THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A demonstration of the collaborative replication and education project

OPEN ACCESS

ORIGINAL RESEARCH REPORT

A Demonstration of the Collaborative Replication and Education Project: Replication Attempts of the Red-Romance Effect

Jordan R. Wagge[*], Cristina Baciu[†], Kasia Banas[‡], Joel T. Nadler[§], Sascha Schwarz[‖], Yanna Weisberg[¶], Hans IJzerman[**], Nicole Legate[†] and Jon Grahe[††]

The present article reports the results of a meta-analysis of nine student replication projects of Elliot et al.'s (2010) findings from Experiment 3, that women were more attracted to photographs of men with red borders (total n = 640). The eight student projects were part of the Collaborative Replication and Education Project (CREP; https://osf.io/wfc6u/), a research crowdsourcing project for undergraduate students. All replications were reviewed by experts to ensure high quality data, and were pre-registered prior to data collection. Results of this meta-analysis showed no effect of red on attractiveness ratings for either perceived attractiveness (mean ratings difference = –0.07, 95% CI [–0.31, 0.16]) or sexual attractiveness (mean ratings difference = –0.06, 95% CI [–0.36, 0.24]); this null result held with and without Elliot et al.'s (2010) data included in analyses. Exploratory analyses examining whether being in a relationship moderated the effect of color on attractiveness ratings also produced null results.

The Collaborative Replications and Education Project: Conducting High-Quality Undergraduate Research

The Collaborative Replications and Education Project (CREP; osf.io/wfc6u) was created to address the need for high-quality direct replications in the field of psychology while training students in psychology courses who complete research projects. The purpose of the CREP more generally is to encourage students and instructors to conduct replications; the resulting data from these projects is crowdsourced into a meta-analysis (such as the present publication). Candidate studies for CREP replications are selected by first identifying the top journals in 9 subdisciplines of psychology, and then identifying the top cited empirical studies from one calendar year. From the studies culled from this process, the CREP advisory team identifies studies that are most feasible for undergraduates to replicate. In selecting

papers for feasibility, practical concerns were considered (e.g. availability of required technology, duration of the study, nonclinical adult populations). The study replicated here, Elliot et al. (2010), was in the list of the top five studies chosen for feasibility and impact (measured by number of subsequent citations following publication) from 2010. Experiment 3 — discussed below — was selected as the most feasible out of the seven studies in the article.

An individual CREP project begins when a group of students, under the advisement of a faculty member, selects one study to replicate from our pre-selected set of studies. The students then prepare and upload to the Open Science Framework (OSF) all related materials and methods for the project, including a videotaped live demonstration of the methodology. Once the proposed replication has been reviewed by an editor and two expert reviewers and IRB approval has been uploaded, the students make necessary revisions, pre-register their project on the OSF, and begin collecting data. When data collection is complete and the students have uploaded raw data and their results, the project is given a final review and, if accepted, the students earn a certificate of completion.

The broader goal of the CREP is to collect enough data across groups so that at least 2.5 times the number of participants are collected in total as compared to the original study (a blanket recommendation suggested by Simonsohn, 2015). Most individual projects are therefore asked to collect data from at least the number

* Avila University, Kansas City, MO, US

† Illinois Institute of Technology, Chicago, IL, US

‡ The University of Edinburgh, Edinburgh, Scotland, UK

§ Southern Illinois University – Edwardsville, Edwardsville, IL, US

‖ Bergische Universität Wuppertal, Wuppertal, DE

¶ Linfield College, McMinnville, OR, US

** Université Grenoble Alpes, Grenoble, FR

†† Pacific Lutheran University, Tacoma, WA, US

Corresponding author: Jordan R. Wagge (Jordan.Wagge@avila.edu)

of participants in the original study to be approved at final review.[1] Once enough data has been collected across multiple sites, a meta-analysis is performed on the data.[2]

The CREP process ensures not only fidelity, but also high quality of replications. Replications completed by student groups are as loyal as possible to the original procedures, as original authors are contacted prior to conducting the study by the CREP board (see also Brandt et al., 2014, for recommendations on how to do replications). The goal of the current CREP project was to determine the robustness of the red-as-romance effect and thereby contribute to estimating the effect size as accurately as possible.

Red, Romance, and Replication
Do women find photographs of men with red borders more attractive? This is what Elliot et al. (2010) tried to answer. In their paper, they present data suggesting that heterosexual women find men more attractive when presented with a red border and they conclude that this association is specific to sexual and physical attraction rather than overall likeability. Specifically, in Experiment 3 of their paper, participants (all heterosexual females) rated a picture of a "moderately attractive young Latino man" (p. 405) on attractiveness, sexual attractiveness, and likeability while the surrounding color of the picture was manipulated (either red or gray). Participants did not differ in their estimates of overall likeability of the man, but those assigned to the "red background" condition rated the man higher on perceived attractiveness and sexual attractiveness. The effect sizes were $d = 0.86$ and $d = 0.85$, respectively, which are typically considered large effects.

This paper is well-cited, but some (e.g. Francis, 2013) have questioned whether the effect might be a result of publication bias. The meta-analysis presented in this paper summarizes CREP projects to replicate Elliot et al. (2010, Exp. 3) across different labs, thereby contributing data to help determine whether the effect size is statistically different from zero.

