Meta-Analysis of the Effect of Red on Perceived Attractiveness

Gabrielle K. Lehmann¹, Andrew J. Elliot², and Robert J. Calin-Jageman¹

Abstract
We conducted meta-analyses of studies that test the red-romance hypothesis, which is that the color red enhances heterosexual attraction in romantic contexts. For men rating women, we found a small, statistically significant effect (\(d = 0.26\ [0.12, 0.40]\), \(p = .0004\), \(N = 2,961\)), with substantial heterogeneity, \(Q(44) = 172.5, p_{Q} < .0001, I^2 = 89\% [82, 94]\), and equivocal results regarding the possibility of upward bias in the estimate. For women rating men, we found a very small effect (\(d = 0.13\ [0.01, 0.25]\), \(p = .03\), \(N = 2,739\)), with substantial heterogeneity, \(Q(35) = 73.0, p_{Q} = .0002, I^2 = 53\% [33, 80]\), and evidence of upward bias in the estimate. Moderator analyses suggest effect sizes may have declined over time (both genders), may be largest when an original shade of red is used (men only), and may be smaller in preregistered studies (women only). We present contrasting interpretations and suggestions for future research.

Keywords
attraction, color, mate selection, red, systematic review

Date received: July 12, 2018; Accepted: August 31, 2018

Visual processing is highly developed in humans. The human cortex contains at least 20 visual field maps, and there is a notable increase in the complexity of cortical visual processing when comparing humans to nonhuman primates (Brewer & Barto, 2012; Wandell, Dumoulin, & Brewer, 2007).

One prominent feature of human vision is its specialization for color processing. Our primate group is distinct from most mammalian species for having a fovea dedicated to color processing (Ahnelt & Kolb, 2000). Within the primates, humans and other Old-World primates are further distinguished for the prevalence of trichromacy (Nathans, 1999).

Given that humans have specific adaptations for color vision, it is not surprising that there is a long history of studying the psychology of color. However, color psychology has encountered several false starts. For example, one well-known finding is that viewing a specific shade of pink can reduce muscle strength and aggression (e.g., Schauss, 1979). The public has long been fascinated by this research—it sparked the use of pink in correctional facilities (Genschow, Noll, Wänke, & Gersbach, 2015), spurred football teams to paint their visiting team locker rooms pink (Buzuvis, 2007), and continues to fuel celebrity fads (Ferrier, 2017). The balance of evidence, however, suggests that pink does not have such a straightforward effect on human behavior. The initial research on muscle tone may have been unduly influenced by demand characteristics (Ingram & Lieberman, 1985; Smith, Bell, & Fusco, 1986), experiments with better controls suggest no effect of pink on muscle tone (Gilliam, 1991; Gilliam & Unruh, 1988), and there seem to be no long-lasting benefits of pink for reducing aggressive behavior in correctional facilities (e.g., Genschow et al., 2015).

Recently, Elliot and Maier (2012) sought to organize color psychology by developing color-in-context theory. This theory contends that both experience and evolution foster reciprocal associations between colors and different psychological states. Importantly, though, these associations do not function like simplistic reflex chains (e.g., pink always reduces aggression) but are learned and conditioned by context.

¹ Department of Psychology, Dominican University, River Forest, IL, USA
² Department of Clinical and Social Sciences in Psychology, University of Rochester, Rochester, NY, USA

Corresponding Author:
Robert J. Calin-Jageman, Department of Psychology, Dominican University, 7900 West Division Street, River Forest, IL 60305, USA.
Email: rcalinjageman@dom.edu
Color-in-context theory has, to date, focused mostly on the color red, which is proposed to have a negative association in achievement contexts, but a positive association in affiliative contexts. Elliot and Maier (2012) define an achievement context as one in which competence is being evaluated, especially when there is a chance of negative evaluation. In this context, red is posited to have a negative influence on performance and motivation. The deleterious effects of red posited in an achievement context are expected to be reversed in affiliation contexts, especially when physical and sexual attraction is evaluated among heterosexuals (Elliot & Maier, 2012). In these contexts, red is thought to elicit positive and appetitive associations, serving to enhance physical attractiveness (Elliot & Niesta, 2008). It is this proposition, which may be referred to as the red-romance hypothesis, that is the focus of the present work.

In proposing the red-romance hypothesis, Elliot and Niesta (2008) suggested that links between red and attraction could represent both evolutionary heritage and social learning. In terms of evolution, red may have special significance among primates, especially as a signal for sexual receptivity (e.g., Deschner, Heistermann, Hodges, & Boesch, 2004). In terms of social learning, many societies associate red with sexual interest and attraction (e.g., red for Valentine’s day and rouging of the lips); these cultural practices could also build associations between red and attraction that are activated in romantic contexts. It is important to note, though, that color associations are complex over evolution and across culture.

One of the first empirical investigations of the red-romance hypothesis was conducted by Elliot and Niesta (2008). In five experiments, male undergraduates were asked to rate the attractiveness of women in photographs. In each sample, perceived attractiveness was substantially higher when the photos featured red (red frame or red shirt) as opposed to other colors (e.g., white, gray, or green). The red-romance hypothesis was then extended to women rating men by some of the same researchers (Elliot et al., 2010), with five additional experiments indicating that red borders and shirts produce large increases in perceived attractiveness (see also Roberts, Owen, & Havlicek, 2010).

Since these initial publications, other experiments have supported the red-romance hypothesis (e.g., Jung, Kim, & Han, 2011; Lin, 2014; Young, 2015). Some studies have suggested that red also influences sexual desire (Schwarz & Singer, 2013) and behavior (Meier, D’Agostino, Elliot, Maier, & Wilkowski, 2012). Others have provided evidence that the effect of red on attraction may be universal across cultures (Elliot, Tracy, Pazda, & Beall, 2013; Han et al., 2018) and have identified potential mediator variables (e.g., Pazda, Elliot, & Greitemeyer, 2014). On the other hand, several experiments have obtained essentially no effect of red on self-reported attraction (e.g., Hesslinger et al., 2015; Seibt & Klement, 2015); this includes a number of relatively high-powered close replications with both male and female participants (Lehmann & Calin-Jageman, 2017).

The mixed results related to the effect of red on self-reported attraction could be due to methodological issues, such as the difficulty of rigorously producing and presenting color stimuli consistently across labs and experimental contexts (Elliot, 2015). Another possibility is moderation, as several factors have been identified which can influence how red affects attraction. For men, red may not influence attraction when the women being rated are unattractive (Young, 2015), have masculine faces (Wen, Zuo, Wu, Sun, & Liu, 2014), and/or are postmenopausal (Schwarz & Singer, 2013). For women, red may not enhance attraction to those expressing shame (Buechner, Maier, Lichtenfeld, & Elliot, 2015) or to those in a racial out-group (Berthold, 2013; Wartenberg, Höpfner, Potthast, & Mirau, 2011). Another possibility is that red effects are weaker than originally estimated. If this is the case, subsequent research is likely to be underpowered relative to the actual effect size, and thus a mixed literature would be expected. Consistent with this possibility, Francis (2012) has argued that the initial study of women rating men is improbably consistent given the sample sizes used, a pattern consistent with effect-size inflation due to publication bias (see Elliot & Maier, 2013, for a response).

One way to facilitate understanding of the red-romance literature is meta-analysis. Meta-analysis provides a quantitative synthesis across disparate research results. Moreover, meta-analysis enables moderator analysis to examine whether different factors might predict observed effect sizes across different experiments. Thus, meta-analysis provides a means to resolve some of the ambiguity related to the red-romance hypothesis.

Based on this reasoning, we undertook a meta-analysis of the empirical literature related to the effect of red on self-reported perceived attractiveness. Notably, this endeavor represents a sort of “adversarial collaboration” (Kahneman, 2003). That is, one of us (second author) is inclined to view the mixed literature as due to the inherent complexity of color research and as indicating a need for experiments testing moderator variables, while also acknowledging the need for more highly powered studies run with rigorous procedures (Elliot, 2015). The other two of us (first and last authors) have conducted a set of replication studies that did not support the red-romance hypothesis (Lehmann & Calin-Jageman, 2017) and are thus inclined to believe that the effect of red on attraction is small, potentially even nonexistent. This meta-analysis represents our attempt to collaboratively and systematically weigh the evidence for these different interpretations.

In conducting this meta-analysis, we focused specifically on the original set of findings, which found a marked increase in self-reported perceived attractiveness with stimuli featuring red relative to control colors of white, gray, blue, and green (Elliot & Niesta, 2008). Thus, this meta-analysis is not a synthesis of all data potentially relevant to the red-romance hypothesis (e.g., it does not include studies showing an effect of red on self-reported sexual desire or attraction-related behavior). We adopted this narrow focus for two reasons. First, collecting similar studies makes moderator analysis less noisy and more fruitful. Second, we wanted to examine the consistency of a specific effect as a foundational step to evaluating the broader evidence linking red to romantic attraction.
Method

The collected data set is posted to the Open Science Framework along with our analysis script (https://osf.io/xy47p/). We encourage the reanalysis, reuse, and extension of this meta-analysis.

Inclusion Criteria

Prior to conducting the literature review, we developed a narrow set of criteria for inclusion:

1. Studies had to involve experimental manipulation of color exposure.
2. One level of the independent variable had to be red.
3. Another level of the independent variable had to be one of the control colors used in the original research on the effect: white, gray, green, or blue (Elliot & Niesta, 2008).
4. Dependent variables had to include perceived attractiveness via self-report, a primary dependent variable used in the original research (Elliot & Niesta, 2008). This could be using the same items as in the original studies or novel items clearly linked to perceived attractiveness (e.g., Seibt & Klement, 2015). We did, however, exclude studies which included only behavioral measures of attraction (e.g., Meier et al., 2012).
5. Ratings had to be of the opposite sex (men rating women or women rating men). Where both types of ratings were made in the same study, these were split by sex. We excluded studies where this was not possible because participant sex was not recorded (Jung et al., 2011). Where sexual preference was recorded, we excluded participants who reported not being attracted to the opposite sex.

Many sources had multiple studies. We applied criteria study by study, collecting each valid experimental contrast within a source. As the review of papers proceeded, we added an additional criterion:

(6) Contrasts were not included that demonstrated effective moderators. For example, Young (2015) found a red effect for moderately attractive stimuli but not for unattractive stimuli. Therefore, we did not include the contrast for unattractive stimuli. We did include conditions with clearly ineffective moderators. For example, Schwarz and Singer (2013) found very similar red effects with both young and old participants, so both contrasts were included.

Finally, we identified two sets of studies that seemed important to evaluate but distinct from the other studies collected:

- Some experiments manipulated facial redness (e.g., Thorstenson, Pazda, Elliot, & Perrett, 2016a). Because this manipulation was integral to physical appearance rather than incidental, it may not be reasonable to expect similar effect sizes for this type of manipulation.
- Some experiments asked participants to rate the tip worthiness, as well as the attractiveness, of waiters or waitresses (e.g., Lynn, Giebelhausen, Garcia, Li, & Patumanon, 2016). This could make the competence of the target person salient. Because of this, these studies may qualify as an achievement context (see Thorstenson et al., 2016a, footnote 1) and might be expected to elicit either no benefit for red or even a negative effect.

We collected these studies but analyzed each set separately.

Search Strategy

Figure 1 shows a flowchart for study selection. We used a three-pronged search strategy to identify candidate studies. First, we used citation mining to identify studies in the published literature. Specifically, we used Google Scholar to examine every study which cited three seminal papers on this topic (Elliot et al., 2010; Elliot & Niesta, 2008; Roberts et al., 2010). For articles identified from this initial search, we then surveyed both cited and citing articles.

The second prong of our search strategy was to identify relevant unpublished studies. To do this, we posted calls for data (published or otherwise) to the Society for Personality and Social Psychology open forum (July 6, 2016) and the Social Psychology Network listserv (July 14, 2016). We also searched the psychology file drawer website (http://psychfiledrawer.org) and the Open Science Framework website (http://osf.io).

Finally, one of us (second author) searched his correspondence for additional studies that had been sent to him.

For the primary meta-analysis, our search strategy identified 41 sources containing 98 distinct contrasts that measured the effect of red on perceived attractiveness (6,682 distinct participants). Of these, 17 were coded as testing an effective moderator (983 participants). This left 45 contrasts for men rating women (2,961 male participants, Table 1) and 36 contrasts for women rating men (2,739 female participants, Table 2).

