

1 Reproducibility Study on the Overperception of Moral Outrage in Online Social Networks
2 and Its Inflation of Beliefs about Intergroup Hostility

3 Shen Meilin¹, Qiao Lizhu¹, Wu Xizhi¹, Bao Jun¹, & Li Hui¹

4 ¹ School of Psychology, Nanjing Normal University

5 Author Note

6 The authors made the following contributions. Shen Meilin: Reproduction Code,
7 Writing Report, Presentation, Integrated Translation; Qiao Lizhu: Reproduction Code,
8 Writing Report, Preparing PPT; Wu Xizhi: Reproduction Code, Writing Report,
9 Preparing PPT; Bao Jun: Reproduction Code, Writing Report, Preparing PPT; Li Hui:
10 Reproduction Code, Writing Report, Preparing PPT.

Abstract

The original study utilized statistical methods such as multilevel modeling, correlation analysis, and regression analysis to conclude that individual-level overperception of moral outrage online leads to collective-level overperception, thereby distorting social cognition of moral and political attitudes. This study reproduces the original findings using publicly available R code and raw data, achieving a consistency rate of 88.52%. Additionally, this study analyzes the causes of discrepancies in the results, which include authorial errors, issues with code modifications, and rounding-induced biases. These findings underscore the importance of data and code transparency and detailed reporting in academic research.

Keywords: R Language; Reproducibility Study; Multilevel Modeling

Reproducibility Study on the Overperception of Moral Outrage in Online Social Networks and Its Inflation of Beliefs about Intergroup Hostility

Introduction

Selected Literature

Reference: Brady, W. J., McLoughlin, K. L., Torres, M. P., Luo, K. F., Gendron, M., & Crockett, M. J. (2023). Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility. *Nature Human Behaviour*.
<https://doi.org/10.1038/s41562-023-01582-0>

Data and Code: <https://osf.io/gtwsk/>

Literature Introduction

Research Background. In the digital age, social networks have a significant impact on individuals' moral and political cognition. Due to the lack of emotional exchange cues present in real-life interactions, such as body language and tone of voice, social networks can lead users to overperceive negative emotions, particularly moral outrage (Lengel & Daft, 1988; Byron, 2008; Walther & D'Addario, 2001; Weisband & Atwater, 1999). Moral outrage is a mixture of anger and disgust triggered by a perceived moral norm violation, playing a key role in motivating collective action and political behavior (Haidt, 2003; Salerno & Peter-Hagene, 2013; Brady & Crockett, 2019; Spring et al., 2018).

The algorithms of social networks tend to promote content that triggers strong emotional responses, which can further amplify users' perceptions of moral outrage (Brady et al., 2020; Brady et al., 2021; Brady et al., 2023). Additionally, expressing outrage online can enhance personal reputation, motivating users to express more outrage even if it does not align with their actual experiences (Jordan et al., 2016; Jordan et al., 2019; Crockett,

2017). This overperception may lead users to erroneously infer the emotions of the entire group, perceiving the group as more extreme than it actually is (Goldenberg et al., 2021).

This misunderstanding can influence our perceptions of groups, making us more politically extreme. Therefore, understanding how emotions are amplified on social networks is crucial for understanding how people form perceptions of intergroup relations in the digital age. By recognizing these biases, we can more accurately assess the information on social networks and avoid unnecessary negative emotions or extreme political positions due to misunderstandings.

Research Hypotheses. **Hypothesis 1:** Observers will overestimate the outrage expressed by authors on social media. **Hypothesis 2:** The more frequently observers use political media, the more likely they are to perceive authors' outrage. **Hypothesis 3:** Overperceived outrage will amplify people's views of intergroup hostility, leading to greater perceived polarization and ideological extremity in society.

Research Findings. **Studies 1-3:** Observers overestimated the outrage expressed by authors on social media (supporting Hypothesis 1). The frequency of observers' political social media use significantly and positively predicted the overestimation of outrage (supporting Hypothesis 2). **Study 4:** Participants in the high outrage group reported significantly higher levels of collective outrage than those in the low outrage group, indicating that individual-level overperception of outrage amplifies collective-level perception of outrage (supporting Hypothesis 3). **Study 5:** Participants in the high overperception of outrage group believed that expressing outrage was more normative in their social network, and they exhibited significantly higher levels of dislike towards political outgroups than those in the low overperception group. Additionally, participants in the high overperception group perceived greater ideological extremity within their social network (supporting Hypothesis 3).

