Psychological Bulletin
2014, Vol. 140, No. 5, 1265–1271

0033-2909/14/$12.00 http://dx.doi.org/10.1037/a0036722

COMMENT

Elusiveness of Menstrual Cycle Effects on Mate Preferences: Comment on Gildersleeve, Haselton, and Fales (2014)

Wendy Wood and Lucas Carden
University of Southern California

This comment uses meta-analytic techniques to reconcile the apparent conflict between Gildersleeve, Haselton, and Fales's (2014) conclusion of "robust" effects of menstrual cycles on women's preferences for men of purported genetic quality and Wood, Kressel, Joshi, and Louie's (2014) assessment that the few, limited effects in this literature appear to be research artifacts. Despite these divergent conclusions, the literature in both reviews shows a broad distribution of effects, with fully one third of findings countering evolutionary psychology predictions. We demonstrate that Gildersleeve et al.'s conclusions were influenced by a small minority of supportive studies. Furthermore, we show that in both reviews, these supportive studies used imprecise estimates of women's cycle phase by failing to validate cycle day (e.g., with hormonal tests) or by including a large number of days in the fertile phase. More recently, as published studies have used more precise methods to estimate menstrual phase, the effect has declined to zero. Additionally, publication status proved important in both reviews, with published but not unpublished studies showing the predicted effects. In general, the limited evidence for evolutionary psychology predictions calls for more sophisticated models of hormonal processes in human mating.

Keywords: menstrual cycle, mate preferences, meta-analysis, researcher degrees of freedom

During the 1970s and '80s, social stereotypes and popular media portrayed women as victims of hormonal fluctuations across the cycle that led to maladjustment and premenstrual syndromes (Chrisler, & Caplan, 2002; Chrisler & Levy, 1990). Psychologists' early research on menstrual disorders may have been informed by these simple hormone-to-behavior depictions (Hyde & Salk, in press). However, research in this area became more sophisticated following demonstrations that women's self-reported menstrual syndromes were influenced in part by artifacts tied to cultural beliefs about cycles (e.g., Ruble, 1977; Ruble & Brooks-Gunn, 1979).

The idea that menstrual cycles have a direct influence on social judgments again has come to the fore in evolutionary psychology theories of mate preferences. In these analyses, women's preferences for potential partners shift across the menstrual cycle to reflect the evolved fitness benefits of mating with different types of men. Wood, Kressel, Joshi, and Louie (2014) conducted a meta-analytic review of prior research testing these theories and reported a general "failure to find consistent effects of women's hormonal cycling on their mate preferences" (p. 17). Their conclusion is in stark contrast to Gildersleeve, Haselton, and Fales's (2014) claim of "robust cycle shifts that were specific to women's preferences for hypothesized cues of (ancestral) genetic quality" (p. 1205).

In this comment, we use meta-analytic techniques to evaluate the accuracy of these two conclusions and to identify the factors that produced the divergent conclusions across reviews. In short, our analysis reveals that the findings within both reviews were highly inconsistent and showed a wide dispersion of effects. Furthermore, Gildersleeve et al.'s (2014) conclusions were influenced by a small subset of studies that did not validate menstrual cycle day.

Distribution of Effects in the Two Reviews

We start our meta-analytic evaluation with an analysis of the overall distribution of findings in this literature. Approximately a third of the effects included in both reviews trended in the direction *opposite* to evolutionary psychology predictions. That is, Gildersleeve et al. (2014) reported 35 findings (out of 96 total) in which fertile women, more than nonfertile ones, preferred men of lower genetic quality (i.e., less masculine, dominant, symmetric). Also worth noting is that in the meta-analytic calculations, only seven of their effects (i.e., 7%) were statistically significant in the predicted direction, while three studies reported statistically significant effects in the reverse direction. Wood et al. (2014) reported a similarly dispersed distribution of effects. Specifically, they noted 30 reversed-direction findings (out of 91 total) in which fertile women preferred men of apparently lower quality. Thus, the distribution of effect sizes in both reviews revealed wide variation

Wendy Wood and Lucas Carden, Department of Psychology, University of Southern California.

We thank Alice H. Eagly, Priyanka Joshi, David Neal, and the members of University of Southern California's social behavior lab for their thoughtful comments on a draft of this article.

Correspondence concerning this article should be addressed to Wendy Wood, Department of Psychology, University of Southern California, Los Angeles, CA 90018. E-mail: wendy.wood@usc.edu

in findings concerning menstrual cycle effects on women's judgments.

