Collaborative registered replication of Griskevicius et al. (2010): Can pro-environmental behavior be promoted by inducing status competition?

Jordan Wagge

Last Updated 2023-12-04

**Corresponding author**:

Ljiljana B. Lazarević, Faculty of Philosophy, University of Belgrade Čika Ljubina 18-20, 11000 Belgrade, Serbia Email: [ljiljana.lazarevic@f.bg.ac.rs](mailto:ljiljana.lazarevic@f.bg.ac.rs) Phone: +381 64 1287 614

**Author contribution:** to be filled in

**Conflict of interest statement:** The authors report no conflict of interest.

**Funding statement:** This project was funded in part by NSF Award 2141930 to J. Wagge.

**Ethics committee approval:** This research was approved under an “umbrella” ethics proposal at the Institutional Review Board of the Department of Psychology, University of Belgrade (Protocol #2021-069). At each site where there was a local ethics board and where ethics approval was required, the local board approved the study. This study complied with the ethics code outlined in the Declaration of Helsinki.

**Acknowledgment:** The authors would like to thank all students, supervisors, and reviewers who contributed to this CREP project. The authors would like to specifically thank Drs. Mark Brandt and Hans IJzerman for their vision and early dedication to the CREP.

## Introduction

The Collaborative Replications and Education Project (Wagge, Brandt, et al., 2019), or CREP, is an initiative seeking to improve undergraduate training in research methods. CREP’s role includes selecting studies to be closely replicated by students, structuring the methods for the replication, reviewing projects both before and after data collection, and facilitating the publication of a pooled analysis (such as the present paper). Authors agree that replications are regarded as a gold standard warranting reliability of published findings (Open Science Collaboration, 2015). In addition, it was found that for students conducting replication of highly-cited studies is perceived as exciting, and it also gives them the real opportunity to make substantial contributions to the field (Grahe et al., 2012; Frank et al., 2012).

CREP participants include students and instructors at institutions worldwide[[1]](#footnote-21). CREP is an alternative to the traditional undergraduate research experience offering several advantages. First, the CREP model, utilized by students and instructors worldwide, allows students to learn about best practices by directly replicating published work by scholars in their field. This active learning experience–also known in teaching literature as mastery experience or high-impact learning practice –can be an important predictor of academic achievement (Loo & Choy, 2013). Second, students get to participate in authentic scientific research (Grahe, 2017) that may eventually be published in a pooled analysis and will also be openly available on the Open Science Framework (which is an open-source management tool that facilitates open and transparent collaboration in scientific research) to meta-analytic researchers in the future. Furthermore, students have the option of performing a “replication+” study to add their own additional research question, condition, or measure to the replication study (assuming it does not interfere with the direct replication). Moreover, students learn the importance of open science practices such as preregistration, replication, open methods, and open data, all of which are foundational to research integrity (Kidwell et al., 2016). Finally, students engage with reviewers (i.e., CREP team members) external to their institution, allowing for opportunities for collaboration and learning. To date, several CREP studies have been published (e.g., Ghelfi et al., 2020; Leighton et al., 2018; Wagge, Baciu, et al., 2019), and others have been included in meta-analyses (Lehmann et al., 2018), which testifies to the project’s ability to make substantial contributions to psychological science.

The present report documents a pooled analysis of data collected by teams who signed up to replicate one of CREP’s earliest selections for replication: Experiment 1 from Griskevicius et al. (2010).

### The original study: Promoting pro-environmental behavior through status competition

Griskevicius and colleagues (2010) reported the results of three experiments suggesting that status competition can promote pro-environmental behavior. Per CREP procedures (see details in Methods - CREP Procedures), only one study in a multi-study paper is typically selected for replication. In this case, the CREP team selected the first experiment from Griskevicius et al. (2010) as it was the focal study that was conceptually replicated and extended in their subsequent experiments (see Methods for justification). Specifically, the original authors showed that activating status motives (desires for social status and prestige) leads people to choose “green” (environmentally-friendly) products over more luxurious non-green products (Study 1). The authors predicted that reading the status story would increase the likelihood of choosing prosocial (green) products. In the absence of status motives (i.e., in the two control conditions), non-green products were viewed as more desirable than their green counterparts. A composite “green” score was calculated by adding up the number of “green” choices the participant made (no green choices - 0; all green choices - 3). However, after status motives were induced, participants were substantially more likely to choose green products (*d* = 0.47). In summary, Griskevicius and colleagues (2010) found that activating status motives inclines people to choose pro-environmental products over more luxurious non-green products, potentially in efforts to signal prosocial, self-sacrificing behavior. Authors interpreted their findings with respect to the “competitive altruism” hypothesis (see Van Vugt et al., 2007), which explains that individuals attempt to appear more altruistic when competing for status. Publicly showing pro-environmental behavior suggests to others that you care and that you are willing and able to bear costs (such as paying more for a less luxurious product) to benefit others; these altruistic tendencies are valued by others and consequently confer prestige and status (Griskevicius et al., 2010).

Experiment 1 examined how activating a status motive would influence respondents to choose between two products that cost the same amount, but one is a relatively luxurious, non-green product and the other is a less-luxurious, pro-environmental product. This multi-lab study aims to determine the robustness of the effect obtained by Griskevicius et al (2010) and to estimate the effect size more precisely. If the original study results are replicated, we would see, in this replication, that respondents in the social status condition would choose the less-luxurious but green product over the more luxurious but non-green product. Despite being published almost 15 years ago, the study still attracts a lot of attention and it has been cited almost 2,500 times.

## Method

### CREP Procedures

This paper was selected for replication by following the standard CREP selection process, outlined by Wagge and colleagues (Wagge, Brandt, et al., 2019). This process that took place in 2013 resulted in the selection of several studies, including Experiment 1 from Griskevicius et al. (2010).

