



Are women more likely to wear red and pink at peak fertility? What about on cold days?

Conceptual, close, and extended replications with novel clothing colour measures

Liana S. E. Hone^{1,2}  and Michael E. McCullough^{2,3*}

¹Clinical and Research Institute on Addictions, University at Buffalo, New York, USA

²Department of Psychology, University of Miami, Coral Gables, Florida, USA

³Department of Psychology, University of California San Diego, California, USA

Evolutionarily minded researchers have hypothesized that women advertise their ovulatory status by wearing red or pink clothing on relatively cold days. Many of these studies have been based on samples of women who have self-reported their clothing choices, a practice that raises questions about accuracy. In two studies, we evaluated the relationship between women's fertility and their clothing choices using four methods for measuring clothing colour: self-reports; trained raters' judgements of garment coloration in outfits that women drew onto mannequins to represent what they would wear to a party with single attractive people in attendance; automated colour coding of the mannequins; and trained raters' judgements of garment coloration as evinced in photographs that women took of themselves. Using these four measures of clothing choice along with measures of women's fertility and outside temperature, we did not find compelling evidence that women are particularly inclined to wear red or pink during peak fertility, even on relatively cold days.

In humans and numerous other species, female fertility is cyclical, and conception hinges on mating around the time of ovulation. In some non-human animals, phenotypic markers and behavioural changes have evolved to indicate when females are ovulating (Dixson, 2012). Such markers include increased sexual receptivity, genital swellings, and oestrus vocalizations or copulation calls (Aujard, Heistermann, Thierry, & Hodges, 1998; Deschner, Heistermann, Hodges, & Boesch, 2004; Dixson, 2012; Gouzoules, Gust, Donaghey, & St. Andre, 1998; Knott, Thompson, Stumpf, & McIntyre, 2010; Maestripieri, 2004; Nunn, 1999; Welling & Puts, 2014; Wolfe, 1979). However, not all species with cyclic fertility exhibit such overt ovulatory cues, and experts do not currently agree on whether human ovulation has come to be concealed over evolutionary time (Havlíček, Cobey, Barrett, Klapilová, & Roberts, 2015; Marlowe, 2004; Welling & Puts, 2014; Wood & Carden, 2014).

Evidence from cross-species comparisons suggests that humans lack overt ovulatory cues, and until the past few decades, it was thought that humans lacked ovulatory cues completely (Alexander & Noonan, 1979; Benshoof & Thornhill, 1979; Burley, 1979; Butler,

*Correspondence should be addressed to Michael E. McCullough, Department of Psychology, UC San Diego, McGill Hall 4328, 9500 Gilman Drive, La Jolla, CA 92093-0109, USA (email: memccullough@ucsd.edu).

1974; Dixson, 2012; Heistermann *et al.*, 2001; Strassmann, 1981; Symons, 1979). However, recent studies suggest that humans might exhibit subtle cues to ovulation (Gangestad & Thornhill, 2008; Haselton & Gildersleeve, 2011), though the evidence for overt cues to ovulation remains rather weak (Arslan, Schilling, Gerlach, & Penke, 2018; Welling & Puts, 2014). Some scholars have speculated that such subtlety in signalling might have evolved to deter unwanted male attention during peak fertility or to confuse paternity certainty after conception (Deschner *et al.*, 2004; Nunn, 1999; Wallen & Zehr, 2004; Wolfe, 1979).

Despite the lack of strong evidence for overt ovulation cues in humans, other evolutionarily minded researchers have hypothesized that women might advertise their ovulatory status by wearing red- and pink-coloured clothing during periods of high conception risk (e.g., Beall & Tracy, 2013; Eisenbruch, Simmons, & Roney, 2015; Tracy & Beall, 2014). This hypothesis rests on two premises. The first is that women are motivated to enhance their attractiveness during ovulation (Haselton, Mortezaie, Pillsworth, Bleske-Rechek, & Frederick, 2007), and the second is that men perceive red as sexually attractive (Elliot & Niesta, 2008; Elliot, Tracy, Pazda, & Beall, 2013; c.f., Peperkoorn, Roberts, & Pollet, 2016). Though recent studies have failed to find a relationship between red and perceived sexual attractiveness (Peperkoorn *et al.*, 2016; Pollet, Costello, Groeneboom, Peperkoorn, & Wu, 2019), some researchers have suggested that men perceive red as sexually attractive because (1) red is associated with love, passion, sex, and romance; (2) red has the potential to attract attention perceptually; (3) red skin tone perhaps indicates sexual arousal; and finally (4) males of other species are attracted to red (Aslam, 2006; Bielert, Girolami, & Jowell, 1989; Elliot & Pazda, 2012; Elliot *et al.*, 2013; Greenfield, 2005). Moreover, researchers have hypothesized that weather should moderate the effect of wearing red or pink at peak fertility (Tracy & Beall, 2014): If wearing red is driven by women's motivation to enhance their attractiveness during ovulation, the effect should emerge when alternative avenues for enhancing attractiveness (i.e., exposing more skin) are not available (i.e., the effect should emerge on relatively cold days; Tracy & Beall, 2014).

To test the hypothesis that women wear red and pink at peak fertility, Beall and Tracy (2013) collected self-report data regarding women's ovulatory cycles and current shirt colours from 100 women in an online sample and 24 women in a college student sample. They reported that women at high conception risk were 3.5 times more likely to be wearing red or pink than women at low conception risk. A successive failure to replicate this finding led them to explore the possibility that ovulating women are more likely than non-ovulating women to wear red on cold days, though not on warm days (Tracy & Beall, 2014).

Replications of Beall and Tracy's (2013) study by independent teams of investigators have produced mixed results. Eisenbruch *et al.* (2015), for instance, sought to replicate Beall and Tracy's (2013) results using a within-subjects design with 43 women and frequent hormone sampling, as well as calendar-counting methods, to assess ovulatory timing. They found that women aged 18–22 were more likely to wear red during the fertile window (assessed via hormone sampling, but not via calendar-counting) than on other cycle days, with within-woman fluctuations in the ratio of oestradiol to progesterone mediating shifts in wearing red. Other researchers, however, have found a less reliable association (e.g., Blake, Dixson, O'Dean, & Denson, 2017) or a lack of an association between conception risk and the probability of wearing red. For example, in a pre-registered, within-subjects study involving 157 women and salivary hormone tests as well as luteinizing hormone tests, women were not more likely to wear red or pink at peak fertility (Stern, Rudolph, & Penke, 2019). Similarly, in a within-subjects study involving 26,000 diary entries from more than 1,000 women, an effect of peak fertility on 'sexy' clothing choices (though not on wearing red specifically) was not found (Arslan *et al.*, 2018).

To date, most studies on women's propensity to wear red during ovulation have been based on women's own judgements of the colours of their clothing, a practice that departs from the standard practices of biologists who study coloration in non-humans (Hill & McGraw, 2006). An exclusive reliance on self-reports is problematic for at least two reasons. First, people evince stable individual differences in self-reports of coloration that are due to both individual differences in colour perception and to individual differences in the application of colour names. These individual differences introduce unsystematic error (Emery, Volbrecht, Peterzell, & Webster, 2017). Second, people's self-reports of physical attributes related to attractiveness often depart from objective measures (Connor Gorber, Tremblay, Moher, & Gorber, 2007). Because clothing colour is an important factor in others' evaluations of attractiveness (Radeloff, 1990), it is conceivable that women's self-reported clothing colour choices sometimes depart from their actual clothing colour for systematic reasons as well.

