Psychological Bulletin
2014, Vol. 140, No. 5, 1272–1280

0033-2909/14/$12.00 http://dx.doi.org/10.1037/a0037714

REPLY

Meta-Analyses and *P*-Curves Support Robust Cycle Shifts in Women's Mate Preferences: Reply to Wood and Carden (2014) and Harris, Pashler, and Mickes (2014)

Kelly Gildersleeve, Martie G. Haselton, and Melissa R. Fales
University of California, Los Angeles

Two meta-analyses evaluated shifts across the ovulatory cycle in women's mate preferences but reported very different findings. In this journal, we reported robust evidence for the pattern of cycle shifts predicted by the ovulatory shift hypothesis (Gildersleeve, Haselton, & Fales, 2014). However, Wood, Kressel, Joshi, and Louie (2014) claimed an absence of compelling support for this hypothesis and asserted that the few significant cycle shifts they observed were false positives resulting from publication bias, *p*-hacking, or other research artifacts. How could 2 meta-analyses of the same literature reach such different conclusions? We reanalyzed the data compiled by Wood et al. These analyses revealed problems in Wood et al.'s meta-analysis—some of which are reproduced in Wood and Carden's (2014) comment in the current issue of this journal—that led them to overlook clear evidence for the ovulatory shift hypothesis in their own set of effects. In addition, we present right-skewed *p*-curves that directly contradict speculations by Wood et al.; Wood and Carden; and Harris, Pashler, and Mickes (2014) that supportive findings in the cycle shift literature are false positives. Therefore, evidence from both of the meta-analyses and the *p*-curves strongly supports genuine, robust effects consistent with the ovulatory shift hypothesis and contradicts claims that these effects merely reflect publication bias, *p*-hacking, or other research artifacts. Unfounded speculations about *p*-hacking distort the research record and risk unfairly damaging researchers' reputations; they should therefore be made only on the basis of firm evidence.

Keywords: ovulation, mate preferences, meta-analysis, *p*-curves, *p*-hacking

Supplemental materials: http://dx.doi.org/10.1037/a0037714.supp

In our meta-analytic review (Gildersleeve, Haselton, & Fales, 2014), we reported support for cycle shifts in women's mate preferences, as predicted by the ovulatory shift hypothesis. Shortly after our meta-analysis appeared online in *Psychological Bulletin*, a separate meta-analysis by Wood, Kressel, Joshi, and Louie (2014) appeared online in *Emotion Review*. In contrast to our findings, Wood et al. claimed that their analysis "failed to support evolutionary psychology predictions about women's evolved preferences for male attributes across the menstrual cycle" (p. 245) and further that "the few significant preference shifts appeared to be research artifacts" (p. 229). Citing the Wood et al. meta-analysis, Wood and Carden (2014) and Harris, Pashler, and Mickes (2014) asserted that the robust cycle shifts in our meta-analysis reflect false positives. Specifically, they speculated that supportive findings in the published literature do not reflect true cycle shifts but, rather, have been artificially produced by publication bias, "*p*-hacking"—wherein researchers capitalize on hidden "degrees of freedom" in data analysis or reporting practices in order to claim statistically significant results (Simonsohn, Nelson, & Simmons, 2014a)—or other research artifacts.

In this reply, we demonstrate that these claims are flawed. We recap the ovulatory shift hypothesis, which Wood and Carden (2014) misrepresent. We document methods used by Wood et al. (2014) that account for the failure of their meta-analysis to reveal genuine cycle shifts. Finally, we present *p*-curves that directly contradict speculations about *p*-hacking and that provide additional evidence for genuine cycle shifts consistent with the ovulatory shift hypothesis.

Kelly Gildersleeve, Department of Psychology and Center for Behavior, Evolution, and Culture, University of California, Los Angeles; Martie G. Haselton, Departments of Psychology and Communication Studies, Institute for Society and Genetics, and Center for Behavior, Evolution, and Culture, University of California, Los Angeles; Melissa R. Fales, Department of Psychology and Center for Behavior, Evolution, and Culture, University of California, Los Angeles.

We thank Britt Ahlstrom, Tran Dinh, Steve Gangestad, Nick Grebe, Bill von Hippel, Christina Larson, Aaron Lukaszewski, Damian Murray, David Pinsof, and Shimon Sapphire-Bernstein for their constructive suggestions during the preparation of this response.

Correspondence concerning this article should be addressed to Kelly Gildersleeve, Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Los Angeles, CA 90095, or to Martie G. Haselton, Department of Psychology, Department of Communication Studies, Institute for Society and Genetics, University of California, Los Angeles, Box 951538, Rolfe Hall, Room 2322, Los Angeles, CA 90095. E-mail: kellygildersleeve@gmail.com or haselton@ucla.edu

The Ovulatory Shift Hypothesis

The central goal of our meta-analysis was to evaluate support in the empirical literature for the ovulatory shift hypothesis (see Gangestad, Thornhill, & Garver-Apgar, 2005). The key prediction of the ovulatory shift hypothesis is that, at high fertility as compared with low fertility, heterosexual women are more attracted to men possessing characteristics that reflected genetic quality in ancestral males. This shift in mate preferences is expected to be most pronounced when women evaluate men's desirability as sex partners (e.g., short-term affair partners) but absent or very weak when women evaluate men's desirability as long-term social partners (see Gangestad et al., 2005; Gildersleeve et al., 2014). Thus, many studies in this literature ask women to evaluate men with reference to a specific relationship type.

