the cycle also may have the benefit of spurring new evolutionary models of women's reproductive activities. One such effort highlights women's evolved capacities to regulate both their mate preferences and their menstrual cycles according to the demands of societal roles (Eagly & Wood, 2013; Wood et al., 2014). Thus, women's preferences for a mate vary with gender roles in their society, as these roles influence the costs and benefits women attach to male attributes. Furthermore, the frequency and patterning of women's menstrual cycles vary with women's productive roles in a society, as these roles influence women's

capacity for extended lactation and frequent childbirth—thereby affecting the frequency of women's cycles. More powerful evolutionary models are being developed that recognize the advent of cultural roles and complex group living in human's evolutionary history, along with women's development of the capacity to tailor their reproductive activities to a range of social roles.

Note

1. Readers are encouraged to check these numbers against the actual data reports. For some reason, Gildersleeve et al. (2014b) substantially overestimated the number of zero effect sizes reported by Wood et al. (2014).

References

- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: NY: Russell Sage Foundation.
- Eagly, A. H., & Wood, W. (2013). The nature–nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspectives on Psychological Science, 8*, 340–357. doi:10.1177/1745691613484767
- Gangestad, S. W. (2016). Comment: Wood et al.'s (2014) speculations of inappropriate research practices in ovulatory cycle studies. *Emotion Review, 8*(1): pp. 85–88.
- Gangestad, S. W., Garver-Apgar, C. E., Simpson, J. A., & Cousins, A. J. (2007). Changes in women's mate preferences across the ovulatory cycle. *Journal of Personality and Social Psychology, 92*, 151–163. doi:10.1037/0022-3514.92.1.151
- Gangestad, S. W., Simpson, J. A., Cousins, A. J., Garver-Apgar, C. E., & Christensen, P. N. (2004). Women's preferences for male behavioral displays change across the menstrual cycle. *Psychological Science, 15*, 203–207. doi:10.1111/j.0956-7976.2004.01503010.x
- Gangestad, S. W., & Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. *Proceedings of the Royal Society B, 265*, 927–933. doi:10.1098/rspb.1998.0380
- Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014a). Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. *Psychological Bulletin, 140*, 1205–1259. doi:10.1037/a0035438
- Gildersleeve, K., Haselton, M. G., & Fales, M. R. (2014b). Meta-analysis and p-curves support robust cycle shifts in women's mate preferences:

- Reply to Wood and Carden (2014) and Harris, Pashler, and Mickes, 2014. *Psychological Bulletin*, 140, 1272–1280. doi:10.1037/a0037714
- Harris, C. R., Chabot, A., & 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
- Harris, C. R., Pashler, H., & Mickes, L. (2014). Elastic analysis procedures: An incurable (but preventable) problem in the fertility effect literature. Comment on Gildersleeve, Haselton, and Fales (2014). *Psychological Bulletin*, 140, 1260–1264. doi:10.1037/a0036478
- Hawkins, C. B., Fitzgerald, C. E., & Nosek, B. A. (2015). In search of an association between conception risk and prejudice. *Psychological Science*. Advance online publication. doi:10.1177/0956797614553121
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi:10.1371/journal.pmed.0020124
- Ioannidis, J. P., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: A large survey. *Canadian Medical Association Journal*, 176(8), 1091–1096. doi:10.1503/cmaj.060410
- Kicinski, M. (2014). How does under-reporting of negative and inconclusive results affect the false-positive rate in meta-analysis? A simulation study. *British Medical Journal Open*, 4(8), e004831. doi:10.1136/bmjopen-2014-004831
- Lau, J., Ioannidis, J., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal*, 333(7568), 597–600. doi:10.1136/bmj.333.7568.597
- Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, 470(7335), 437. doi:10.1038/470437a
- Scott, I. M., Clark, A. P., Josephson, S. C., Boyette, A. H., Cuthill, I. C., Fried, R. L., ... Penton-Voak, I. S. (2014). Human preferences for sexually dimorphic faces may be evolutionarily novel. *Proceedings of the National Academy of Sciences*, 111(40), 14388–14393. doi:10.1073/pnas.1409643111
- Scott, I. M., & Pound, N. (in press). Menstrual cycle phase does not predict political conservatism. *PLOS ONE*.
- Thornhill, R., & Gangestad, S. W. (1999). The scent of symmetry: A human sex pheromone that signals fitness? *Evolution and Human Behavior*, 20, 175–201. doi:10.1016/S1090-5138(99)00005-7
- Thornhill, R., Gangestad, S. W., Miller, R., Scheyd, G., McCollough, J. K., & Franklin, M. (2003). Major histocompatibility complex genes, symmetry, and body scent attractiveness in men and women. *Behavioral Ecology*, 14, 668–678. doi:10.1093/beheco/arg043
- Wilcox, A. J., Dunson, D. B., & Baird, D. D. (2000). The timing of the “fertile window” in the menstrual cycle: Day specific estimates from a prospective study. *British Medical Journal*, 321, 1259–1262. doi:10.1136/bmj.321.7271.1259
- Wilcox, A. J., Dunson, D. B., Weinberg, C. R., Trussell, J., & Baird, D. D. (2001). Likelihood of conception with a single act of intercourse: Providing benchmark rates for assessment of post-coital contraceptives. *Contraception*, 63, 211–215. doi:10.1016/S0010-7824(01)00191-3
- Wood, W., & Carden, L. (2014). Elusiveness of menstrual cycle effects on mate preferences: Comment on Gildersleeve, Haselton, and Fales (2014). *Psychological Bulletin*, 140, 1265–1271. doi:10.1037/a0036722
- Wood, W., Kressel, L., Joshi, P., & Louie, B. (2014). Meta-analysis of menstrual cycle effects on women’s mate preferences. *Emotion Review*. Advance online publication. doi:10.1177/1754073914523073