If women really are motivated to choose specific colours of clothing when they are at high risk of conception, the link between ovulation and wearing red or pink garments should manifest itself on all valid measurements of women's colour choices (Borsboom, Mellenbergh, & van Heerden, 2004). Thus, in the two studies reported here, we explored four different methods for measuring garment colour: (1) trained raters' judgements of garment coloration in outfits that women drew onto mannequins to represent what they would wear to a large, night-time party with lots of single, attractive people in attendance (Study 1); (2) automated colour coding of the mannequins (Study 1); (3) women's self-reports (Study 2, Analysis I); and (4) trained raters' judgements of garment coloration as evinced in photographs that women took of themselves (Study 2, Analysis II).

The present studies

First, in Study 1, we sought to conceptually replicate the finding that women at peak fertility are more likely to wear red or pink on relatively cold days. We did so using data collected in a laboratory between September 2012 and May 2013. These data were collected as part of an unrelated study prior to the publication of Beall and Tracy (2013) in August 2013 or Tracy and Beall (2014) in February 2014. We used trained raters' judgements of garment coloration in outfits that women drew onto mannequins to represent the clothing they would wear to a large, night-time party with lots of attractive people in attendance, and automated colour coding of the mannequins.

Having found inconclusive evidence in Study 1 for an effect of conception risk on the likelihood of wearing red or pink on relatively cold days, in Study 2 (Analysis I), we used data from a large sample collected online between February 2015 and May 2015 to more closely replicate Beall and Tracy (2013) and Tracy and Beall (2014). Additionally, in Study 2 (Analysis II), we attempted to extend Beall and Tracy (2013) and Tracy and Beall (2014). For the close replication (i.e., our attempt to directly replicate the original 2013 online study to the best of our ability), we used Beall and Tracy's (2013) published methods, including the use of women's self-reports of their shirt colour (Study 2, Analysis I). For the extension, we used updated measures of conception risk and a novel measure of shirt colour: Trained raters' judgements of women's garment coloration in photographs that participants took of themselves (Study 2, Analysis II). In line with Tracy and Beall (2014), we tested whether outside temperature moderated the effect of conception risk on the likelihood of wearing red or pink in both Study 2, Analysis I (close replication) and Study 2, Analysis II (extension).

Power considerations

The post-hoc computed power of Beall and Tracy's (2013) two studies (online sample $N = 100$; college sample $N = 24$) is 0.64 and 0.50 with effect sizes of $r = .23$ and $r = .40$, respectively, yielding a weighted effect size of $r = .32$. We therefore estimated statistical power using $r = .32$ as an estimate of the expected effect size. To achieve statistical power of .80, .90, .99, and .999 (assuming $\alpha = .05$), replications with an expected medium effect size of $r = .32$ require sample sizes of $N = 79$, $N = 105$, $N = 183$, or $N = 254$, respectively (Erdfelder, Faul, & Buchner, 1996).

STUDY I OVERVIEW

Using existing data, we first sought to test the hypothesis that women at peak fertility are more likely to wear red or pink (Beall & Tracy, 2013) on relatively cold days (Tracy & Beall, 2014) using trained raters' judgements of garment coloration in outfits that women drew onto mannequins to represent what they would wear to a large, night-time party with lots of attractive people in attendance. We also used automated colour coding of the mannequins. These data were collected before the publication of Beall and Tracy (2013), and Tracy and Beall (2014), the latter of which reported that temperature moderated the effect of conception risk on wearing red or pink.

STUDY I METHOD***Participants***

Participants were 440 women who participated in a laboratory experiment on a large university campus between September 2012 and May 2013. Like Beall and Tracy (2013), we excluded women who were older than 40, or who used hormonal birth control, smoked cigarettes (Windham, Elkin, Swan, Waller, & Fenster, 1999), or were pregnant (e.g., were not experiencing regular menstrual cycles or had recently given birth). Unlike Beall and Tracy (2013), we did not request that women not participate if they within five days of the onset of their period (to avoid effects attributable to premenstrual symptoms) because we collected data prior to the publication of this exclusion criterion (Beall & Tracy, 2013). This yielded useable data from 161 women aged 17-32 ($M = 18.96$, $SD = 1.65$). Among the 161 women selected for Study 1, conception risk data were missing for three women who chose not to report the first day of their last period, or reported a date that fell after the date of their study participation (i.e., they estimated the first day of their next period instead of recalling the first day of their last period). In line with Wilcox, Dunson, and Baird (2000) methods for assigning conception risk, we excluded women whose cycle length was shorter than 27 days ($n = 35$), longer than 30 days ($n = 21$), or not calculable ($n = 2$). Application of our exclusion criteria resulted in a subsample of 100 women for Study 1. Our power to detect an association of $r = .32$ was .89.

Measures***Conception risk***

Participants were asked to indicate their age (open-ended response) and asked to answer several questions about their ovulatory cycle, including the following questions with response options of 'Yes' or 'No': 'Are you currently using hormonal

contraceptives (birth control pill, skin patch, injection, implant, vaginal ring, etc.)?’ ‘Do you regularly smoke cigarettes?’ ‘Have you skipped a menstrual period in the past three months?’ and ‘Have you recently given birth?’ They were asked, ‘How long is your menstrual cycle? A cycle starts on the first day of your menstrual period and ends on the first day of your next menstrual period. Please feel free to check your calendar, or use the calendar provided below, or use a calculator for accuracy’ with the response options ranging from less than 21 days to greater than 35 days. They were also asked to ‘Please also circle the first day of your last menstrual period on the calendar below’ followed by a picture of a six-month calendar. Finally, they were asked ‘When was the first day of your last menstrual period?’ followed by a fill in the blank: ‘Day_____, Month_____, Year_____’.

Based on these questions, we estimated conception risk using two different methods: First, following Beall and Tracy (2013), we divided women into high- and low-conception-risk groups based on the first day of women’s last periods. Because Beall and Tracy (2013) did not adjust for cycle length when assigning conception risk group, which may have been a source of error in conception risk assignment, we grouped women with 27-, 28-, 29-, and 30-day cycle lengths into high- or low-conception-risk groups based on methods in Wilcox et al. (2000). Second, we assigned each woman a probability of pregnancy following a single act of unprotected intercourse on day they participated in the study based on the conception risk associated with each day of a woman’s cycle (Wilcox, Dunson, Weinberg, Trussell, & Baird, 2001). This approach enabled us to consider a continuous measure of conception risk as well.

Shirt colour

We instructed participants to, ‘Imagine that you are attending a social gathering at a friend’s apartment tonight around 10:30 PM. From what your friend tells you, it will be a large party where there will be lots of single attractive people. Using the colored pencils provided, indicate on the outline of the human figure (on the next page) what you will be wearing to this party by drawing an outfit. Be sure to show the outlines of each item of clothing—your shirt, pants, shorts, skirt, etc.—clearly’ (adapted from Durante, Li, & Haselton, 2008).

Using participants’ drawings on the mannequins, we assigned shirt colour using two different methods. For the first method, two independent raters, blind to participants’ conception risk, coded the shirts each participant drew onto the mannequin as ‘red or pink’ or ‘not red or pink’. For the second method, we digitally scanned participants’ drawings and opened a *.jpg* image in Adobe Photoshop CS5. We used the Quick Selection tool to select the participant’s shirt, dress, or upper garments above the waist, copied the selection, pasted it in a new document, and auto-cropped it with Adobe’s trim tool. Next, we used an Average Color Seeker (IDimager, 2007) to determine the average colour of the trimmed image. The Average Color Seeker considers all of the pixels in the image and creates a colour patch that is the average of all the pixels, to which it assigns a Hexadecimal Color Code. We used Adobe’s Color Picker to determine the HSV/HSB values of the Hexadecimal Color Coded patch created by the Average Color Seeker. If the Color Picker determined that the hue value of the Hexadecimal Color Coded patch fell between 290° and 10°, we coded the average colour as red. However, if hue values between 290° and 10° had saturation values under 15% (or saturation values under 20% and brightness values under 60%) we coded the average colour as white, grey, or black.