For the analysis of the effect of facial redness, we identified four sources containing 10 distinct contrasts involving the effect of facial redness on perceived attractiveness for men rating women (5 contrasts, 614 male participants) and women rating men (5 contrasts, 717 female participants; Table 3).

For the effect of red on ratings of waiters and waitresses, we identified two sources (Lynn & Giebelhausen, 2013; Lynn, Giebelhausen, Garcia, Li, & Patumanon, 2016) reporting eight contrasts, seven of which involved rating waitresses (1,176 male participants) and one of which involved rating waiters (220 female participants).

Effect Size Measures

We calculated effect sizes from summary statistics (group means, standard deviations [SDs], and sample sizes). When these were not reported, we contacted original authors. In the rare
instances where data could not be directly obtained, we used WebPlotDigitizer (http://arohatgi.info/WebPlotDigitizer/app/) to recover summary statistics. Checking this approach against figures with known summary data showed less than 5% error.

As a measure of effect size, we used Cohen’s $d$ calculated as $(M_{\text{red}} - M_{\text{control}})/s_{\text{pooled}}$. All effect sizes were corrected for the small upward bias in this sample statistic (Hedges, 1981).

For studies in which more than one control color was used, we calculated the sample-size weighted average of perceived attractiveness across the control colors. This had the benefit of not cherry-picking-specific contrasts. It also improved the precision of effect size estimates by aggregating data across control conditions. As mentioned above, only original control colors were used.

Effect sizes for within-subjects designs were calculated in the same way as between-subjects designs, ensuring comparability of effect size estimates (Morris & DeShon, 2002). The calculation of expected variance, however, was different for
### Table 1. Table of Contrasts in Which Men Rated Women.

| Contrast | Year | Journal/Publication | Preregistered? | Control Color | Control Form | Color Type | Participant Type | Study Design | Ethnic Majority Matched | Ethnicity Matched | Red | Original | Control Color Matched | Presentation Control | Baseline Attractiveness (%) | N_ref | M_ref | S_ref | N_control | M_control | S_control | t | \(d\) | 95% CI |
|----------|------|----------------------|----------------|---------------|--------------|------------|-----------------|------------|-----------------------|-----------------|-----|---------|-----------------------|----------------------|-----------------------------|-------|-------|-------|-----------|-----------|-----------|----|-------|-------|
| Elliot and Niesta (2008)—Exp 1—JESP | 2008 | Y | N | White | Background | Bust | North America | Students | Between | White subjects | Matched | 20.5 | Y | N | Y | 66 | 15.00 | 7.33 | 0.90 | 12.00 | 6.25 | 1.06 | 1.08 [0.27, 1.89] |
| Elliot and Niesta (2008)—Exp 2—JESP | 2008 | Y | N | White | Background | Bust | North America | Students | Between | White subjects | Matched | 19.3 | Y | N | Y | 66 | 16.00 | 7.07 | 0.78 | 16.00 | 6.13 | 1.53 | 0.75 [0.04, 1.47] |
| Elliot and Niesta (2008)—Exp 3—JESP | 2008 | Y | N | Gray | Background | Bust | North America | Students | Between | White subjects | Matched | 20.0 | Y | Y | Y | 61 | 20.00 | 6.65 | 1.07 | 17.00 | 5.91 | 1.07 | 0.67 [0.00, 1.33] |
| Elliot and Niesta (2008)—Exp 4—JESP | 2008 | Y | N | Green | Background | Bust | North America | Students | Between | White subjects | Matched | 20.0 | Y | Y | Y | 58 | 16.00 | 6.29 | 0.89 | 15.00 | 5.65 | 1.07 | 0.62 [0.10, 0.134] |
| Elliot and Niesta (2008)—Exp 5—JESP | 2008 | Y | N | Blue | Clothing | Bust | North America | Students | Between | White subjects | Matched | 19.8 | Y | Y | Y | 64 | 12.00 | 7.21 | 1.49 | 11.00 | 6.09 | 1.49 | 0.84 [0.01, 1.70] |
| Purdy (2009)—Exp 1—High Arousal—UNC Asheville Journal of Undergraduate Research | 2009 | N | N | Gray | Background | Bust | North America | Students | Between | White Matched | — | 21.0 | N | N | Y | 66 | 13.00 | 7.26 | 3.27 | 13.00 | 6.24 | 3.27 | 0.30 [0.47, 0.108] |
| Purdy (2009)—Exp 1—Low Arousal—UNC Asheville Journal of Undergraduate Research | 2009 | N | N | Gray | Background | Bust | North America | Students | Between | White Matched | — | 21.0 | N | N | Y | 67 | 13.00 | 6.37 | 3.31 | 13.00 | 6.33 | 3.27 | 0.01 [-0.76, 0.78] |
| Roberts, Owen, and Havlicek (2010)—Exp 1—Evolutionary Psychology | 2010 | Y | N | Blue/ green/ white | Clothing | Bust | Europe | Students | Within | White subjects | Matched | — | N | N | Y | 38 | 30.00 | 4.60 | 1.10 | W-S | 4.40 | 0.92 | 0.920 | 0.19 [0.03, 0.34] |
| Roberts et al. (2010)—Exp 2—Evolutionary Psychology | 2010 | Y | N | Blue/ green/ white | Clothing | Bust | Europe | Students | Within | White subjects | Matched | — | N | N | Y | 27 | 15.00 | 3.65 | 0.45 | - | W-S | 3.46 | 0.48 | 0.40 [0.15, 0.65] |
| Roberts et al. (2010)—Exp 3—Evolutionary Psychology | 2010 | Y | N | White | Clothing | Bust | Europe | Students | Within | White subjects | Matched | — | N | N | Y | 43 | 25.00 | 4.82 | 0.85 | W-S | 4.88 | 0.85 | 0.920 | 0.08 [-0.23, 0.08] |
| Gutguen (2012)—Exp 1—JESP | 2012 | Y | N | Blue/ green/ white | Clothing | Bust | Europe | Students | Between | — | 19.2 | N | N | Y | 51 | 30.00 | 5.95 | 1.24 | 9.00 | 5.04 | 1.22 | 0.74 [0.31, 1.16] |
| Pardas, Elliot, and Gratzemeyer (2012)—Exp 1—JESP | 2012 | Y | N | Green | Clothing | Bust | Europe | Students | Between | White subjects | Matched | 23.5 | N | Y | Y | 50 | 27.00 | 6.07 | 1.17 | 22.00 | 5.00 | 2.13 | 0.63 [0.05, 1.21] |
| Bigelow et al. (2013)—Exp 1—UNC Asheville Journal of Undergraduate Research | 2013 | N | N | Blue | Background | Head shot | North America | Students | Between | subjects | Matched | — | — | N | N | Y | 66 | 4.00 | 4.65 | 2.05 | 4.00 | 6.28 | 1.81 | -0.73 [-2.16, 0.70] |
| Elliot, Tracy, Pardas, and Beall (2013)—Exp 1—JESP | 2013 | Y | N | Blue | Background | Head shot | Africa | Adults | Between | Black subjects | Matched | 26.8 | N | Y | Y | 79 | 21.00 | 4.62 | 0.59 | 21.00 | 4.14 | 0.85 | 0.64 [0.02, 1.26] |
| Schwartz and Singer (2013)—Exp 1—Adults Rate Young—JESP | 2013 | Y | N | White | Background | Bust | Europe | Adults | Between | subjects | — | 53.5 | N | N | Y | 64 | 30.00 | 6.53 | 1.96 | 30.00 | 6.13 | 2.17 | 0.19 [-0.32, 0.70] |
| Schwartz and Singer (2013)—Exp 1—UGrads Rate Young—JESP | 2013 | Y | N | White | Background | Bust | Europe | Students | Between | — | 24.7 | N | N | Y | 68 | 30.00 | 6.20 | 1.42 | 30.00 | 6.40 | 1.18 | -0.15 [-0.66, 0.36] |
| Stefan and Guuguenu (2013)—Exp 1—Positive Personal communication to A.E. | 2013 | N | N | White | Clothing | Full body | Europe | Students | Between | White subjects | Matched | 22.1 | N | N | Y | 67 | 18.00 | 8.133 | 16.30 | 16.00 | 67.50 | 26.87 | 0.62 [-0.07, 1.31] |
| Williams and Neelon (2013)—Exp 1—All—Psi Chi Journal | 2013 | Y | N | Blue | Background | Bust | North America | Students | Between | White subjects | Matched | 22.5 | N | N | Y | 64 | 16.00 | 6.81 | 1.03 | 15.00 | 6.15 | 1.20 | 0.58 [-0.14, 1.30] |
| Williams and Neelon (2013)—Exp 2—Positive and Neutral—Psi Chi Journal | 2013 | Y | N | Blue | Background | Bust | North America | Students | Between | White subjects | Matched | 21.5 | N | N | Y | 69 | 26.00 | 6.40 | 1.28 | 26.00 | 6.56 | 1.06 | -0.13 [-0.67, 0.04] |

(continued)
Table 1 (continued)

| Contrast Year | Academic Journal Publication? | Preregistered? | Control Color | Color Form | Photo Type | Participant Type | Study Design | Ethnic Majority Match | Ethnicity | Red Original | Control Color | Presentation Control | Baseline Attractiveness (%) | N_red | M_red | S_red | N_control | M_control | S_control | r | d 95% CI |
|---------------|-----------------------------|---------------|---------------|------------|------------|-----------------|-------------|-----------------------|-----------|--------------|---------------|------------------------|---------------------------|--------|-------|------|-----------|----------|-----------|---|--------|
| Lin (2014)—Exp 1 Display | 2014 | Y | N | Blue | Item | Bust | Asia | Students | Between | Asian subjects | Matched | — | N | N | Y | 39 | 2000 | 3.50 | 0.69 | 20.00 | 2.55 | 0.31 | 1.53 [0.83, 2.24] |
| Pazda, Elliot, and Grasmeyer (2014)—Exp 1 | 2014 | Y | N | White | Clothing | Full body | Unknown | Online worker | Between | White subjects | Matched | 24.5 | N | N | N | 56 | 109.00 | 6.03 | 1.78 | 125.00 | 5.46 | 1.77 | 0.32 [0.06, 0.58] |
| Wen, Zuo, Wu, Sun, and Liu (2014)—Exp 1 | 2014 | Y | N | Blue/white | Clothing | Bust | Asia | Students | Between | Asian Matched | Y | Y | Y | 22.00 | 0.24 | 0.93 | 44.00 | 0.10 | 0.81 | 0.16 [-0.35, 0.67] |
| Kirsch (2015)—Exp 1 | 2015 | N | N | White | Clothing | Bust | Europe | Students | Between | White Matched | — | — | — | 68 | 5700 | 6.85 | 1.44 | 48.00 | 6.42 | 1.22 | 0.32 [-0.07, 0.71] |
| Seibt and Klement (2015)—Exp 1 | 2015 | N | N | Green | Background | — | Europe | Students | Between | — Matched | Y | Y | Y | 2000 | 4.29 | 0.86 | 19.00 | 4.17 | 0.81 | 0.14 [0.01, 0.26] |
| Sullivan et al. (2016)—Exp 1 | 2015 | N | N | Gray | Background | Head | North America | Students | Between | Mixed | Matched | N | Y | Y | 53 | 1900 | 4.29 | 0.86 | 19.00 | 4.17 | 0.81 | 0.14 [0.01, 0.26] |
| Gilson and Privitera (2016)—Exp 1 | 2016 | Y | N | White | Clothing | Head | North America | Students | Between | — | Matched | N | N | Y | 53 | 4600 | 4.30 | 1.11 | 46.00 | 4.19 | 1.07 | 0.975 [0.01, 0.17] |
| Gilson and Privitera (2016)—Exp 2 | 2016 | Y | N | Blue | Background | Head | North America | Students | Between | Mixed | Matched | N | Y | Y | 53 | 4600 | 4.30 | 1.11 | 46.00 | 4.19 | 1.07 | 0.975 [0.01, 0.17] |