In their commentary, Wood and Carden (2014) claimed that "No precedent exists for predicting cycle shifts when relationship length is unspecified" (p. 1269) and that Gildersleeve et al. (2014) "are the first researchers, to our knowledge, to make such a prediction, and it counters evolutionary psychology reasoning" (p. 1269). However, theory and the published literature reveal these claims to be seriously misinformed. Rather than asking women to evaluate men with reference to a specific relationship type, many studies in this area have simply asked women to evaluate men's "sexiness" (e.g., Gangestad & Thornhill, 1998; Rikowski & Grammer, 1999; Thornhill & Gangestad, 1999; Thornhill et al., 2003), "physical attractiveness" (e.g., Penton-Voak et al., 1999, Study 1), or "attractiveness" (e.g., Johnston et al., 2001; Penton-Voak & Perrett, 2000; see also Bressan & Stranieri, 2008; Caryl et al., 2009; Feinberg et al., 2006; Harris, 2011; Havlicek et al., 2005; Hromatko et al., 2006; Jones et al., 2005; Koehler et al., 2006; Little et al., 2007, 2008; Moore et al., 2011; Oinonen & Mazmanian, 2007; Provost et al., 2008; Rantala et al., 2006; Roney & Simmons, 2008; Roney et al., 2011; Rupp, James, et al., 2009; Rupp, Librach, et al., 2009; Thornhill et al., 2013; Vaughn et al., 2010; Welling et al., 2007). Penton-Voak et al. (1999, Study 2) were the first to explicitly differentiate between evaluations of men as short-term sex partners versus long-term relationship partners. Although this was a methodological advance, evaluations of "sexiness," "physical attractiveness," and the like nonetheless capture desirability as a sex partner (see Gildersleeve et al., 2014, p. 1207). Therefore, we structured our meta-analysis to evaluate whether cycle shifts were strongest in an explicitly specified short-term relationship context, intermediate when women evaluated men's sexiness or attractiveness without reference to a specified context, and absent or very weak in an explicitly specified long-term relationship context. We found evidence for precisely this pattern.

Why Did Wood et al.'s (2014) and Gildersleeve et al.'s (2014) Meta-Analyses Produce Different Findings?

Our focal analysis revealed statistically significant relationship context-dependent cycle shifts, as predicted by the ovulatory shift hypothesis. In addition, despite relatively low power, follow-up analyses revealed the predicted pattern of cycle shifts in preferences for some specific male characteristics (e.g., body masculinity and behavioral dominance). By contrast, Wood et al.'s (2014) analysis revealed one significant relationship context-independent cycle shift in women's preference for symmetry and a nearly significant context-independent cycle shift in women's preference for masculinity, both of which the authors attributed to research artifacts.

How is it that, based on reviews of largely the same sets of findings, these two groups of authors came to such different conclusions? In fact, there were many differences between the two meta-analyses. Here, we focus on those that are most important for explaining discrepancies in their findings (see the supplemental materials for additional differences).

Differences in Samples of Effects

The literature on cycle shifts in women's mate preferences has rapidly expanded over the past two decades, and studies have varied in the strength with which they have tested the predictions of the ovulatory shift hypothesis. We designed inclusion criteria to balance trade-offs between capturing the diversity of mate preference measures used in this literature and limiting our sample to those studies that would provide a strong test of the hypothesis. Specifically, we created two nested samples of effects. The "Broad" sample included measures both of "revealed" and "stated" preferences for all male characteristics researchers have examined in relation to the ovulatory shift hypothesis (Gildersleeve et al., 2014, see pp. 1235–1236).[1] In contrast, the "Narrow" sample excluded measures of "stated" preferences, which might tend to elicit reports of crystallized, general preferences and therefore fail to capture transient shifts in preferences (see p. 1236). It also excluded characteristics for which links with genetic quality are dubious (see p. 1235). We found the predicted pattern of cycle shifts in both samples, although as expected, effects were stronger in the Narrow sample.[2] In contrast, Wood et al. (2014) used even more relaxed inclusion criteria than in our Broad sample. As a result, their sample included confounded effects and effects for which predictions are weak (see the supplemental materials for details).

Problems in Effect Size Coding in the Wood et al. (2014) Analysis

The Wood et al. (2014) sample included an unrealistically large number of effect sizes (g) coded as exactly 0.00, indicating no difference whatsoever between women's preference for a given characteristic at high versus low fertility (17 effects—14%—of their total sample; eight effects—11%—of the subsample expected to be positive under the ovulatory shift hypothesis). For seven of these effects, we were ourselves unable to obtain sufficient information from study authors to compute a precise effect size. To avoid introducing error into overall effect size estimates, we coded these effects as missing. In other such cases, we computed an effect different from zero (e.g., Teatero, 2009). In one case, we located a figure that could be used to estimate the effect (the cycle shift in women's preference for vocal masculinity;

[1] Wood and Carden (2014) claimed that we *excluded* measures of "stated" preferences, though clearly we did not.

[2] In their commentary, Wood and Carden (2014) reported reanalyses of our Broad sample only, despite our having presented clear arguments for why the Narrow sample offers a stronger test of the ovulatory shift hypothesis.

Figure 2 from Feinberg et al., 2006, p. 220). Using a data extraction tool (Rohatgi, 2011), we estimated Cohen's $d \approx .29$ to .36 (depending on assumptions about the correlation between preference strength at high and low fertility). Therefore, this effect is clearly nonzero.[3] Whatever the reason for the many 0.00 effects in Wood et al.'s analysis, their remarkable frequency and discrepancies between Wood et al.'s data set and ours raise questions about the accuracy of the effect sizes they reported (also see the supplemental materials).

Differences in Analysis Strategies

Although differences between the two meta-analyses in content are noteworthy, differences in their analytic approaches are perhaps even more important.

The importance of relationship context. The ovulatory shift hypothesis predicts relationship context-dependent cycle shifts in women's mate preferences. Wood et al. (2014) nonetheless presented a number of analyses that estimated cycle shifts across *all* relationship contexts, including a long-term context, *in which no cycle shifts are predicted*. Theory clearly indicates that examining cycle shifts across all relationship contexts will seriously underestimate predicted cycle shifts (e.g., by 33% if 33% of the effects in a given analysis were measured in a long-term mating context and if the effect size in that context were in fact zero). Furthermore, all of the moderation and publication bias analyses that Wood et al. presented examined effects across all relationship contexts. Wood and Carden (2014) repeated this error in many of the moderation and publication bias analyses they conducted on our meta-analysis sample (Gildersleeve et al., 2014). It makes no sense to examine associations between study characteristics and effect size or to examine publication bias across effects for which no cycle shifts are predicted.