In addition to highlighting the highly inconsistent support for menstrual cycle effects in this literature, Wood et al. (2014) argued that the few obtained effects could be explained as research artifacts. Specifically, they noted that "the effects declined over time in published work, were limited to studies that used broader, less precise definitions of the fertile phase, and were found only in published research" (p. 1). In the remainder of our comment, we evaluate whether Gildersleeve et al.'s (2014) evidence of cycle shifts could also be attributed to research artifacts, especially to imprecise estimates of the fertile phase. We address this question by conducting analyses on comparable data from the two reviews. That is, we analyzed Gildersleeve et al.'s broad set of measures of genetic quality[1] and Wood et al.'s preferences for symmetry, dominance, and masculinity (i.e., the small number of testosterone studies precluded these analyses). Wood et al. originally structured their review to evaluate each attribute separately in order to maximize the likelihood of detecting and interpreting cycle shifts. Before discussing the artifact analyses, it is helpful to compare the sets of studies included in the two reviews.

Study Samples

The two reviews largely share the same published database, with 31 articles in common. Gildersleeve et al. (2014) included an additional four unique published studies,[2] and Wood et al. (2014) included 11 unique published studies. This similarity in published data is to be expected, given that articles were identified using standard computerized search tools of published literature databases. In contrast, unpublished studies, which tend to be identified through diverse search strategies, differed more between the two reviews. Five unpublished reports were common to the two reviews, with Gildersleeve et al. including six additional unique unpublished reports and Wood et al. including eight unique ones. Thus, the "file drawer" literature in this area consists at minimum of 19 unpublished studies.[3]

To be included in either review, a study had to meet certain basic quality criteria. In both reviews, these included the following: assessing naturally cycling women who were not using birth control, assessing ovulatory cycle phase, providing an objective assessment of mate preferences, and providing sufficient information to compute an effect size. Gildersleeve et al. (2014, pp. 1234–1235) applied additional selection rules to exclude particular findings from their broad set of studies, along with even more restrictive selection rules to exclude findings for their narrow set of studies.

A challenge in using such restrictive rules is that other meta-analysts may disagree with the selection logic. For example, from our perspective, Gildersleeve et al.'s (2014) additional selection rules are inconsistent (e.g., including measures of "social respect and influence," as explained on p. 1235, but excluding measures of "social status," as explained in Criterion 4) and based on theoretically extraneous concerns (e.g., excluding measures of reported preferences while including measures of revealed preferences). Especially given the strikingly wide dispersion in effects in this literature, it is possible to develop selection rules post hoc to highlight particular studies with predicted effects and to exclude those with unpredicted effects. It is worth noting that the 10 unique published studies (14 total effects) that were included by Wood et al. (2014) but do not appear in Gildersleeve et al.'s broad set of findings yielded a null mean effect ($g = 0.04$, 95% CI [−0.07, 0.15]). Although the selection rules make it difficult to interpret Gildersleeve et al.'s findings, the analyses reported below use their broad sample of findings, selected on their Criteria 1–7. However, their narrow subset of findings that reflect more restricted decision rules are even more difficult to interpret, and thus we do not discuss these further.

To understand what generated such high variability in the findings in this literature, we used meta-analytic techniques to evaluate potential moderators. As we explain, the most consistent, prominent predictors involved the precision with which the original researchers defined the fertile phase.

Menstrual Cycle Shifts Not Found in Studies With More Precise Estimates of the Fertile Phase

The reviewed studies typically used women's self-reported cycle day to identify whether they were in fertile (follicular) or nonfertile (luteal, early follicular/menses) phases. Self-reports may be unreliable because women may not accurately track their own cycle patterns (see Gildersleeve et al., 2014, supplemental materials). Accordingly, some studies validated cycle phase through hormonal tests (e.g., luteinizing hormone, progesterone) or through assessments of the date of menstrual onset (e.g., menstrual daily diaries). For studies that used hormonal validation, DeBruine et al. (2010) estimated that about a third of women needed to be excluded due to lack of evidence that ovulation occurred during a monthly cycle. Given these sources of unreliability, self-reports of cycle day are an uncertain indicator of hormonal fluctuations.

In the reviewed literature, Wood et al. (2014) identified 24 effect sizes that validated women's cycle phase. It is noteworthy that these studies revealed no evidence of shifts across the cycle in mate preferences (see Wood et al., 2014, supplemental Table B). We cannot interpret Gildersleeve et al.'s (2014) coding of validation because they failed to note numbers of studies in their review that used these procedures (e.g., Cárdenas & Harris, 2007; Little, Jones, Burt, & Perrett, 2007, Study 1; Oinonen & Mazmanian, 2007; Pawlowski & Jasienska, 2005). Gildersleeve et al. instead tested whether stronger effects were found in studies with stronger "manipulations" of fertility (i.e., configured self-reported days so as to differentiate conception risk between fertile and nonfertile phases). However, this analysis apparently failed to show any significant results, suggesting again that studies that more accurately detected the fertile phase did not show menstrual cycle shifts in preferences.