Methods
CREP Procedures
In the case of the Elliot et al. (2010) Experiment 3 replications presented here, the CREP board first contacted the original first author who provided information and materials; the materials provided were recreations of the photographs used in the original 2010 study. The original photograph parameters for the red and grey photos were, respectively, LCh (50.0, 59.6, 31.3) and (50.0, –, 69.1). Dr. Elliot sent us photographs from a subsequent replication of the red/gray experiment, and reported LCh values of (44.0, 49.3, 18.2) and (44.0, –, 293.2). Because small differences in spectrophotometer calibration and adjustment can create big differences, a CREP board member had both the red and grey materials assessed using the same spectrophotometer run by the same person, in the same conditions. This color expert found only very small differences between the pictures sent by Dr. Elliot (red LCh[57.7, 63.3, 29.3], grey LCh[54,–0.1, 1.2]) and recreations printed by our team (red LCh[55.8, 65, 26.8], grey LCh[52.5, –0.3, 1.2]). Both sets were used in subsequent replications.

CREP pages for each individual replication can be found here: https://osf.io/flaue/. Once a team's project had been approved for data collection by the review team, the students pre-registered their OSF page and began collecting data, notifying the Director when data collection was complete. The OSF page was again reviewed by a review team and, if the project met CREP requirements for completion (a completion pledge, shared data, reported results with an $n \geq$ the original study) the students were provided with a certificate. In this early phase of the CREP, students also received a monetary reward of $300 upon completion.

Red and Romance
For this meta-analysis we included all high-fidelity studies with available data (and included in a footnote where we only had access to the summary results, i.e. Frazier, 2014) that were completed prior to 13 Nov 2015, with one exception. The first author of this paper became part of the CREP board. To familiarize herself with the process, she collected data in late 2017.

For the purposes of this research, we were interested in the overall effect of the red or gray background.[3] We thus included all replications that were publicly posted as part of the Collaborative Replication and Education Project (Frazier, 2013; Schwarz, 2013; Banas, 2014; Boelk & Madden, 2014; Johnson, Meltzer, & Grahe, 2016; Legate et al., 2015; Maves & Nadler, 2015; Khislavsky, 2016; Wagge et al., 2017). Despite their general similarity to the original Elliot study and to each other, the minor differences in execution, location, and methodology that emerged across the various labs are described in detail later in this section. However, descriptions of the methodology, copies of materials, videos of procedures, and descriptions of data analysis for each study can be found on each OSF project page. Several projects were not pre-registered (Schwarz, 2013; Johnson et al., 2016; Khislavsky, 2016; Maves & Nadler, 2015); however, given that these teams could not collect data until receiving the photographs in the mail and by that point had already submitted their project for review, we supported the inclusion of these data in our analysis.

Participants
In all studies, graduate (Wagge et al., 2017) and undergraduate (all other replications) student researchers invited adult women to participate in their individual studies at their home universities. Seven of the replications were conducted within the continental United States (Boelk & Madden, 2014, $n = 72$; Johnson et al., 2016, $n = 73$; Legate et al., 2015, $n = 50$; Frazier, 2014,[4] $n = 59$; Maves & Nadler, 2015, $n = 130$; Khislavsky, 2015, $n = 187$; Wagge et al., 2017, $n = 21$), one was conducted in the United Kingdom (Banas, 2014, $n = 43$), and one was conducted in Germany (Schwarz, 2013, $n = 38$) for a total n of 673 prior to exclusions. As per Elliot et al. (2010)'s instructions for replication, researchers limited participation to heterosexual or bisexual women (while also excluding color-blind participants); lesbian ($n = 10$) and colorblind ($n = 3$) participants have been excluded from all analyses, in addition to participants that guessed the true purpose of the study[5] ($n = 15$) as well as participants with missing

data (n = 4) or identified as having a sexual preference of "other" (n = 1), leaving a total of 581 participants in the eight replications with raw data provided[6] (M age = 20.53, SD age = 3.18), and 640 total. Sample characteristics, including ethnic composition, are summarized in Tables 1 and 2 of the supplemental material.

Materials
All researchers used the same photos of a Latino-American, college-aged male. These 4 in. × 6 in. photos had either a red background or a gray background on an 8.5″ by 11″ piece of paper.

Participants completed the same assessments as those in Experiment 3 of Elliot et al. (2010), beginning with Maner et al.'s (2003) 3-item perceived attractiveness measure to assess attractiveness of the man in the photo (e.g. "How pleasant is this person to look at?"; scored 1 *not at all* to 9 *very much*; α = .89; Ωtotal = .9; ΩHierarchical = .07), followed by two items from Greitemeyer's (2005) five-item sexual receptivity measure (to assess sexual attractiveness; α = .90) and Jones et al.'s (2004) six-item likeability measure (to assess perceived likeability, α = .86; Ωtotal = .92; ΩHierarchical = .79).[7]

Procedures
Researchers tested participants in a closed room without any natural sunlight, as per instructions by the original researcher. Depending on condition, each participant viewed a grayscale paper copy of a male's photograph mounted on a red or grey background for the duration of approximately five seconds — this procedure was double-blind, where one research assistant prepared the photographs prior to the session and another provided an envelope containing the photograph without seeing its contents. After viewing the photo, participants completed Maner et al.'s (2003) perceived attractiveness measure, two items from Greitemeyer's (2005) five-item sexual receptivity measure, and Jones et al.'s (2004) likeability measure. Upon completion, researchers asked all subjects to provide relevant demographic information about themselves including sexual orientation, gender, and whether or not they were color-blind, as well as their best guess regarding the study purpose. Each experiment took approximately 10 minutes to complete in its entirety.