Research Conclusions. This study found that people tend to overestimate others' feelings of moral outrage on social media, which biases perceptions of collective outrage and affects perceptions of social norms and group polarization. These results suggest that social media platforms may distort people's understanding of social morality and political attitudes by amplifying biases in moral emotion perception, further exacerbating affective polarization and ideological extremity between groups.

Methods

Overview of Original Research Methods

Studies 1-3.

1. **Research Objective:** Investigate the overperception of moral outrage on social media.
2. **Research Procedure:** Studies 1 to 3 used a series of field surveys and behavioral experiments to explore the overperception of moral outrage on social media. Researchers first monitored and collected tweet data on controversial American political topics in real-time using the Twitter API. Then, they used machine learning tools to identify users expressing moral outrage from this data. The researchers contacted these users via direct messages, inviting them to report their actual emotional experiences when composing the tweets. Meanwhile, another independent group of participants was recruited to evaluate the authors' emotions in the same tweets, to measure whether there was a bias in the participants' perceptions of the authors' emotions compared to the authors' self-reported emotions. Participants were also asked to report the frequency of their daily social media use to obtain political information and to quantify this usage on a sliding scale from 0 to 100, where 0 means never using and 100 means very high frequency. Additionally, participants estimated the amount of time they spent daily following political content on social media.

94 **3. Reproduction Approach:** Follow the provided code to clean data, perform
95 descriptive statistics, and apply multilevel models to conclude that there is
96 overperception of moral outrage on social media. Use correlation and regression
97 analysis, as per the original code, to find that political media use predicts the
98 overperception of outrage.

99 **Study 4.**

- 100 **1. Research Objective:** Study 4 tested whether the overperception of individual
101 outrage found in Studies 1-3 amplifies the perception of collective moral outrage.
- 102 **2. Research Procedure:** First, the researchers used previously collected Twitter user
103 data, which included users' self-reported outrage levels and at least ten observers'
104 judgments of these users' outrage levels. Then, they constructed simulated newsfeeds.
105 Using the above data, they created two groups of simulated Twitter newsfeeds. One
106 group was the high-overperception newsfeed, containing tweets where observers
107 reported significantly higher levels of outrage than the authors did. The other group
108 was the low-overperception newsfeed, containing tweets where the observers'
109 judgments of outrage were close to the authors' self-reports. The two newsfeeds were
110 matched on the authors' self-reported outrage levels. Subsequently, participants were
111 randomly assigned to view either the high-overperception or low-overperception
112 newsfeed and were asked to judge the average moral outrage of the social network
113 members.
- 114 **3. Reproduction Approach:** Use the publicly available code to reproduce the analysis
115 with the raw data "study4_data_raw.csv" and "stim_descriptives_data.csv". The
116 "study4_data_raw.csv" includes information such as participants' gender, age, and
117 collective outrage judgments, while the "stim_descriptives_data.csv" includes
118 outrage judgment values for high and low overperception newsfeeds. The original
119 code includes seven parts: loading R packages, importing data, creating functions,

data cleaning, descriptive analysis, exploratory analysis, and data visualization.

Study 5.

1. **Research Objective:** Building on Study 4, the researchers aimed to further investigate whether the conditions of high and low overperception newsfeeds affect the social norm judgments of outrage, affective polarization, and ideological extremity.
2. **Research Procedure:** Participants were randomly assigned to either the high-overperception or low-overperception newsfeed. After viewing the newsfeeds, participants first made judgments about the norms of outrage in a norm judgment task to see how different groups of participants perceived outrage. Each participant viewed all ten tweets, one tweet per trial, and was asked the following question for each trial: “How socially appropriate or inappropriate would it be for someone to post this tweet to the social media network you recently viewed?” Participants responded on a scale from -3 (very socially inappropriate) to 3 (very socially appropriate). The dependent variable of interest was the difference score between the mean appropriateness ratings for high-outrage tweets and neutral tweets, representing the appropriateness of outrage tweets relative to non-outrage tweets. After the norm judgment task, participants also made judgments about group affective polarization. To measure how the conditions affected perceptions of group affective polarization, participants were asked to judge how the social network they viewed felt about the political ingroup and outgroup using a 0-100 feeling thermometer scale, a standard measure of affective polarization. Specifically, participants were asked, “On average, how do you think the people in this network feel about Republicans (Democrats)?” Finally, participants made judgments about ideological extremity. To measure how the conditions affected perceptions of ideological extremity, participants were asked, “What do you think the political ideology of the typical person in the network is?” on a scale from -3 (extremely liberal) to 3 (extremely conservative). Ideological

extremity was defined as the absolute value of the judgments, with higher values indicating greater perceived ideological extremity for both parties.