[1] We reconstructed Gildersleeve et al.'s (2014) data from their Table 1 and Figure 3. For the three findings with different effects across these two sources, the correct data were taken from Table 1.

[2] One of these articles was published subsequent to Wood et al.'s (2014) review; two did not report data in a form that enabled effect size calculations, and the author did not supply this when contacted; and one tested men's skin color darkness, an attribute that likely signified outgroup status for the Caucasian participants.

[3] Of course, the actual number of unpublished studies is larger, especially given that a number of researchers alerted Wood et al. (2014), after the end date of their literature search, to file-drawer unpublished menstrual cycle studies they had conducted.

The definition of the fertile phase by the original researchers is critical to understanding menstrual cycle effects. Women have approximately a 6-day window of high conception risk during each cycle, outside of which the probability of pregnancy given a single act of intercourse appears to be less than 1% (Wilcox, Dunson, Weinberg, Trussell, & Baird, 2001). Given that women's cycles are composed of many nonfertile days and just a few fertile ones, menstrual cycle effects can only be identified with sensitive measures of the relatively brief fertile period. If researchers use an overly broad high-fertile window, they risk weakening the predicted effects on mate preferences, as women who are actually in a low fertile phase are misclassified as being high in fertility.[4] It is surprising, then, that the original researchers in this literature used highly variable definitions, ranging from 3 days to 15 days of the month (see Harris, Chabot, & Mickes, 2013).

To understand how the original researchers' definitions of the fertile phase influenced their findings, we computed regression models predicting mate preference effect sizes from phase length. In Wood et al.'s (2014) review, stronger effects emerged in studies with larger, less precise fertile phases, and null effects emerged in studies with smaller, more precise fertile phases for symmetry preferences (see Wood et al., 2014, Figure 2a) and masculinity preferences (see Wood et al., 2014, Figure 2b), but not dominance preferences, which failed to reveal any effects for cycle phase. When we conducted these analyses on Gildersleeve et al.'s (2014) data, this pattern was significant for the theoretically important short-term relationship findings, regression coefficient $= 0.02$, $Z = 2.25$, $Q(1, 10) = 5.06$, $p = .024$, $k = 12$ (see Figure 1 and the Type of Relationship Effects: Linked to Imprecise Methods of Assessing Menstrual Phase section).[5] Similar, but not significant, effects emerged in analyses on the full set of data. It is striking that the effects approximated zero in studies with smaller fertile phases.

We can only speculate about why studies with larger fertile windows produced larger preference shifts. It is not likely that they accurately captured the fertile phase, given that the predicted cycle shifts were not obtained in studies that validated phase, and that less precise windows should weaken, not strengthen mate preference effects. As Gildersleeve et al. (2014) noted, the variability in fertile phase definition might reflect that the original researchers failed to set definitions in advance of data collection, and instead flexibly defined the fertile phase based on study findings. Researchers who used this strategy might have begun with a window of 6 fertile days, and in the case of finding nonsignificant results, conducted exploratory analyses that successively broadened the number of days and maximized the distinction between fertile and nonfertile women's preferences in a given sample. This confirmatory testing strategy would have inflated Type I error and produced spurious evidence of cycle shifts (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

Gildersleeve et al. (2014) did not report testing for the influence of length of fertile phase on effect sizes. Instead, they claimed that the predicted cycle shifts emerged in a subset of studies that represented fertility through a continuous measure of the conception risk associated with each day of the cycle. However, the studies that used continuous measures of fertility risk actually yielded highly inconsistent effects. As Gildersleeve et al. reported, in the theoretically important short-term relationship context, women in more fertile days did not significantly prefer men with putative good genes ($g = 0.17$ [95% CI $= -0.04$, 0.38], $T = 0.17$, $k = 6$). Thus, continuous measures did not yield supportive results in this critical context. In contrast, unspecified relationship contexts revealed markedly large effects ($g = 0.62$ [95% CI $= 0.28$, 0.95], $T = 0.24$, $k = 6$). Four of these large-effect studies were produced by a single lab, and it is unclear why they produced such strong effects in a context not predicted by evolutionary psychology theorizing (Gangestad & Thornhill, 1998; Thornhill, Chapman, & Gangestad, 2013; Thornhill & Gangestad, 1999; Thornhill et al., 2003). Nonetheless, a similar cluster of studies produced outlier effects in Wood et al.'s (2014) meta-analysis. However, because that meta-analysis relied on the dichotomous fertile versus nonfertile comparison, the large effects in this set of studies were attributed to the large, imprecise fertile phases used by these researchers when reporting the dichotomous comparison.