Once the focal study was selected, the CREP leadership team sent an email to the corresponding author of the original work. The CREP team communicated its process and goals and asked for input on possible extension hypotheses and guidance for replication teams. The key components of this correspondence are available here: <https://osf.io/vdo0i/wiki/home/>. There are two key points from the correspondence: first, the main author noted that the effect may not replicate if participants did not equate “green” choices with prosocial behavior; second, the original first author noted that the connection between “green” choices and status was unique to politically liberal groups. The former will provide us with a lens through which we will interpret our results, while the latter provides an explanation for why many student teams included political ideology in their extension hypotheses.

In 2014, groups signed up for the project and prepared OSF project pages for pre-data collection review. These pre-data collection project pages had to include materials, a planned analytical strategy, a video documenting the procedure, and local Institutional Review Board (IRB) approval. The project pages were examined by two reviewers and the executive reviewer, at which point, projects were either approved or sent back for revisions until approval could be granted. Groups were asked to pre-register project pages prior to data collection. After data collection was completed, project pages were again reviewed by two reviewers and the executive reviewer. At this stage, project pages were revised to include the dataset, a short report describing the obtained results, and a signed completion pledge confirming that the research followed ethical guidelines and the predefined study protocol. Following a positive review, individual projects were deemed complete.

#### Target sample size

Per early CREP policies, the required sample size was set as a rule of thumb to be at least half of the original N. For this study, the targeted sample size should have been 84 (half of the original N of 168), but unfortunately, in the documentation file, there was a typo (N = 82). Thus, the target sample size for each group was set to 82 respondents.

#### Differences from the original study

The original study was conducted in a lab, where participants were tested in small groups. Replications were conducted both in the lab and online. Some groups conducted direct replications, and some groups included extension variables (e.g., pro-environmental literacy, political ideology), but after the original protocol was completed. Characteristics of the replications can be found at <https://osf.io/gt39p>.

### Disclosures

#### Preregistrations

Each lab preregistered its materials, protocol, and analytical strategy on the OSF prior to data collection. Additionally, this meta-analysis was pre-registered: <https://osf.io/ach3n>.

#### Data, Materials, and Resources

The authors of the original study provided materials for replication. All groups who completed projects uploaded their data, analyses, and a short description of their results to their OSF project page. All datasets, materials, analytical scripts, and other materials can be found on the central project OSF page (<https://osf.io/rh2nw/>). The complete, anonymized dataset is available here: <https://osf.io/3xrdf>. A codebook for our merged data set with all variables collected as part of the project is available at <https://osf.io/7nkhq/>.

#### Reporting

A description of the sample size, data exclusions, manipulations, and measures are reported in the study.

#### Ethical approval

Data were collected in accordance with the 1964 Declaration of Helsinki. Overarching IRB approval for this study was obtained from the IRB of the Department of Psychology, University of Belgrade, Serbia (<https://osf.io/pcwhg/>). In addition, as part of the CREP procedure, all groups obtained local IRB approval when necessary.

### Sample

From 2014 to 2020, 49 groups from nine different countries expressed interest in conducting a CREP replication of Griskevicius et al. (2010). Of these 49 groups, 25 did not contribute data to the final analysis: specifically, three did not create their OSF page, four did not begin data collection, five did not complete data collection, eleven did not provide their data, one did not provide their codebook, and one did not follow the CREP procedure. The final sample of completed projects included data collected by 24 groups from 6 different countries: USA, UK, Germany, Canada, the Netherlands, and Iceland. The overview of groups participating in this project can be found at <https://osf.io/vc5rh>. Overall, the present project included approximately 30 reviewers, 3 CREP assistants, and 3 executive reviewers.

The final data set consisted of data collected from 3774 participants, by 24 different student groups. All individual data sets are available here: **[INSERT LINK WHEN READY]**. Each student group recruited at least 82 respondents, with the exception of one group that recruited 70 participants (due to the COVID-19 outbreak that interfered with the project). Nine groups failed to collect data on at least one of the following: age, gender, or race/ethnicity. Details about the missing data are available here: <https://osf.io/vc5rh>.

A total of 21 participants were excluded prior to publicly posting the compiled dataset because they reported ages of 16 (*n* = 3) or 17 (*n* = 18); these exclusions resulted in a total final dataset of 3774 participants. The final sample consisted of 2302 individuals who identified as women and 973 who identified as men; 18 participants identified as gender variant/non-conforming, “other”, “prefer not to say,” or transgender/non-binary. An additional 481 participants did not indicate their gender. The mean reported age for participants was 23.98 years (*SD* = 10.21), though age was not collected for 956 of the participants.

Demographic information about participant race/ethnicity was reported by some sites and not others, and within each site, the question varied in how it was asked[[2]](#footnote-40). Therefore, we recoded the original race information to include the following categories: White (*n* = 1339, 71.34%), Black (*n* = 129, 6.93%), Asian (*n* = 131, 6.98%), Hispanic/Latino (*n* = 151, 8.04%), Middle Eastern (*n* = 28, 1.49%), Multiracial (*n* = 8, 0.43%), Native Hawaiian or Pacific Islander (*n* = 5, 0.27%), and “other” (*n* = 71, 3.78%). These reported percentages do not include the participants who either did not report their race/ethnicity or were not asked this question (*n* = 1897, 50.26% of the total).

### Procedure

The experiment was designed as a between-subject and included two conditions: one in which participants read a “cover” story designed to elicit “a desire for social status” (Griskevicius et al., 2010, p. 395), and a control condition, that is one with a control story that elicited similar affect, but not status motives. To eliminate potential suspicions, participants were instructed that they will participate in a study investigating memory and that they would be asked to recall information later in the session. To ensure that some particular aspect of the story second control condition in which participants did not read the story was included. In all conditions, participants were asked to choose between three pairs of products (a car, a dishwasher, and soap), in which each product was offered in proenvironmental (green) and a conventional (the more luxurious nongreen) form. For each choice, the two products were equal in price, were manufactured by the same company and also were described by key features of the product. Stories were the same as in the original study, see Griskevicius et al. (2010) for details.