Weather

Tracy and Beall (2014) collected data on an unusually cold day and on an unusually warm day to assess whether the effect of conception risk on wearing red or pink was moderated by outdoor temperature. To consider the effect of temperature in our Study 1, we recorded the temperature on the day that each participant completed the laboratory study using the Farmer's Almanac. Because we collected data prior to the publication of Tracy and Beall (2014), we did not obtain measures of perceived temperature or warmth. The median temperature was 24.25°C (range = 15.61°C–28.44°C). We computed a median split based on temperature, and the difference between 'cold' ($M = 21.97$, $SD = 2.14$) and 'warm' ($M = 25.81$, $SD = 1.05$) days was significant, $t(71.44) = 11.40$, $p < .001$.

Procedure

After providing informed consent to participate in an Institutional Review Board-approved laboratory study, participants stood in the laboratory for a photograph. Standardized full-length photographs of participants were later coded by trained independent raters blind to conception risk. Raters recorded whether participants' shirts were red or pink, or not red or pink. However, only eight women were coded as wearing red or pink. Statistical tests for associations between shirt colour and ovulatory status would have been grossly underpowered so we did not consider photograph-based assessments of women's shirt coloration further.

After standing for a photograph, participants were asked to answer demographic questions and questions about their ovulatory cycle. Based on these questions, we assigned conception risk using two different methods: For the first method, we grouped women into high- or low-conception-risk groups adjusting for cycle length (Wilcox et al., 2000). For the second method, we assigned each woman a probability of pregnancy following a single act of unprotected intercourse on day she participated in the study (Wilcox et al., 2001).

After completing demographic and ovulatory cycle questions, participants were asked to draw on a mannequin what they would wear to a hypothetical party later that evening (Durante et al., 2008). We assigned shirt colour from these mannequins using two different methods: (1) trained raters' judgements; and (2) automated colour coding. We obtained measurements of the outside temperature on the day that each participant completed the laboratory study using the Farmer's Almanac (<https://www.farmersalmanac.com/weather-history>).

STUDY 1 RESULTS

Using the modified version of Beall and Tracy's (2013) methods for assigning conception risk (Wilcox et al., 2000), 69 women were assigned to the low-conception-risk group and 31 women were assigned to the high-conception-risk group. Using the continuous measure of conception risk (Wilcox et al., 2001), the probability of pregnancy following a single act of unprotected intercourse on the day a woman participated in the study was $M = .03$ ($SD = .03$).

Inter-rater reliability between the two independent raters (blind to participants' conception risk) who coded the shirts each participant drew onto a mannequin as 'red or pink' or 'not red or pink' was adequate, $\kappa = 0.88$, $p < .001$. Raters were in agreement on 95 of the 100 drawings. The five discrepancies were resolved through

discussion with a third rater. Of the 100 drawings, our raters categorized 30 of the mannequins as wearing red or pink and 70 as wearing another colour. With regard to automated colour coding of the mannequins, five participants shaded their mannequins too lightly for the Average Color Seeker software to assign an average colour, but we were able to categorize the shirt as red or pink by visually determining which coloured pencils the participants used. On this basis, we determined that four of the five participants used the pink-coloured pencil. A determination could not be made for one of the five participants. Of the 99 digitally scanned drawings, 28 were coded as wearing red or pink and 71 as wearing another colour. The correspondence between shirt colour coding using these approaches was significant, $\chi^2(1) = 67.84$, $p < .001$, $\Phi = 0.83$, with seven (7.07%) discordant categorizations (see Table 1).

Women in the high-conception-risk group were not more likely to use red or pink to colour their mannequin than were women in the low-conception-risk group (see Figure 1). This was the case both when we used raters to judge participants' drawings, $\chi^2(1) = 0.38$, $p = .54$, and when we used automated colour coding, $\chi^2(1) = 0.01$, $p = .91$. In a multivariate logistic regression in which we predicted shirt color with conception risk (high vs. low), temperature (continuous), and their interaction, none of the main effects were statistically significant ($ps > .16$). However, the interactions of conception risk and temperature had p values of .03 (rater-based assessment of shirt color) and .09 (automated assessment of shirt color). Even so, the simple effects of conception risk on the probability of wearing red or pink were non-significant for both the low-temperature ($ps = .43$ and $.23$, respectively) and high-temperature ($ps = .11$ and $.32$, respectively) groups. Using a continuous measure of conception risk, mean conception risk did not differ between women who used red or pink to colour their mannequin ($Ms = 0.03$, $SDs = 0.03$) and women who did not use red or pink to colour their mannequin ($Ms = 0.03$, $SDs = 0.03$). This was the case both when we used trained raters to judge participants' drawings, $t(98) = -0.07$, $p = .95$, and when we used automated colour coding, $t(97) = -0.27$, $p = .79$.

In a multivariate logistic regression, the effects of conception risk (continuous; $B = .78$, $Wald [1] = .01$, $p = .92$), temperature (continuous; $B = .20$, $Wald [1] = 0.04$, $p = .84$), and their interaction ($B = -4.64$, $Wald [1] = 2.65$, $p = .10$) on shirt colour were non-significant when we used automated colour coding as our method for assessing shirt colour. However, when we used raters' judgements of participants' drawings as our method for assessing shirt colour, the interaction of the continuous measure of

Table 1. Cross-tabulation of trained raters' judgements of women's shirt coloration (red/pink vs. not red/pink) as evinced in outfits that women drew onto a mannequin to represent the clothing they would wear to a party with attractive men in attendance and automated shirt colour coding based the mannequins

		Automated coding of shirt colour		
		Not red/pink	Red/pink	Total
Rater's Judgements of Shirt Colour	Not Red/Pink	67	3	70
	Red/Pink	4	25	29
	Total	71	28	99

Note. Table reports counts.

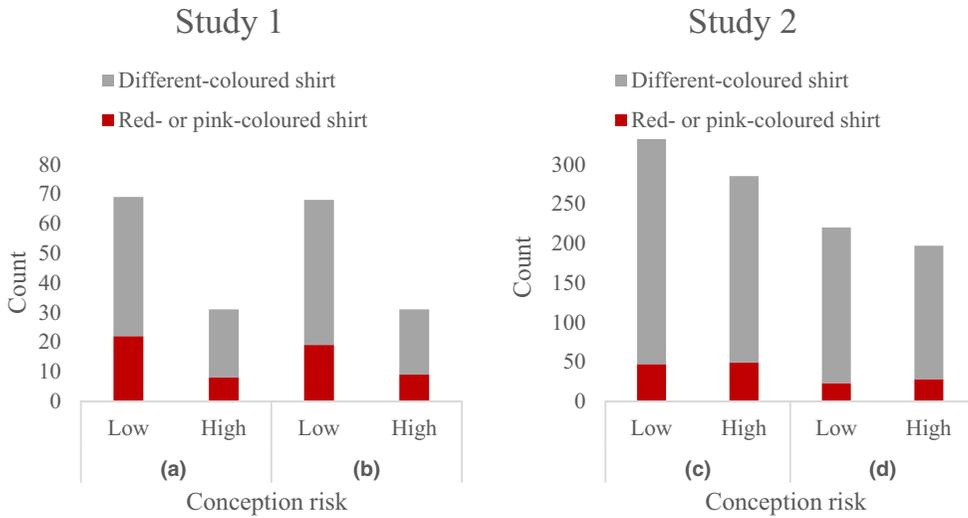


Figure 1. Women's shirt colour preferences as a function of conception risk using four separate approaches to measuring shirt colour preference. Note. (a) Trained raters' colour assessments based on outfits that women drew onto a mannequin to represent the clothing they would wear to a party with attractive men in attendance (Study 1); (b) automated colour coding of outfits that women drew onto a mannequin to represent the clothing they would wear to a party with attractive men in attendance (Study 1). (c) Self-reported shirt colour (Study 2, Analysis I); (d) trained raters' judgements of women's shirt colour as evinced in photographs that participants took of themselves (Study 2, Analysis II).

conception risk and temperature ($B = -7.15$, $Wald [1] = 5.42$, $p = .02$) on shirt colour was significant. In this analysis, the effect of conception risk (continuous; $B = -1.12$, $Wald [1] = 0.02$, $p = .89$) was not significant and the effect of temperature (continuous; $B = .05$, $Wald [1] = 0.25$, $p = .62$) was not significant.