Known Differences Between Original and Replication Studies
Although we recreated the study very faithfully, there are a few (minor) known differences between the included replication attempts and the original study conducted by Elliot et al. (2010). These differences are as follows:

· The original study tested participants one at a time in a closed room. At least two of the replication studies allowed for two to three participants at a time, but in a way that ensured that none of the participants could view the other participants' photographs (Boelk & Madden, 2014; Johnson et al., 2016).
· The original study only utilized photographs with red or grey background. Some researchers digitally applied color variations to add a yellow condition to

the original materials used by Elliot et al. (2010) as a separate condition, thus not changing the nature of the study itself. Comparisons with this additional background color were excluded from the present meta-analysis.
· One potential "hidden moderator" of the effect could be relationship status, to test for this, three teams also asked for participants to list their relationship status (Johnson et al., 2016; Legate et al., 2015; Banas, 2014).
· Finally, one replication study ran this study in tandem with another color-related investigation (Banas, 2014). In every session, the Elliot et al. (2010) replication was always run first in its entirety.

Results
Our goal was to attempt to replicate the original findings, and therefore we used the same statistical analyses as those used by Elliot et al. (2010), an analysis of the ratings differences (cf. Anderson & Maxwell, 2016). Using the Exploratory Software for Confidence Intervals (ESCI) (Cumming, 2016), we used a random-effects meta-analysis for the ratings differences between the red and gray conditions in each replication. For each category (perceived attractiveness, perceived likeability, sexual attractiveness) we completed a meta-analysis comparing ratings differences between red and gray backgrounds, both with and without Elliot et al.'s (2010) original data.

For all analyses, a positive ratings difference indicates that participants who viewed the picture surrounded by red rated that picture higher (e.g. more attractive) than participants who viewed the picture in gray. Conversely, negative ratings differences indicate a preference for those surrounded by gray. The ratings differences for each replication as well as the overall mean effect are depicted in forest plots in **Figures 1−3**. All analyses have been completed excluding the participants discussed in the methods section (i.e. colorblind, lesbian or "other" sexual preference, guessed purpose, missing data).

Replication Results
Independent sample t-tests were completed to determine if condition (red or gray) affected ratings of perceived attractiveness, sexual attractiveness, and likeability. No significant differences between conditions were revealed (ps of .53, .60, and .67, respectively). See **Figures 1−3** for a summary of means and standard deviations by group.