(continued)
| Contrast | Year | Academic Journal | Publication? Preregistered? | Control Color | Color Form | Photo Type | Continent | Participant Type | Study Design | Ethnic Majority | Match | Red Original | Matched | Presentation Control | Baseline Attractiveness (%) | N_red | M_red | S_red | N_control | M_control | S_control | r | d 95% CI |
|----------|------|------------------|-----------------------------|---------------|------------|-----------|-----------|---------------|-------------|----------------|-------|--------------|---------|-------------------|-----------------------------|-------|-------|-------|-----------|-----------|-----------|---|---------|
| O’Mara et al. (2016)—Exp 3—Long Shirt—Unpublished data; response to call | 2016 | N | N | White | Clothing | Bust | North America | Students Between subjects | Matched | 19.5 | N | N | Y | 77 | 63.00 | 6.968 | 1.27 | 57.00 | 7.12 | 0.98 | -0.14 [-0.50, 0.02] |
| O’Mara et al. (2016)—Exp 3—Tank Top—Unpublished data; response to call | 2016 | N | N | White | Clothing | Bust | North America | Students Between subjects | Matched | 19.5 | N | N | Y | 73 | 51.00 | 6.696 | 1.13 | 55.00 | 6.827 | 1.005 | -0.12 [-0.50, 0.26] |
| Peperkoorn, Roberts, and Pollet (2016)—Exp 1—Short Term—Evolutionary Psychology | 2016 | Y | N | White | Clothing | Bust | Europe | Students Between subjects | Matched | 23.7 | N | N | Y | 67 | 34.00 | 14.26 | 3.31 | 34.00 | 15.47 | 2.18 | -0.43 [-0.91, 0.05] |
| Peperkoorn et al. (2016)—Exp 2—Short Term—Evolutionary Psychology | 2016 | Y | N | White | Clothing | Bust | Europe | Students Between subjects | Matched | 19.5 | N | N | N | 67 | 33.00 | 15.06 | 4.14 | 36.00 | 15.47 | 3.61 | -0.10 [-0.58, 0.37] |
| Sullivan et al. (2016)—Exp 2—OSF project page https://osf.io/pm7x | 2016 | N | N | Gray | Background | Bust | North America | Students Between subjects | Matched | — | N | N | Y | 65 | 21.00 | 6.24 | 1.29 | 14.00 | 6.21 | 1.45 | 0.02 [-0.66, 0.69] |
| Costello, Groeneboom, and Pollet (2017)—Exp 1—Unpublished masters degree University of Laiden | 2017 | N | N | Blue | Item | Bust | Europe | Students Between subjects | Matched | 22.6 | N | N | Y | 34 | 65.00 | 2.25 | 1.04 | 64.00 | 2.38 | 0.95 | -0.12 [-0.47, 0.32] |
| Lehmann and Calin-Jageman (2017)—Exp 1—Social Psychology | 2017 | Y | Y | White | Background | Bust | North America | Students Between subjects | Matched | 20.8 | Y | N | Y | 69 | 21.00 | 6.52 | 1.16 | 11.00 | 6.52 | 1.56 | 0.00 [-0.73, 0.73] |
| Lehmann and Calin-Jageman (2017)—Exp 2—Social Psychology | 2017 | Y | Y | Gray | Background | Bust | North America | Students Between subjects | Matched | 35.8 | N | N | N | 67 | 104.00 | 6.5 | 1.32 | 106.00 | 6.38 | 1.32 | 0.10 [-0.17, 0.37] |
| Pazda, Thorstrom, and Elliot (2017)—Exp 1—Unpublished data; response to call | 2017 | N | N | Green | Clothing | Full body | North America | Students Within subjects | Matched | — | N | N | Y | 82 | 115.00 | 7.81 | 1.21 | 7.58 | 1.41 | 0.840 | 0.17 [0.07, 0.28] |
| Pazda et al. (2017)—Exp 2—Unpublished data; response to call | 2017 | N | Y | Green | Clothing | Full body | North America | Students Within subjects | Matched | 37.4 | N | N | N | 87 | 228.00 | 8.08 | 1.28 | 7.95 | 1.39 | 0.803 | 0.10 [0.02, 0.18] |
| Overall | | | | | | | | | | | | | | | | | | | | 0.26 [0.12, 0.40] |

Note. Columns 2–17 indicate how each contrast was coded for different moderators. Each moderator classification is explained in the text. Subsequent columns provide the basic descriptive statistics for each contrast: sample size (N), mean (M), and standard deviation (S) for the red group and the control group. For within-subjects designs, the correlation between repeated measures is provided (r). Finally, for each contrast, Cohen’s d and its 95% confidence interval is reported with d calculated so that positive values indicate higher attraction in the red condition. When a moderator could not be coded, the cell is marked with an emdash (—). Finally, for within-subjects designs, the total sample size is provided in the column for red sample size and the column for the control group is marked "W-S." Y = yes; N = no; Exp = experiment; PSP = Journal of Personality and Social Psychology; EJSP = European Journal of Social Psychology; JEP = Journal of Experimental Psychology; JSP = Journal of Social Psychology; DV = dependent variable.
Table 2. Table of Contrasts in Which Women Rated Men.

| Contrast | Year | Academic Journal | Preregistered? | Control Color | Color Form | Photo Type | Continent | Participant Type | Study Design | Ethnic Majority | Ethnicity Match | M_{lep} | Control Color | Matched | Presentation Control | Baseline Attractiveness (%) | N_{red} | M_{red} | S_{red} | N_{control} | M_{control} | S_{control} | t | d | 95% CI |
|----------|------|------------------|----------------|---------------|------------|------------|-----------|--------------|-------------|----------------|---------------|--------|--------------|---------|----------------------|-----------------------------|--------|--------|--------|------------|-------------|-------------|----|----|--------|
| Elliot et al. (2013)—Exp 1—JEP: General | 2010 | Y | N | White | Background | Bust | North America | Students | B-S | White | Matched | 202 | Y | N | Y | 58 | 10 | 6.79 | 1.00 | 11 | 5.67 | 1.34 | 0.90 [0.00, 1.80] |
| Elliot et al. (2010)—Exp 2—JEP: General | 2010 | Y | N | White | Background | Bust | North America | Students | B-S | White | Matched | 205 | Y | N | Y | 65 | 20 | 7.15 | 0.74 | 12 | 6.20 | 1.08 | 1.05 [0.29, 1.81] |
| Elliot et al. (2010)—Exp 3—JEP: General | 2010 | Y | N | Gray | Background | Bust | North America | Students | B-S | White | Mismatch | 196 | Y | Y | Y | 53 | 16 | 6.69 | 1.22 | 17 | 5.27 | 2.04 | 0.82 [0.11, 1.53] |
| Elliot et al. (2010)—Exp 4—JEP: General | 2010 | Y | N | Green | Clothing | Bust | Asia | Students | B-S | Asian | Matched | 206 | Y | Y | Y | 56 | 27 | 6.32 | 1.09 | 28 | 5.50 | 1.50 | 0.61 [0.07, 1.16] |
| Elliot et al. (2010)—Exp 7—JEP: General | 2010 | Y | N | Blue | Clothing | Bust | Europe | Students | B-S | White | Matched | 194 | Y | Y | Y | 64 | 12 | 7.50 | 1.17 | 15 | 6.13 | 1.89 | 0.82 [0.03, 1.61] |
| Roberts et al. (2010)—Exp 1—Evolutionary Psychology | 2010 | Y | N | White | Background | Bust | North America | Students | B-S | White | Matched | 20.2 | Y | N | Y | 58 | 10 | 6.79 | 1.00 | 11 | 5.67 | 1.34 | 0.90 [0.00, 1.80] |
| Elliot et al. (2010)—Exp 2—JEP: General | 2010 | Y | N | White | Clothing | Bust | Europe | Students | B-S | White | Matched | 20.5 | Y | N | Y | 65 | 20 | 7.15 | 0.74 | 12 | 6.20 | 1.08 | 1.05 [0.29, 1.81] |
| Elliot et al. (2010)—Exp 3—JEP: General | 2010 | Y | N | Gray | Background | Bust | North America | Students | B-S | White | Mismatch | 196 | Y | Y | Y | 53 | 16 | 6.69 | 1.22 | 17 | 5.27 | 2.04 | 0.82 [0.11, 1.53] |
| Elliot et al. (2010)—Exp 4—JEP: General | 2010 | Y | N | Green | Clothing | Bust | Asia | Students | B-S | Asian | Matched | 206 | Y | Y | Y | 56 | 27 | 6.32 | 1.09 | 28 | 5.50 | 1.50 | 0.61 [0.07, 1.16] |
| Elliot et al. (2010)—Exp 7—JEP: General | 2010 | Y | N | Blue | Clothing | Bust | Europe | Students | B-S | White | Matched | 194 | Y | Y | Y | 64 | 12 | 7.50 | 1.17 | 15 | 6.13 | 1.89 | 0.82 [0.03, 1.61] |
| Roberts et al. (2010)—Exp 1—Evolutionary Psychology | 2010 | Y | N | White | Background | Bust | North America | Students | B-S | White | Matched | 20.2 | Y | N | Y | 58 | 10 | 6.79 | 1.00 | 11 | 5.67 | 1.34 | 0.90 [0.00, 1.80] |
| Elliot et al. (2010)—Exp 2—JEP: General | 2010 | Y | N | White | Clothing | Bust | Europe | Students | B-S | White | Matched | 20.5 | Y | N | Y | 65 | 20 | 7.15 | 0.74 | 12 | 6.20 | 1.08 | 1.05 [0.29, 1.81] |
| Elliot et al. (2010)—Exp 3—JEP: General | 2010 | Y | N | Gray | Background | Bust | North America | Students | B-S | White | Mismatch | 196 | Y | Y | Y | 53 | 16 | 6.69 | 1.22 | 17 | 5.27 | 2.04 | 0.82 [0.11, 1.53] |
| Elliot et al. (2010)—Exp 4—JEP: General | 2010 | Y | N | Green | Clothing | Bust | Asia | Students | B-S | Asian | Matched | 206 | Y | Y | Y | 56 | 27 | 6.32 | 1.09 | 28 | 5.50 | 1.50 | 0.61 [0.07, 1.16] |
| Elliot et al. (2010)—Exp 7—JEP: General | 2010 | Y | N | Blue | Clothing | Bust | Europe | Students | B-S | White | Matched | 194 | Y | Y | Y | 64 | 12 | 7.50 | 1.17 | 15 | 6.13 | 1.89 | 0.82 [0.03, 1.61] |
| Roberts et al. (2010)—Exp 1—Evolutionary Psychology | 2010 | Y | N | White | Background | Bust | North America | Students | B-S | White | Matched | 20.2 | Y | N | Y | 58 | 10 | 6.79 | 1.00 | 11 | 5.67 | 1.34 | 0.90 [0.00, 1.80] |
| Elliot et al. (2010)—Exp 2—JEP: General | 2010 | Y | N | White | Clothing | Bust | Europe | Students | B-S | White | Matched | 20.5 | Y | N | Y | 65 | 20 | 7.15 | 0.74 | 12 | 6.20 | 1.08 | 1.05 [0.29, 1.81] |
| Elliot et al. (2010)—Exp 3—JEP: General | 2010 | Y | N | Gray | Background | Bust | North America | Students | B-S | White | Mismatch | 196 | Y | Y | Y | 53 | 16 | 6.69 | 1.22 | 17 | 5.27 | 2.04 | 0.82 [0.11, 1.53] |
| Elliot et al. (2010)—Exp 4—JEP: General | 2010 | Y | N | Green | Clothing | Bust | Asia | Students | B-S | Asian | Matched | 206 | Y | Y | Y | 56 | 27 | 6.32 | 1.09 | 28 | 5.50 | 1.50 | 0.61 [0.07, 1.16] |
| Elliot et al. (2010)—Exp 7—JEP: General | 2010 | Y | N | Blue | Clothing | Bust | Europe | Students | B-S | White | Matched | 194 | Y | Y | Y | 64 | 12 | 7.50 | 1.17 | 15 | 6.13 | 1.89 | 0.82 [0.03, 1.61] |