To follow up on the significant interaction of the continuous measures of conception risk and temperature on raters' judgements of shirt colour, we tested the simple effects. As was the case when we classified subjects' conception risk with a binary measure, no significant simple effects at either high temperatures ($B = -24.46$, $Wald [1] = 3.26$, $p = .07$) or low temperatures ($B = 18.74$, $Wald [1] = 3.28$, $p = .07$) were found. This suggests that although the coefficient of the regression of shirt colour on conception risk at low temperatures was different from the coefficient of the regression of shirt colour on conception risk at high temperatures, neither effect was significantly different from zero.

STUDY I DISCUSSION

Study 1, which was a conceptual replication of Beall and Tracy's (2013) work, did not yield evidence that women are significantly more likely to wear red or pink when they are likely fertile than when they are likely not fertile. We note that both our 'cold-' and 'warm-weather' subgroups evinced a mean temperature greater than mean temperature (13.79°C) of the 'warm-weather' subgroup (and far above the 'cold-weather' subgroup; 3.76°C) in Tracy and Beall's (2014) study. Despite this limitation, we assessed whether temperature moderated the effect of conception risk on the

likelihood of wearing red or pink on the premise that an effect on relatively cold days should emerge regardless of regional temperature range. This did not seem to be the case, although the associations of conception risk with the probability of wearing red or pink were in the theoretically expected directions both for women in the cold-weather and warm-weather groups. It is not unreasonable to surmise that with greater temperature variation (so that women from cold-weather conditions faced colder temperatures), the significant positive association between conception risk and the probability of wearing red or pink could have emerged.

STUDY 2 OVERVIEW

To address limitations of Study 1 (including small sample size and restricted temperature range), we conducted a study on this question using a large online sample that enabled us to more closely replicate and extend the methods of Beall and Tracy (2013) and Tracy and Beall (2014). Study 2 comprised two sets of analyses: a close replication (Analysis I) and an extension (Analysis II). In Analysis I, we attempted to exactly replicate the methods of Beall and Tracy (2013) using women's self-reports of their shirt colour and a binary measure of temperature. In Analysis II, we sought to extend Beall and Tracy's (2013) findings by using the same data used in Analysis I and making several alterations to data exclusion and measurement of key variables.

STUDY 2 METHOD

Participants

To obtain an adequate sample size and range of temperatures, we surveyed 1,344 women in the United States online via Amazon's Mechanical Turk between 20 February 2015 and 13 May 2015. Like Beall and Tracy (2013), we excluded women who were older than 40, or who used hormonal birth control, smoked cigarettes (Windham *et al.*, 1999), were pregnant, were not experiencing regular menstrual cycles, or were within five days of the onset of their period (to avoid proposed effects attributable to premenstrual symptoms; Beall & Tracy, 2013). This yielded usable data from 780 women aged 18-40 ($M = 26.92$, $SD = 5.62$) for Study 2.

Analysis I

To replicate Beall and Tracy's (2013) methods as closely as possible, we further excluded women (1) who were not 100% confident that their estimate of their last period of menses was accurate within ± 5 days ($n = 21$); (2) whose first day of menses during their last period had occurred more than 28 days previously ($n = 16$); or (3) who could not be placed in the high- or low-conception-risk group with 100% confidence ($n = 126$). See Beall and Tracy (2013; pp. 3) for an example of how exclusions of women who could not be placed in the high- or low-conception-risk group with 100% confidence are made. Applying these additional exclusion criteria resulted in a subsample of 617 women for our close replication. Our power to detect an effect of $r = .32$ with 617 participants (assuming $\alpha = 0.05$, two-tailed) exceeded 0.999.

Analysis II

Among the 780 women eligible for Study 2, which we conducted to conceptually replicate and extend Beall and Tracy (2013), conception risk was coded as missing for 55 women because they chose not to report the first day of their last period, they reported a date that fell after the survey timestamp, or the number of days since the first day of their last period exceeded their cycle length. In line with our Study 1 methods for assigning conception risk, we excluded 126 additional women whose cycles were shorter than 27 days and 102 women whose cycles exceeded 30 days (i.e., whose cycles are too short or too long to be accurately classified using this method). After these exclusions, our subsample comprised 497 women for our Analysis II. Of these, 464 women provided useable photographs; several women uploaded photographs that we could not categorize (typically due to bad lighting). Our power to detect an effect of $r = .32$ with 464 participants exceeded 0.999 (Erdfelder *et al.*, 1996).

Analysis I measures

Conception risk

As in Beall and Tracy (2013), participants were asked, 'How many days has it been since the onset of your last period of menses?' followed by 'Within how many days are you 100% confident in your above estimate?' and response options ranging from '0 (I'm 100% confident)' to 'More than 5 (I'm not very confident)'. Using participants' responses to questions taken from Beall and Tracy (2013), we divided women into high (days 6–14) and low (days 1–5; days 15–28) conception-risk groups following Beall and Tracy's (2013) procedure.

Shirt colour

As in Beall and Tracy (2013), participants were also asked, 'What color is the shirt you are currently wearing? (If your shirt is multicolored, please select the color which is most prevalent)', with response options of 'black', 'blue', 'gray', 'green', 'pink', 'red', 'white', 'yellow', and 'other'. We categorized shirt colour via self-report response to the item used by Beall and Tracy (2013) following their procedure.

Weather

Tracy and Beall (2014) collected data on an unusually cold day and on an unusually warm day to assess whether the effect of conception risk on wearing red or pink was moderated by temperature. To replicate their approach as closely as possible in our Analysis I, we computed a median split based on temperature (median = 7.78°C) via responses to the statement, 'Please verify the actual temperature outside today by using your preferred search engine, weather website, or thermometer'. Perceived temperature was measured with the item, 'What do you think the temperature is today?' with a drop-down menu to select the perceived temperature. Additionally, perceived warmth was measured with the item, 'How cold or warm do you think it is today?' on a scale ranging from 1 extremely cold to 9 extremely warm.

Analysis II measures

Conception risk

Beall and Tracy (2013) assigned conception risk using participants' responses to the item, 'How many days has it been since the onset of your last period of menses?' The cognitive burden of this method, which required women to count backward from the day of the study to their last period, might have introduced high amounts of error to our estimates of participants' conception risk. To address this issue, we instead asked participants, 'For the following questions, please feel free to check your calendar or use a calculator for accuracy. When was the first day of your last menstrual period?' and provided a drop-down menu to select a day, month, and year. We then counted from the dates reported for women's last periods to survey timestamps of the dates they participated in the study. As in our Study 1, we grouped women with 27-, 28-, 29-, and 30-day cycle lengths into high- or low-conception-risk groups based on methods in Wilcox *et al.* (2000) to reduce error in conception risk assignment. We also created a continuous measure of conception risk, as in Study 1, by assigning each woman a probability of pregnancy following a single act of unprotected intercourse on day they participated in the study (Wilcox *et al.*, 2001).