Meta-Analysis Including Original Results
For perceived attractiveness, we found a mean rating difference of −0.07, 95% CI [−0.31, 0.16]; when we also included the original data we found a mean rating difference of −0.01, 95% CI [−0.24, 0.22]. For sexual attractiveness, we found a mean rating difference of −0.06, 95% CI [−0.36, 0.24]; with original data, we found a mean rating difference of .11, 95% CI [−0.28, 0.49]. The proximity of these effects to zero and the range of the CIs are counter to the red-romance hypothesis; we would expect an effect above and not overlapping zero given Elliot et al.'s original results.
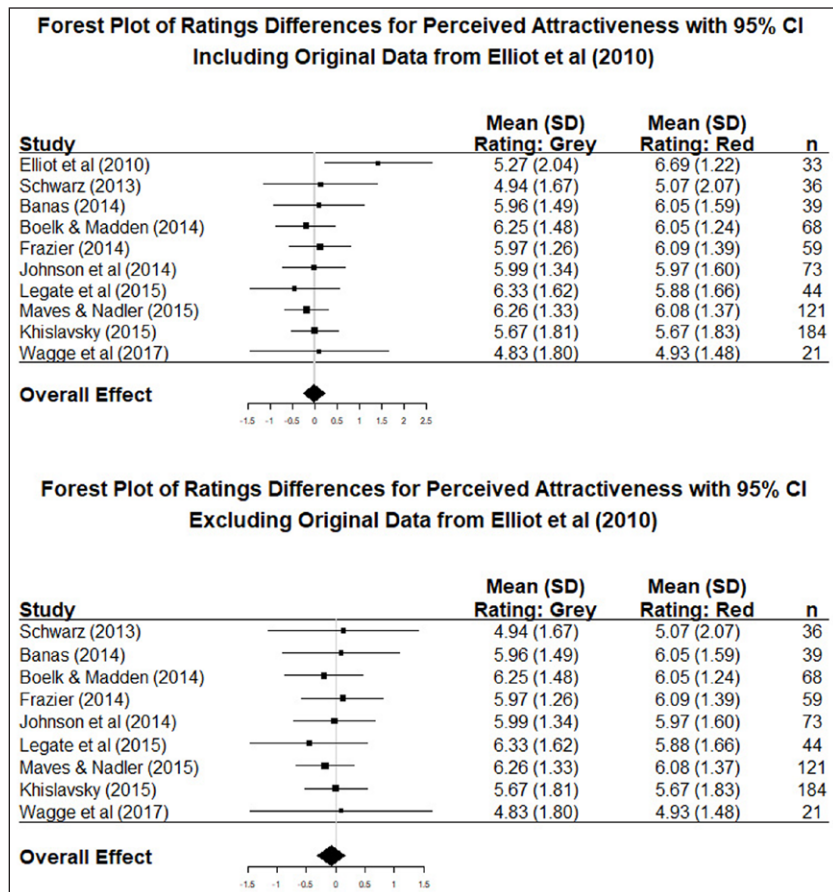
Figure 1: Forest plots for perceived attractiveness. As the plots indicate, there is no effect of a red background on perceived attractiveness including or excluding Elliot et al.'s original data.
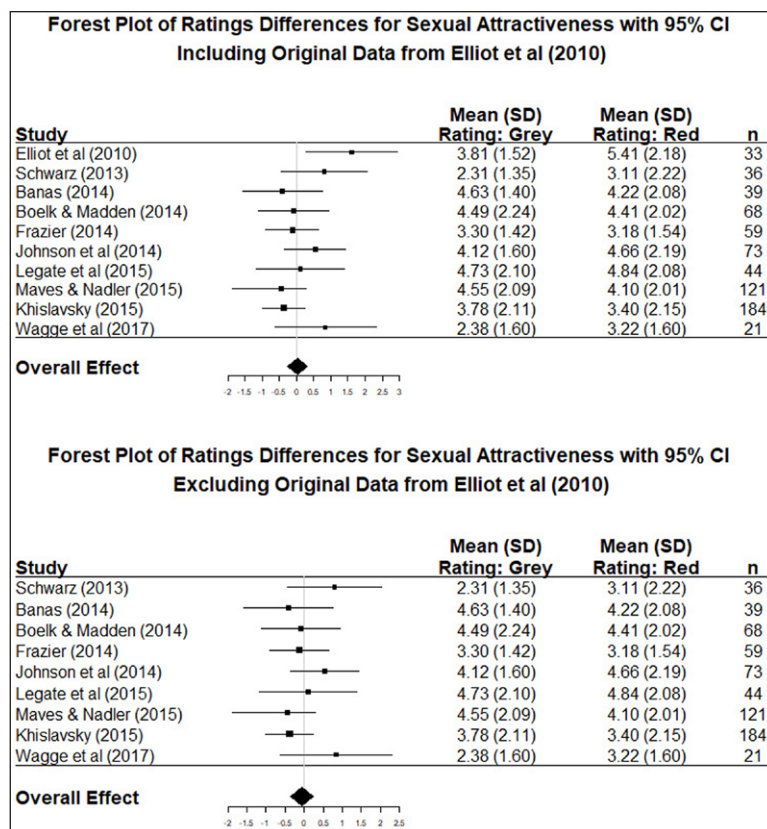

Figure 2: Forest plots for sexual attractiveness, with (top panel) and without (bottom panel) including the original Elliot et al. (2010) data. As the forest plot indicates, there is no effect of a red background on sexual attractiveness.
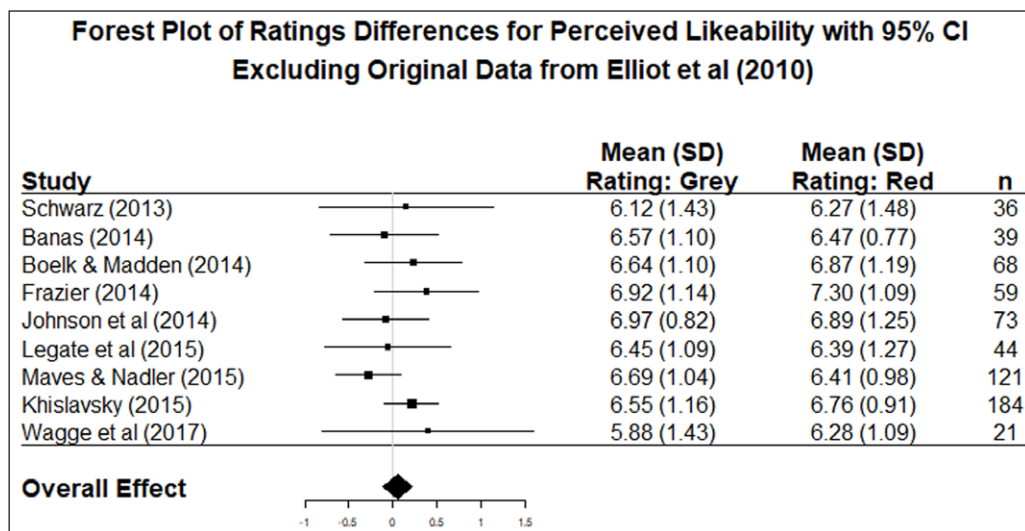
Figure 3: Forest plots for perceived attractiveness without the original Elliot et al. (2010) data; this data was unavailable but Elliot et al. (2010) report null effects for perceived likeability. As the forest plot indicates, we found no effect of a red background on perceived likeability.