(continued)
Table 2. (continued)

| Contrast                        | Year | Academic Journal | Preregistered? | Control Color | Control Form | Photo Type | Continent | Participant Type | Study Design | Ethnic Majority | Ethnicity Match | Red Original | Control Color Matched | Presentation Control | Baseline Attractiveness (%) | N_red | M_red | S_red | N_control | M_control | S_control | r     | d      | 95% CI     |
|---------------------------------|------|------------------|----------------|---------------|--------------|------------|-----------|------------------|--------------|-----------------|---------------------|--------------|----------------------|----------------------|----------------------------|--------|-------|-------|----------|-----------|-----------|-------|--------|---------|-----------|
| Khislavedy (2016)—Exp 1—OSF     | 2015 | N                | N              | Gray          | Background   | Bust       | North     | Students          | B-S          | Latino          | Matched             | 199           | Y       | Y     | Y      | 58        | 95        | 5.73     | 1.76  | 92      | 5.61     | 1.84      |    0.06 [-0.22, 0.35] |
| Kirsch (2015)—Exp 1—Heterosexual—Published monograph | 2015 | N                | N              | White         | Clothing    | Bust       | Europe    | Students          | B-S          | Latino          | Matched             | 225           | N       | N     | N      | 64        | 76        | 5.33     | 1.83  | 85      | 6.13     | 1.54      | -0.47 [-0.78, -0.16] |
| Legea et al. (2015)—Exp 1—OSF  | 2015 | N                | Y              | Gray          | Background   | Bust       | North     | Students          | B-S          | White           | Mismatch            | 20.7          | Y       | Y     | Y      | 67        | 25        | 5.76     | 1.61  | 23      | 6.33     | 1.58      | -0.35 [-0.92, 0.22] |
| Lehmann and Calin-Jageman (2015)—Class Exp—OSF project page: https://osf.io/f2udj | 2015 | N                | N              | White         | Background   | Bust       | North     | Students          | B-S          | White           | Matched             | 31.7          | N       | N     | N      | 34        | 29        | 4.34     | 1.63  | 28      | 3.69     | 1.65      | 0.39 [-0.13, 0.92] |
| Lehmann and Calin-Jageman (2017)—Social Psychology | 2017 | Y                | Y              | Gray          | Background   | North     | America  | Students          | B-S          | Latino          | Matched             | 20.4          | Y       | Y     | Y      | 49        | 56        | 4.67     | 2.05  | 60      | 4.89     | 1.80      | -0.11 [-0.48, 0.25] |
| Overall                         |      |                  |                |               |              |            |           |                  |              |                 |                     | 26.7          | N       | N     | N      | 26        | 54        | 3.19     | 0.53  | 31      | 3.07     | 0.55      | 0.22 [-0.22, 0.66] |

Note. Columns 2–17 indicate how each contrast was coded for different moderators. Each moderator classification is explained in the text. Subsequent columns provide the basic descriptive statistics for each contrast: sample size (N), mean (M), and standard deviation (S) for the red group and the control group. For within-subjects designs, the correlation between repeated measures is provided (r). Finally, for each contrast, Cohen’s d and its 95% confidence interval is reported with d calculated so that positive values indicate higher attraction in the red condition. When a moderator could not be coded, the cell is marked with an emdash (—). For within-subjects designs, the total sample size is provided in the column for red sample size and the column for the control group is marked “W-S.” Y = yes; N = no; Exp = experiment; B-S = Between Subjects; W-S = Within Subjects.
| Contrast | Year | Control | Red Color Matched | Participant Type | Study Design | Ethnic Minority | Study Ethnicity | Match | Control Color | Attractiveness (%) | DV | Ntotal | P_red | Mean_red | SD_red | Mean_control | SD_control | 95% CI | τ | 95% CI |
|----------|------|---------|-------------------|------------------|-------------|----------------|----------------|-------|---------------|-------------------|----|---------|--------|-----------|--------|-------------|------------|-------|------|----------|
| **Women rating men** | | | | | | | | | | | | | | | | | | | | | | |
| Thorstenson et al. (2016)—Exp 1—Perception | 2016 | Y | N | Less red | Face | Unknown | Online worker | W-S | — | — | N | N | N | — | Forced choice | 108 | .69 (.32) | | | 0.58 [0.38, 0.79] |
| Thorstenson et al. (2016)—Exp 2—Perception | 2016 | Y | N | Neutral | Face | Unknown | Online worker | W-S | — | — | N | N | N | — | Forced choice | 119 | .63 (.31) | | | 0.41 [0.22, 0.59] |
| Thorstenson et al. (2016)—Exp 3—Perception | 2016 | Y | N | Neutral | Face | North America | Students | W-S | — | — | N | Y | Y | 53 | Perceived attractiveness | 126 | 4.84 1.14 4.25 1.23 .47 | | | 0.49 [0.30, 0.68] |
| Thorstenson et al. (2016)—Exp 4—Perception | 2016 | N | N | Neutral | Face | North America | Students | W-S | White | Matched | 20.0 | N | Y | Y | 64 | Perceived attractiveness | 197 | 3.99 1.25 3.70 1.16 .62 | | | 0.24 [0.12, 0.36] |
| Overall | | | | | | | | | | | | | | | | | | | | | | 0.38 [0.20, 0.57] |
| **Men rating women** | | | | | | | | | | | | | | | | | | | | | | |
| Pazda et al. (2016)—Exp 1—Perception | 2016 | Y | N | Less red | Face | Unknown | Online worker | W-S | — | — | N | N | N | — | Forced choice | 133 | .72 (.34) | | | 0.65 [0.47, 0.84] |
| Pazda et al. (2016)—Exp 2—Perception | 2016 | Y | N | Neutral | Face | Unknown | Online worker | W-S | — | — | N | N | N | — | Forced choice | 127 | .62 (.30) | | | 0.39 [0.21, 0.57] |
| Pazda et al. (2016)—Exp 3—Perception | 2016 | Y | N | Less red | Face | Unknown | Online worker | W-S | — | — | N | N | N | 43 | Perceived attractiveness | 152 | 5.82 1.53 4.87 1.63 .57 | | | 0.60 [0.44, 0.76] |
| Pazda et al. (2016)—Exp 4—Perception | 2016 | Y | N | Neutral | Face | North America | Students | W-S | — | — | 20.1 | N | Y | Y | 29 | Perceived attractiveness | 101 | 5.01 1.30 4.52 1.33 .66 | | | 0.37 [0.20, 0.54] |
| Pazda et al. (2016b)—Exp 1—unpublished data; response to call | 2016 | N | N | Neutral | Face | North America | Students | W-S | White | Mis-match | 20.0 | N | Y | Y | 12 | Perceived attractiveness | 101 | 4.82 1.20 4.29 1.20 .71 | | | 0.44 [0.28, 0.60] |
| Overall | | | | | | | | | | | | | | | | | | | | | | 0.49 [0.33, 0.65] |

Note. Contrasts are grouped by gender of the participant. Columns 2–16 provide moderator classifications using the same coding scheme as the main meta-analysis; details in the text. Column 17 (DV) indicates if the dependent variable was perceived attractiveness of a forced choice between a more and a less red face. Subsequent columns provide descriptive statistics. All contrasts were within subjects: Ntotal reports sample size. For contrasts with a forced-choice design, P_red indicates the proportion of times the more red face was selected. For contrasts with a rating, means (M) and standard deviations (S) are reported for the red and control conditions. The τ column indicates the correlation across repeated measures. Finally, for each contrast, Cohen’s d and its 95% confidence interval is reported with d calculated so that positive values indicate higher attraction in the red condition. Y = yes; N = no; Exp = experiment; B-S = Between Subjects; W-S = Within Subjects.
within-subjects studies, incorporating the paired nature of the design and the correlation between paired measures to obtain a smaller expected variance than for equivalent data from a between-subjects design.

**Coding of Potential Moderators**

We coded studies for a range of possible moderators:

- **Publication year**—unpublished studies were coded as the year in which data collection ended;
- **Publication status**—published in a peer-reviewed journal or not. Any journal with a peer-review process was classified as peer reviewed, including undergraduate research journals (e.g., Bigelow, Taylor, & Underwood, 2013, published online in the UNC Asheville Journal of Undergraduate Research);
- **Preregistration status**—preregistered study or not;
- **Study design**—between- or within-subjects;
- **Control color**—white, gray, blue, or green;
- **Color form**—the aspect of the photograph that was altered: clothing, something on clothing, an item held, a border, or a background;
- **Photo type**—head shot, bust, or entire body;
- **Participant type**—students, online workers, or nonstudent adults;
- **Continent on which the study was conducted**—Africa, Asia, Europe, or North America. Online studies without geographical restriction were not coded.
- **Participant ethnic majority**—ethnicity of group representing >50% of participants or coded as mixed where no ethnicity was >50% of participants;
- **Ethnicity match**—coded as a match if the majority participant ethnicity matched the ethnicity of one or more of the individuals being rated; otherwise coded as nonmatch;
- **Mean participant age**;
- **Red original**—coded as yes if the red used was within five units of the original L and C values for red. Coded as no if the red used was beyond these parameters or if no color parameters were provided;
- **Control color matched**—coded as yes if the control color was matched to the red color used in aspects other than hue (within five units of L value for gray control color, within five units of L and C value for blue and green). Coded as no if the control color was not well matched or if no color parameters were provided;
- **Presentation control**—coded as yes if stimuli were presented in a way that maintained consistency within each group (e.g., on paper or a lab computer); coded as no otherwise (e.g., participant’s own computer);
- **Baseline attractiveness**—the mean level of perceived attractiveness in the control group. As different studies used different scales, data were normalized as a percentage of each study’s measurement scale: \( \frac{M_{\text{control}} - \text{scale}_{\text{minimum}}}{\text{scale}_{\text{max}} - \text{scale}_{\text{min}}} \).

The coding scheme and list of moderators was not established prior to the literature search but evolved as studies were examined.

**Analysis**

Analyses were conducted using version 2.0-0 of the metafor package (Viechtbauer, 2010) for R (Ihaka & Gentleman, 1996). All analyses reported are for random-effects models using restricted maximum-likelihood estimation and the Knapp and Hartung (2003) adjustment of standard errors.

For categorical moderators, any level with only one contrast was dropped from the analysis. For categorical moderators with only two levels, a meta-regression was conducted directly contrasting the two levels. For those with more than two levels, an overall meta-regression was first conducted with each level dummy coded. Significant results were then followed up by making all pairwise comparisons with Bonferroni correction for multiple comparisons. For heterogeneity of effect size, we report not only a formal test for heterogeneity (Q statistic and its p value) but also \( I^2 \) and its 95% confidence interval (CI). Decisions on whether and how to identify and handle outliers were not made a priori. Analyses are reported with and without outliers included.

**Results**

**The Effect of Red on Perceived Attractiveness**

**Men rating women.** For men rating women, we identified 45 contrasts of perceived attractiveness between a target featuring red versus a control color (2,961 distinct male participants). Of these, 28 were currently published. The other 17 contrasts were from unpublished sources (1,429 male participants; 48% of all qualifying data for this sex).

Analysis of these data with a random-effects meta-analysis yields a statistically significant effect: \( d = 0.26 \) [0.12, 0.40], \( p = .0004 \), \( N = 2,961 \) (Figure 2). The CI is consistent with anywhere from a very small effect up to a medium effect in the predicted direction. This interpretation must be qualified, however, due to extreme variation in effect size, \( Q(44) = 172.5, p < .0001, I^2 = 89\% \) [82, 94]. The \( I^2 \) value indicates that almost 90% of the variation in effect size is beyond what would be expected due to sampling error.

Part, though not all, of this heterogeneity was due to the study by Gilston and Privitera (2016). This study obtained a very large effect size (\( d = 1.92 \), compared to a median effect size \( d = 0.18 \) in the other studies), requiring considerable caution in interpretation. Due to its within-subjects design and relatively large sample size, this contrast had an outsized impact on the meta-analysis. Removing this outlier study gives an overall estimated effect size that is smaller (though still statistically significant) and much less heterogeneous (though still quite mixed): \( d = 0.19 \) [0.09, 0.29], \( p = .0004 \), \( Q(43) = 89.3, p_Q < .001, I^2 = 66\% \) [53, 90].
Meta-analysis estimates are only accurate if the data analyzed represent an unbiased sample of all the studies collected on the topic. To test for bias, we created a funnel plot (Figure 3) and tested for a relationship between effect size and standard error. The analysis did not show a statistically significant relationship ($r = -.07 \ [-.36, .22]$, $p = .63$, $k = 45$; $p = .82$, with the Gilston & Privitera study excluded). This analysis was strongly influenced by an outlier for expected sampling error. 