Shirt colour

Whereas Beall and Tracy (2013) used women's own judgements of the colours of their clothing, we asked participants to 'Please upload a picture of the shirt you are currently wearing. The photograph should not include any identifying information (e.g., no faces, tattoos, birthmarks, etc.)'. These photographs enabled us to use two independent raters, blind to participants' conception risk, to code each shirt in each uploaded photograph as 'red or pink' or 'not red or pink'. If a dress was worn, the colour of the dress was coded.

Weather

In our Analysis II, we used a continuous measure of temperature via responses to the statement, 'Please verify the actual temperature outside today by using your preferred search engine, weather website, or thermometer'.

Procedure

After providing informed consent to participate in an Institutional Review Board-approved online study, participants completed a questionnaire comprising questions from Beall and Tracy (2013), enabling us to conduct Analysis I, followed by additional questions about their ovulatory cycle, enabling us to conduct Analysis II. Participants were also asked to upload a picture of the shirt they were currently wearing (Analysis II). Because participants might have chosen to upload a photograph they conveniently had available rather than taking a new picture, which would take more time, participants were prevented from hitting the 'back' button. This prevented participants from changing their self-reported shirt colour answer after uploading a photograph and matching their self-reported shirt colour to their uploaded photograph retrospectively. We collected data on a range of warm and cold days and asked participants to rate the perceived temperature so that we could evaluate temperature as a moderator of the relationship between ovulatory status and shirt colour (Tracy & Beall, 2014).

STUDY 2 RESULTS

Analysis I

Using Beall and Tracy's (2013) methods for assigning conception risk, 332 women were assigned to the low-conception-risk group and 285 women were assigned to the high-conception-risk group. Using Beall and Tracy's (2013) methods for assigning shirt colour (self-report), 96 women reported wearing red or pink and 521 reported wearing some other colour. With regard to temperature, a median split yielded a cold-weather subgroup with a mean temperature of -1.00°C ($SD = 5.47^{\circ}\text{C}$) and a warm-weather subgroup with a mean temperature of 20.32°C ($SD = 6.80^{\circ}\text{C}$). Cold- and warm-weather subgroups differed in both perceived temperature, $t(775) = 40.81$, $p < .001$ ($M = -1.29^{\circ}\text{C}$, $SD = 7.82^{\circ}\text{C}$; $M = 21.05^{\circ}\text{C}$, $SD = 7.45^{\circ}\text{C}$), and perceived warmth, $t(697) = 28.39$, $p < .001$, ($M = 2.94$, $SD = 1.09$; $M = 5.69$, $SD = 1.57$), indicating that in line with the original study (Tracy & Beall, 2014), women in the cold-weather subgroup perceived the temperature as colder than did women in the warm-weather subgroup.

Using Beall and Tracy's (2013) methods for assessing shirt colour (self-report), women in the high-conception-risk group were not significantly more likely to be wearing red or pink than were women in the low-conception-risk group, $\chi^2(1) = 1.08$, $p = .30$ (see Figure 1). In a multivariate logistic regression, there was no effect of conception risk (high vs. low; $B = .24$, $Wald [1] = 0.56$, $p = .45$), temperature (cold vs. warm; $B = .02$, $Wald [1] = 0.01$, $p = .94$), or its interaction with conception risk ($B = -.0001$, $Wald [1] = 1.0029\text{E-}7$, $p = 1.00$), on shirt colour.

Analysis II

Using Wilcox *et al.*'s (2000) measure of conception risk, 264 women were assigned low conception risk and 233 women were assigned high conception risk. Although the correspondence between conception risk categorization using Wilcox *et al.*'s (2000) approach and Beall and Tracy's (2013) approach was statistically significant, $\chi^2(1) = 103.61$, $p < .001$ (Table 2), the two approaches assigned 135 women (27.16%) into contradictory conception risk groups. The point-biserial correlation between the continuous measure of probability of pregnancy (Wilcox *et al.*, 2001) and Beall and Tracy's (2013) approach was $r = .56$, $p < .001$. The point-biserial correlation between the continuous measure of probability of pregnancy and Wilcox *et al.*'s (2000) approach was $r = 0.78$, $p < .001$.

With regard to shirt colour, inter-rater reliability of the two independent raters (blind to participants' conception risk) who coded the shirts in each uploaded photograph as

Table 2. Cross-tabulation of women's conception risk (high vs. low) calculated using methods reported in Beall and Tracy (2013) versus methods reported in Wilcox *et al.* (2000)

		Wilcox <i>et al.</i> 's (2000) method		Total
		Low	High	
Beall and Tracy's (2013) method	Low	191	62	253
	High	73	171	244
Total		264	233	497

Note. Table reports counts.

Table 3. Cross-tabulation of women's self-reported shirt colour (red/pink vs. not red/pink) and trained raters' judgements of women's shirt colour (red/pink vs. not red/pink) as evinced in photographs that participants took of themselves

		Self-reported shirt colour		
		Not red/pink	Red/pink	Total
Raters' judgements of shirt colour	Not red/pink	394	12	406
	Red/pink	7	51	58
Total		401	63	464

Note. Table reports counts.

'red or pink' or 'not red or pink' was adequate, $\kappa = 0.87, p < .001$. Raters agreed on 451 of the 464 photographs. The thirteen discrepancies were resolved through discussion with a third rater. Using this measure, 58 women were coded as wearing red or pink and 406 were coded as wearing another colour. The correspondence between our approach to categorization of shirt colour and Beall and Tracy's (2013) approach was significant, $\chi^2(1) = 312.30, p < .001, \Phi = 0.82$, with 19 (4.10%) discordant categorizations (see Table 3).

Using trained raters' judgements of uploaded photographs for assessing shirt colour and Wilcox *et al.*'s (2000) method for assigning conception risk group, women in the high-conception-risk group were not more likely to be wearing red or pink than were women in the low-conception-risk group, $\chi^2(1) = 1.37, p = .24$ (see Figure 1). In a multivariate logistic regression, there was no effect of conception risk (high vs. low; $B = -.09, Wald [1] = 0.04, p = .84$), temperature (cold vs. warm; $B = -.18, Wald [1] = 0.16, p = .69$), or its interaction with conception risk ($B = .85, Wald [1] = 1.98, p = .16$) on shirt colour.

Conversely, using trained raters' judgements of uploaded photographs for assessing shirt colour and the continuous measure of conception risk, mean conception risk was, consistent with Beall and Tracy (2013), significantly different among women who were coded as wearing red or pink ($M = 0.05, SD = 0.03$) than among women who were coded as wearing another colour ($M = 0.04, SD = 0.03$), $t(420) = -1.97, p = .049$. In a multivariate logistic regression, the effects of conception risk (continuous; $B = 9.09, Wald [1] = 2.85, p = .09$) and temperature (continuous; $B = -0.75, Wald [1] = 3.36, p = .51$) on shirt colour were both non-significant. The interaction of the continuous measures of conception risk and temperature on shirt colour, however, was nearly statistically significant ($B = .75, Wald [1] = 3.36, p = .07$).

To follow up the nearly significant interaction of the continuous measures of conception risk and temperature on raters' judgements of shirt colour, we tested the simple effects. Contrary to Tracy and Beall's (2014) finding that the effect of conception risk on wearing red was significant on cold days, but not on warm days, the trending interaction obtained here suggested the opposite: The significant bivariate effect of conception risk on the probability of wearing red was evidently due to a significant simple effect of conception risk on likelihood of wearing red or pink at high temperatures ($B = 19.40, Wald [1] = 6.77, p = .01$), but not at low temperatures ($B = .78, Wald [1] = 0.01, p = .91$). Even though these simple effects ran counter to Tracy & Beall's (2014) findings, we re-iterate that the interaction was not statistically significant by conventional standards.