Finally, for perceived likeability we found a mean rating difference of 0.05, 95% CI [–0.12, 0.22]; Elliot et al. (2010) also found a null effect for this measure and therefore only reported the p value as > .63, so we did not have the information available to calculate the mean rating difference including the original data. Altogether, with and without the original data included, we did not find a discernible effect of red (versus grey) background color on attractiveness.

We performed equivalence tests using the R package TOST (Lakens, Scheel, & Isager, 2018) to test whether our results were significantly smaller than Elliot et al. (2010). We rejected the null hypotheses for both perceived attractiveness [$t(636.26) = 10.401$, $p < .001$)] and sexual attractiveness [$t(637.16) = 10.16$, $p < .001$], concluding that the observed effects in our meta-analysis are significantly lower than the point estimates reported in the original study.

Exploratory Analyses
We ran a set of analyses to address whether differences in the replication studies impacted the results. First, we assessed whether there were any differences when participants were run in groups (2–3 at a time; $n = 138$) compared to alone ($n = 442$) using a 2 × 2 Factorial ANOVA where the second independent variable was background color — see **Figure 4** for a visual summary of this data. For perceived attractiveness, there was no interaction, $F(1,579) = 0.005$, $p = .94$, and there were no main effects of condition or group/solo setting. Mean perceived attractiveness ratings for the red background ($M = 5.75$, $SD = 1.70$) did not differ from the ratings for the gray background ($M = 5.82$, $SD = 1.68$, $F(1,579) = .39$, $p = .53$, and ratings completed in a solo setting ($M = 5.79$, $SD = 1.69$) did not differ from ratings completed in groups ($M = 6.06$, $SD = 1.41$), $F(1,579) = 2.98$, $p = .08$. For sexual attractiveness there was no interaction [$F(1,579) = 1.38$, $p = .24$] or effect of background color [$F(1,579) = 0.29$, $p = .59$], ($M_{red} = 3.86$, $SD_{red} = 2.10$; $M_{gray} = 3.76$, $SD_{gray} = 2.13$), but there was an

effect of groups such that participants who completed the questionnaire in groups rated the man as significantly more sexually attractive ($M = 4.40$, $SD = 2.00$) than participants who completed the questionnaire by themselves ($M = 3.86$, $SD = 2.11$), $F(1,579) = 7.07$, $p = .01$, $\eta^2 = 0.01$. We observed a similar outcome with likeability. Again there was no interaction [$F(1,579) = 0.05$, $p = .82$] or effect of background color [$F(1,579) = 0.06$, $p = .81$], ($M_{red} = 6.51$, $SD_{red} = 1.15$; $M_{gray} = 6.54$, $SD_{gray} = 1.03$), but participants who completed the task in groups rated the man as significantly more likeable ($M = 6.80$, $SD = 0.96$) than participants who completed the questionnaire by themselves ($M = 6.51$, $SD = 1.15$), $F(1,579) = 8.51$, $p = .004$, $\eta^2 = 0.01$.

Next, we completed exploratory analyses to determine whether attractiveness and likeability ratings differed by relationship status, and if this interacted with condition. Out of the 334 participants who were asked their relationship status, we excluded the category "other" (N = 8) along with missing responses, merged the rest of the categories into two levels of a new variable ("married" and "committed relationship" into one level, "single" and "casually dating" into another) and conducted a 2 × 2 Factorial ANOVA where the first IV was condition and the second IV was relationship status ("in a relationship", "not in a relationship"). We found no main effects or interactions in any of these analyses (see **Figure 5** for bar graphs).