Data were collected in the lab and online, using various survey platforms (e.g., Qualtrics). Each group documented the mode of data collection that they used. Respondents were recruited from both student and general populations. Descriptions of individual datasets, along with all deviations are available at <https://osf.io/t8h3k>.

Any additional variables were included after the main part of the study, that is, at the end of the survey.

### Treatment of missing data, data preparation, and data analysis

We excluded participants without responses for the dependent variables. To create a merged dataset, we accessed all individual datasets through each team’s OSF page if they were available, and contacted teams directly for their data if it was not. When necessary, we also contacted groups to obtain information on whether the dataset was raw or transformed and to obtain codebooks. Any missing data was left as is.

We conducted the same analyses as the original paper: for each product (car, soap, and dishwasher) we reported percentages by condition, Chi-square tests, and Phi coefficients. We also computed a composite green score and ran an ANOVA across conditions (story, no story, and control story) and reported effect size and post-hoc comparisons using the HSD test.

One advantage of the multi-lab approach is that natural variation in the data collection process across different labs can be leveraged to test for moderators or boundary conditions (Kline et al., 2018). Accordingly, we also performed several exploratory analyses. We ran 2 x 2 factorial ANOVAs to determine whether political orientation (liberal/conservative) or political party (democrat/republican) interacts with condition (control/status). We ran an ANOVA to explore whether different settings (online/lab) yielded different results. Second, as per the recommendation of the original authors, we tested the moderating role of political orientation.Third, we explored whether the main effects replicated in other countries in which we collected data, and tested potential differences between the US and other countries. Fourth, the long-lasting nature of our project (data collection was done between 2014 and 2020) enabled testing whether the effects differed across time. Lastly, the original authors stated that it would be worth testing for potential gender differences, thus, as part of our exploratory analyses we explored gender differences in status motives using a 2 x 2 factorial ANOVA (gender: woman or man; condition: control vs. status). All analyses were done using R markdown and R Studio (RStudio Team, 2020, R Core Team, 2021).

## Results

### Confirmatory Analyses

#### Green car

The original paper found that participants selected the green car more often in the status condition (54.5%) than in the control condition (37.2%), (1, N = 168) = 4.56, *p* = .033, = .165. In our sample, 55.5% of control participants selected the green car, compared to 54.83% of the status participants. We did not detect an effect of condition on green car selection, (1, N = 3726) = 0.13, *p* = 0.72, = -0.006.

#### Green cleaner

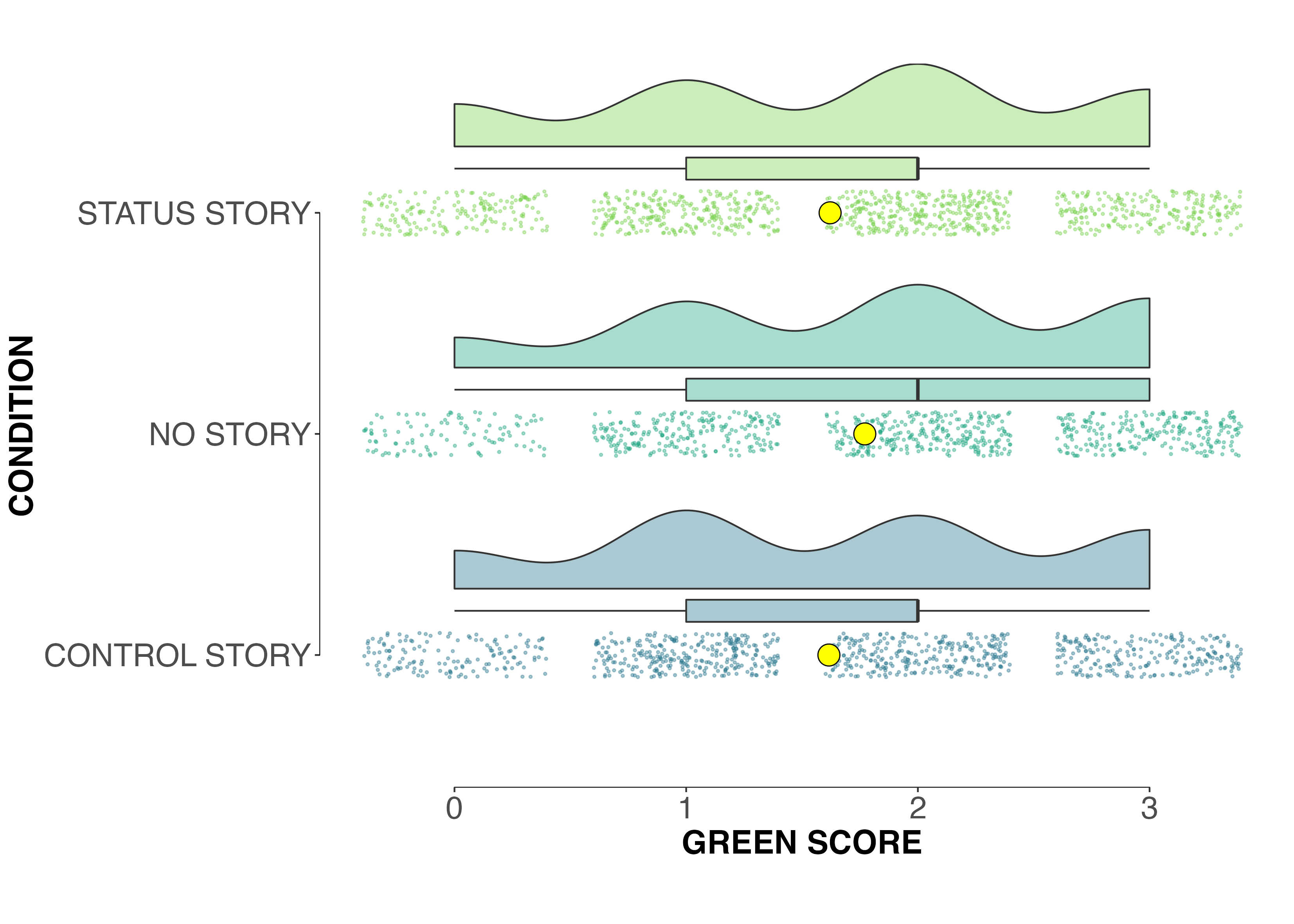
Similarly, the original paper found that participants in the status condition selected the green soap more frequently than participants in the control condition (41.8% and 25.7%, respectively), (1, N = 168) = 4.52, *p* = .034, = .164. In our sample, 56.75% of the control and 54.95% of the status participants selected the green soap. As with the “car” selection, we did not detect an effect of condition on the choice of cleaner, (1, N = 3725) = 1.06, *p* = 0.3, = -0.018.