As Gildersleeve et al. (2014) also noted, using continuous measures does not circumvent the problem of researcher flexibility. The original researchers might have chosen this method of defining fertility—which was used by less than a third of studies in their review, when it yielded the predicted effects. In addition, to use conception risk estimates, researchers had to make a number of decisions that provided additional flexibility in estimates. For example, they had to identify the starting and ending points to women's cycles, which can extend up to 40 days or more in length, and designate a date of ovulation (Vitzthum, 2009). For these reasons, researchers using continuous measures of conception risk retained considerable flexibility in defining fertility.

In summary, both Gildersleeve et al.'s (2014) and Wood et al.'s (2014) data revealed that support for evolutionary psychology predictions in short-term relationship contexts was evident primarily in studies with less precise estimates of the fertile phase, especially studies that used larger fertile windows. Furthermore, studies that validated cycle phase through hormonal or other assessments did not report a cycle shift in Wood et al.'s review. It is thus possible that the original researchers extended fertile phase estimates in ways that capitalized on chance findings in their data. This pattern highlights the need for more sophisticated measures of cycle phase that reliably capture monthly hormonal fluctuations associated with ovulation.

Decline Effect Over Time: Linked to Imprecise Methods of Assessing Menstrual Phase

When research findings capitalize on chance, they tend to decline over time as subsequent research fails to document the initial pattern (Ioannidis, 2005; Lehrer, 2010). Gildersleeve et al. (2014) did not test for decline effects, although our reanalysis of their data revealed strong effects for date of publication. That is, earlier studies in their review reported stronger tenden-

[4] Specifically, Gildersleeve et al. (2014, supplemental materials, p. 5) argued that an "overly broad high-fertility window can result in women completing putative 'high-fertility' sessions on true low-fertility days of the cycle. This is analogous to failing to expose participants in a test condition to the manipulation and, therefore, comparing controls to controls."

[5] With the exception of this figure, which is based on the fixed-effects model used by Gildersleeve et al. (2014), we report random-effects categorical models and maximum likelihood regression models that are appropriate with a heterogeneous literature in which studies are not exact replicates of each other (Borenstein, Hedges, Higgins, & Rothstein, 2009).

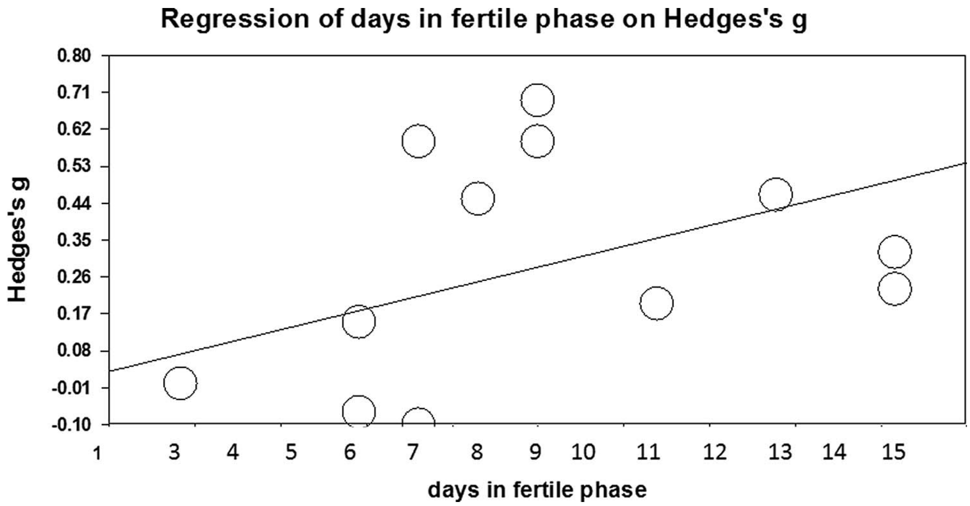

Figure 1. Relation between precision of the fertile phase and effect sizes in Gildersleeve et al. (2014). Larger positive effect sizes reflect stronger preferences among fertile (vs. nonfertile) women for men of high purported genetic quality in short-term relationships.

cies for fertile women to prefer men with attributes presumably indicative of genetic quality, regression coefficient = −0.03, $Z = -2.71$, $Q(1, 36) = 7.38$, $p = .007$ (see Figure 2). Specifically, in more recent publications, effect sizes approached zero. A comparable pattern emerged when these analyses were conducted in the subset of studies that specified short-term relationships or that did not specify a relationship. The decline effect also characterized Wood et al.'s (2014) data (see Wood et al., 2014, Figures 3a and 3b). In that review, menstrual cycle shifts declined to zero in more recently published research testing women's preferences for symmetry and for masculinity (although not dominance, which generally failed to reveal any significant effects of cycle phase).