More powerful, overall analysis. Wood et al. (2014) examined cycle shifts in women's preferences for specific categories of male characteristics (e.g., symmetry) but did not examine cycle shifts across all categories for which the ovulatory shift hypothesis predicts effects (i.e., across symmetry, masculinity, dominance, etc.). Wood and Carden (2014) claimed that Wood et al. (2014) "structured their review to evaluate each attribute separately in order to maximize the likelihood of detecting and interpreting cycle shifts" (p. 1266). In fact, they did the opposite: power to detect cycle shifts would of course have been greater had they combined across these categories (the number of effects included in their analyses was as few as $k = 5$). The focal analysis in our meta-analysis was such an overall analysis, and it revealed a robust pattern of cycle shifts consistent with the ovulatory shift hypothesis.

A reanalysis of Wood et al.'s (2014) effects. Had Wood et al. (2014) conducted the more powerful, overall analysis across male characteristics, would they have found support for the predicted pattern of cycle shifts? To address this question, we reanalyzed Wood et al.'s data, combining their symmetry, masculinity, dominance, and testosterone samples (i.e., all effects for which cycle shifts are predicted). We left all effect sizes and confidence intervals exactly as reported in Wood et al.'s supplement. We used the same multilevel analytic approach as in our meta-analysis. This approach, unlike that used by Wood et al., properly accounts for the nonindependence of multiple effects nested within a single study (see Gildersleeve et al., 2014, p. 1238). Analyses revealed a small, marginally significant weighted mean cycle shift in a short-term context ($g = 0.11$, $SE = 0.06$, $p = .07$); a small, highly significant weighted mean cycle shift in an unspecified context ($g = 0.15$, $SE = 0.04$, $p = .001$); and no shift in a long-term context ($g = 0.03$, $SE = 0.06$, $p = .57$). Clearly, then, had Wood et al. conducted the more powerful, overall analysis in their data set, they would have observed far more support for the ovulatory shift hypothesis than they claimed to find.

Are Cycle Shifts "Research Artifacts"?

Despite problems in Wood et al.'s (2014) effect size coding that were likely to downwardly bias effect size estimates, our reanalysis of Wood et al.'s data set revealed predicted cycle shifts. They argued that any such effects are "artifacts of research practices" (p. 245). In support of this claim, they presented analyses showing that studies in their sample tended to produce larger effect sizes when they were (a) published, (b) published earlier, or (c) used wider high-fertility windows (which they problematically assert are less precise). Wood and Carden (2014) replicated these associations in our Broad sample of effects and also pointed to variability in both meta-analysis samples to support their argument that positive findings in this literature are mere false positives. Are such claims justified?

Does Publication Bias Imply No True Effects?

Wood et al. (2014) and Wood and Carden (2014) made a point of noting that published studies tended to yield larger predicted cycle shifts than do unpublished studies. As Simonsohn (2013) explained, however, the declaration that publication bias exists in a literature is not newsworthy. As long as statistical significance improves articles' chances of making it into the published literature, published studies will *of course* report more robust findings than unpublished studies. Publication bias is to be expected *no matter what the area of psychology*, and its mere existence is not diagnostic of whether true effects exist. As we discuss below, p-curves can be used to examine the impact of publication bias, and p-curves are strongly inconsistent with the claim that publication bias accounts for the positive findings in the cycle shifts literature.

Has Effect Size Declined Over Time?

Wood and Carden (2014) claimed to have found decline effects in both meta-analysis samples. Wood and Carden interpret these effects as evidence that false positives in this area are increasingly being replaced with true null effects.

Yet Wood and Carden (2014) neglected other plausible explanations for changes in effect size over time. For example, study characteristics change over time as researchers develop new methods, test new predictions, attempt to establish the boundaries of previously documented effects, and so on. Such systematic

[3] Because we were lacking the correlation needed to compute a precise effect size (the author was unable to provide that information), we coded this effect as missing and excluded it from our meta-analysis (Gildersleeve et al., 2014, Table 1, p. 1212).

changes can reduce mean effect size over time, which can masquerade as a decline effect. As an illustration, all studies examining cycle shifts in women's preferences for scents associated with symmetry appeared from 1998 to 2003. In contrast, studies examining cycle shifts in preferences for facial symmetry appeared in 2002 and have continued into the current decade (perhaps, in part, because these studies, unlike scent studies, can be conducted online). Cycle shifts in preferences for scents associated with symmetry tend to be stronger than those for facial symmetry. Accordingly, controlling for preference type within Wood et al.'s (2014) symmetry sample renders the decline effect null (Gangestad, 2014). Given the many possible confounds not controlled for in Wood et al.'s analyses, we see no reason to favor Wood and Carden's research artifact interpretation of the association between publication year and effect size over alternative explanations. Moreover, *p*-curves presented below contradict the artifact interpretation.

Do Narrower High-Fertility Windows Better Estimate Cycle Shifts?

Many studies examining cycle shifts have categorized women as "high-fertility" or "low-fertility" according to their current cycle day, typically as estimated based on their recalled date of last menstrual onset. This method requires that researchers choose which days to count as high and low fertility. Wood et al. (2014) and Wood and Carden (2014) asserted that broader high-fertility windows are less precise than narrower windows, citing Wilcox et al. (2001) as support. They also reported that, in both meta-analysis samples, studies using broader high-fertility windows (e.g., 9 days) tended to produce larger effect sizes than studies using narrower windows (e.g., 6 days). They interpreted these associations as inconsistent with true cycle shifts and consistent with *p*-hacking. For example, Wood and Carden speculated that researchers

> flexibly defined the fertile phase based on study findings . . . [they] might have begun with a window of six 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. (p. 1267)

Reinforcing the point, Harris et al. (2014) stated that they "are unable to think of any completely benign explanation for this pattern" (p. 1263).

But are broader high-fertility windows less precise? It is true that, *for an individual cycle*, the period of highest fertility generally includes just the 6 days ending with ovulation, whereas the remainder of the cycle is effectively nonfertile (Wilcox et al., 1995). Notably, however, Wood and Carden (2014) and Harris et al. (2014, p. 1263) failed to appreciate that the timing of ovulation varies across cycles both within and between women (see Figure 1), and studies in this literature have rarely verified ovulation. Critically, in a sample of observations across multiple cycles for which the day of ovulation is both variable and unknown, the high-fertility window will be distributed across a range wider than just 6 days (see also Figure 1).