General Discussion
Elliot et al. (2010) found in their Experiment 3 that a red background caused heterosexual participants to find a Latino-American male more attractive than on a grey background. We did not reach the same conclusions, even though we loyally reproduced the original experiment with extensive feedback from the original author. If we included the data from the original study, the results also failed to reach significance. We only replicated the null effect on perceived likeability. There are several possible explanations for our results. First, the
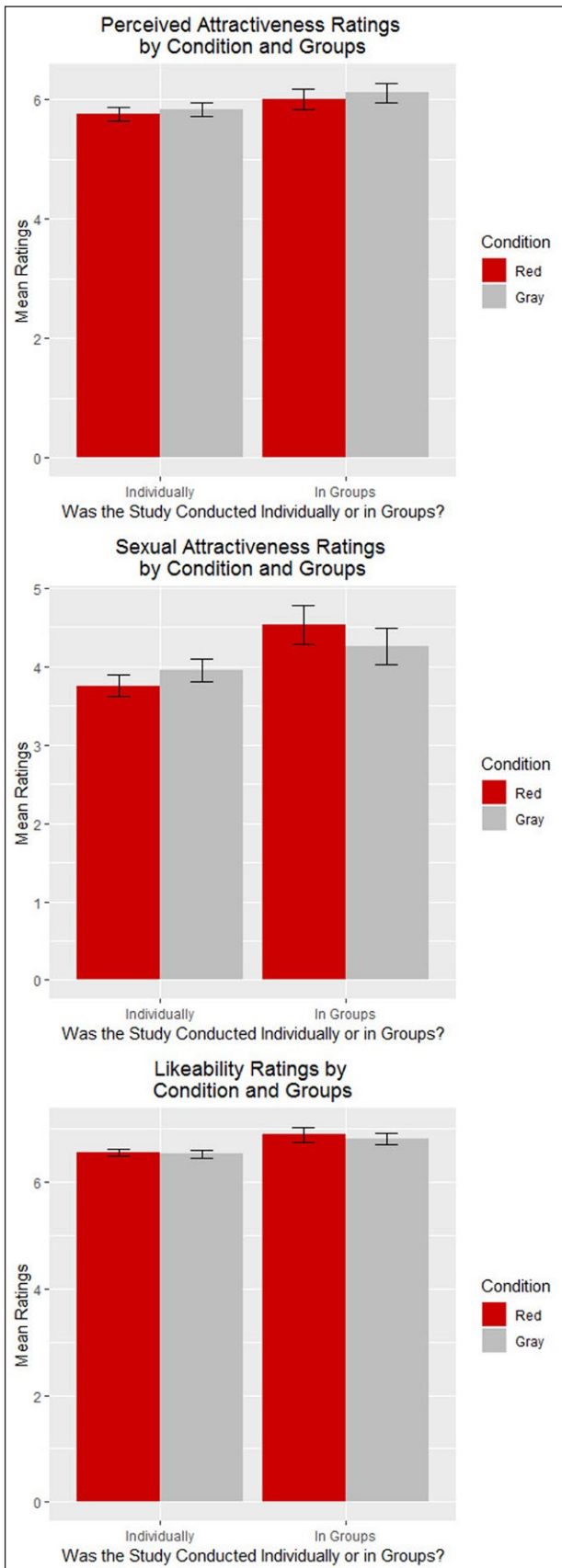
Figure 4: Bar graphs depicting mean ratings for perceived attractiveness, sexual attractiveness, and likeability by condition (red or gray) and whether the study was conducted individually or in groups.
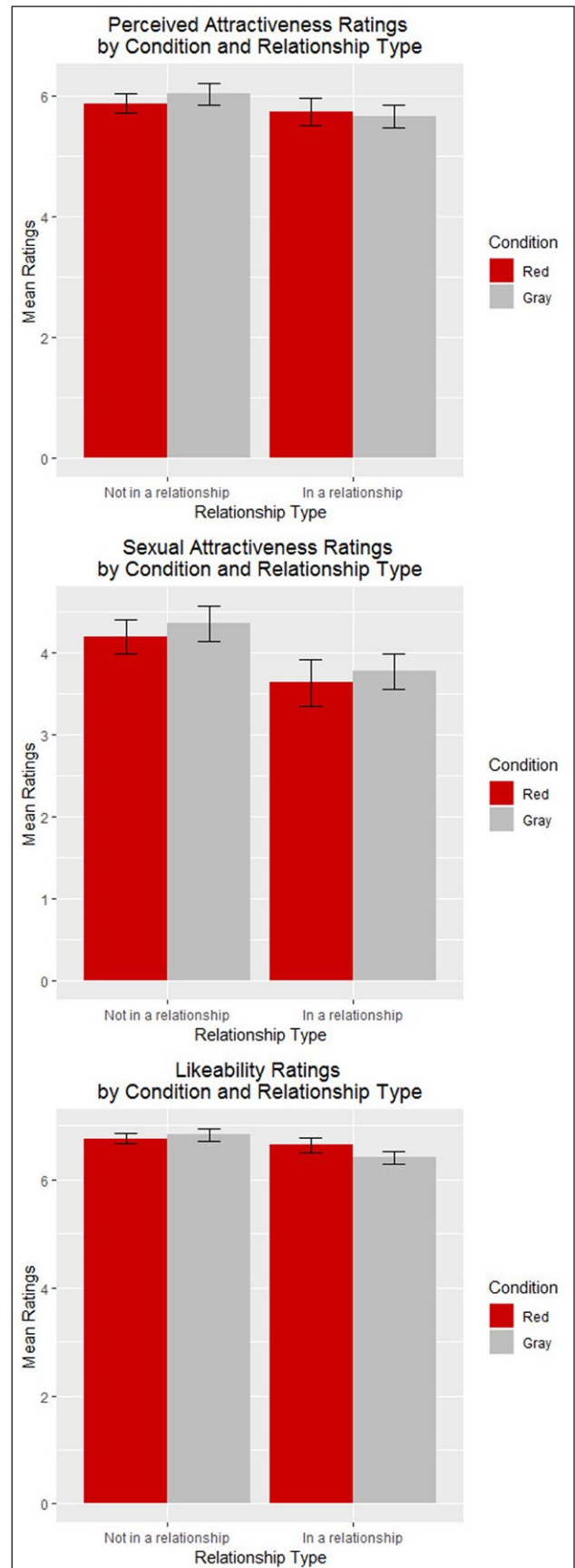

Figure 5: Bar graphs depicting mean ratings for perceived attractiveness, sexual attractiveness, and likeability by condition and participant relationship type.

original results may have reflected a Type I error. To us, this seems to be the most likely interpretation as it is consistent with other research investigating this effect (e.g. Hesslinger et al., 2015). After all, we loyally replicated the experiment and replications were run with several different independent teams. In addition, the manipulation was not complex to run, and it is unlikely that interpretations of the color red have changed over the years since the study was conducted.