3. Reproduction Approach: Follow the provided code to reproduce data cleaning, descriptive statistics, two t-tests, and one F-test. The researchers compared the social norm judgments of outrage, affective polarization, and ideological extremity between participants under high and low overperception newsfeed conditions.

R Packages Used

The R packages used by the researchers are introduced as follows:

1. **kableExtra**: An R package for creating complex tables, which can beautify tables and add extra formatting, commonly used for presenting data in academic papers and reports.

2. **psych**: A package containing various psychological statistical methods, suitable for psychological and social science datasets, offering descriptive statistics, factor analysis, reliability analysis, and more.

3. **Hmisc**: Developed by Frank Harrell Jr., this package includes various statistical methods and tools for data summary, transformation, regression analysis, etc.

4. **lsr**: Provides the implementation of Linear Symbolic Regression, a computational method for discovering underlying patterns in data.

5. **rstatix**: A simple package for common statistical tests, such as t-tests and ANOVA, and exploratory data analysis, based on the tidyverse ecosystem.

6. **tidyverse**: A collection of modern, consistent R packages, including ggplot2, dplyr, and tidyr, used for data science, emphasizing tidy data handling and visualization.

7. **car**: Developed by John Fox, this package provides various R functions for data analysis, including ANOVA, regression, and classification analysis.

8. **lmerTest**: Provides estimation and testing functions for linear mixed-effects models, a powerful tool for analyzing hierarchical or grouped data.

Results

Study 1

Descriptive Statistics. Green represents the authors’ reported levels of outrage/happiness, and red represents the participants’ ratings of outrage/happiness. The data distribution is shown using violin plots. The left side displays the violin plot from the original literature, and the right side shows the violin plot generated by our code, which matches the shape of the original plot.

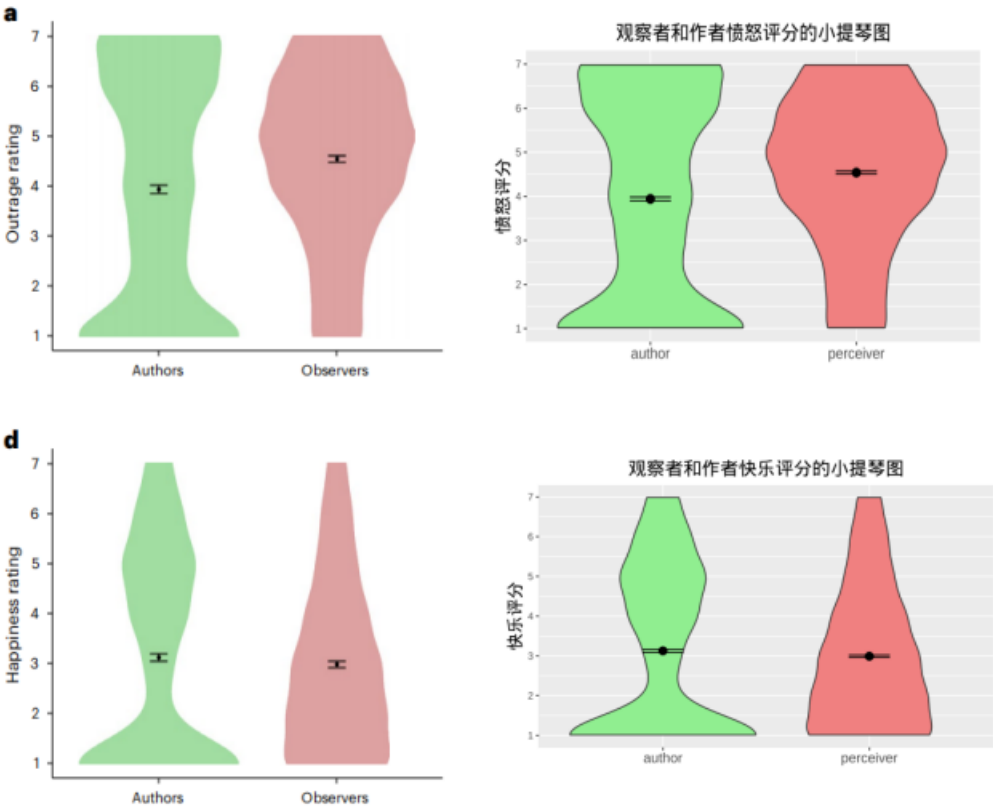


Figure 1 Reproducibility Results of Descriptive Statistics in Study 1

Inferential Statistics. The original study used multilevel modeling to test whether participants tend to overperceive the authors’ outrage/happiness. In the multilevel