Wilcox et al. (2001) used actuarial methods to estimate the conception probability (fertility) associated with any given day of

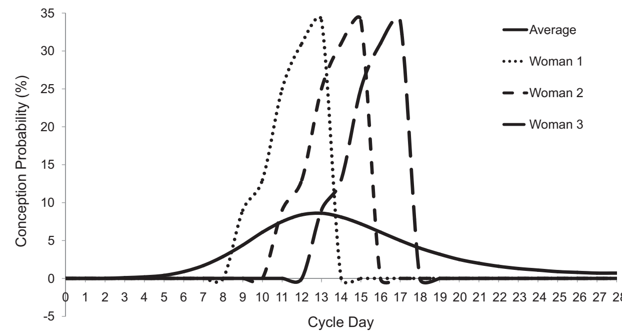

Figure 1. Probability of conceiving from a single act of sexual intercourse as a function of cycle day: Values across women and for hypothetical individual women. Average conception probability values are based on regularly cycling women with no verification of ovulation (Wilcox et al., 2001). Conception probabilities for hypothetical Woman 1, Woman 2, and Woman 3 are based on women with hormonally verified ovulation on cycle days 13, 15, and 17, respectively (Wilcox, Weinberg, & Baird, 1995). Given variability in the timing of ovulation within the cycle, the high-fertility window—which spans approximately 6 days for a given cycle—is spread across a wider range of days for women considered in aggregate.

the cycle in light of this variation (see their Table 1, p. 213). Using their estimates, one can split the cycle at the midpoint of conception probability to identify "high-fertility" and "low-fertility" days—those closer to the maximum and minimum conception probability values, respectively. This creates a *9*-day high-fertility window. Thus, researchers who correctly understood the implications of Wilcox et al.'s article (or earlier reports; e.g., Jochle, 1973), and who had information only about a woman's last menstrual onset, would have little reason to use a 6-day high-fertility window. In fact, it would be reasonable to expect studies using broader (e.g., 9-day) high-fertility windows to generate larger effects in the presence of true cycle shifts.[4] Given this, Wood et al.'s decision to ask researchers to recompute effects for their meta-analysis based on a narrower high-fertility window (resulting in an average window of 6.85 days in their short-term context) is expected, all else equal, to downwardly bias effect size estimates. Note that we are not advocating the use of a 9-day high-fertility window. Rather, we urge researchers to empirically evaluate different methods of assessing fertility (see Gildersleeve et al., 2014, pp. 1254–1255).

Does Variability Indicate Inconsistent Evidence?

Wood and Carden (2014) pointed to variability in the magnitude, direction, and statistical significance of the effects in both meta-analysis samples to support their contention that predicted effects are not robust. Such variability is, of course, perfectly consistent with the existence of true effects, as meta-analysts—including Wood—have acknowledged (see, e.g., Eagly & Wood, 1991; Wood et al., 1994). Sampling error alone produces a distribution of effect sizes, with some overestimating and some under-

[4] This will be true when researchers include all cycle days in analyses. Predictions are complicated when researchers exclude certain days (e.g., excluding menstrual days in order to minimize extraneous variation tied to associated physical symptoms).

estimating the true effect. Furthermore, because statistical significance depends on effect size *and* sample size, even large effects will fail to reach statistical significance in small samples. Indeed, a key rationale for conducting meta-analyses is to evaluate the magnitude and robustness of mean effects in light of this variation (see Borenstein, Hedges, Higgins, & Rothstein, 2010). Finally, as we have highlighted, Wood and Carden examined this variation across all relationship contexts, counting negative and nonsignificant effects in a long-term relationship context as inconsistent with the ovulatory shift hypothesis, even though no cycle shift is predicted in that context. This inappropriate method created the illusion of more variability than actually exists (see also the supplemental materials).

P-Curves Reveal Evidence for Real Effects, Not *P*-Hacking

To return to where we began: Our meta-analysis yielded robust evidence for cycle shifts predicted by the ovulatory shift hypothesis, whereas Wood et al.'s (2014) produced little evidence for such effects. Furthermore, Wood et al. speculated that the few significant cycle shifts they detected were false positives resulting from publication bias, *p*-hacking, or other research artifacts. Wood and Carden (2014) and Harris et al. (2014) made similar arguments in their commentaries. For example, Harris et al. speculated that "many *or perhaps all* (emphasis added) results [in the cycle shifts literature] may have been created by exploitation of unacknowledged degrees of freedom ('*p*-hacking')" (p. 1260) and concluded their commentary by questioning whether positive findings in this literature "have been invented out of the whole cloth" (p. 1263).[5]

Although all of these groups of authors speculated that positive findings in the cycle shift literature resulted from *p*-hacking, none of them presented any direct evidence that *p*-hacking actually has occurred or that it can explain positive findings in the absence of true effects. In fact, they could have performed such tests using publicly available evidence: published *p*-values. The same group of researchers who initially raised awareness about *p*-hacking and its potential to increase Type I error rates recently developed a method for assessing whether a set of positive research findings has *evidential value*—shows hallmarks of true, robust effects—or appears to reflect selective reporting (e.g., due to *p*-hacking or publication bias; Simonsohn et al., 2014a). Thus, this method can be used to examine which of the two meta-analyses—Wood et al.'s (2014) or ours (Gildersleeve et al., 2014)—is more likely to be correct.