Figure 2. Overall meta-analysis of the effect of red on perceived attractiveness when men rate women. Each contrast is shown with its effect sizes and 95% confidence interval (CI). The diamond shows the overall estimated effect size and its 95% CI using a random-effects model ($d = 0.26\ CI\ [0.12, 0.40]$, $p = .0004$, $k = 45$, $N = 2,961$ distinct participants). Contrasts are shown in chronological order. Those marked with a † were unpublished; for these contrasts, the year shown is when data collection ended. The Gilston and Privitera (2016) study is marked with an empty square to flag the unusually large effect size obtained.
variation. This outlier is a study by Bigelow, Taylor, and Underwood (2013) that had only four participants/group (it is the study on the bottom left of Figure 3). With this outlier excluded, there is a relationship between effect size and standard error ($r = .29 \ [0.00, 0.55], p = .048, k = 44$).

**Women rating men.** For women rating men, we identified 36 different contrasts between red and a control color, encompassing 2,739 distinct participants. Of these, 16 were currently published. The other 20 were from unpublished sources (1,761 female participants, 64% of all qualifying data for this sex).

From these data, a random-effects meta-analysis indicates a very small but statistically significant effect: $d = 0.13 \ [0.01, 0.25], p = .03, N = 2,739$ (Figure 4). The CI indicates that the data are consistent with anywhere from almost no effect up to a small effect in the predicted direction. This interpretation must be qualified, however, due to substantial variation in effect size, $Q(35) = 73.0, p = .0002, I^2 = 53\% [33, 80]$.

We again tested for potential bias in the data obtained by creating a funnel plot (Figure 5) and examining the relationship between effect size and standard error. This analysis yielded a statistically significant relationship, $r = .52 \ [.23, .72], p = .001, k = 36$. The funnel plot highlights the study by Bigelow et al. (2013) as an outlier for sampling error ($n = 6–8/group; it is the white circle at the very bottom of Figure 5). With this outlier excluded, the relationship remained significant, $r = .63 \ [.38, .80], p < .001, k = 35$.

**Moderator Analysis of the Effect of Red on Perceived Attractiveness**

To explore heterogeneity of effect size, we conducted moderator analyses for both men rating women (Table 4) and women rating men (Table 5). Moderator analyses were not preregistered and are not corrected for multiple comparisons across moderators. Moreover, the moderators we explored are not orthogonal.

**Culture/ethnicity.** The red-romance hypothesis predicts that red enhances attraction across cultures (e.g., Elliot et al., 2013). We examined this issue by coding studies by continent (Europe, North America, or Asia; Africa was not included because of insufficient data). Then, we examined the relationship between study location and red effect. Consistent with the red-romance hypothesis, continent was not related to effect size for men (Supplemental Figure 1, $p = .30; p = .18$, with outlier excluded) or women (Supplemental Figure 2, $p = .88$). We also coded the majority ethnicity of participants in each study. Ethnicity was also not a predictor of effect size for men (Supplemental Figure 3, $p = .20$; outlier study was not included because it did not report ethnicity) or for women (Supplemental Figure 4, $p = .42$). Most contrasts were with majority Caucasian participants, so differences with other participant pools were not precisely estimated.

A more complex issue is how red effects might be moderated when the participant and targets differ in their culture or ethnicity. One published study reports a strong effect when participants rated those from another ethnicity (Elliot et al., 2010), but two unpublished studies found that red effects are attenuated when participants rate an ethnic out-group (Berthold, 2013; Wartenberg et al., 2011). We examined whether a match between participant and target ethnicity might moderate the red effect. We did not observe a significant relationship between ethnicity match and effect size for men (Supplemental Figure 5, $p = .31$; outlier study was not included because it did not report ethnicity) or for women (Supplemental Figure 6, $p = .17$).

**Other participant characteristics.** As sexual behavior can vary tremendously across the life span, we examined mean participant age as a potential moderator. Consistent with the findings
of Schwarz and Singer (2013), age did not emerge as a significant predictor of effect size for men (Supplemental Figure 7, $p = .23$; $p = .41$, with outlier excluded) or for women (Supplemental Figure 8, $p = .17$). This finding is tentative, though, due to restriction of range in ages.

Another potentially important factor is the pool used to recruit participants, particularly in the contrast between online workers and in-person student samples. For men, participant type was not a statistically significant moderator (Supplemental Figure 9, $p = .60$; $p = .66$, with outlier excluded).
Studies (Figure 7, statistically significant decline in effect size with preregistered .50, with outlier excluded). For female participants, there was a tion and effect size for male participants (Figure 6, 2015). We did not observe a relationship between preregistra-

Design considerations. Next, we considered design factors that might relate to effect size. First, we analyzed preregistration status. Preregistration is thought to help limit research flexibilit-
y that can bias effect size estimates (e.g., Kaplan & Irvin, 2015). We did not observe a relationship between preregistra-
tion control: If the stimuli were presented in a way that con-
trolled viewing parameters within a condition (e.g., as a photograph or on a lab computer).

The red original analysis for men was not statistically significant (Supplemental Figure 13, p = .11) but reached significance with the outlier study excluded (p = .03, $d_{\text{difference}} = 0.38 [0.03, 0.72]$). In this restricted analysis, studies with a well-matched red obtained a medium effect size estimate ($d = 0.53 [0.24, 0.82], k = 7$) while those that did not match or did not specify their red parameters obtained a small effect ($d = 0.15 [0.07, 0.23], k = 37$). The red original analysis was not significant for female participants (Supplemental Figure 14, $p = .57$).

Matching of the control color did not emerge as a predictor of effect size for men (Supplemental Figure 15, $p = .43; p = .28$, with outlier excluded) or women (Supplemental Figure 16, $p = .90$).

Finally, we examined whether maintaining consistent control over presentation parameters might be related to effect sizes. Most studies without presentation control were with online workers, so this analysis mirrors the analysis of participant type reported above. There was not a statistically significant relationship between presentation control and effect size for male participants (Supplemental Figure 17, $p = .26; p = .48$, with outlier excluded). For women, this analysis was not significant but showed a trend for studies with presentation control to obtain larger effect sizes (Supplemental Figure 18, $p = .07, d_{\text{difference}} = 0.24 [-0.02, 0.50]$).

Stimulus differences. We also explored whether differences in the stimuli used might relate to effect size. For both sexes, we did not find evidence for a relationship between effect size and type of photo (head shot, full body, or bust; Supplemental Figures 19 and 20; $p = .32$ for men [$p = .95$ with outlier excluded], $p = .80$ for women), how color was manipulated (border, background, or clothing; Supplemental Figures 21 and 22; $p = .68$ for men [$p = .74$ with outlier excluded], $p = .54$ for women), or the control color used (Supplemental Figures 23 and 24; $p = .95$ for men [$p = .46$ with outlier excluded], $p = .22$ for women).

Young (2015) found that unattractive stimuli do not benefit from a red effect. We analyzed whether baseline attractiveness is related to the red effect (attraction ratings were standardized to a scale from 0% to 100%). For men, this analysis was not

Study design (between or within subjects) was not signifi-
cantly related to effect size for men (Supplemental Figure 11, $p = .58; p = .48$, with outlier excluded) or for women (Supplemental Figure 12, $p = .80$).

Color parameters. Attending to color parameters (or not) could also be a potential factor influencing study results (Elliot & Maier, 2014). To address this, we analyzed three factors: (1) red original: If the red used was specified to be close to the range of values in the study by Elliot and Niesta (2008) or Elliot et al. (2010); (2) control color matched: If the control color used was specified to be well matched to the red used (coded as a no if not close or not specified); and (3) presenta-
tion control: If the stimuli were presented in a way that con-
trolled viewing parameters within a condition (e.g., as a photograph or on a lab computer).

The red original analysis for men was not statistically significant (Supplemental Figure 13, $p = .11$) but reached significance with the outlier study excluded ($p = .03, d_{\text{difference}} = 0.38 [0.03, 0.72]$). In this restricted analysis, studies with a well-matched red obtained a medium effect size estimate ($d = 0.53 [0.24, 0.82], k = 7$) while those that did not match or did not specify their red parameters obtained a small effect ($d = 0.15 [0.07, 0.23], k = 37$). The red original analysis was not significant for female participants (Supplemental Figure 14, $p = .57$).

Matching of the control color did not emerge as a predictor of effect size for men (Supplemental Figure 15, $p = .43; p = .28$, with outlier excluded) or women (Supplemental Figure 16, $p = .90$).

Finally, we examined whether maintaining consistent control over presentation parameters might be related to effect sizes. Most studies without presentation control were with online workers, so this analysis mirrors the analysis of participant type reported above. There was not a statistically significant relationship between presentation control and effect size for male participants (Supplemental Figure 17, $p = .26; p = .48$, with outlier excluded). For women, this analysis was not significant but showed a trend for studies with presentation control to obtain larger effect sizes (Supplemental Figure 18, $p = .07, d_{\text{difference}} = 0.24 [-0.02, 0.50]$).

Stimulus differences. We also explored whether differences in the stimuli used might relate to effect size. For both sexes, we did not find evidence for a relationship between effect size and type of photo (head shot, full body, or bust; Supplemental Figures 19 and 20; $p = .32$ for men [$p = .95$ with outlier excluded], $p = .80$ for women), how color was manipulated (border, background, or clothing; Supplemental Figures 21 and 22; $p = .68$ for men [$p = .74$ with outlier excluded], $p = .54$ for women), or the control color used (Supplemental Figures 23 and 24; $p = .95$ for men [$p = .46$ with outlier excluded], $p = .22$ for women).

Young (2015) found that unattractive stimuli do not benefit from a red effect. We analyzed whether baseline attractiveness is related to the red effect (attraction ratings were standardized to a scale from 0% to 100%). For men, this analysis was not

Study design (between or within subjects) was not signifi-
cantly related to effect size for men (Supplemental Figure 11, $p = .58; p = .48$, with outlier excluded) or for women (Supplemental Figure 12, $p = .80$).

Color parameters. Attending to color parameters (or not) could also be a potential factor influencing study results (Elliot & Maier, 2014). To address this, we analyzed three factors: (1) red original: If the red used was specified to be close to the range of values in the study by Elliot and Niesta (2008) or Elliot et al. (2010); (2) control color matched: If the control color used was specified to be well matched to the red used (coded as a no if not close or not specified); and (3) presenta-
tion control: If the stimuli were presented in a way that con-
trolled viewing parameters within a condition (e.g., as a photograph or on a lab computer).

The red original analysis for men was not statistically significant (Supplemental Figure 13, $p = .11$) but reached significance with the outlier study excluded ($p = .03, d_{\text{difference}} = 0.38 [0.03, 0.72]$). In this restricted analysis, studies with a well-matched red obtained a medium effect size estimate ($d = 0.53 [0.24, 0.82], k = 7$) while those that did not match or did not specify their red parameters obtained a small effect ($d = 0.15 [0.07, 0.23], k = 37$). The red original analysis was not significant for female participants (Supplemental Figure 14, $p = .57$).

Matching of the control color did not emerge as a predictor of effect size for men (Supplemental Figure 15, $p = .43; p = .28$, with outlier excluded) or women (Supplemental Figure 16, $p = .90$).

Finally, we examined whether maintaining consistent control over presentation parameters might be related to effect sizes. Most studies without presentation control were with online workers, so this analysis mirrors the analysis of participant type reported above. There was not a statistically significant relationship between presentation control and effect size for male participants (Supplemental Figure 17, $p = .26; p = .48$, with outlier excluded). For women, this analysis was not significant but showed a trend for studies with presentation control to obtain larger effect sizes (Supplemental Figure 18, $p = .07, d_{\text{difference}} = 0.24 [-0.02, 0.50]$).

Stimulus differences. We also explored whether differences in the stimuli used might relate to effect size. For both sexes, we did not find evidence for a relationship between effect size and type of photo (head shot, full body, or bust; Supplemental Figures 19 and 20; $p = .32$ for men [$p = .95$ with outlier excluded], $p = .80$ for women), how color was manipulated (border, background, or clothing; Supplemental Figures 21 and 22; $p = .68$ for men [$p = .74$ with outlier excluded], $p = .54$ for women), or the control color used (Supplemental Figures 23 and 24; $p = .95$ for men [$p = .46$ with outlier excluded], $p = .22$ for women).