#### Green dishwasher

Finally, in the original paper, participants selected the green dishwasher more in the status condition than the control condition (49.1% and 34.5%, respectively), (1, N = 168) = 3.30, *p* = .069, = .140. We found no such effect when comparing the control participants who selected the green dishwasher (52.9%) to the status condition participants who did the same (51.65%), (1, N = 3729) = 0.49, *p* = 0.48, = -0.012.

#### Composite Green Score

Internal consistency of the three items (i.e., products) was low (Cronbach’s alpha = 0.38) which was expected (Rammstedt & Beierlein, 2014; Malkewitz et al., 2023). Despite this, we followed the procedure used in the original study and computed a composite green score by assigning a score of “1” to all “green” selections and a score of “0” to all non-“green” selections. Because participants made a dichotomous choice for each of the three products, scores ranged from 0 (no “green” products) to 3 (all “green” products). The mean composite score was 1.64 (*SD* = 0.996). The distribution of composite scores can be seen in Figure 1.



**Figure 1.** The rain cloud diagram shows the distribution of responses under different conditions. Yellow dots represent mean values surrounded by individual data points with jitter. Vertical lines in the centers of the boxplots indicate the medians. Shaded areas around them correspond to interquartile ranges, and whiskers encompass 1.5 times the interquartile ranges. The shaded areas above the boxplots indicate the density of the data.

The original paper by Griskevicius and colleagues (2010) reported a significant effect of status on the composite score when the status prime condition was compared to both control conditions (control story and no story) together, *F*(1, 166) = 8.53, *p* = .004, *d* = 0.47. The same test on our data did not reveal a significant effect, , , = .000, *d* = 0.04. Notably, this composite score has low reliability with a Cronbach’s alpha of 0.38, therefore it is unsurprising that we did not find an effect. The original paper did not report a measure of reliability for the composite score. However, when we performed this analysis without combining the control conditions, we found a small effect, , , = .002. Post-hoc testing using Tukey’s HSD correction for multiple pairwise comparisons resulted in no pairwise comparisons with a *p* < .05 (no story/control = 0.06, status story and control story conditions = 0.97, status story/no story = 0.09).

The rest of our analyses combine the “control story” and “no story” conditions into one control group, consistent with the analyses performed by Griskevicius et al. (2010).

### Exploratory Analyses

*Lab vs. Online*. To determine whether there were differences in results across different testing settings, we first collapsed all reported lab settings (group, individual, and just “lab”) into one variable level. We then conducted chi-square tests of independence using a dichotomous setting variable (lab [*N* = 2024] versus online [*N* = 1581]). We found no relationship between green choice and setting for any of the three types of products (*p*s of 0.26, 0.13, and 0.67 for the green car, soap, and dishwasher, respectively).

To see if there was an interaction between condition and setting, we conducted a 2 (condition: status versus control) x 2 (lab versus online) ANOVA using composite score as a dependent variable. We were unable to detect a mean difference in composite score between those who completed the study in the lab (*M* = 1.66), *SD* = 0.98) and those who completed the study online (*M* = 1.64), *SD* = 1.01), , , = .000. As we found earlier, there was also no main effect of condition; the mean scores for participants in the grouped control condition (*M* = 1.65, *SD* = 0.99) did not differ from those in the status condition (*M* = 1.61, *SD* = 1), , , = .000. There was also no interaction between the two variables, , , = .000.

*Political orientation*. When the CREP team first contacted the original authors about the replication, they recommended examining political orientation as a possible extension variable. Therefore, many groups decided to add a question about political orientation (liberal/conservative) and others added a question about the political party (Republican/Democrat/Independent); some groups added both questions, and some groups added questions with responses that were more fitting to their particular country of origin (e.g., Canada) (see Table X on our supplement for more information). We decided to use these data to test whether green selections differed for liberals and conservatives. Some institutions measured this with a scale (from very liberal to very conservative, for example), while others measured it with categorical response types (e.g., liberal, conservative, neutral). We collapsed over these responses to create a novel variable with two levels: a “liberal” level that included all responses indicating any degree of being liberal, and a “conservative” level that included all responses indicating any degree of being conservative. We excluded any “neutral” or “other” responses from this variable.

Using this method of collapsing data, we found that n = 551 participants were classified as conservative, while n = 744 participants were classified as liberal.

#### Liberal/conservative

Because we wanted to examine political-party-esque effects in other countries, we looked at liberal/conservative… Additionally, Because the distinction between liberal and conservative does not always capture the same information as political party (CITATION),… We conducted a 2 x 2 factorial ANOVA (political orientation: liberal vs. conservative; condition: control vs. experimental) to determine whether political orientation interacted with the condition in predicting its association with the composite green score. We found a main effect of political orientation such that participants who identified as liberal selected significantly more green products on average (*M* = 1.87, *SD* = 0.98) than participants who identified as conservative (*M* = 1.25, *SD* = 0.97), , , = .092. As we found earlier, there was also no main effect of condition; the mean scores for participants in the grouped control condition (*M* = 1.65, *SD* = 0.99) did not differ from those in the status condition (*M* = 1.61, *SD* = 1), , , = .000. There was also no interaction between the two variables, , , = .000.