The method involves examining the skewness of the frequency distribution of statistically significant p values ($p < .05$) for a set of published research findings—the "*p*-curve" (Simonsohn et al., 2014a). If no true effect exists, and no *p*-hacking has occurred, a *p*-curve will be uniform (flat), with equal frequencies of p values between .04 and .05, .03 and .04, .02 and .03, and so on. If a nonzero true effect exists, a *p*-curve will be right-skewed, with a relatively high frequency of p values closer to 0 (e.g., Hung, O'Neill, Bauer, & Kohne, 1997). If no effect exists, but *p*-hacking has produced the illusion of an effect, a *p*-curve will be left-skewed, with a relatively high frequency of p values closer to .05. This is because, in the absence of a true effect, it is easier to "hack" a p value to just under .05 than to obtain a p closer to 0. Also, a *p*-hacker might plausibly have the limited ambition to stop *p*-hacking once he or she has surpassed the threshold for publication ($<.05$).[6] Indeed, the "replication crisis" in psychology was triggered in part by the suspiciously high frequency of p values just under .05 in the field at large (Masicampo & Lalande, 2012; see also Leggett, Thomas, Loetscher, & Nicholls, 2013).

If Wood, Harris, and colleagues' suspicions are correct—that no true cycle shifts exist, and positive findings merely reflect selective reporting—*p*-curves of published theory-supportive findings in this literature should be *flat* (consistent with publication bias and $d = 0$) *or left-skewed* (consistent with *p*-hacking and $d = 0$). In contrast, if our conclusions are correct, and true cycle shifts exist, *p*-curves should be right-skewed. We constructed several *p*-curves to examine the pattern of p values for cycle shifts in mate preferences and theoretically related findings that Harris et al. (2014) also claimed are likely to have been *p*-hacked (e.g., Pillsworth & Haselton, 2006). Figures 2 and 3 present two key *p*-curves. We present additional *p*-curves, selection criteria, and a *p*-curve disclosure form (as recommended by Simonsohn et al., 2014a) in the supplemental materials.

The *p*-curve presented in Figure 2 contains all p values $< .05$ from effects included in our meta-analysis that evaluated one of the following predictions of the ovulatory shift hypothesis (excluding statistically nonindependent *p*s; see the supplemental materials): the *Cycle Shift Prediction*, the prediction that women's attraction to cues of genetic quality will increase at high relative to low fertility within a short-term or unspecified relationship context (i.e., the simple or main effect of fertility on such preferences), or the *Context Moderation Prediction*, the prediction that this increase at high relative to low fertility (cycle shift) will be stronger in a short-term than in long-term relationship context (i.e., the Fertility $\times$ Context interaction). The *p*-curve presented in Figure 3 contains the *p*s just described, as well as additional p values that evaluated the *Partner Qualities Moderation Prediction*. This prediction is that women in relationships experience cycle shifts in their feelings about their current romantic partner and other men, but the precise nature of these shifts depends on the extent to

[5] In addition to claims about *p*-hacking, Harris et al.'s (2014) commentary consists mostly of claims that we committed "statistical and logical errors" (e.g., p. 1260). All of these claims rest on misreadings or mischaracterizations of what we actually said and of the cycle shift literature. To provide a few illustrative examples (there are many additional examples in Harris et al.'s, 2014, commentary), Harris et al. stated that Gildersleeve et al. (2014) "maintained that the use of continuous fertility methods . . . gets around the problem of flexibility" (p. 1261). We actually said, "We cannot definitively rule out the possibility that researchers chose, post hoc, to use a continuous fertility variable because doing so produced the predicted cycle shifts" (Gildersleeve et al., 2014, p. 1249). Likewise, about one of our figures, Harris et al. said, "We find it odd that Gildersleeve et al. would suggest that readers can rely on 'eyeball' judgments of complex multidimensional data" (p. 1263). We actually said, "a visual analysis cannot replace rigorous statistical tests" (Gildersleeve et al., 2014, p. 1249). Last and perhaps most important, Harris et al. repeatedly claimed, "recent attempts at exact replication of fertility results have mostly failed" (p. 1260). However, this claim appears to be based on a single finding from a study examining cycle shifts in preferences for facial masculinity that Harris (2011) conducted as a close, but not exact, replication of early studies in this area (Penton-Voak & Perrett, 2000; Penton-Voak et al., 1999).

[6] This is not merely speculation. Simonsohn et al. (2014a, 2014b) also documented this expected shape of *p*-curves using extensive simulations of *p*-hacking behavior outlined previously in Simmons et al. (2011).

which their current partner possesses characteristics hypothesized to be particularly desirable to women at high fertility (e.g., symmetry, facial masculinity, and sexual attractiveness).[7] We included *p* values corresponding to this prediction because it is closely related to the two predictions above, and Harris et al. (2014) discussed findings in this literature at length, speculating that they resulted from *p*-hacking.[8] The *p*-curves in Figures 2 and 3 present exact two-tailed *p* values that we calculated from test statistics and degrees of freedom reported in the published articles or by study authors. The supplemental materials present additional *p*-curves based on values as reported in the published articles (and therefore includes *p*s based on directed and one-tailed tests).

All of the *p*-curves we plotted, including those presented here and in the supplemental materials, are significantly right-skewed, often strongly so. Therefore, these *p*-curves contradict speculations that cycle shifts in women's mate preferences can be explained by publication bias or *p*-hacking[9] and offer further support for our conclusions that cycle shift effects are genuine and robust.[10]

One might reasonably ask whether *p*-hackers really do stop once they have obtained a *p* value just under .05 or continue until they have obtained an even smaller *p* value. For example, researchers might try out multiple high-fertility windows and then select the analysis that produced the *smallest p* value. Would that practice, with no true effects, generate *p*-curves that are right-skewed, like those we have presented? Simulations of such a practice indicate that this is extremely unlikely.[11] In nearly all cases, sampling from multiple plausible fertile windows when no true effect exists produces *left*-skewed *p*-curves. Only when hypothetical researchers began by trying out 10 or more high-fertility windows and selecting the lowest *p* value from these analyses, is the *p*-curve even slightly right-skewed. It strikes us as highly unrealistic that nearly *every* researcher publishing a significant cycle shift began by examining 10 or more high-fertility windows and unfailingly selected the window producing the smallest *p* value, without regard to precedent in the literature or other factors. Furthermore, if Wood and Carden (2014) were correct that researchers in this area have generally started with a 6-day window and then tried out successively larger windows until they obtained $p < .05$, this strategy would almost certainly have resulted in a *left*-skewed *p*-curve—exactly the opposite of what we observed.