model, the random effects were the tweet authors and participants, the predictor variables were the dummy codes for authors and participants, and the clustering variables were the tweet author IDs and participant IDs. The results showed that the participants' reported levels of outrage were significantly higher than the authors' self-reported levels, indicating that participants had an overperception of outrage, $b = 0.59$, $P = 0.011$, 95% CI [0.14, 1.04]. However, participants did not overperceive the authors' happiness, $b = -0.13$, $P = 0.538$, 95% CI [-0.54, 0.28]. The reproduction results were completely consistent with the original study.

Table 1 Reproducibility test results of study 1 Inferential statistics

Study1	Multilevel model (Anger)			Multilevel model (Happiness)		
	b	P	95%CI	b	P	95%CI
Original research	0.59	0.011	(0.14, 1.04)	-0.13	0.538	(-0.54, 0.28)
This research	0.59	0.011	(0.14, 1.04)	-0.13	0.538	(-0.54, 0.28)
δ	0%	0%	0%	0%	0%	0%
rating	identical	identical	identical	identical	identical	identical

Study 2

Descriptive Statistics. The descriptive statistics of Research 2 are presented in the form of violin plots, as verified for reproducibility using R code. As shown below, figures b and e on the left represent the violin plots from the original literature, while those on the right are generated using self-written R code. From the figures below, it is evident that the shape of the plots generated by the code matches that of the original literature.

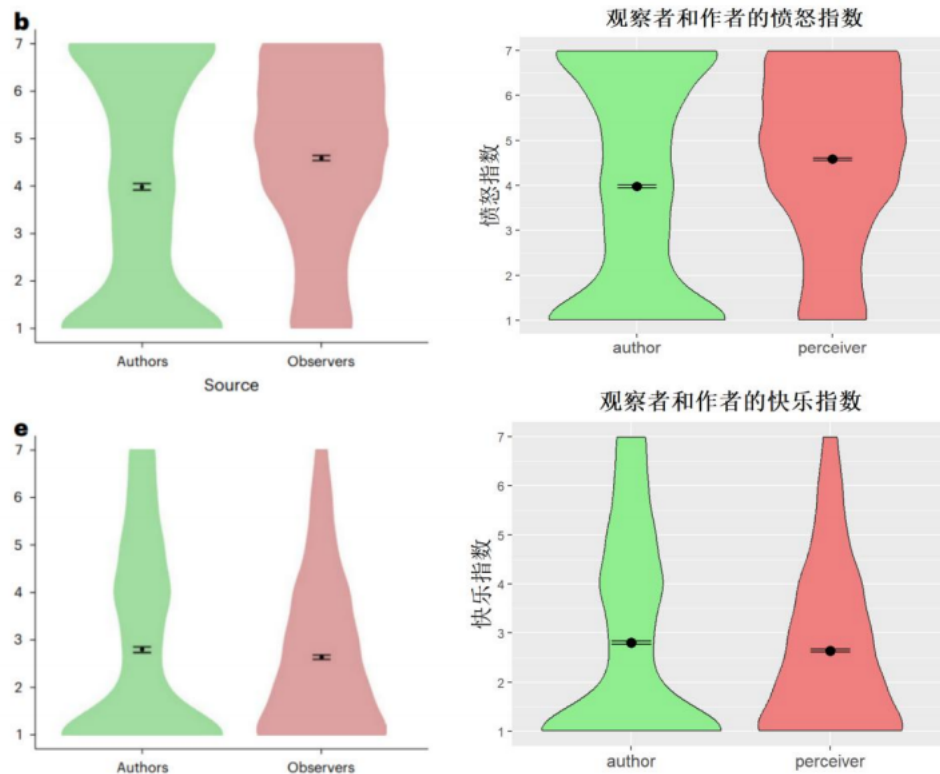


Figure 2 Reproducibility Results of Descriptive Statistics in Study 2

Inferential Statistics. Research 2 utilized a multilevel modeling approach to investigate whether observers tend to overperceive moral outrage in authors. Random effects included the tweet author and observer, with predictor variables encompassing participant sources (authors and observers) and clustering variables comprising tweet ID and observer ID. Upon comparing the inferential statistical analysis results from the original literature with those reproduced using R code, it was found that the model analysis results closely resembled those of the original literature.