STUDY 2 DISCUSSION

Study 2, which involved a close replication (Analysis I) and extension (Analysis II) of Beall and Tracy (2013) and Tracy and Beall (2014), did not indicate that women are more likely to wear red or pink at peak fertility on cold days. That is, in Analysis I we did not find an effect when using Beall and Tracy's (2013) methods for assessing shirt colour (self-report), conception risk (binary), and temperature (binary). It is worth noting that when we used trained raters' judgments of uploaded photographs as the method for assessing shirt colour in Analysis II, bivariate analyses revealed that mean conception risk was *indeed* significantly higher among women who were coded as wearing red or pink than among women who were coded as wearing another colour. Conception risk did not predict shirt color in a multivariate model that included conception risk (continuous), temperature (continuous), and their interaction as predictors, but the interaction of conception risk and temperature was nearly significant ($p = .07$). However, this nearly significant interaction pointed to a pattern of results that contradicted the results of Tracy and Beall (2014). In contrast to their findings, the nearly significant interaction here would suggest that women are more likely to wear red or pink at peak fertility on warm days, but not on cold days.

GENERAL DISCUSSION

Humans appear to conceal their ovulatory status, even though many of our closest non-human relatives appear to advertise it through behavioural displays of sexual receptivity, genital swellings, or specialized vocalizations. Natural selection therefore may have favoured the concealment of oestrus in ancestral humans because concealment enabled ancestral women to deter unwanted male attention at peak fertility or to initiate sexual interactions outside of the fertile window (Deschner *et al.*, 2004; Nunn, 1999; Wallen & Zehr, 2004; Wolfe, 1979). Even so, researchers have recently considered whether only partially concealed ovulation might have been favoured in our evolutionary history. Such speculation is not unreasonable: There is a large literature on subtle changes that coincide with ovulation, which suggests that ovulation is only partially concealed. For example, women are rated as more attractive at ovulation (Welling & Puts, 2014) and ovulation increases women's sexual desire and efforts to obtain it (Arslan *et al.*, 2018).

Our studies, using novel clothing colour methods, do not robustly support the conjecture that women reveal their ovulatory status by wearing red or pink for several reasons: First, so few women wore red or pink (eight of 100 women in Study 1) that we were unable to assess the correlation between conception risk and photographs of what women actually wore to the laboratory. Second, although we found marginally significant ($p = .07$) evidence that women favoured red or pink garments during peak fertility on cold (and on warm) days in our modest sample (Study 1; $n = 100$) using one of our two methods for assessing women's colour preferences (i.e., trained raters' judgements but not computer automated coding), this effect did not replicate in a second and much larger sample in which the range of temperatures was broader. In fact, in our large sample (Study 2, Analysis II; $n = 497$) using trained raters' judgements for assessing women's colour preferences, we found the reverse: A nearly significant interaction effect of conception risk and temperature on shirt colour ($p = .07$) suggested that women favoured red or pink garments during peak fertility on warm days but not on cold days. Because this effect was not found in Analysis I of Study 2, was non-significant in Analysis II of Study 2 ($p = .07$), and was in the opposite direction from what has been previously

reported (Tracy & Beall, 2014) – we place little faith in its reliability. Finally, when using established, self-report methods for assessing women’s colour preferences in our close replication of the original studies (Study 2, Analysis I; $n = 617$), we found no effect in either direction.

Limitations

Our conclusions should be interpreted within the context of several limitations. First, our studies were completed prior to the publication of key methodological advances. For example, Gangestad *et al.* (2016) have since made recommendations for measuring conception probability in ovulatory cycle research and Blake, Dixson, O’Dean, and Denson (2016) have published standardized protocols for characterizing women’s fertility. This work has shown that calendar-counting methods for assessing ovulatory status are only moderately correlated with hormonally assessed ovulatory status (which is now taken to be the most valid measure of ovulatory status currently available; Gangestad *et al.*, 2016) and the approximate validity of calendar-counting measures is 0.55. Though our Study 1 conducted in 2012 and our Study 2 conducted in 2015 both followed standard best practices at the time, we could not consider the crucial points raised by Gangestad *et al.* or Black *et al.* published later, in 2016, potentially limiting our ability to detect an effect. Even so, our two studies provide independent conceptual, close, and extended replications of the original studies and present new methods of choice of dress assessment among several hundred women, which can be considered a strength.

Second, to closely replicate Beall and Tracy (2013), we limited ourselves to using a between-subjects design and indirect methods to categorize women into high- and low-conception-risk groups. Although we made efforts to limit researcher degrees of freedom associated with indirect methods by committing to Wilcox *et al.*’s (2000) method before analysing our data, and although we conducted studies with power exceeding 0.85, all indirect methods have validity problems, even when using continuous measures, rather than binary measures (Blake *et al.*, 2017; Gangestad *et al.*, 2016). Indeed, women often fail to correctly recall the first day of their last period or their cycle lengths, with accuracy as low as 57% (Wegienka & Baird, 2005). The results reported here draw added attention to the methodological issues faced by social psychologists and suggest that the costs associated with urinary and blood ovulation kits are justified (Wideman, Montgomery, Levine, Beynon, & Shultz, 2013).

Ideally, studies investigating ovulatory status and garment colour should now use hormonal markers of ovulation to appropriately assess conception risk. Consistent with this shift in ovulatory status methodology, in their effort to conceptually replicate Beall and Tracy (2013), Eisenbruch *et al.* (2015) circumvented counting method issues by using ovarian hormones to assign fertility and found that with within-woman fluctuations in the ratio of oestradiol to progesterone mediated shifts in wearing red. Additionally, Blake *et al.* (2017) were able to partially replicate Eisenbruch *et al.* (2015) findings among young women. However, the effect they found was not significant (though positive), not found in other samples of varying ages, and may not be specific to the colour red.

Third, as in the original studies, our studies depended on the accuracy of women’s self-reports of age, hormonal contraceptive use, health, and menstrual cycles. Furthermore, like the original studies, our studies did not consider the use of medication (other than hormonal contraceptives) or presence of endocrine disorders that might interact with

hormone levels. It is also possible that our self-report, photograph-based, and mannequin-based assessments of women's shirt colours suffered from validity problems. However, our determinations of shirt colour showed relatively high correspondence in both Study 1 and Study 2 ($\Phi_s > 0.80$), and we did not find consistent evidence for an association between ovulatory status and shirt colour for any of our four measures – each of which has unique features that could potentially undermine its validity. The method independence of our conclusion gives further reason to suspect that our results were not a product of invalid methods for assessing shirt colour preferences.

Fourth, in line with the original studies, we did not request that women excuse themselves from participating based on their sexual orientation and so our studies were not limited to straight-identified women. However, we did collect data on women's sexual orientation by asking, 'Please check the single option that best describes your sexual orientation' in Study 1 and 'Which of the following best describes your sexual orientation?' with seven response options ranging from 'exclusively heterosexual' to 'equally heterosexual and homosexual', to 'exclusively homosexual' (Study 2 included an eighth option of 'asexual'). In Study 1, nine women reported being exclusively homosexual. In Study 2, 15 women reported being exclusively homosexual and five women reported being asexual. Because our results were the same when we excluded these women, we retained them in our samples to mirror the methods of the original study and we believe our results were not influenced by the inclusion of these non-heterosexual women. Future research could consider role of ovulatory cycle effects on social behaviour among non-heterosexual women.