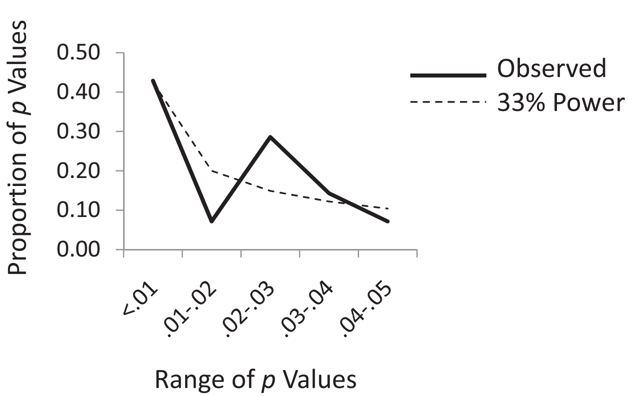

Figure 2. *P*-curve of exact two-tailed *p* values evaluating the Cycle Shift Prediction and Context Moderation Prediction. $N = 14$ *p* values, Total *N* across studies = 1,375. *P*-curve is significantly right skewed, $\chi^2(28) = 47.40$, $p = .01$.

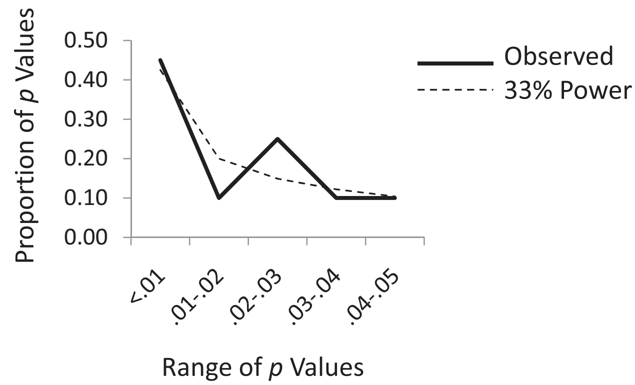

Figure 3. *P*-curve of exact two-tailed *p* values evaluating the Cycle Shift Prediction, Context Moderation Prediction, and Partner Qualities Moderation Prediction. $N = 20$ *p* values, Total *N* across studies = 1,640. *P*-curve is significantly right skewed, $\chi^2(40) = 75.98$, $p = .0005$.

P-curves can also speak to the appropriate interpretation of associations between effect size and publication status, publication year, and high-fertility window size. As we noted above, a variety of factors could account for such associations. By virtue of providing evidence consistent with genuine nonzero effects, *p*-curves

[7] For example, one prediction is that women with partners who are relatively low in sexual attractiveness experience an increase in attraction to *other men* at high relative to low fertility (Pillsworth & Haselton, 2006). For further details see the supplemental materials.

[8] Notably, nearly all effects in this literature used luteinizing hormone tests to verify ovulation (Guermandi et al., 2001). Therefore, this set of studies provides some of the most rigorous tests of predictions that follow from the ovulatory shift hypothesis.

[9] *P*-curves are also significantly right-skewed when sets of effects in our meta-analysis (those testing the Cycle Shift Prediction) and in the related literature on relationship dynamics (those testing the Partner Qualities Moderation Prediction) are examined separately (see the supplemental materials; note that there were too few effects—two and four effects in the exact and reported sets of *p*-values, respectively—to examine the Context Moderation Prediction separately).

[10] *P*-curves also offer a straightforward new method for estimating effect size (Simonsohn et al., 2014b). Skewness of *p*-curves is a function of sample size and effect size (as both increase, right skew increases). Therefore, if one knows the *N*s and *p*s, one can calculate the best fitting *d* characterizing a set of effects (Simonsohn et al., 2014b). The range of estimated effect sizes for all *p*-curves we present (using the K - S estimation procedure) is $d = .28–.78$, with larger effects in curves examining the Partner Qualities Moderation Prediction. For all effects in aggregate, $d = .47$; see Figure 3. Note that one can compute *d* using only exact two-tailed *p* values. An important proviso is that the validity of these estimations depends on researchers having reported the correct *N*s and test statistics (required to compute exact *p*s). We have no reason to believe that widespread reporting errors exist in these literatures; however, verifying all *N*s and test statistics will be required to achieve the most precise estimates of *d* using this new method.

[11] We conducted a simulation of 10,000 replicated samples ($N = 50$, drawn from a population with a normally distributed outcome variable and 29-day cycles; Haselton, Grebe, Gildersleeve, & Gangestad, 2014). Thirteen different plausible fertile phase windows, ranging from 6 to 10 days, were defined, and a number of different "rules" for examining and selecting window size explored, differing with respect to how many windows were examined initially (one, two, five, eight, 11, and all 13), where researchers always took the lowest *p* value out of that initial set and moved on to additional windows if no *p* was less than .05, stopping once a *p* less than .05 was achieved.

are inconsistent with Wood et al.'s (2014) and Wood and Carden's (2014) interpretation of these associations as evidence for false positives.

Conclusions

As we previously reported, our meta-analysis revealed strong support for the ovulatory shift hypothesis. New analyses using the *p*-curve method again revealed strong support for genuine cycle shifts as predicted by the ovulatory shift hypothesis. Claims by Wood et al. (2014), Wood and Carden (2014), and Harris et al. (2014) that the abundance of positive findings in the cycle shifts literature merely reflects publication bias, *p*-hacking, or other research artifacts did not anticipate and cannot explain these new findings.

We emphasize that Wood et al.'s (2014) conclusions from their meta-analysis are at odds not only with the conclusions of our meta-analysis *but also with our reanalysis of their data and with the p-curves*. We have identified several crucial factors responsible for these discrepancies: Wood et al. made multiple methodological and analytic decisions that weakened their ability to detect real effects. They also inappropriately cited possible evidence of publication bias and variability in research findings as inconsistent with true effects; claimed to have identified a decline effect in the literature, when in fact, confounds may well explain the observed association between effect size and publication year; and mistakenly interpreted an association between effect size and high-fertility window size as evidence of *p*-hacking, but in doing so, ultimately revealed their own misunderstanding of conception probabilities across the cycle.