Specifically, observers' perception of outrage was significantly higher than authors, $b = 0.58$, $P = 0.001$, 95% CI = (0.25, 0.92). Conversely, there was no significant difference in the overperception of happiness between observers and authors, $b = -0.17$, $P = 0.295$, 95% CI = (-0.49, 0.15). However, there was a minor difference in the confidence interval for happiness scores, where the rounded happiness score computed from the code changed from

0.14 in the original literature to 0.15.

Table 2 Reproducibility test results of study 2 Inferential statistics

Study2	Multilevel model (Anger)			Multilevel model (Happiness)		
	<i>b</i>	<i>P</i>	95%CI	<i>b</i>	<i>P</i>	95%CI
Original research	0.58	0.001	(0.25, 0.92)	-0.17	0.295	(-0.49, 0.14)
This research	0.58	0.001	(0.25, 0.92)	-0.17	0.295	(-0.49, 0.15)
δ	0%	0%	0%	0%	0%	7.14%
rating	identical	identical	identical	identical	identical	bias due to rounding

Study 1&2

The original study utilized combined data from Research 1 _ 2 to verify whether participants' political media usage could predict overperception of anger through correlation and regression analyses. It was found that there was a significant positive correlation between the two variables, $r(222) = 0.19$, $P = 0.004$, with a 95% confidence interval of (0.06, 0.31). Additionally, participants' political media usage significantly predicted overperception of anger, $b = 0.17$, $P = 0.009$. The reproduced results were entirely consistent with the original literature, and scatterplots with regression lines generated using self-written code matched those from the original study (original on the left, self-generated on the right).

Table 3 Reproducibility test results of study 2 Inferential statistics

Study1&2	correlation			regression	
	<i>r</i>	<i>P</i>	95%CI	<i>b</i>	<i>P</i>
Original research	0.19	0.004	(0.06, 0.31)	0.17	0.009
This research	0.19	0.004	(0.06, 0.31)	0.17	0.009
δ	0%	0%	0%	0%	0%
identical	identical	identical	identical	identical	identical

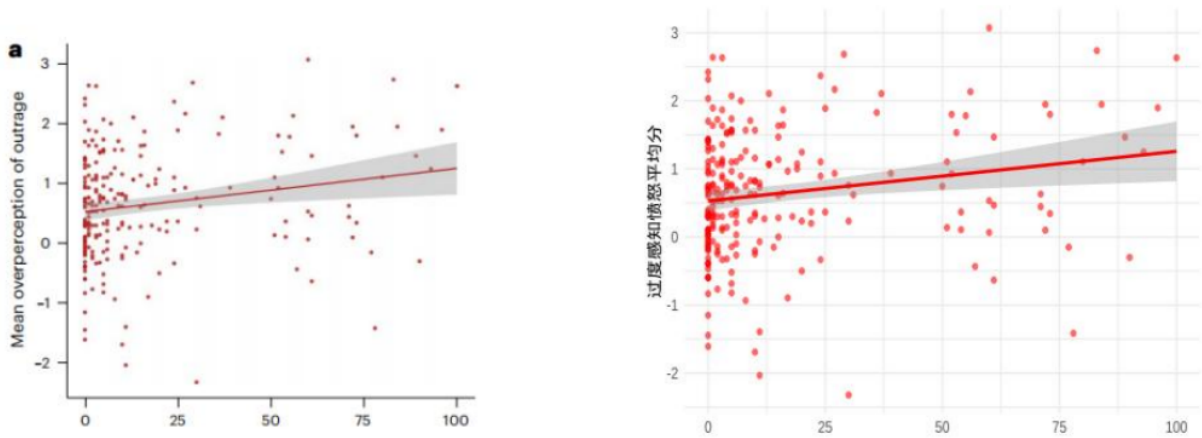


Figure 3 Reproducibility Results of Descriptive Statistics

Study 3

Descriptive Statistics. Research 3 utilized the same sample of authors reporting emotions as Research 2, with an increased participant sample size of 350 individuals. Descriptive statistics are presented in violin plots, depicting the distribution of emotions rated by authors (green) and participants (red), similar to Research 1 and 2. Figures on the left display violin plots from the original literature, while those on the right are generated using self-written code. The shape of the plots generated by the code matches that of the original literature, as shown below.