The decline effect in this literature appears to be tied to research practices and reporting conventions, especially to recent studies' increased precision in defining the fertile phase. In Gildersleeve et al.'s (2014) data, studies published more recently had smaller fertile phases, comprising fewer days, $r(25) = -.37$, $p = .057$. Similarly, Wood et al.'s (2014) data revealed greater methodological precision in more recently published studies. Thus, in more recently published studies, the fertile phase was defined more narrowly for symmetry preferences, $r(15) = -.50$, $p = .041$, and masculinity preferences, $r(31) = -.44$, $p = .010$ (but not dominance). In like manner, studies that used validation techniques were published more recently for symmetry preferences ($p = .007$) and masculinity

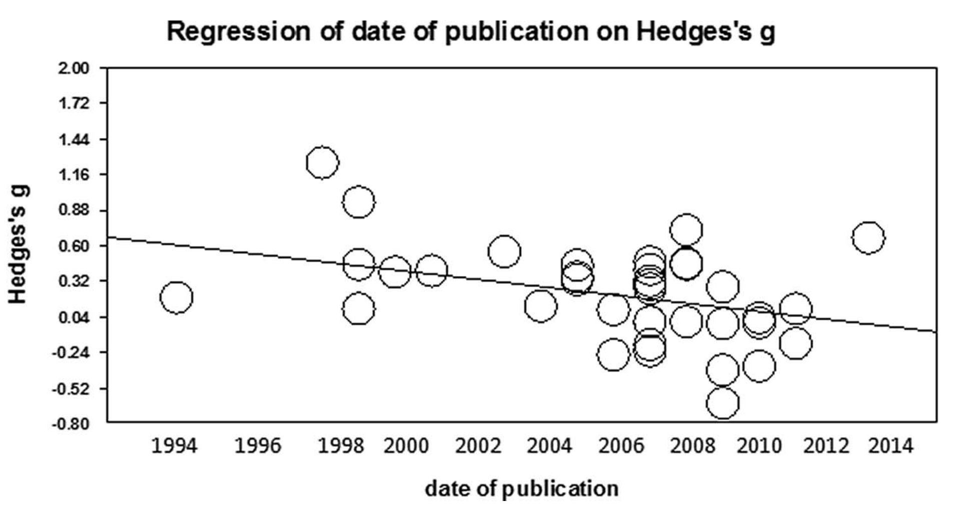

Figure 2. Relation between date published and effect sizes in Gildersleeve et al. (2014). Larger positive effect sizes reflect stronger preferences among fertile (vs. nonfertile) women for men of high purported genetic quality ($k = 38$).

preferences ($p = .019$), with a similar trend for dominance preferences ($p = .067$).

In summary, the decline effect in recent years appears tied to researchers' increasing use of more precise methods to evaluate menstrual cycle phases. These recent methodological advances appear to be controlling for false-positive conclusions about the effects of cycle phase on women's judgments, leading effects to approximate zero in the recent literature.

Type of Relationship Effects: Linked to Imprecise Methods of Assessing Menstrual Phase

A central prediction in evolutionary psychology theories is that fertile women prefer to have short-term affairs with masculine and symmetric men, due to the genetic benefits they can provide. In Gildersleeve et al.'s (2013) words, "preference shifts are often observed primarily when women evaluate men's desirability as *short-term* sex partners" (p. 517, emphasis added). Similarly, Tybur and Gangestad (2011) noted that "preference shifts across the cycle are specific to women's judgments of men's sexiness (or desirability as a *short-term* partner), not their attractiveness as long-term, investing mates" (p. 3382, emphasis added).

Short-Term Affairs

Gildersleeve et al. (2014) reportedly confirmed women's cycle shifts in short-term preferences for men of genetic quality. However, as we explained above, our reanalysis of their data revealed the predicted effects only in studies with long, imprecise estimates of the fertile phase (see Figure 1B). In addition, the 6 short-term relationship studies that used a continuous measure of conception risk did not yield effects different from zero. Thus, Gildersleeve et al.'s significant effects in short-term relationships were limited to studies with less accurate definitions of menstrual phase.