We found a similar pattern of results when the control conditions (no story and control story) were ungrouped: a main effect of liberal/conservative (, , = .092), no main effect of status condition (, , = .002), and no interaction between the two (, , = .001).

#### Democrat/republican

Similarly, for political party (“Democrat” or “Republican”), we conducted a 2 x 2 factorial ANOVA (political party: Democrat n = [332] vs. Republican [182]; condition: control vs. status) including only participants from the United States.

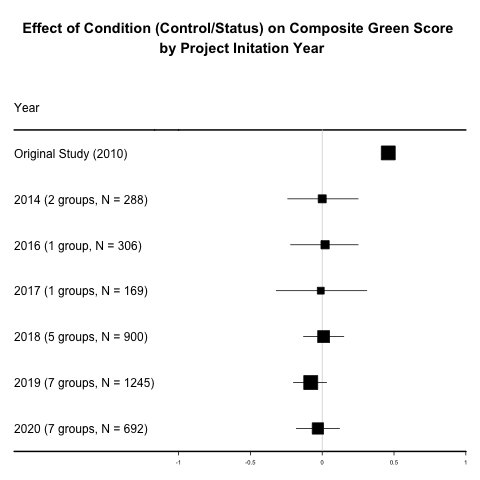
Importantly, we did not have adequate power to detect the interactions tested here. Using the *Superpower* package (Lakens et al., 2021) in R, we determined that we would have needed at least 418 participants *per cell* to detect an interaction effect between political party and condition. In the four cells created by political party and condition (Republican/Status (N = 69) or Democrat/Status (N = 123), (Republican/Control (N = 113) or Democrat/Control (N = 209) our cell sizes did not reach this threshold. Because of these issues with power, we did not further reduce power by conducting a test with ungrouped control conditions.

We found a main effect of political party such that participants who identified as Democrat selected significantly more green products (*M* = 1.77, *SD* = 1.03) than participants who identified as Republican (*M* = 1.12, *SD* = 0.86), , , = 0.09. We did not detect an interaction between political party and condition, , , = 0, and as expected from our power analysis mentioned in the previous paragraph, we detected no main effect of condition (control: *M* = 1.65, *SD* = 0.99); status: *M* = 1.61, *SD* = 1), , , = 0.

*US vs. others*. The original study was conducted only with individuals in the United States. Here, we conducted an exploratory analysis to determine whether the tested effects differed for US-based participants versus participants from other countries. For US-based participants only (*N* = 2151), we found no association between condition (control versus status) and green car selection ((1, N = 2104)) = 0.12, *p* = 0.73, = 0.01), green cleaner selection ((1, N = 2103)) = 0.02, *p* = 0.88, = 0), or green dishwasher ((1, N = 2107)) = 0, *p* = 0.95, = 0) selection.

We then conducted a 2 x 2 ANOVA to test whether there was a main effect of condition (control or status) or interaction with geographic setting (US or non-US). We detected no significant main effect of condition (*p* = 0.56) or interaction (*p* = 0.41). We found a small main effect of location, such that individuals from non-US countries selected significantly more green products (*M* = 1.83, *SD* = 0.95) than individuals from the United States (*M* = 1.48, *SD* = 1.01), , , = 0.03.

*Stability of effect over time*. To determine if any effects differed across years, we examined effects across projects from different years, as determined by their project codes assigned at the time they signed up for the study (e.g. project #14-3 was initiated in 2014). With year treated as a factor, we examined whether we could detect any differences in composite green scores across conditions (control and status). As demonstrated by Figure 2, there does not appear to be any meaningful change in effect size over time.



**Figure 2.** Forest plot caption stuff

*Gender differences*. Gender differences. In their original paper, Griskevicius and colleagues (2010) stated that exploring gender differences would be interesting in future studies. Although they did not find any gender differences, Griskevicius et al., (2010) hypothesized that men could be more likely to engage in pro-environmental “show-off” behavior than women, since men are generally more concerned about status motives. Their findings suggested that in both men and women status motives influenced product selection similarly, but suggested that this should be explored further. Because of our additional statistical power, we sought to test these potential gender differences as part of our exploratory analyses.

To test this question, we conducted a 2 x 2 factorial ANOVA (gender: woman or man; condition: control vs. status). We found a main effect of gender, such that participants who identified as women selected significantly more green products (*M* = 1.69, *SD* = 0.99) than participants who identified as men (*M* = 1.46, *SD* = 1.01), , , = 0.01. We did not detect an interaction between gender and condition, , , = 0.

When we conducted a 2 x 3 factorial ANOVA with the independent conditions (control, status, and no story). We found an effect of condition, , , = 0. Post-hoc testing with Tukey’s HSD correction revealed significant differences between the “no story” condition and the other two conditions (95% CI of the difference for no story v. control: [0.01, 0.21], *p* = 0.03); 95% CI of the difference for no story v. status: [-0.21, -0.01], *p* = 0.03). There was no difference between the control story and the status story conditions, *p* = 0.99). Participants in the “no story” condition made more “green” choices (*M* = 1.7, *SD* = 0.98) than participants in the control condition (*M* = 1.58, *SD* = 1.01) or status condition (*M* = 1.59, *SD* = 1). This effect was not detected when the ungrouped conditions were examined in the entire data set; it was detected only when we excluded participants who did not respond to the question about gender or listed a gender other than “Male” or “Female.”