A meta-analysis on the link between red and romance has recently been completed (including Dr. Elliot) to examine the effect across gender and implementation of redness (e.g. background, facial redness, red clothes, Lehmann, Elliot, & Calin-Jageman, 2018); this meta-analysis includes some (but not all) of the data included here as well as additional CREP data with additional conditions or where the research has been conducted with variability in settings (such as online v. in person) or materials. The authors report a very small effect of red background when women view men, $d = 0.13$, 95% CI [0.01, 0.25], $p = .03$, $n = 2,739$.

However, the current results do not mean that the red-is-romance effect does not exist. It could indeed be that stimulus selection mediates effects of the color red on attraction; indeed, studies that have found effects of red on attractiveness ratings have employed a different type of stimulus (e.g. clothes in Roberts, Owen, & Havlicek, 2010; lipstick in Stephen & McKeegan, 2010) while the effects of different backgrounds in photographs seem to elicit no differences (Hesslinger, Goldbach, & Carbon, 2015). Finally, selection of the stimuli may also matter. Young (2015) found that when men were rating pictures of women in a within-subjects design with red backgrounds (compared to grey), the effect of background on attractiveness ratings was moderated by the woman's attractiveness (determined by pre-ratings of the photographs).

One other possibility is that our stimulus materials faded over time. We report on a range of studies completed between 2013 and 2017 using the same printed materials that were sent between experimenters and the CREP. Visual inspection of **Figures 1** and **2** does not support this as a significant limitation — there is not a pattern that demonstrates a strong red effect to start that then declines.[8]

We do think it is important to make some final notes on the crowdsourcing of student projects, as the involvement of novice (student) researchers may lead to concern about the quality of the research. In this study (as in all CREP studies) there are various ways in which we applied stringent quality control. First, we selected studies for feasibility for undergraduate research. It is unlikely that offering pictures on different color backgrounds cannot be done by undergraduate researchers. Furthermore, researchers frequently involve student researchers in original research. Our procedure also ensured much stricter oversight (through a faculty member, two reviewers, a CREP board member, *and* the original author) – and thus greater quality – than most research procedures, resulting in a very accurate documentation of and high degree in the research process.

Conclusion

In a meta-analysis of nine replications performed by student teams, we could not replicate the effect of a red (versus gray) background on perceptions of male attractiveness. This research can be seen as a "proof-of-concept" of crowdsourced undergraduate research and thus as a key tool to help reduce the consequences of the "replication crisis" (Grahe et al., 2012; see also Earp & Trafimow, 2015). Though people may express concerns about the quality of relying on undergraduate researchers for replication research, these concerns can be countered through careful selection criteria, strict quality control by advanced researchers, and precise documentation. Overall, providing undergraduate students with research opportunities also provides important pedagogical opportunities, which will teach them to not focus on positive results, but instead on solid methods (Cetkovic-Cvrlje et al., 2013). We think that the future is bright: having undergraduate students actively contribute to our knowledge database thus allows for more accurate results, while they become better trained researchers.

Data Accessibility Statement

All the materials, data, graphs, and analysis scripts can be found on this paper's project page on the Open Science Framework (DOI: https://doi.org/10.17605/OSF.IO/9ZTWY).

Notes

1 CREP teams are typically required to collect data from 100 participants for all studies with an original $n > 100$.

2 The CREP currently has four projects with completed data sets, including the one presented here.

3 Studies that were completed online or using different materials were not considered high-fidelity and therefore were not included in this meta-analysis.

4 As noted before, only summary data is available for Frazier (2014), so this study is not included in any analyses involving raw data.

5 Defined by their response to the question "What do you think the purpose of this study is?" If participants indicated that the study had anything to do with the color of the picture's background, they were excluded.

6 There is no overlap here; the removed participants each only belonged to one of the listed categories.

7 Cronbach's α was reported in Elliot et al. (2010) as .94, .88, and .87 for perceived attractiveness, sexual attraction, and perceived likeability, respectively. We did not report Ω for the sexual attractiveness items, as Ω requires at least three items.

8 Additionally, we found no correlation between year of study (1–5) or any of the scales (attractiveness, sexual attractiveness, and likeability). This held for both "red" and "grey" conditions.

Additional Files
The Additional Files for this article can be found as follows:

- **Table S1.** Demographic Characteristics of Each Replication Included in Analysis. DOI: https://doi.org/10.1525/collabra.177.s1
- **Table S2.** Participant Ethnicity for Each Replication Included in Analysis. DOI: https://doi.org/10.1525/collabra.177.s2

Competing Interests
The authors have no competing interests to declare.

Author Contributions
JRW added data, revised the analyses, rewrote the original manuscript to reflect the new analyses and updated literature, and coordinated manuscript preparation. CB assisted with the administration of the replication projects and the writing of the original manuscript. KB collected data, provided feedback on manuscript drafts and conducted selected analyses. JTN collected data and assisted with the original manuscript. SS collected data and worked on the current manuscript. YW collected data and worked on the current manuscript. HIJ helped write the first version, conducted analyses, reviewed and oversaw studies, contacted the original author to reach agreement about the replication studies, provided critical commentary on later versions of the paper, and did code review. NL collected data and assisted with manuscript construction. JG reviewed and oversaw studies and provided feedback on various drafts of the manuscript.