Wood et al. (2014) did not obtain any evidence of cycle shifts in short-term relationship contexts. The short-term relationship findings thus are a clear point of divergence between the two reviews. We tested whether this conflict in findings for short-term relationships was due to flexibility in defining the fertile phase at the level of meta-analytic reviewers. Whenever possible, Wood et al. asked authors to recompute their effects based on shorter, more precise fertile windows. Thus, in Wood et al.'s review, the mean length of the fertile phase for short-term relationship contexts was 6.85 days ($k = 22$; collapsed across symmetry, dominance, and masculinity). Gildersleeve et al. (2014) reported findings based on longer fertile phases, averaging 8.61 days ($k = 12$). These different lengths presumably reflect flexibility in the ways that Wood et al. and Gildersleeve et al. defined the fertile phase to the original study authors. In general, the results strikingly demonstrate that the predicted effects emerged only in studies using less precise definitions of the fertile phase, whether these definitions were used by the original study authors or were imposed by the meta-analytic reviewers.

Relationships of Unspecified Type or Length

Gildersleeve et al. (2014) deviated from standard evolutionary psychology logic to introduce the prediction that fertile women prefer quality men when relationships are not specified. They are the first researchers, to our knowledge, to make such a prediction, and it counters evolutionary psychology reasoning about the fitness benefits of fertile women choosing quality men for short-term affairs. Nonetheless, studies that did not specify a relationship revealed shifts in preferences across the cycle in Gildersleeve et al.'s data and also in Wood et al.'s (2014) studies evaluating preferences for symmetric (but not masculine or dominant) men. Wood et al. pointed out that these findings were driven by a handful of studies and that evolutionary psychology theories have not predicted cycle shifts when the length of relationship is unclear.

The results of studies that did not specify a relationship were tied to several artifacts associated with poor quality research methods. That is, in both reviews, the decline effect in this set of studies was significant, with the effects approximating zero in more recent publications. Furthermore, in Wood et al.'s (2014) review, the predicted shifts in preferences for male symmetry in studies that did not specify a relationship were only obtained when menstrual phases were defined imprecisely—in studies that failed to validate cycle day and that used broader estimates of the fertile phase. Also, in Gildersleeve et al.'s (2014) review, shifts in unspecified relationship contexts were due primarily to a handful of large-effect sized studies from a single lab.

Long-Term Relationships

Another common evolutionary psychology prediction is that women in the nonfertile, luteal phase of the cycle, whose progesterone levels are high as in pregnancy, "show stronger preferences for characteristics in a partner or an associate that might be beneficial at this time, such as social and material support" (Jones et al., 2008, p. 78). Along similar lines, Little, Jones, and Burriss (2007) predicted that women have "increased preferences for both feminine faced men and women during the luteal phase of the menstrual cycle" (p. 634). Surprisingly, Gildersleeve et al. (2014) did not make this prediction with respect to attributes that would facilitate long-term partner investment (e.g., generosity, kindness; see Prediction 3). Nonetheless, neither their data set nor ours yielded evidence for such effects. Thus, we can be relatively certain that women's preferences for traits associated with partner investment do not shift across the cycle.

In summary, the effects of relationship length in the two reviews do not clearly support evolutionary psychology predictions. Only Gildersleeve et al. (2014) obtained the predicted cycle shifts within short-term relationships. However, these effects proved to be closely tied to imprecise measurement by the original researchers as well as by the meta-analysts. Furthermore, no precedent exists for predicting cycle shifts when relationship length is unspecified, and in both reviews the observed shifts in this context could be attributed to artifacts and to the influence of a small subset of studies.

Publication Bias

Both of the meta-analyses tested for publication bias using funnel plots and associated statistical tests. Gildersleeve et al.'s (2014) sample revealed little evidence for such bias. Wood et al.'s

(2014) review revealed considerable publication bias in research testing women's preferences for symmetric men (similar to prior meta-analyses on fluctuating asymmetry; e.g., Palmer, 1999; van Dongen & Gangestad, 2011), although analyses on preferences for dominance and masculinity did not show these effects. Funnel plot methods, however, provide only a best guess of the form and extent to which publication bias constrains meta-analytic results. The standard protection against such bias is a thorough literature search to identify the full population of unpublished work (Sutton, 2009).