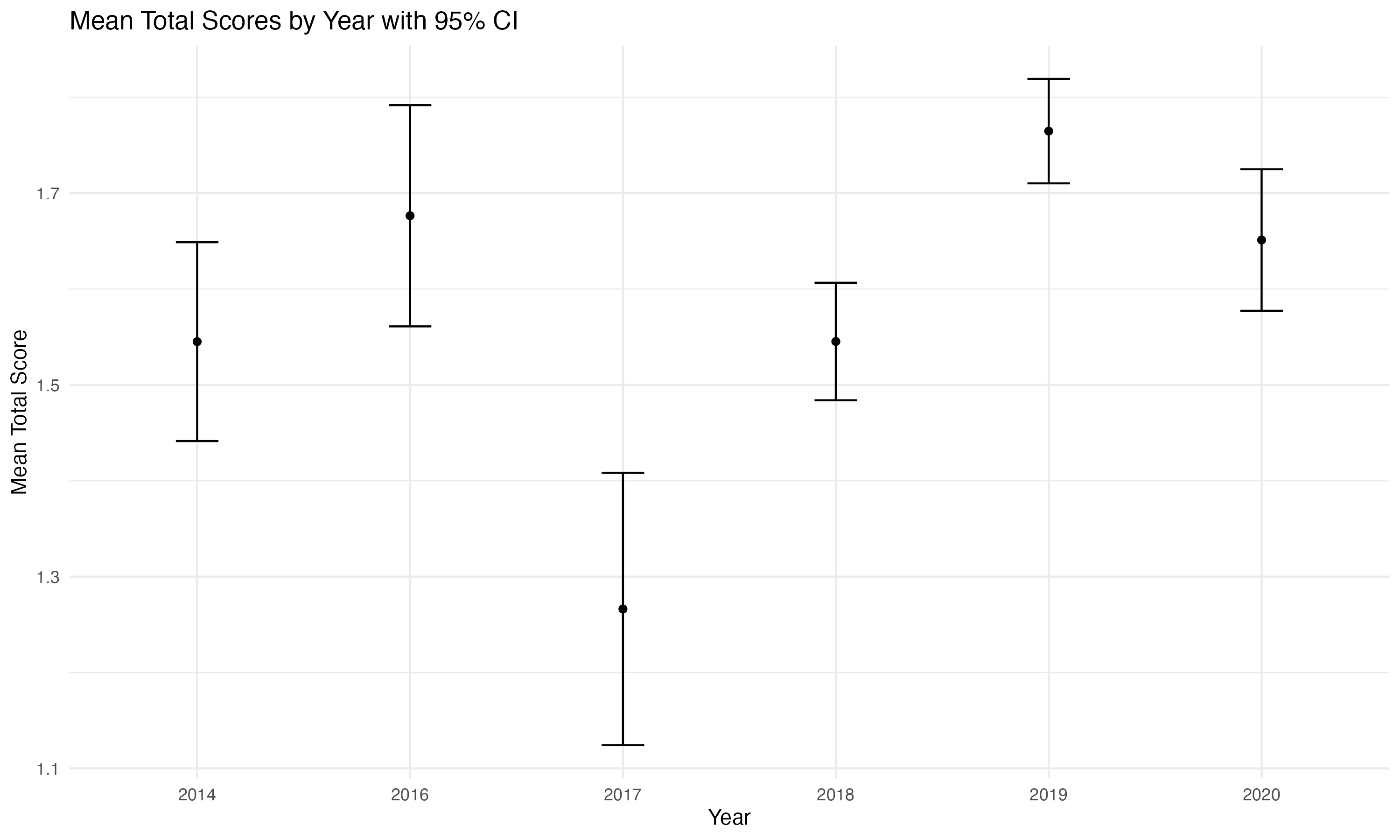
## Discussion

Griskevicius and colleagues’ original study (2010) suggested that pro-environmental behavior can be promoted using status competition. Griskevicius et al. (2010) link this finding to the “competitive altruism” hypothesis, which proposes a connection between altruistic behavior and public good (Hardy & Van Vugt, 2006). That is, earning a good reputation via prosocial and pro-environmental behavior and action, and demonstrating an ability and willingness to bear costs for the public good can increase an individual’s status in a group. The original study showed that the selection of all three green products (green car, green cleaner, green dishwasher) was higher in the status condition compared to the control condition.

In this multi-lab replication study, we recruited a large sample of participants across six different countries. We employed the same analytical strategy as the original study. As part of our pre-registered confirmatory analysis, we tested whether an effect existed when the green score was calculated as a composite; this analysis did not result in a statistically significant effect. We also analyzed separately if an effect existed for each of the green vs. non-green products. For all three products, we failed to detect a significant effect. Similar to the original study, we tested whether the two control conditions differed from one another, and found no reliable differences between them, although there appeared to be potential differences between both “story” conditions (status and control) and the “no story” control such that exposure to any story resulted in fewer “green” choices. Thus, overall, we did not replicate the main conclusion of Experiment 1 from Griskevicius et al. (2010), and with a much more highly powered analysis (original n = 168, present n = 3774) find no evidence in support of the original hypothesis. In addition to a much larger sample size, a strength of this study is that data were collected in 6 different countries, providing additional support for our conclusions. Therefore, it may be concluded that our findings do not corroborate the results obtained in the original study.

One possible explanation for our difference in results is that our control participants selected the green products at much higher rates than in the original study. Our control participants’ rates of selecting the green car, soap, and dishwasher were 55.5%, 56.75%, and 52.9%, respectively, while the original control participants’ rates were much lower at 37.2%, 25.7%, and 34.5%. There could be a ceiling effect with green product choices, such that a certain proportion of the population will never choose to buy green products; likewise, perhaps green products were so new in 2010 that a smaller group of the population would buy them when given the choice. In this case, perhaps status primes moved some folks over the fence, but by the time our data collection occurred the people who *could* have been persuaded by a status prime were already making these choices.

Given this possible alternative explanation, we examined whether there appeared to be any overall pattern in the total composite score over time. As shown in the figure below, some years do differ from others but the early years (2014 and 2016) are not significantly different from the last two years (2021 and 2022).