Finally, we note that our Study 1 took place in a geographical region where the median temperature was 24.25°C. As a result, both our 'cold-' and 'warm-weather' subgroups evinced a mean temperature greater than mean temperature of the 'warm-weather' subgroup in Tracy and Beall's (2014) study. It is also of note that whereas Tracy and Beall (2014) sampled participants on an unseasonably cold versus an unseasonably warm day, in our Study 2, we sampled participants across a range of cold and warm days. It is possible that differences in geographic location in Study 1 and method in Study 2 could account for the different results.

Conclusion

Understanding similarities and differences between humans and our non-human relatives enables us to better map the adaptations that evolved in hominins, particularly our own species. It can also provide clues to the features of the physical world and the social world in which our psychology evolved. Currently, experts do not agree on whether human ovulation has come to be concealed over evolutionary time (Havlíček *et al.*, 2015; Marlowe, 2004; Welling & Puts, 2014; Wood & Carden, 2014), and there is considerable debate as to whether women advertise their ovulatory status through physiological or behavioural changes that alter their appearance (Arslan *et al.*, 2018; Beall & Tracy, 2013; Durante *et al.*, 2008; Eisenbruch *et al.*, 2015; Haselton *et al.*, 2007; Stern *et al.*, 2019; Tracy & Beall, 2014). The studies we presented here, which incorporated conceptual replications, close replications, and extended replications of the original studies reporting that women are more likely to wear red or pink at peak fertility (Beall & Tracy, 2013; Tracy & Beall, 2014), failed to reliably confirm the conjecture that women advertise their ovulatory status by wearing red, even on relatively cold days. These results support the idea that women may not advertise their ovulatory status overtly – there plausibly are costs associated with overtly advertising one's ovulatory status indiscriminately (such as

undesired male attention). However, there are also costs associated with not advertising it at all (which might discourage desired male attention). Thus, it seems to us that additional attention to these research questions, motivated by strong theoretical rationale and large-sample experiments with high-quality measurement methods, remains a potentially productive direction for research.

Acknowledgements

Preparation of this manuscript was supported in part by grant from the NIAAA, F32 AA025830 (PI: Liana S.E. Hone).

Conflicts of interest

All authors declare no conflict of interest.

Author contributions

Liana Senneth Elliott Hone (Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Validation; Visualization; Writing – original draft; Writing – review & editing) Michael Earl McCullough (Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Visualization; Writing – original draft; Writing – review & editing).

Data availability statement

The data that support the findings of this study are openly available in OSF at <https://osf.io/8t3n6/>, reference number 8t3n6.

References

- Alexander, R. D., & Noonan, K. M. (1979). Concealment of ovulation, parental care, and human social evolution. In N. Chagnon & W. Irons (Eds.), *Evolutionary biology and human social behavior: An anthropological perspective* (pp. 436–453). North Scituate, MA: Duxbury Press.
- Arslan, R. C., Schilling, K. M., Gerlach, T. M., & Penke, L. (2018). Using 26,000 diary entries to show ovulatory changes in sexual desire and behavior. *Journal of Personality and Social Psychology*, <https://doi.org/10.1037/pspp0000208>
- Aslam, M. M. (2006). Are you selling the right colour? A cross-cultural review of colour as a marketing cue. *Journal of Marketing Communications*, *12*(1), 15–30. <https://doi.org/10.1080/13527260500247827>
- Aujard, F., Heistermann, M., Thierry, B., & Hodges, J. (1998). Functional significance of behavioral, morphological, and endocrine correlates across the ovarian cycle in semifree ranging female Tonkean macaques. *American Journal of Primatology*, *46*, 285–309. [https://doi.org/10.1002/\(SICI\)1098-2345\(1998\)46:4<285:AID-AJP2>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1098-2345(1998)46:4<285:AID-AJP2>3.0.CO;2-8)
- Beall, A. T., & Tracy, J. L. (2013). Women are more likely to wear red or pink at peak fertility. *Psychological Science*, *24*, 1837–1841. <https://doi.org/10.1177/0956797613476045>
- Benshoof, L., & Thornhill, R. (1979). The evolution of monogamy and concealed ovulation in humans. *Journal of Social and Biological Structures*, *2*, 95–106. [https://doi.org/10.1016/0140-1750\(79\)90001-0](https://doi.org/10.1016/0140-1750(79)90001-0)

- Bielert, C., Girolami, L., & Jowell, S. (1989). An experimental examination of the colour component in visually mediated sexual arousal of the male chacma baboon (*Papio ursinus*). *Journal of Zoology*, *219*, 569–579. <https://doi.org/10.1111/j.1469-7998.1989.tb02601.x>
- Blake, K. R., Dixson, B. J. W., O'Dean, S. M., & Denson, T. F. (2016). Standardized protocols for characterizing women's fertility: A data-driven approach. *Hormones and Behavior*, *81*, 74–83. <https://doi.org/10.1016/j.yhbeh.2016.03.004>
- Blake, K. R., Dixson, B. J. W., O'Dean, S. M., & Denson, T. F. (2017). No compelling positive association between ovarian hormones and wearing red clothing when using multinomial analyses. *Hormones and Behavior*, *90*, 129–135. <https://doi.org/10.1016/j.yhbeh.2017.03.005>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Burley, N. (1979). The evolution of concealed ovulation. *The American Naturalist*, *114*, 835–858. <https://doi.org/10.1086/283532>
- Butler, H. (1974). Evolutionary trends in primate sex cycles. *Contributions to Primatology*, *3*, 2–35.
- Connor Gorber, S., Tremblay, M., Moher, D., & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review. *Obesity Reviews*, *8*(4), 307–326. <https://doi.org/10.1111/j.1467-789X.2007.00347.x>
- Deschner, T., Heistermann, M., Hodges, K., & Boesch, C. (2004). Female sexual swelling size, timing of ovulation, and male behavior in wild West African chimpanzees. *Hormones and Behavior*, *46*, 204–215. <https://doi.org/10.1016/j.yhbeh.2004.03.013>
- Dixson, A. F. (2012). *Primate sexuality*. Oxford, UK: Oxford University Press.
- Durante, K. M., Li, N. P., & Haselton, M. G. (2008). Changes in women's choice of dress across the ovulatory cycle: Naturalistic and laboratory task-based evidence. *Personality & Social Psychology Bulletin*, *34*, 1451–1460. <https://doi.org/10.1177/0146167208323103>
- Eisenbruch, A. B., Simmons, Z. L., & Roney, J. R. (2015). Lady in Red: Hormonal predictors of women's clothing choices. *Psychological Science*, *26*, 1332–1338. <https://doi.org/10.1177/0956797615586403>
- Elliot, A. J., & Niesta, D. (2008). Romantic red: Red enhances men's attraction to women. *Journal of Personality and Social Psychology*, *95*, 1150–1164. <https://doi.org/10.1037/0022-3514.95.5.1150>
- Elliot, A. J., & Pazda, A. D. (2012). Dressed for sex: Red as a female sexual signal in humans. *PLoS ONE*, *7*, 1–5. <https://doi.org/10.1371/journal.pone.0034607>
- Elliot, A. J., Tracy, J. L., Pazda, A. D., & Beall, A. T. (2013). Red enhances women's attractiveness to men: First evidence suggesting universality. *Journal of Experimental Social Psychology*, *49*(1), 165–168. <https://doi.org/10.1016/j.jesp.2012.07.017>
- Emery, K. J., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017). Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research*, *141*, 66–75. <https://doi.org/10.1016/j.visres.2016.12.007>
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Computational Statistics & Data Analysis*, *28*(1), 1–11. [https://doi.org/10.1016/s0167-9473\(99\)90014-2](https://doi.org/10.1016/s0167-9473(99)90014-2)
- Gangestad, S. W., Haselton, M. G., Welling, L. L. M., Gildersleeve, K., Pillsworth, E. G., Burriss, R. P., ... Puts, D. A. (2016). How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evolution and Human Behavior*, *37*, 85–96. <https://doi.org/10.1016/j.evolhumbehav.2015.09.001>
- Gangestad, S. W., & Thornhill, R. (2008). Human oestrus. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 991–1000. <https://doi.org/10.1098/rspb.2007.1425>
- Gouzoules, H., Gust, D. A., Donaghey, B., & St. Andre, E. (1998). Estrus vocalizations in two primate species (*Cercocebus torquatus atys* and *Macaca nemestrina*): Evidence for an effect of intrasexual competition. *Evolution of Communication*, *2*, 189–215. <https://doi.org/10.1075/eoc.2.2.03gou>
- Greenfield, A. B. (2005). *A perfect red: Empire, espionage, and the quest for the color of desire*. New York, NY: Harper Collins.