Publication effects can be tested directly by comparing the findings of published versus unpublished research. Both reviews conducted this test. Wood et al. (2014) reported that only published research revealed predicted shifts across the cycle in preferences for masculinity, dominance, and symmetry. Gildersleeve et al. (2014, p. 1241) reported that after controlling for the effect of relationship context, larger effects were obtained in published than unpublished studies. They failed to mention that unpublished work actually revealed a null effect ($g = 0.04$, 95% CI [−0.03, 0.10], $T = 0.0$, $k = 12$), and that fertile women preferred men of purported genetic quality only in published studies ($g = 0.21$, 95% CI [0.12, 0.30], $T = 0.19$, $k = 38$), $Q_B(1) = 9.17$, $p = .002$. The same result emerged in both reviews when we conducted analyses on subsets of the data (e.g., the combined short-term and unspecified relationship studies, the short-term relationship findings only).

Publication status is a proxy variable that itself reflects the variety of causally relevant factors responsible for an effect being reported in the literature (Wood & Eagly, 2009). Consistent with the present publication effects, statistical significance is known to influence the publishability of research (Sutton, 2009). In research registry evaluations that record initial research protocols and then follow the planned research over time, studies with larger, statistically significant findings are more likely to be published (e.g., Dwan, Gamble, Williamson, & Kirkham, 2013). Another factor that possibly could affect publishability is the quality of studies' methods. However, in this literature, published and unpublished studies seemed to have comparable quality methods. Wood et al. (2014) reported that unpublished studies had similar strength manipulations of male stimulus attributes as published ones, and they were equally or more likely to validate menstrual cycle phase.

In summary, Wood et al.'s (2014) and Gildersleeve et al.'s (2014) reviews both found null effects in the unpublished literature, despite their different samples of unpublished studies. In both reviews, only the published studies supported evolutionary psychology predictions. These strong effects of publication status provide an additional explanation for the considerable inconsistency in the findings in this literature.

Summary and Conclusion

Our comment challenges Gildersleeve et al.'s (2014) claim to have found a "robust" effect. In fact, both meta-analytic reviews of this literature reported highly inconsistent evidence that fertile women especially prefer short-term affairs with men of apparent genetic quality. Fully a third of the studies in both reviews countered this evolutionary psychology prediction, and the effects in both reviews were driven by a few large-effect-size findings. Furthermore, contrary to all prior evolutionary psychology theorizing, the most consistent evidence for preference shifts across the menstrual cycle occurred in studies that did not specify a relationship length.

Research methods are a key to understanding the wide dispersion of effects in this literature. In both meta-analyses, studies with less precise methods for detecting cycle day and associated conception risk seemed to drive the evidence supporting evolutionary psychology predictions. That is, the predicted cycle shifts in women's mate preferences were not found in studies that validated cycle day or in short-term relationship studies that used narrower, more precise definitions of the fertile phase. Furthermore, continuous measures of conception risk did not detect shifts in the critical short-term context. However, research practices in this area appear to be improving, as researchers in recent years are coming to define fertile phases more narrowly and to use hormonal and other validation procedures to target ovulation and cycle phase. Along with these improvements, the size of study findings is declining, approximating zero in the most recently published studies.

Gildersleeve et al. (2014) failed to recognize the impact of research methods that we report in this comment, and their conclusions were influenced by studies with less precise methods of defining cycle day and the fertile phase. Because this single factor—quality of the methods used to identify cycle day—proved to have such strong effects on the findings of individual studies and on the reports of the two meta-analytic reviews, future menstrual cycle research will need to carefully implement the best procedures for defining fertility. At minimum, studies on menstrual cycles and social judgments should report hormonal validation of ovulation and cycle phase. With even more precise analyses, studies can isolate and assess the levels of specific hormones (e.g., progesterone, estrogen) believed to influence women's judgments.

Wood et al. (2014) published their review in the hope that a report of the inconsistent findings and research artifacts in the existing literature would spur future researchers to develop more sophisticated evolutionary accounts of the role of hormones in social behavior. Potential directions this research could take are suggested by the evidence that the two central variables in this area—women's mate preferences and the patterning of their menstrual cycles—are responsive to women's societal roles and thus vary across cultures and historical periods (see Wood et al., 2014). Also promising are social neuroendocrinology investigations of the hormonal fluctuations that correspond with the performance of certain social behaviors (Van Anders, in press). We believe that theoretical models that recognize the socially situated nature of human biology are likely to most effectively capture the evolutionary pressures that guide human reproduction.

References

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.

Cárdenas, R. A., & Harris, L. J. (2007). Do women's preferences for symmetry change across the menstrual cycle? *Evolution and Human Behavior, 28,* 96–105. doi:10.1016/j.evolhumbehav.2006.08.003

Chrisler, J. C., & Caplan, P. (2002). The strange case of Dr. Jekyll and Ms. Hyde: How PMS became a cultural phenomenon and a psychiatric disorder. *Annual Review of Sex Research, 13,* 274–306.