**Figure 3.** Mean composite scores by year. There were no noticeable differences in mean composite scores between the earliest projects (2014 and 2015) and the latest projects (2021 and 2022).

Anticipating this issue, one group added an extension variable to determine whether participants found owners of the green products in the study to be more nice, caring, or altruistic than owners of the non-green products. They had participants rate these characteristics from 1 (totally agree that the owner has this quality) to 9 (totally disagree that the owner has this quality), and found that participants rated the owners of green products as nicer (*M* = 9.19, *SD* = 4.21) than owners of non-green products (*M* = 12.19, *SD* = 4.19), *t*(226) = -10.85, *p* < .001, *d* = 0.72. Likewise, participants rated green product owners as more caring (*M* = 7.95, *SD* = 3.91) than non-green product owners (*M* = 13.47, *SD* = 4.32), *t*(225) = -16.35, *p* < .001, *d* = 1.09, and more altruistic (*M* = 9.83, *SD* = 4.43) than non-green product owners (*M* = 13.87, *SD* = 4.19) as well, *t*(224) = -10.16, *p* < .001, *d* = 0.68. This site did not replicate the main findings of the original study, thereby weakening the argument that the effects might not replicate because participants might not find these products to be associated with prosocial behavior.

Per the recommendation of the original authors, we performed exploratory analyses to test whether political orientation significantly moderates the effect of status motives on pro-environmental choices. In this study, some testing sites collected data on political orientation and/or political party as extension variables. Whether we defined political orientation as liberal-conservative or Democrat-Republican, we found that liberals and Democrats selected more green products than conservatives and Republicans. This effect did not, however, interact with any effect of status motives.

One difference between the original study and many team replications was that in the original study, participants came into the lab in small groups and were seated at computers with partitions between them. We were curious whether this quasi-social setting could contribute to our different results. Therefore, we also explored whether the setting in which participants were tested (lab – presumably at least somewhat social – versus online) influenced the results. Online data collection has become increasingly popular over the last decade and offers several advantages over traditional lab settings, such as reduced costs, more effective testing, increased generalizability of results, and more efficient (i.e., automated) data collection (Dandurand et al., 2008; Gosling et al., 2004; Nayak & Narayan, 2019). Consistent with findings demonstrating that online studies tend to yield similar results to lab studies (e.g., Riva et al., 2003), we found no differences between data collected online versus in a laboratory. However, we have no way of knowing whether participants who completed the study online were truly alone in their setting, and have no way of knowing whether participants who completed the study in the laboratory just one or many people around them. Therefore, this explanation for the difference in our results could be a direction for future research.

It is important to note that the original study was conducted in the United States, while this multi-site replication project collected data in five other countries: Canada, Netherlands, UK, Germany, and Iceland. We found no significant differences in findings between the samples from different countries. Still, we believe this finding is expected as all countries belong to the group of the so-called WEIRD (Western, Educated, Industrialized, Rich, and Democratic; Henrich et al., 2010) states, and highly sustainable countries with sound environmental policies according to the Fragile State Index (Fund for Peace, 2022).

Another feature of this multi-site replication study is that data collection lasted several years, which gave us the opportunity to explore whether the tested effects changed over time. We did not find substantial differences in effect sizes obtained from replication studies completed between 2014 and 2020. Finally, in their paper, Griskevicius and associates stated that gender differences in status motives would be worth exploring. Our findings suggest that women are more inclined toward green products than men, but a moderating effect of gender on the relationship between experimental condition and product preference was not obtained.

### Limitations of the study

The present study is not without limitations. First, as stated previously, all of the countries in which the study was conducted are WEIRD and highly sustainable countries. Our findings suggest that status motives do not lead people to pursue pro-environmental products. However, we cannot assume that this finding is generalizable to all populations. One possibility is that participants from non-WEIRD and fragile countries, many of which have recently increased efforts to promote pro-environmental behaviors (e.g., Diaz et al., 2020), could be more sensitive to any potential relation between social status and pro-environmental action. Therefore, it will be worth exploring in future work whether the results of the original study would replicate in non-WEIRD and countries that are considered as fragile.

Second, we tested moderation effects only for those variables for which we had sufficient statistical power. Thus, most of the extension variables that were measured could not be used in analyses. To more thoroughly explore the role of extension variables in the effects being replicated, future CREP studies could pose stricter guidelines on the usage of extension variables in student groups.

### Conclusion

One goal for this project was to provide a medium for students to engage in high-quality replication research while contributing to the field. Student-led projects can and do offer adequate rigor, the quality of the studies was also insured by continuous and careful supervision of senior researchers (faculty members, two reviewers, one CREP board member), recommendations from the original first author, and documentation of the research process (e.g., Grahe et al., 2020; Wagge et al., 2019; Wagge et al., 2023).

The primary goal of this project, however, was to determine whether pro-environmental behavior could be promoted through status competition. This study presents a large sample multi-site replication of the Griskevicius et al (2010) study, performed by student teams at various institutions. Overall, our study did not replicate previous work and found no evidence that pro-environmental behavior can be promoted by inducing status motives. Though these findings do not negate the theoretical assertions of Griskevicius et al, they do suggest extensive boundary conditions. Given our lack of demonstrating reliable findings, this theory should be reconsidered to identify what populations might still demonstrate the effect, or why the effect has diminished. In the time that passed between the original study and the replications (ranging from 4-10 years) since the original finding, more focus has been directed toward climate change and environmental choices. It is possible that the changes in awareness regarding this topic moderated the manifestation of the effect.