Given recent doubts about the evidential value of published research findings, many researchers have called for "cleaning up" psychological science. We fully support this effort. However, just as claims regarding the existence of hypothesized effects should be supported with strong empirical evidence, so should claims regarding whether *p*-hacking or other practices have produced the illusion of positive evidence. In the case of the literature on cycle shifts in women's mate preferences, speculations about a widespread "false positive problem" are unwarranted. Had Wood et al. (2014), Wood and Carden (2014), and Harris et al. (2014) examined the *publicly available evidence*—published *p* values—they would have found little evidence that the literature on cycle shifts is plagued by false positives. Their speculations were unfounded and very serious. Such claims distort the scientific record and risk unfairly damaging the reputations of scholars in this area. Indeed, they have the potential to undermine the endeavor to clean up psychological science. Real progress will occur only when criticisms of science are as rigorous as the science itself.

References

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). *Introduction to meta-analysis*. Chichester, United Kingdom: Wiley.

Bressan, P., & Stranieri, D. (2008). The best men are (not always) already taken: Female preference for single versus attached males depends on conception risk. *Psychological Science, 19,* 145–151. doi:10.1111/j.1467-9280.2008.02060.x

Caryl, P. G., Bean, J. E., Smallwood, E. B., Barron, J. C., Tully, L., & Allerhand, M. (2009). Women' preference for male pupil size: Effects of conception risk, sociosexuality, and relationship status. *Personality and Individual Differences, 46,* 503–508. doi:10.1016/j.paid.2008.11.024

Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin, 17,* 306–315. doi:10.1177/0146167291173011

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

Gangestad, S. W. (2014). *Wood et al.'s (2014) suspicions of p-hacking in the ovulation cycle literature are contradicted by publicly-available evidence*. Manuscript submitted for publication.

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.

Gangestad, S. W., Grebe, N. M., & Gao, A. (2014). *Of ovulatory cycle effects and P curves: Evidence for real phenomena, not p-hacking*. Manuscript submitted for publication.

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.

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

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

Gangestad, S. W., Thornhill, R., & Garver-Apgar, C. E. (2010). Men's facial masculinity predicts changes in their female partners' sexual interests across the ovulatory cycle, whereas men's intelligence does not. *Evolution and Human Behavior, 31,* 412–424.

Garver-Apgar, C. E., Gangestad, S. W., Thornhill, R., Miller, R. D., & Olp, J. J. (2006). Major histocompatibility complex alleles, sexual responsivity, and unfaithfulness in romantic couples. *Psychological Science, 17,* 830–835.

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

Guermandi, E., Vegetti, W., Bianchi, M. M., Uglietti, A., Ragni, G., & Crosignani, P. (2001). Reliability of ovulation tests in infertile women. *Obstetrics & Gynecology, 97,* 92–96. doi:10.1016/S0029-7844(00)01083-8

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., 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

Haselton, M. G., & Gangestad, S. W. (2006). Conditional expression of women's desires and men's mate guarding across the ovulatory cycle. *Hormones and Behavior, 49,* 509–518.Haselton, M. G., Grebe, N., Gildersleeve, K., & Gangestad, S. W. (2014). *Aggressive p-hacking: Does it produce the illusion of evidential value?* Manuscript in preparation.

Havlícek, J., Roberts, S. C., & Flegr, J. (2005). Women's preference for dominant male odour: Effects of menstrual cycle and relationship status. *Biology Letters, 1,* 256–259. doi:10.1098/rsbl.2005.0332

Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature, 21,* 406–427. doi:10.1007/s12110-010-9101-5

Hromatko, I., Tadinac, M., & Prizmic, H. (2006). Women's hormonal status and mate value influence relationship satisfaction and perceived male attractiveness. *Psychological Topics, 15,* 315–330.

Hung, H. M. J., O'Neill, R. T., Bauer, P., & Köhne, K. (1997). The behavior of the *P*-value when the alternative hypothesis is true. *Biometrics, 53,* 11–22. doi:10.2307/2533093

Jöchle, W. (1973). Coitus induced ovulation. *Contraception, 7,* 523–564. doi:10.1016/0010-7824(73)90023-1

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

Jones, B. C., Little, A. C., Boothroyd, L., DeBruine, L. M., Feinberg, D. R., Law Smith, M. J.,. . . Perrett, D. I. (2005). Commitment to relationships and preferences for femininity and apparent health in faces are strongest on days of the ovulatory cycle when progesterone level is high. *Hormones and Behavior, 48,* 283–290. doi:10.1016/j.yhbeh.2005.03.010

Koehler, N., Rhodes, G., & Simmons, L. W. (2006). Do cyclic changes in women's face preferences target cues to long-term health? *Social Cognition, 24,* 641–656. doi:10.1521/soco.2006.24.5.641

Larson, C. M., Haselton, M. G., Gildersleeve, K. A., & Pillsworth, E. G. (2013). Changes in women's feelings about their romantic relationships across the ovulatory cycle. *Hormones and Behavior, 63,* 128–135.

Larson, C. M., Pillsworth, E. G., & Haselton, M. G. (2012). Ovulatory shifts in women's attractions to primary partners and other men: Further evidence of the importance of primary partner sexual attractiveness. *PLos ONE, 7,* e44456.

Leggett, N. C., Thomas, N. A., Loetscher, T., & Nicholls, M. E. (2013). The life of p: "Just significant" results are on the rise. *The Quarterly Journal of Experimental Psychology, 66,* 2303–2309. doi:10.1080/17470218.2013.863371

Li, N. P., Bailey, J. M., Kenrick, D. T., & Linsenmeier, J. A. W. (2002). The necessities and luxuries of mate preferences: Testing the tradeoffs. *Journal of Personality and Social Psychology, 82,* 947–955. doi:10.1037/0022-3514.82.6.947

Li, N. P., Pillsworth, E. G., & Haselton, M. G. (2006). [Cycle shifts in women's mate preferences assessed using a mate dollars paradigm]. Unpublished raw data.

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.