Chrisler, J. C., & Levy, K. B. (1990). The media construct a menstrual monster: A content analysis of PMS articles in the popular press. *Women & Health, 16,* 89–104. doi:10.1300/J013v16n02_07

DeBruine, L., Jones, B. C., Frederick, D. A., Haselton, M. G., Penton-Voak, I. S., & 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." *Evolutionary Psychology, 8,* 768–775.

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—An updated review. *PloS ONE, 8,* e66844. doi:10.1371/journal.pone.0066844

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: Biological Sciences, 265,* 927–933. doi:10.1098/rspb.1998.0380

Gildersleeve, K., DeBruine, L., Haselton, M. G., Frederick, D. A., Penton-Voak, I. S., Jones, B. C., & Perrett, D. I. (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., & Fales, M. R. (2014). Do women's mate preferences change across the ovulatory cycle? A meta-analytic review. *Psychological Bulletin, 140,* 1205–1259. doi:10.1037/a0035438

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

Hyde, J. S., & Salk, R. H. (in press). Comment: Menstrual cycle fluctuations in women's mate preferences. *Emotion Review*.

Ioannidis, J. P. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association, 294,* 218–228. doi:10.1001/jama.294.2.218

Jones, B. C., DeBruine, L. M., Perrett, D. I., Little, A. C., Feinberg, D. R., & Law Smith, M. J. (2008). Effects of menstrual cycle phase on face preferences. *Archives of Sexual Behavior, 37,* 78–84. doi:10.1007/s10508-007-9268-y

Lehrer, J. (2010, December 13). The truth wears off: Is there something wrong with the scientific method? *The New Yorker*. Retrieved from http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer?currentPage=all

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

Little, A. C., Jones, B. C., Burt, D. M., & Perrett, D. I. (2007). Preferences for symmetry in faces change across the menstrual cycle. *Biological Psychology, 76,* 209–216. doi:10.1016/j.biopsycho.2007.08.003

Oinonen, K. A., & Mazmanian, D. (2007). Facial symmetry detection ability changes across the menstrual cycle. *Biological Psychology, 75,* 136–145. doi:10.1016/j.biopsycho.2007.01.003

Palmer, A. R. (1999). Detecting publication bias in meta-analysis: A case study of fluctuating asymmetry and sexual selection. *The American Naturalist, 154,* 220–233. doi:10.1086/303223

Pawlowski, B., & Jasienska, G. (2005). Women's preferences for sexual dimorphism in height depend on menstrual cycle phase and expected duration of relationship. *Biological Psychology, 70,* 38–43. doi:10.1016/j.biopsycho.2005.02.002

Ruble, D. N. (1977, July 15). Premenstrual symptoms: A reinterpretation. *Science, 197,* 291–292. doi:10.1126/science.560058

Ruble, D. N., & Brooks-Gunn, J. (1979). Menstrual symptoms: A social cognition analysis. *Journal of Behavioral Medicine, 2,* 171–194. doi:10.1007/BF00846665

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435–452). New York, NY: Russell Sage Foundation.

Thornhill, R., Chapman, J. F., & Gangestad, S. W. (2013). Women's preferences for men's scents associated with testosterone and cortisol levels: Patterns across the ovulatory cycle. *Evolution and Human Behavior, 34,* 216–221. doi:10.1016/j.evolhumbehav.2013.01.003

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

Tybur, J. M., & Gangestad, S. W. (2011). Mate preferences and infectious disease: Theoretical considerations and evidence in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366,* 3375–3388. doi:10.1098/rstb.2011.0136

Van Anders, S. (in press). Comment: The social neuroendocrinology example: Incorporating culture resolves biobehavioral evolutionary paradoxes. *Emotion Review*.

Van Dongen, S., & Gangestad, S. W. (2011). Human fluctuating asymmetry in relation to health and quality: A meta-analysis. *Evolution and Human Behavior, 32,* 380–398. doi:10.1016/j.evolhumbehav.2011.03.002

Vitzthum, V. J. (2009). The ecology and evolutionary endocrinology of reproduction in the human female. *American Journal of Physical Anthropology, 140*(Suppl. 49), 95–136. doi:10.1002/ajpa.21195

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7,* 632–638. doi:10.1177/1745691612463078

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., & Eagly, A. H. (2009). Advantages of certainty and uncertainty. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 455–472). New York, NY: Russell Sage Foundation.

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

Received December 14, 2013
Revision received February 23, 2014
Accepted March 17, 2014 ■