## References

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. Behavior Research Methods, 40(2), 428–434. <https://doi.org/10.3758/BRM.40.2.428>

Díaz, M. F., Charry, A., Sellitti, S., Ruzzante, M., Enciso, K., & Burkart, S. (2020). Psychological factors influencing pro-environmental behavior in developing countries: Evidence from Colombian and Nicaraguan students. Frontiers in Psychology, 11, 580730. <https://doi.org/10.3389/fpsyg.2020.580730>

Frank, M. C., & Saxe, R. (2012). Teaching Replication. Perspectives on Psychological Science, 7(6), 600–604. <https://doi.org/10.1177/1745691612460686>

Fund for Peace (2022). Fragile State Index 2022 - Annual Report. Retrieved from <https://fragilestatesindex.org/2022/07/13/fragile-states-index-2022-annual-report/>

Ghelfi, E., Christopherson, C. D., Urry, H. L., Lenne, R. L., Legate, N., Ann Fischer, M., Wagemans, F. M. A., Wiggins, B., Barrett, T., Bornstein, M., de Haan, B., Guberman, J., Issa, N., Kim, J., Na, E., O’Brien, J., Paulk, A., Peck, T., Sashihara, M., … Sullivan, D. (2020). Reexamining the effect of gustatory disgust on moral judgment: A multilab direct replication of Eskine, Kacinik, and Prinz (2011). Advances in Methods and Practices in Psychological Science, 3(1), 3–23. <https://doi.org/10.1177/2515245919881152>

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. American Psychologist, 59(2), 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>

Grahe, J. E. (2017). Authentic research projects benefit students, their instructors, and science. In How we teach now: The GSTA guide to student-centered teaching. (pp. 352–368). Society for the Teaching of Psychology.

Grahe, J. E., Cuccolo, K., Leighton, D. C., & Cramblet Alvarez, L. D. (2020). Open science promotes diverse, just, and sustainable research and educational outcomes. Psychology Learning & Teaching, 19(1), 5–20. <https://doi.org/10.1177/1475725719869164>

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. Perspectives on Psychological Science, 7(6), 605-607. <https://doi.org/10.1177/1745691612459057>

Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: Status, reputation, and conspicuous conservation. Journal of Personality and Social Psychology, 98(3), 392–404. <https://doi.org/10.1037/a0017346>

Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: The competitive altruism hypothesis. Personality and Social Psychology Bulletin, 32(10), 1402–1413. <https://doi.org/10.1177/0146167206291006>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. Behavioral and Brain Sciences, 33(2-3), 61-83. <https://doi.org/10.1017/s0140525x0999152x>

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-Cost, effective method for increasing transparency. PLOS Biology, 14(5), e1002456. <https://doi.org/10.1371/journal.pbio.1002456>

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., … & Sowden, W. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. Advances in Methods and Practices in Psychological Science, 1(4), 443-490.

Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. Advances in Methods and Practices in Psychological Science, 4(1), 251524592095150. <https://doi.org/10.1177/2515245920951503>

Lehmann, G. K., Elliot, A. J., & Calin-Jageman, R. J. (2018). Meta-analysis of the effect of red on perceived attractiveness. Evolutionary Psychology, 16(4), 1474704918802412. <https://doi.org/10.1177/1474704918802412>

Leighton, D. C., Legate, N., LePine, S., Anderson, S. F., & Grahe, J. (2018). Self-esteem, self-disclosure, self-expression, and connection on Facebook: A collaborative replication meta-analysis. Psi Chi Journal of Psychological Research, 23(2), 98–109. <https://doi.org/10.24839/2325-7342.JN23.2.98>

Loo, C. W., & Choy, J. L. F. (2013). Sources of self-efficacy influencing academic performance of engineering students. American Journal of Educational Research, 1(3), 86-92.

Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach’s α, McDonald’s ωt and the greatest lower bound. Social Sciences & Humanities Open, 7(1), 100368.

Nayak, M. S. D. P., & Narayan, K. A. (2019). Strengths and weaknesses of online surveys. Technology, 6(7), 0837-2405053138.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

Rammstedt, B. & Beierlein, C. (2014). Can’t We Make It Any Shorter?. Journal of Individual Differences, 35(4), 212-220.

R Core Team. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>

Riva, G., Teruzzi, T., & Anolli, L. (2003). The use of the Internet in psychology research: Comparison of online and offline questionnaires. CyberPsychology & Behavior, 6(1), 73–80. <https://doi.org/10.1089/109493103321167983>

RStudio Team. (2020). RStudio: Integrated Development Environment for R. RStudio, PBC. <http://www.rstudio.com/>

Van Vugt, M., Roberts, G., & Hardy, C. (2007). Competitive altruism: Development of reputation-based cooperation in groups. In R. Dunbar & L. Barrett (Eds.), Handbook of evolutionary psychology (pp. 531–540). Oxford, England: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198568308.013.0036>

Wagge, J. R., Baciu, C., Banas, K., Nadler, J. T., Schwarz, S., Weisberg, Y., IJzerman, H., Legate, N., & Grahe, J. (2019). A demonstration of the Collaborative Replication and Education Project: Replication attempts of the red-romance effect. Collabra: Psychology, 5(1), 5. <https://doi.org/10.1525/collabra.177>

Wagge, J. R., Brandt, M. J., Lazarevic, L. B., Legate, N., Christopherson, C., Wiggins, B., & Grahe, J. E. (2019). Publishing research with undergraduate students via replication work: The Collaborative Replications and Education Project. Frontiers in Psychology, 10, 247. <https://doi.org/10.3389/fpsyg.2019.00247>

Wagge, J. R., Hurst, M. A., Brandt, M. J., Lazarevic, L. B., Legate, N., & Grahe, J. E. (2023). Teaching research in principle and in practice: What do psychology instructors think of research projects in their courses? Psychology Learning & Teaching, 22(1), 4–19. <https://doi.org/10.1177/14757257221101942>

1. As of October 2023, 480 projects at 174 institutions in 29 countries have been initiated. [↑](#footnote-ref-21)
2. Regarding demographics, the original study reported only information about gender of the participants (male and female). [↑](#footnote-ref-40)