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

Little, A. C., Jones, B. C., & DeBruine, L. M. (2008). Preferences for variation in masculinity in real faces change across the menstrual cycle: Women prefer more masculine faces when they are more fertile. *Personality and Individual Differences, 45,* 478–482. doi:10.1016/j.paid.2008.05.024

Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of *p* values just below .05. *Quarterly Journal of Experimental Psychology, 65,* 2271–2279. doi:10.1080/17470218.2012.711335

Moore, F. R., Cornwell, R. E., Law Smith, M. J., Al Dujaili, E. A. S., Sharp, M., & Perrett, D. I. (2011). Evidence for the stress-linked immunocompetence handicap hypothesis in human male faces. *Proceedings of the Royal Society B: Biological Sciences, 278,* 774–780. doi:10.1098/rspb.2010.1678

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

Penton-Voak, I. S., & Perrett, D. I. (2000). Female preference for male faces change cyclically: Further evidence. *Evolution and Human Behavior, 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, K., & Minamisawa, R. (1999). Menstrual cycle alters face preference. *Nature, 399,* 741–742. doi:10.1038/21557

Peters, M., Rhodes, G., & Simmons, L. W. (2008). Does attractiveness in men provide clues to semen quality? *Journal of Evolutionary Biology, 21,* 572–579. doi:10.1111/j.1420-9101.2007.01477.x

Peters, M., Simmons, L. W., & 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

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. doi:10.1016/j.evolhumbehav.2005.10.002

Provost, M. P., Troje, N. F., & Quinsey, V. L. (2008). Short-term mating strategies and attraction to masculinity in point-light walkers. *Evolution and Human Behavior, 29,* 65–69. doi:10.1016/j.evolhumbehav.2007.07.007

Puts, D. A. (2004). *Female choice, male dominance, and the evolution of low voice pitch in men* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (AAT 3139711)

Puts, D. A. (2005). Mating context and menstrual phase affect women's preferences for male voice pitch. *Evolution and Human Behavior, 21,* 388–397.

Rantala, M. J., Eriksson, C. J. P., Vainikka, A., & Kortet, R. (2006). Male steroid hormones and female preference for male body odor. *Evolution and Human Behavior, 27,* 259–269. doi:10.1016/j.evolhumbehav.2005.11.002

Rice, W. R., & Gaines, S. D. (1994). 'Heads I win, tails you lose': Testing directional alternative hypotheses in ecological and evolutionary research. *Trends in Ecology & Evolution, 9,* 235–237. doi:10.1016/0169-5347(94)90258-5

Rikowski, A., & Grammer, K. (1999). Human body odour, symmetry, and attractiveness. *Proceedings of the Royal Society B: Biological Sciences, 266,* 869–874. doi:10.1098/rspb.1999.0717

Rohatgi, A. (2011). WebPlotDigitizer [Computer software]. Available from http://arohatgi.info/WebPlotDigitizer/

Roney, J. R., & Simmons, Z. L. (2008). Women's estradiol predicts preference for facial cues of men's testosterone. *Hormones and Behavior, 53,* 14–19. doi:10.1016/j.yhbeh.2007.09.008

Roney, J. R., Simmons, Z. L., & Gray, P. B. (2011). Changes in estradiol predict within-women shifts in attraction to facial cues of men's testosterone. *Psychoneuroendocrinology, 36,* 742–749. doi:10.1016/j.psyneuen.2010.10.010

Rupp, H. A., James, T. W., Ketterson, E. D., Sengelaub, D. R., Janssen, E., & Heiman, J. R. (2009). Neural activation in women in response to masculinized male faces: Mediation by hormones and psychosexual factors. *Evolution and Human Behavior, 30,* 1–10. doi:10.1016/j.evolhumbehav.2008.08.006

Rupp, H., Librach, G. R., Feipel, N. C., Ketterson, E. D., Sengelaub, D. R., & Heiman, J. R. (2009). Partner status influences women's interest in the opposite sex. *Human Nature, 20,* 93–104. doi:10.1007/s12110-009-9056-6

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. doi:10.1177/0956797611417632

Simonsohn, U. (2013). It really just does not follow, comments on Francis (2013). *Journal of Mathematical Psychology, 57,* 174–176. doi:10.1016/j.jmp.2013.03.006

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). *P*-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143,* 534–547. doi:10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). *P-curve and effect size: Correcting for publication bias using only significant results.* Available from http://papers.ssrn.com/sol13/papers.cfm?abstract_id=2377290

Soler, C., Núñez, M., Gutiérrez, R., Núñez, J., Medina, P., Sancho, M., & Núñez, A. (2003). Facial attractiveness in men provides clues to semen quality. *Evolution and Human Behavior, 24,* 199–207. doi:10.1016/S1090-5138(03)00013-8

Teatero, M. (2009). *Mating strategies across the menstrual cycle: Preferences, jealousy, and masculinity* (Unpublished master's thesis). Lakehead University, Thunder Bay, Ontario, Canada.

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

Vaughn, J. E., Bradley, K. I., Byrd-Craven, J., & Kennison, S. M. (2010). The effect of mortality salience on women's judgments of male faces. *Evolutionary Psychology, 8,* 477–491.

Welling, L. L. M., Jones, B. C., DeBruine, L. M., Conway, C. A., Law Smith, M. J., Little, A. C., & Al-Dujaili, E. A. S. (2007). Raised salivary testosterone in women is associated with increased attraction to masculine faces. *Hormones and Behavior, 52,* 156–161. doi:10.1016/j.yhbeh.2007.01.010

Wilcox, A. J., Dunson, D. B., Weinberg, C. R., Trussell, J., & Day Baird, 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

Wilcox, A. J., Weinberg, C. R., & Baird, D. D. (1995). Timing of sexual intercourse in relation to ovulation: Effects on the probability of conception, survival of pregnancy, and sex of the baby. *New England Journal of Medicine, 333,* 1517–1521. doi:10.1056/NEJM199512073332301

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

Wood, W., Lundgren, S., Ouellette, J. A., Busceme, S., & Blackstone, T. (1994). Minority influence: A meta-analytic review of social influence processes. *Psychological Bulletin, 115,* 323–345.

Received May 30, 2014
Revision received July 21, 2014
Accepted July 21, 2014 ■