Young (2015) found that unattractive stimuli do not benefit from a red effect. We analyzed whether baseline attractiveness is related to the red effect (attraction ratings were standardized to a scale from 0% to 100%). For men, this analysis was not
**Table 4. Moderator Analysis for Men Rating Women.**

| Moderator                  | Contrasts | \(N_{red}/N_{control}\) | Test Statistic | \(p^2\) 95% CI | Test Statistic\(_{no}\) | \(p_{no}\) 95% CI | Interpretation |
|---------------------------|-----------|---------------------------|----------------|----------------|--------------------------|----------------|---------------|
| Overall meta-analysis     | 45        | 1.725/1,768               | \(F(1, 43) = 3.87\) | .056 88 [79, 79] | \(F(1, 42) = 8.74\) | .005 64 [38, 38] | If outlier excluded, smaller effects over time, unstandardized slope = \(-0.05\) CI \([-0.10, 0.00]\) |
| Year                      | 45        | 1.725/1,768               | \(F(1, 43) = 3.49\) | .069 87 [79, 79] | \(F(1, 41) = 1.03\) | .316 67 [52, 52] | If outlier included, a trend for more attractive stimuli related to smaller effects, unstandardized slope = \(-0.008\) CI \([-0.016, -0.00001]\) |
| Control_attractiveness    | 44        | 1.703/1,724               | \(F(1, 43) = 2.62\) | .113 89 [81, 94] | \(F(1, 42) = 4.84\) | .033 60 [43, 89] | If outlier excluded, studies using a red well matched to the original studies had larger effects, \(d_{difference} = 0.33\) CI \([-0.08, 0.75]\) |
| Eth_majority              | 32        | 1.272/1,267               | \(F(2, 29) = 1.69\) | .203 71 [53, 92] | \(F(2, 29) = 1.69\) | .203 71 [53, 92] | |
| Red_age                   | 38        | 1.582/1,632               | \(F(1, 36) = 1.49\) | .231 86 [75, 75] | \(F(1, 35) = 0.71\) | .405 49 [20, 20] | |
| Presentation_control      | 45        | 1.725/1,768               | \(F(1, 43) = 1.32\) | .257 87 [79, 93] | \(F(1, 42) = 0.51\) | .479 66 [50, 89] | |
| Continent                 | 43        | 1.595/1,622               | \(F(2, 40) = 1.25\) | .299 89 [82, 94] | \(F(2, 39) = 1.79\) | .181 65 [48, 90] | |
| Eth_match                 | 33        | 1.293/1,288               | \(F(1, 31) = 1.07\) | .308 73 [58, 92] | \(F(1, 31) = 1.07\) | .308 73 [58, 92] | |
| Photo_type                | 43        | 1.667/1,713               | \(F(2, 40) = 1.17\) | .320 86 [77, 93] | \(F(2, 39) = 0.05\) | .951 69 [53, 89] | |
| Color_match               | 45        | 1.725/1,768               | \(F(1, 43) = 0.64\) | .430 88 [79, 93] | \(F(1, 42) = 1.21\) | .278 67 [49, 88] | |
| Preregistered             | 45        | 1.725/1,768               | \(F(1, 43) = 0.61\) | .441 87 [79, 93] | \(F(1, 42) = 0.46\) | .503 65 [49, 89] | |
| Design                    | 45        | 1.725/1,768               | \(F(1, 43) = 0.32\) | .575 89 [82, 94] | \(F(1, 42) = 0.51\) | .478 68 [53, 90] | |
| Participants              | 45        | 1.725/1,768               | \(F(2, 42) = 0.52\) | .601 88 [80, 93] | \(F(2, 41) = 0.42\) | .657 66 [51, 89] | |
| Color_form                | 45        | 1.725/1,768               | \(F(2, 42) = 0.39\) | .678 88 [80, 93] | \(F(2, 41) = 0.31\) | .735 69 [52, 89] | |
| Color_contrast            | 43        | 1.657/1,678               | \(F(4, 38) = 0.17\) | .953 85 [75, 92] | \(F(4, 37) = 0.93\) | .455 60 [41, 86] | |

Note. Reported in this table is the name of the moderator, the number of contrasts involved in the analysis, the total sample size in red and control conditions for the analysis, the \(F\) value (with degrees of freedom) for the test of moderation, the resulting \(p\) value, and the \(p^2\) and its 95% CI which indicates the range of variation in effect size remaining after accounting for the moderator. The moderator tests were performed with and without the outlier study by Gilston and Privitera, so results are shown first with the moderator included and then again with the outlier excluded (comment no for no-outlier). Finally, for moderators that were statistically significant in either analysis, a plain-English interpretation is provided; B-S = Between Subjects; W-S = Within Subjects.

Statistically significant (Figure 8, \(p = .07\), unstandardized slope = \(-0.008 [-0.018, 0.0007]\); \(p = .32\), with outlier excluded) but showed a trend in the opposite direction. For women, this analysis also was not statistically significant (Figure 9, \(p = .51\), unstandardized slope = \(-0.008 [-0.016, 0.00]\)) but showed the same trend. Across genders, there may be a weak ceiling effect: More attractive stimuli may elicit somewhat smaller effect sizes. This analysis suffers from restriction of range, as the stimuli in all but a few of the studies fall within the 25–75% attractiveness range.

**Decline effects.** Some scientific phenomena seem to become weaker with repeated tests, a trend known as the “decline effect” (Olfson & Marcus, 2013; Schooler, 2011). To test for a decline effect, we analyzed the relationship between year of publication and effect size (unpublished data sets were coded by the year data collection ended). For men, this analysis was not statistically significant (Figure 10, \(p = .056\), unstandardized slope = \(-0.05 [-0.10, 0.00]\); \(p = .005\), with outlier excluded) but showed a trend for later studies to yield a smaller effect. For women, this analysis was statistically significant (Figure 11, \(p = .005\)), the unstandardized slope was \(-0.07 [-0.12, -0.02]\), which indicates a decline of \(\sim 1/15\)th of an \(SD\) in effect per year (though note this is for a fit to a linear trend and the CI for the slope is very long).

**Publication status.** Finally, we tested whether publication status is related to effect size. For men rating women, the relationship was not statistically significant (\(p = .08\); \(p = .18\), with outlier excluded) but showed a trend for studies from published sources to have larger effect sizes (\(d = 0.38 [0.18, 0.58]\)) than those from unpublished sources (\(d = 0.12 [0.06, 0.18]\)). For women rating men, the relationship also was not statistically significant (\(p = .06\)) but showed a trend for published studies to report larger effects (\(d = 0.27 [0.10, 0.45]\)) than unpublished studies (\(d = 0.04 [-0.09, 0.17]\)).
Manipulations of Facial Redness

We conducted a separate meta-analysis of contrasts that manipulated facial redness (faces with additional red coloring compared to either neutral or less red faces). We did not include studies that asked participants to alter facial color to achieve maximal attractiveness (e.g., Stephen, Law Smith, Stirrat, & Perrett, 2009), nor studies that examined correlations between attraction and facial coloration in unaltered images (e.g., Stephen et al., 2012). In addition, one qualifying study was not analyzed because it did not provide summary statistics by color condition (Fisher, Hahn, Debruijn, & Jones, 2014). We identified four sources containing 10 distinct contrasts (Table 3). Of these, five were for men rating women (614 total participants), one of which was not currently published (N = 101, 16.4% of the total participants). The other five contrasts were for women rating men (717 total participants), one of which was not currently published (N = 197, 27.5% of the total participants). For some of these contrasts, the dependent variable was perceived attractiveness rated on a Likert-type scale; for others, it was attractiveness preference in a forced-choice paradigm (selecting between two stimuli).

For men rating women, the meta-analysis indicates a medium-sized increase in attractiveness with increased facial redness (Figure 12, $d = 0.49$ [0.33, 0.65], $p = .001$, $Q(4) = 8.3$, $p = .08$, $I^2 = 52\%$ [0, 94], $k = 5$, $N = 614$). The CI is short and suggests at least a medium effect and possibly a strong effect. The test for heterogeneity was not statistically significant. For women rating men, the meta-analysis indicates a medium-sized increase in attractiveness with increased facial redness (Figure 13, $d = 0.38$ [0.20, 0.57], $p = .005$, $Q(4) = 12.2$, $p = .02$, $I^2 = 68\%$ [9, 96], $k = 5$, $N = 717$ distinct participants). The CI is relatively short, suggesting at least a small effect but possibly a medium effect. Note, though, that there is considerable heterogeneity of effect size.

Red Effects for Rating Waiters and Waitresses

We conducted a separate meta-analysis of contrasts examining the effect of red on the perceived attractiveness of waiters and waitresses (Lynn et al., 2016). In these studies, online participants viewed photos of a restaurant server. Participants were asked to imagine that the person being viewed had provided “good but not exceptional service.” Participants estimated the tip they would leave and then rated the server’s attractiveness. Across the two sources (Lynn & Giebelhausen, 2013; Lynn et al., 2016), there were seven distinct contrasts for men rating women (1,176 distinct participants, Table 6), only one of which was published (N = 220; Lynn et al., 2016).

Meta-analysis of these studies suggests that for men rating waitresses, there is not a statistically significant effect of red on perceived attractiveness ($d = -0.16$ [-0.33, 0.01], $p = .06$, $Q(6) = 8.2$, $p = .22$, $F = 24\%$ [0, 87], $k = 7$), but there is a trend indicating a tendency toward a weak repulsive effect. For
| Contrast                    | Year | Preregister? | Color | Form | Photo Type | Continent | Participant Type | Study Design | Ethnic Majority Match | Ethnicity | Red | Original | Matched | Presentation Control | Baseline Attractiveness (%) | N<sub>red</sub> | M<sub>red</sub> | S<sub>red</sub> | N<sub>control</sub> | M<sub>control</sub> | S<sub>control</sub> | d<sub>95% CI</sub> |
|-----------------------------|------|--------------|-------|------|------------|-----------|------------------|-------------|----------------------|-----------|-----|----------|--------|----------------------|--------------------------|---------------|----------------|----------------|---------------------|---------------------|-------------------|--------------------|
| Women rating men            |      |              |       |      |            |           |                  |             |                      |           |     |          |        |                     |                          |               |                |                |                     |                     |                   |        |
| Lynn et al. (2016)—Exp 1—  | 2016 | Y            | White | Clothing | Bust    | Online   | workers         | B-S White   | Matched              | —         | N   | N        | N      | matched             |                          | 61             | 72              | 5.42           | 2.33                | 77                  | 6.49              | 2.03               | -0.49 [-0.82, -0.16] |
| Exp 1—Journal of Hospitality & Tourism Research |      |              |       |      |            |           |                  |             |                      |           |     |          |        |                     |                          |               |                |                |                     |                     |                   |        |
| Men rating women            |      |              |       |      |            |           |                  |             |                      |           |     |          |        |                     |                          |               |                |                |                     |                     |                   |        |
| Lynn and Giebelhausen (2013)—Exp 1—Unpublished manuscript | 2013 | N            | White | Clothing | Full body | Online   | workers         | B-S White   | Mismatch              | 28.08     | N   | N        | N      | matched             |                          | 74             | 62              | 4.82           | 1.20                | 64                  | 5.42              | 1.04               | -0.53 [-0.89, -0.18] |
| Lynn and Giebelhausen (2013)—Exp 2—Unpublished manuscript | 2013 | N            | White | Clothing | Full body | Online   | workers         | B-S White   | Mismatch              | 28.08     | N   | N        | N      | matched             |                          | 81             | 67              | 5.81           | 1.05                | 61                  | 5.87              | 0.87               | -0.06 [-0.41, 0.29]  |
| Lynn and Giebelhausen (2013)—Exp 3—Unpublished manuscript | 2013 | N            | White | Clothing | Bust    | Online   | workers         | B-S White   | Mismatch              | 30.57     | N   | N        | N      | matched             |                          | 82             | 59              | 5.56           | 1.19                | 54                  | 5.89              | 0.88               | -0.31 [-0.68, 0.06]  |
| Lynn and Giebelhausen (2013)—Exp 4—Unpublished manuscript | 2013 | N            | White | Clothing | Bust    | Online   | workers         | B-S White   | Mismatch              | 28.57     | N   | N        | N      | matched             |                          | 66             | 171             | 4.81           | 0.99                | 156                 | 4.96              | 1.01               | -0.15 [-0.37, 0.07]  |
| Lynn and Giebelhausen (2013)—Exp 5—Unpublished manuscript | 2013 | N            | White | Clothing | Full body | Online   | workers         | B-S White   | Mismatch              | 28.3      | N   | N        | N      | matched             |                          | 68             | 102             | 5.01           | 1.09                | 86                  | 5.09              | 1.44               | -0.06 [-0.35, 0.22]  |
| Lynn and Giebelhausen (2013)—Exp 6—Unpublished manuscript | 2013 | N            | White | Clothing | Full body | Online   | workers         | B-S White   | Mismatch              | 29.4      | N   | N        | N      | matched             |                          | 79             | 114             | 5.54           | 0.95                | 97                  | 5.76              | 1.14               | -0.21 [-0.48, 0.06]  |
| Lynn et al. (2016)—Exp 1—Journal of Hospitality & Tourism Research | 2016 | Y            | White | Clothing | Bust    | Online   | workers         | B-S White   | Mismatch              | —         | N   | N        | N      | matched             |                          | 60             | 110             | 6.43           | 1.78                | 110                 | 6.36              | 1.62               | 0.04 [-0.22, 0.31]   |
| Overall                     |      |              |       |      |            |           |                  |             |                      |           |     |          |        |                     |                          |               |                |                |                     |                     |                   | -0.16 [-0.28, -0.04] |

Note. Columns 2–17 provide moderator classifications using the same coding scheme as the main meta-analysis; details in the text. Column 17 (DV) indicates if the dependent variable was perceived attractiveness of a forced choice between a more and a less red face. Subsequent columns provide the basic descriptive statistics for each contrast: sample size (N), mean (M), and standard deviation (S) for the red group and the control group. For within-subjects designs, the correlation between repeated measures is provided (r). Finally, for each contrast, Cohen's d and its 95% confidence interval is reported with d calculated so that positive values indicate higher attraction in the red condition. When a moderator could not be coded, the cell is marked with an emdash (—). Y = yes; N = no; Exp = experiment; B-S = Between Subjects; W-S = Within Subjects.
women rating waiters, there is one contrast, from a published report by Lynn and Giebelhausen (2013). As reported in that manuscript, waiters dressed in red were rated as significantly less attractive than those dressed in white ($d = -0.48 [-0.82, -0.16], p = .003, N = 149$).