- Haselton, M. G., & Gildersleeve, K. (2011). Can men detect ovulation? *Current Directions in Psychological Science*, 20, 87–92. <https://doi.org/10.1177/0963721411402668>
- Haselton, M. G., Mortezaie, M., Pillsworth, E. G., Bleske-Rechek, A., & Frederick, D. A. (2007). Ovulatory shifts in human female ornamentation: Near ovulation, women dress to impress. *Hormones and Behavior*, 51(1), 40–45. <https://doi.org/10.1016/j.yhbeh.2006.07.007>
- Havlíček, J., Cobey, K. D., Barrett, L., Klapilová, K., & Roberts, S. C. (2015). The spandrels of Santa Barbara? A new perspective on the peri-ovulation paradigm. *Behavioral Ecology*, 26, 1249–1260. <https://doi.org/10.1093/beheco/arv064>
- Heistermann, M., Ziegler, T., van Schaik, C. P., Launhardt, K., Winkler, P., & Hodges, J. K. (2001). Loss of oestrus, concealed ovulation and paternity confusion in free-ranging Hanuman langurs. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268, 2445–2451. <https://doi.org/10.1098/rspb.2001.1833>
- Hill, G. E., McGraw, K. J. (Eds.) (2006). *Coloration: Mechanisms and measurements* (Vol. 1). Cambridge, MA: Harvard University Press.
- IDimager. (2007). Average Color Seeker (Version 0.41). Retrieved from <http://www.snapfiles.com/get/avgcolor.html>
- Knott, C. D., Thompson, M. E., Stumpf, R. M., & McIntyre, M. H. (2010). Female reproductive strategies in orangutans, evidence for female choice and counterstrategies to infanticide in a species with frequent sexual coercion. *Proceedings of the Royal Society B: Biological Sciences*, 277, 105–113. <https://doi.org/10.1098/rspb.2009.1552>
- Maestripieri, D. (2004). Primate copulation calls and postcopulatory female choice. *Behavioral Ecology*, 16(1), 106–113. <https://doi.org/10.1093/beheco/arh120>
- Marlowe, F. W. (2004). Is human ovulation concealed? Evidence from conception beliefs in a hunter-gatherer society. *Archives of Sexual Behavior*, 33, 427–432. <https://doi.org/10.1023/B:ASEB.0000037423.84026.1f>
- Nunn, C. L. (1999). The evolution of exaggerated sexual swellings in primates and the graded-signal hypothesis. *Animal Behaviour*, 58, 229–246. <https://doi.org/10.1006/anbe.1999.1159>
- Peperkoorn, L. S., Roberts, S. C., & Pollet, T. V. (2016). Revisiting the red effect on attractiveness and sexual receptivity: No effect of the color red on human mate preferences. *Evolutionary Psychology*, December, 1–13. <https://doi.org/10.1177/1474704916673841>
- Pollet, T. V., Costello, J., Groeneboom, L., Peperkoorn, L. S., & Wu, J. (2019). Do red objects enhance sexual attractiveness? No evidence from two large replications. *Displays*, 56(September 2018), 23–29. <https://doi.org/10.1016/j.displa.2018.10.008>
- Radeloff, D. J. (1990). Role of color in perception of attractiveness. *Perceptual and Motor Skills*, 71(1), 151–160. <https://doi.org/10.2466/pms.1990.71.1.151>
- Stern, J., Rudolph, S., & Penke, L. (2019). Probing ovulatory cycle shifts in women's make-up and clothing style. *Theoretical & empirical status of cycle phase effects. Symposium conducted at the 31th annual meeting of the Human Behavior and Evolution Society*. Boston, MA.
- Strassmann, B. I. (1981). Sexual selection, paternal care, and concealed ovulation in humans. *Ethology and Sociobiology*, 2(1), 31–40. [https://doi.org/10.1016/0162-3095\(81\)90020-0](https://doi.org/10.1016/0162-3095(81)90020-0)
- Symons, D. (1979). *The evolution of human sexuality*. New York, NY: Oxford University Press.
- Tracy, J. L., & Beall, A. T. (2014). The impact of weather on women's tendency to wear red or pink when at high risk for conception. *PLoS ONE*, 9(2), e88852. <https://doi.org/10.1371/journal.pone.0088852>
- Wallen, K., & Zehr, J. L. (2004). Hormones and history: The evolution and development of primate female sexuality. *Journal of Sex Research*, 41(1), 101–112. <https://doi.org/10.1080/00224490409552218>
- Wegienka, G., & Baird, D. D. (2005). A comparison of recalled date of last menstrual period. *Journal of Women's Health*, 14(3), 248–252. <https://doi.org/10.1089/jwh.2005.14.248>
- Welling, L. L. M., & Puts, D. A. (2014). Female adaptations to ovulation. In V. Weekes-Shackelford & T. Shackelford (Eds.), *Evolutionary perspectives on human sexual psychology and behavior* (pp. 243–260). New York, NY: Springer Science + Business Media.

- Wideman, L., Montgomery, M. M., Levine, B. J., Beynnon, B. D., & Shultz, S. J. (2013). Accuracy of calendar-based methods for assigning menstrual cycle phase in women. *Sports Health: A Multidisciplinary Approach*, 5(2), 143–149. <https://doi.org/10.1177/1941738112469930>
- Wilcox, A. J., Dunson, D., & Baird, D. D. (2000). The timing of the “fertile window” in the menstrual cycle: Day specific estimates from a prospective study. *BMJ*, 321, 1259–1262. <https://doi.org/10.1136/bmj.321.7271.1259>
- Wilcox, A. J., Dunson, D. B., Weinberg, C. R., Trussell, J., & Baird, D. D. (2001). Likelihood of conception with a single act of intercourse: Providing benchmark rates for assessment of post-coital contraceptives. *Contraception*, 63, 211–215. [https://doi.org/10.1016/S0010-7824\(01\)00191-3](https://doi.org/10.1016/S0010-7824(01)00191-3)
- Windham, G. C., Elkin, E. P., Swan, S. H., Waller, K. O., & Fenster, L. (1999). Cigarette smoking and effects on menstrual function. *Obstetrics and Gynecology*, 93(1), 59–65. [https://doi.org/10.1016/S0029-7844\(98\)00317-2](https://doi.org/10.1016/S0029-7844(98)00317-2)
- Wolfe, L. (1979). Behavioral patterns of estrous females of the Arashiyama West troop of Japanese macaques (*Macaca fuscata*). *Primates*, 20, 525–534. <https://doi.org/10.1007/BF02373434>
- 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. <https://doi.org/10.1037/a0036722>

Received 14 December 2018; revised version received 20 January 2020