References
Anderson, S. F., & **Maxwell, S. E.** (2016). There's more than one way to conduct a replication study: Beyond statistical significance. *Psychological Methods*, *21*(1), 1–12. DOI: https://doi.org/10.1037/met0000051

Banas, K. (2014, May 15). Replication of Elliot et al. (2010) for CREP at the University of Edinburgh. Retrieved from: osf.io/cvdpw.

Boelk, K., & **Madden, W.** (2014, August 5). Fork of Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & Liu, H. (2010). Retrieved from: osf.io/zf7c9.

Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., & **van 't Veer, A.,** et al. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224. DOI: https://doi.org/10.1016/j.jesp.2013.10.005

Cetkovic-Cvrlje, M., Ramakrishnan, L., Dasgupta, S., Branam, K., & **Subrahmanyan, L.** (2013). A multi-disciplinary analysis of intensive undergraduate research. *Council on Undergraduate Research Quarterly*, *33*(4), 16–22.

Cumming, G. (2016). ESCI (Exploratory Software for Confidence Intervals) [Computer software]. Available from: www.routledge.com/cw/cumming.

Earp, B. D., & **Trafimow, D.** (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*(621). DOI: https://doi.org/10.3389/fpsyg.2015.00621

Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & **Lu, H.** (2010). Red, rank, and romance in women viewing men. *Journal of Experimental Psychology: General*, *139*(3), 399–417. DOI: https://doi.org/10.1037/a0019689

Francis, G. (2013). Publication bias in "Red, rank, and romance in women viewing men," by Elliot et al. (2010). *Journal of Experimental Psychology: General*, *142*(1), 292–296. DOI: https://doi.org/10.1037/a0027923

Frazier, A. (2014, November 13). Fork of Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & Liu, H. Retrieved from: osf.io/u0mig.

Grahe, J. E., Reifman, A., Hermann, A. D., Walker, M., Oleson, K. C., Nario-Redmond, M., & **Wiebe, R. P.** (2012). Harnessing the undiscovered resource of student research projects. *Perspective on Psychological Science*, *7*(6), 605–607. DOI: https://doi.org/10.1177/1745691612459057

Greitemeyer, T. (2005). Receptivity to sexual offers as a function of sex, socioeconomic status, physical attractiveness, and intimacy of offer. *Personal Relationships*, *12*(3), 373–386. DOI: https://doi.org/10.1111/j.1475-6811.2005.00121.x

Hesslinger, V. M., Goldbach, L., & **Carbon, C. C.** (2015). Men in red: A reexamination of the red-attractiveness effect. *Psychonomic Bulletin and Review*, *22*, 1142–1148. DOI: https://doi.org/10.3758/s13423-015-0866-8

Johnson, K., Meltzer, A., & **Grahe, J.** (2016, June 15). Replication and CREP Meta-Analysis of Elliot et al. 2010 for. Retrieved from: osf.io/ictud.

Jones, J. T., Pelham, B. W., Carvallo, M., & **Mirenberg, M. C.** (2004). How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology*, *87*(5), 665–683. DOI: https://doi.org/10.1037/0022-3514.87.5.665

Khislavsky, A. (2016, March 14). Replication of Elliot et al. (2010). Red, rank, and romance in women viewing men. Retrieved from: osf.io/f2udj.

Lakens, D., Scheel, A. M., & **Isager, P. M.** (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. DOI: https://doi.org/10.1177/2515245918770963

Legate, N., Baciu, C., Horne, L. M., Fiol, S., Paniagua, D., Muqeet, M., & **Zachocki, E.** (2015, January 21). Replication of Elliot et al. (2010) at IIT. Retrieved from: osf.io/zih7c.

Lehmann, G. K., Elliot, A. J., & **Calin-Jageman, R. J.** (2018). Meta-analysis of the effect of red on perceived

attractiveness. *Evolutionary Psychology*. DOI: https://doi.org/10.1177/1474704918802412

Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., & Leuberg, S. J. (2003). Sexually selective cognition: Beauty captures the mind of the beholder. *Journal of Personality and Social Psychology*, 85, 1107–1120. DOI: https://doi.org/10.1037/0022-3514.85.6.1107

Maves, M., & Nadler, J. T. (2015, November 13). Replication of "Red, Rank, and Romance." Retrieved from: osf.io/9bm8v.

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

Schwarz, S. (2013, November 8). Fork of Fork of Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., & Liu, H. (2010). Retrieved from: osf.io/fuw14.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559–569. DOI: https://doi.org/10.1177/0956797614567341

Stephen, I. D., & McKeegan, A. M. (2010). Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception*, 39, 1104–1110. DOI: https://doi.org/10.1068/p6730

Wagge, J. R., Mathis, N., & Hartley, T. (2017, July 7). Replication of Elliot et al. (2010) from Avila University (CREP #17–18). Retrieved from: osf.io/5hqyt.