**Discussion**

Both color-in-context theory and the red-romance hypothesis are supported by other data; this is not a definitive assessment of either. Still, our meta-analyses help synthesize the evidence that red enhances perceived attractiveness in romantic contexts.

For men rating women, we obtained an overall effect size in the predicted direction that is statistically significant ($d = 0.26/0.19$ with/without an outlier study). This effect size is small, with the CI suggesting anything from a very small effect ($d = 0.12/0.09$) to a medium effect ($d = 0.40/0.29$). For women rating men, the overall effect size was also in the predicted direction and statistically significant ($d = 0.13$). This effect size is very small, with the CI suggesting anything from an intrattractably small effect ($d = 0.01$) to a small effect ($d = 0.25$).
Interpreting these estimates requires consideration of possible bias in the studies we obtained. We evaluated this issue by examining the relationship between effect size and standard error. For men rating women, the relationship was not statistically significant in the overall data set (nor with an effect-size outlier excluded) but was with a sample-size outlier excluded (with or without an effect-size outlier excluded). For women rating men, there was a statistically significant relationship. Tests for bias are difficult to interpret. On the one hand, they can miss bias due to low power, especially when there are less than 60 studies available for analysis (Macaskill, Walter, & Irwig, 2001). On the other hand, false positives are also possible because a relationship between effect size and sampling error can occur for reasons other than bias (Egger, Smith, Schneider, & Minder, 1997), especially if there are systematic differences between the small and large studies (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). Out of an abundance of caution, it would probably be wise to consider the effect size estimates we obtained at least somewhat optimistic, especially for women rating men.

The effect size estimates we obtained must also be qualified by the aforementioned heterogeneity that was substantial and apparent for participants from both sexes. We used moderator analysis to try to identify possible sources for this heterogeneity. For men rating women, only two statistically significant moderators were detected, but each relationship was only present with an effect-size outlier omitted: (1) Studies published more recently tended to have smaller effects \( (p = .056/.005 \text{ with outlier included/excluded}) \) and (2) studies using an original red tended to have larger effects \( (p = .11/.03) \). There was also a nonsignificant trend for a weak baseline attractiveness ceiling effect \( (p = .07/.32) \). For women rating men, only two statistically significant relationships were detected: (1) Studies published more recently had smaller effects and (2) preregistered studies had smaller effects. There were also nonsignificant trends for a weak baseline attractiveness ceiling effect, and
for stronger effects for in-person student participants, studies with greater presentation control, and published studies. In general, moderator analyses were underpowered. This means nonsignificant findings did not rule out the possibility of moderation and even significant findings left uncertainty about the magnitude of moderation.

In a supplementary analysis, facial color studies showed that men and women find photos more attractive when red is added to the target’s face (relative to having red removed and to unaltered photos). This finding is robust: Red enhances attraction when it is intrinsic to the stimulus. This pattern was also reported in an experiment that we were unable to incorporate into the meta-analysis (Fisher et al., 2014) and in additional experiments published after our initial literature review (Han et al., 2018; Kandrik et al., 2017). Moreover, several studies have indicated that participants asked to adjust photos to maximize attractiveness increase facial redness (e.g., Stephen et al., 2009). It should be noted, though, that the factors influencing facial attractiveness are complex. Yellowness also seems to enhance attraction for Caucasian participants (e.g., Fisher et al., 2014), there may be cultural differences in
preference for color cues (Han et al., 2018), and in unmanipulated images, the degree of facial redness does not always correlate with attractiveness ratings (e.g., Stephen et al., 2012).

Studies in which participants estimate a tip and rate the attractiveness of a waiter or waitress did not show an effect of red on perceived attractiveness. There was a nonsignificant trend for men to rate waitresses wearing red as less attractive. This suggests a way to test the prediction from color-in-context theory that context determines the direction of the red effect. This could be done by directly comparing the effect of red on attraction to neutral targets and waiters/waitresses.

Although we were able to agree on how to conduct and report these meta-analyses, we could not come to a consensus on interpretation. Therefore, the next two sections present our contrasting interpretations.

**First and Third Authors: The True Effect Is Very Small, Potentially Nonexistent**

The simplest conclusion from our results is that the true effect of incidental red on attraction is very small, potentially nonexistent. The meta-analytic results suggest small effects. These may be overestimates, given that we found some evidence of bias, especially for female participants. Moreover, preregistered studies, which one would expect to be most accurate, have overall found effects too small to be of practical significance ($d = 0.10$ for men; $d = -0.10$ for women; these are equivalent to $r^2 = .002$). Finally, the decline effect observed is an indication that the original research substantially overestimated the true effect.

We cannot rule out the possibility that there are strong moderators which could unlock robust effects of intrinsic red on attraction. In particular, our moderator analysis shows that larger effects have been obtained when stimulus presentation is carefully controlled and when the red used is carefully matched to what was used in the original studies. But if these moderators are powerful, that would mean that red enhances attraction only under a very narrow range of conditions. That would be puzzling. First, it would raise the question of how the original research, which used a number of different shades of red, always selected an effective color. Second, it would raise the question of how the same researchers then went on to document strong red effects with other stimuli that were not well matched to the original red (e.g., Elliot et al., 2013; Pazda, Elliot, & Greitemeyer, 2012; Pazda, Prokop, & Elliot, 2014).

![Figure 12. Overall meta-analysis of the effect of facial redness on perceived attractiveness (measured via Likert-type scale and forced choice paradigm) for men looking at photos of women. Each contrast is shown with its effect sizes and 95% confidence interval (CI). The diamond shows the overall estimated effect size and its 95% CI using a random-effects model ($d = 0.49 \ [0.33, 0.65], p = .001, N = 614$ distinct participants). Contrasts are shown in chronological order. Those marked with a † were unpublished; for these contrasts, the year shown is when data collection ended.](image1)

![Figure 13. Overall meta-analysis of the effect of facial redness on perceived attractiveness (measured via Likert-type scale and forced choice paradigm) for women looking at photos of men. Each contrast is shown with its effect sizes and 95% confidence interval (CI). The diamond shows the overall estimated effect size and its 95% CI using a random-effects model ($d = 0.38 \ [0.20, 0.57], p = .005, N = 717$ distinct participants). Contrasts are shown in chronological order. Those marked with a † were unpublished; for these contrasts, the year shown is when data collection ended.](image2)
and under a range of different conditions (e.g., Elliot et al., 2013). Third, this explanation makes it difficult to understand how recent replications that used careful color matching with in-person participants found little-to-no effect of red on attraction for both male (Lehmann & Calin-Jageman, 2017) and female participants (the currently unpublished studies from the Collaborative Replications and Education Project (CREP) project [osf.io/wfc6u]; Grahe et al., 2012). Finally, it is difficult to conceive why the expression of red-romance links would depend on highly specific shades of red when the evolutionary and cultural histories thought to underlie this association were probably not highly specific (red body coloration is not highly consistent within primate species and Valentines don’t come in only one shade of red). The most parsimonious explanation of the meta-analysis results is that intrinsic red has only a very small effect on self-reported perceived attractiveness, perhaps too small to be of any practical significance. We happily concede, though, that manipulations of facial redness are robust and meaningful; it will be interesting to determine whether they are specific to red and/or sensitive to context as predicted by the color-in-context theory.

Second Author: The Findings for Male Perceivers Are Clearer Than Those for Female Perceivers, and Methodological Improvement Is Needed

For both men rating women and women rating men, the overall analysis indicates a positive effect of red on perceived attractiveness, but these two effects are best interpreted separately. For men rating women, there is clear evidence of a small, statistically reliable effect. There is also clear evidence that the effect size varies across studies. This is unsurprising given that the meta-analysis collapses across studies that substantively differ in terms of participant characteristics, protocol, color stimuli, rigor, and so on; it points to the possibility of a reliably strong effect under some conditions and a reliably weak or nonexistent effect under others. Moderator analyses did not show any consistent effects (i.e., same results with/without outlier), but these analyses themselves collapsed across many substantive differences and often were unable to precisely represent the focal moderator (especially for color-based variables). Subsequent work would do well to identify and manipulate potentially fruitful moderators (e.g., Young, 2015), including some not considered herein (e.g., variables derived from ecological valence theory, such as school color, season, or size of color stimulus; Schloss & Heck, 2017; Schloss & Palmer, 2017). For women rating men, there is evidence of a very small, statistically reliable effect. There is also clear evidence that the effect size varies across studies and that the effect size might be an overestimate. Significant moderation by publication year and preregistration status may also suggest an overestimated effect. Thus, overall, the existing evidence seems weaker for women rating men than for men rating women. In subsequent work, one variable that could be manipulated to provide a more sensitive test of the effect for women rating men is the type of red presented. Light, pinkish red may be associated with femininity and decrease attractiveness; intense, vivid red may be linked to aggressiveness and decrease attractiveness; while more prototypic red may be associated with health/status and increase attractiveness (Elliot & Maier, 2014). Such differential associations within the same hue may be grounded in evolutionary processes, social learning processes, or both (see Changizi, 2010; Higham & Winters, 2015).

A meta-analysis is only as strong as the methodological quality of the studies that contribute to it, and it is important to acknowledge that the quality of the studies in the red-romance literature (like the color psychology literature more broadly) is not strong. Two primary weaknesses are that nearly all existing studies are underpowered and fail to attend to important color science procedures, especially regarding color production (e.g., spectral assessment, matching color attributes) and presentation (e.g., ambient illumination, background contrast; Elliot, 2015; Fairchild, 2015). Indeed, not a single published study that contributed to our main meta-analysis would be considered exemplary based on these two criteria alone. As such, the information provided by the current meta-analysis must be considered tentative. What is needed is empirical work that carefully attends to sample size and the complexity inherent in color science methodology (see Elliot, 2015). Additionally attending to the other aspects of the evidentiary value movement (Finkel, Eastwick, & Reis, 2015), including keeping clear, comprehensive records of studies conducted (via public preregistration and personal archive), will further bolster the quality of downstream meta-analytic work.

Suggestions for Future Research

Although we differ in our interpretations of this meta-analysis, we agree that it offers some important information for planning subsequent research on the red-romance hypothesis:

- The overall effect size estimates for men rating women and women rating men are not representative due to heterogeneity but probably provide reasonable inputs for sample size planning. These effect size estimates indicate that large samples are needed in between-subjects designs: 234 per group for men rating women (or 436 using the estimate that excludes an outlier) and 930 per group for women rating men. Within-subjects designs can reduce sample size requirements considerably.
- Future research will be most convincing if publicly preregistered.
- A surprising quantity of data we obtained is not published and there were nonsignificant trends suggesting that publication status may be related to effect size. An accurate literature requires all quality experiments to be published.
- Even though the trends were not statistically significant, studies that have online workers and/or a lack of control over presentation parameters tend to find
smaller effects for female participants. It would be best to use online studies sparingly (e.g., in preliminary fashion) and to attend carefully to issues regarding stimulus presentation.

- The online studies collected here involved 3,381 distinct participants (across all meta-analyses). Many of these studies were conducted on Mechanical Turk. Future studies with this participant pool should take steps to ensure naivete to the research question.

- Other dependent variables may be promising (e.g., sexual attraction). We did not examine other dependent variables here, but note that some relatively well-powered replication studies have indicated little to no effect of red on other dependent variables (Lehmann & Calin-Jageman, 2017; Peperkoorn, Roberts, & Pollet, 2016), though some replications have found effects (Schwarz & Singer, 2013).

- We observed some indication of possible ceiling effects with very attractive stimuli. It is important to calibrate stimuli for maximal sensitivity for the research question. This does not always mean using original stimuli. For example, two participants in a recent replication study described the stimulus materials used in the original work as “out of date” (Lehmann & Calin-Jageman, 2017).

- Pazda, Thorstenson, and colleagues have started using forced-choice paradigms and have focused on attractiveness preferences for degrees of facial redness (Pazda, Thorstenson, & Elliot, 2016; Thorstenson et al., 2016a). This approach seems fruitful, especially if specificity for red can be demonstrated.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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

ORCID iD

Robert J. Calin-Jageman  http://orcid.org/0000-0002-9837-6529

Supplemental Material

Supplemental material for this article is available online.

References

Ahnelt, P. K., & Kolb, H. (2000). The mammalian photoreceptor mosaic adaptive design. Progress in Retinal and Eye Research, 19, 711.

Banas, K. (2014). Replication of Elliot et al. (2010) for CREP at the University of Edinburgh. Retrieved from osf.io/cvdwp

Berthold, A. (2013). Unpublished data. Retrieved from https://osf.io/dmj4s/

Bigelow, M. G., Taylor, G., & Underwood, M. (2013). Context-moderated effect of color on physiological and self-report measures of emotional response. UNC Asheville Journal, Undergraduate Research Program, Asheville, NC.

Blech, C. (2014). Replication of Elliot et al. (2010). Red, rank, and romance in women viewing men. Retrieved from osf.io/t2u5

Blech, C. (2015). Unpublished data from a class experiment. Retrieved from https://osf.io/dmj4s/

Boelk, K., & Madden, W. (2014). Fork of Elliot, A.J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R.H., Maier, M.A., & Lie, H. (2010). Retrieved from osf.io/ztfe9

Brewer, A. A., & Barto, B. (2012). Visual field map organization in human visual cortex. In Visual cortex—Current status and perspectives. InTech. doi:10.5772/51914

Buechner, V. L., Maier, M. A., Lichtenfeld, S., & Elliot, A. J. (2015). Emotion expression and color: Their joint influence on perceived attractiveness and social position. Current Psychology, 34, 422–433. doi:10.1007/s12144-014-9266-x

Buzuis, E. (2007). Reading the pink locker room: On football culture and Title IX. William and Mary Journal of Women and the Law, 14, 1–53.

Changizi, M. (2010). The vision revolution: How the latest research overturns everything we thought we knew about human vision. Dallas, TX: BenBalla Books.

Costello, J., Groeneboom, L., & Pollet, T. (2017). Romantic red: Do red products enhance the attractiveness of the consumer? (Unpublished masters degree). University of Leiden, Leiden, the Netherlands.

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. doi:10.1016/j.yhbeh.2004.03.013

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ, 315, 629–634. doi:10.1136/bmj.315.7109.629

Elliot, A. J. (2015). Color and psychological functioning: A review of theoretical and empirical work. Frontiers in Psychology, 6. doi:10.3389/fpsyg.2015.00368

Elliot, A. J., Kayser, D. N., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., . . . Liu, H. (2010). Red, rank, and romance in women viewing men. Journal of Experimental Psychology: General, 139, 399–417. doi:10.1037/a0019689

Elliot, A. J., & Maier, M. A. (2012). Color-in-context theory. Advances in Experimental Social Psychology, 45, 61–125. doi:10.1016/B978-0-12-394286-9.00002-0

Elliot, A. J., & Maier, M. A. (2013). The red-attractiveness effect, applying the Ioannidis and Trikalinos (2007b) test, and the broader scientific context: A reply to Francis (2013). Journal of Experimental Psychology General, 142, 297–300. doi:10.1037/a0029592

Elliot, A. J., & Maier, M. A. (2014). Color psychology: Effects of perceiving color on psychological functioning in humans. Annual Review of Psychology, 65, 95–120. doi:10.1146/annurev-psych-010213-115035

Elliot, A. J., & Niesta, D. (2008). Romantic red: Red enhances men’s attraction to women. Journal of Personality and Social Psychology, 95, 1150–1164. doi:10.1037/0022-3514.95.5.1150
Lynn, M., Giebelhausen, M., Garcia, S., Li, Y., & Patumanon, I. (2016). Clothing color and tipping. *Journal of Hospitality & Tourism Research, 40*, 516–524. doi:10.1177/1096348013504001

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine, 20*, 641–654. doi:10.1002/sim.698

Maves, M., & Nadler, J. T. (2016). Data and analysis. Retrieved from osf.io/99mbv8

Meier, B. P., D’Agostino, P. R., Elliot, A. J., Maier, M. A., & Wilkowski, B. M. (2012). Color in context: Psychological context moderates the influence of red on approach- and avoidance-motivated behavior. *PLoS One, 7*, e40333. doi:10.1371/journal.pone.0040333

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105–125. doi:10.1037/1082-989X.7.1.105

Nathans, J. (1999). The evolution and physiology of human review color vision: Insights from molecular genetic studies of visual pigments. *Neuron, 24*, 299–312. doi:10.1016/S0896-6273(00)80845-4

Olsson, M., & Marcus, S. C. (2013). Decline in placebo-controlled trial results suggests new directions for comparative effectiveness research. *Health Affairs, 32*, 1116–1125. doi:10.1377/hlthaff.2012.1353

O’Mara, E. M., Kershaw, C., Receveur, A., Hunt, C., Askar, S., Ballas, T., ... Wilhoit, S. (2016). Unpublished data. Retrieved from https://osf.io/ecb8s/

O’Mara, E. M., & Trujillo, A. (2015). Unpublished data. Retrieved from https://osf.io/dmj4s/

Pazda, A. D., Elliot, A. J., & Greitemeyer, T. (2012). Sexy red: Perceived sexual receptivity mediates the red-attraction relation in men viewing woman. *Journal of Experimental Social Psychology, 48*, 787–790. doi:10.1016/j.jesp.2011.12.009

Pazda, A. D., Elliot, A. J., & Greitemeyer, T. (2014). Perceived sexual receptivity and feminaheness: Separate paths linking red and black to perceived attractiveness. *Color Research & Application, 39*, 208–212. doi:10.1002/col.21804

Pazda, A. D., Prokop, P., & Elliot, A. J. (2014). Red and romantic rivalry. *Personality and Social Psychology Bulletin, 40*, 1260–1269. doi:10.1177/0146167214539709

Pazda, A. D., Thorstenson, C. A., & Elliot, A. J. (2016a). Unpublished data. Retrieved from https://osf.io/dgjx2/

Pazda, A. D., Thorstenson, C. A., Elliot, A. J., & Perrett, D. I. (2016b). Unpublished data. Retrieved from https://osf.io/gwtp6/

Pazda, A. D., Thorstenson, C. A., & Elliot, A. J. (2017). Unpublished data—Lingerie studies. Retrieved from https://osf.io/dmj4s/

Pazda, A. D., Thorstenson, C. A., Elliot, A. J., & Perrett, D. I. (2016). Women’s facial redness increases their perceived attractiveness: Mediation through perceived healthiness. *Perception, 45*, 739–754. doi:10.1177/0301006616633386

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, 14*, 1–13. doi:10.1177/1474704916673841

Pollet, T. (2013). Unpublished data. Retrieved from https://osf.io/dmj4s/

Purdy, M. A. (2009). The influence of the amygdala and color on judgments of attractiveness. In UNC Asheville Journal, Undergraduate Research Program, Asheville, NC.

Roberts, S. C., Owen, O., & Havlicek, J. (2010). Distinguishing between perceiver and wearer effects in clothing color-associated attributions. *Evolutionary Psychology, 8*, 350–364. doi:10.1177/147470491000800304

Schauss, A. G. (1979). Tranquilizing effect of color reduces aggressive behavior and potential violence. *Journal of Orthomolecular Psychiatry, 8*, 218–221.

Schloss, K. B., & Heck, I. A. (2017). Seasonal changes in color preferences are linked to variations in environmental colors: A longitudinal study of fall. *I-Perception, 8*, 204166951774217. doi:10.1177/2041669517742177

Schloss, K. B., & Palmer, S. E. (2017). An ecological framework for temporal and individual differences in color preferences. *Vision Research, 141*, 95–108. doi:10.1016/j.visres.2017.01.010

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature, 470*, 437–437. doi:10.1038/470437a

Schwarz, S., & Singer, M. (2013). Romantic red revisited: Red enhances men’s attraction to young, but not menopausal women. *Journal of Experimental Social Psychology, 49*, 161–164. doi:10.1016/j.jesp.2012.08.004

Seibt, T. (2015). Romantic red effect in the attractiveness perception. In Hassaac (pp. 31–34). doi:10.18638/hassacc.2015.3.1.186

Seibt, T., & Klement, V. (2015). The impact of the colour red on attractiveness perception. In *4th Advanced Research in Scientific Areas* (pp. 20–24). doi:10.18638/arsa.2015.4.1.799

Smith, J. M., Bell, P. A., & Fusco, M. E. (1986). The influence of color and demand characteristics on muscle strength and affective ratings of the environment. *The Journal of General Psychology, 113*, 289–297. doi:10.1080/00221309.1986.9711040

Stefan, J., & Gueguen, N. (2013). Unpublished data. Retrieved from https://osf.io/dmj4s/

Stephen, I. D., Law Smith, M. J., Stirrat, M. R., & Perrett, D. I. (2009). Facial skin coloration affects perceived health of human faces. *International Journal of Primatology, 30*, 845–857. doi:10.1007/s10764-009-9380-z

Stephen, I. D., Scott, I. M. L., Coetzee, V., Pound, N., Perrett, D. I., & Penton-Voak, I. S. (2012). Cross-cultural effects of color, but not morphological masculinity, on perceived attractiveness of men’s faces. *Evolution and Human Behavior, 33*, 260–267. doi:10.1016/j.evolhumbehav.2011.10.003

Sullivan, J., Lavoie, E. A., Bays, R. B., Fontana, S., Goodkind, R., Johnson, R., ... Lavoie, M. (2016). Replication of Elliot et al., 2010: Red, Rank, and Romance. Retrieved from osf.io/pm7fx

Thorstenson, C. A., Pazda, A. D., Elliot, A. J., & Perrett, D. I. (2016a). Facial redness increases men’s perceived healthiness and attractiveness. *Perception, 46*, 650–664. doi:10.1177/030073221661680124
Thorstenson, C. A., Pazda, A. D., Elliot, A. J., & Perrett, D. I. (2016b). Unpublished data. Retrieved from https://osf.io/dmj4s/

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–43. doi: 10.18637/jss.v036.i03

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron, 56*, 366–383. doi:10.1016/j.neuron.2007.10.012

Wartenberg, W., Höpfner, T., Potthast, P., & Mirau, A. (2011). If you wear red on a date, you will please your mate. In *Proceedings of Empiriepraktikumskongress* (pp. 26–27). Jena, Germany: University of Jena.

Wen, F., Zuo, B., Wu, Y., Sun, S., & Liu, K. (2014). Red is romantic, but only for feminine females: Sexual dimorphism moderates red effect on sexual attraction. *Evolutionary Psychology, 12*, 719–735. doi:10.1177/147470491401200404

Williams, C. L., & Neelon, M. (2013). Conditional beauty: The impact of emotionally linked images on the red effect in sexual attraction. *Psi Chi Journal of Psychological Research, 18*, 10–19.

Young, S. G. (2015). The effect of red on male perceptions of female attractiveness: Moderation by baseline attractiveness of female faces. *European Journal of Social Psychology, 45*, 146–151. doi: 10.1002/ejsp.2098