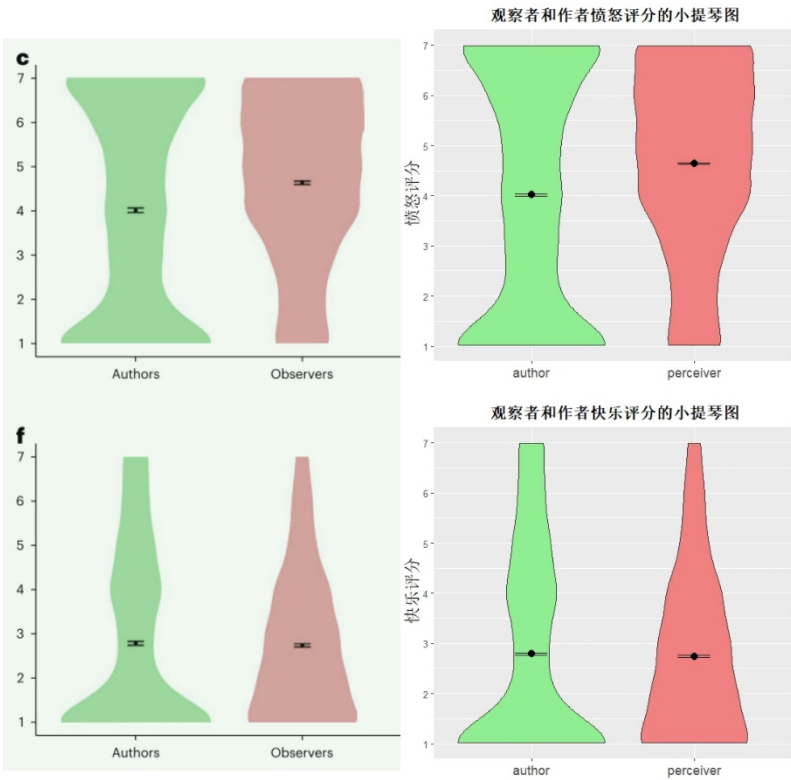


Figure 4 Reproducibility Results of Descriptive Statistics in Study 3

Inferential Statistics. Research 3 employed a multilevel modeling approach to investigate whether observers tend to overperceive authors' moral outrage. Random effects included tweet author and observer, while predictor variables encompassed participant sources (authors and observers), with tweet ID and observer ID serving as clustering variables. Upon comparing the inferential statistical analysis results from the original literature with those reproduced using R code, it was found that the model analysis results were consistent with the original literature.

Specifically, observers' perception of anger was significantly higher than authors, $b = 0.62$, $P < 0.001$, 95% CI = (0.37, 0.88). Conversely, there was no significant difference in the overperception of happiness between observers and authors, $b = -0.06$, $P = 0.656$, 95% CI = (-0.30, 0.19). However, there were differences in the confidence intervals compared to the original literature, as detailed in the table below.

Table 4 Reproducibility test results of study 3 Inferential statistics

Study3	Multilevel model (Anger)			Multilevel model (Happiness)		
	<i>b</i>	<i>P</i>	95%CI	<i>b</i>	<i>P</i>	95%CI
Original research	0.62	<0.001	(0.27, 0.93)	-0.06	0.656	(-0.50, 0.11)
This research	0.62	<0.001	(0.37, 0.88)	-0.06	0.656	(-0.30, 0.19)
δ	0%	0%	29.41%	0%	0%	24.49%
rating	identical	identical	significant deviation	identical	identical	significant deviation

Research 3 also replicated the correlation between overperception of anger and political social media usage, $r(248) = 0.20$, $P = 0.001$, 95% CI = (0.08, 0.32). Multiple regression analysis adjusted for observers' ideological extremism, partisan identity strength, and tendency to overperceive happiness showed that political social media usage is a significant predictor of overperception of anger. The reproduced results were entirely consistent with the original literature, and scatterplots with regression lines generated using self-written code matched those from the original study (original on the left, self-generated on the right).

Table 5 Repeatability test results of correlation and regression analysis for study 3

Study3	correlation			regression	
	<i>r</i>	<i>P</i>	95%CI	<i>b</i>	<i>P</i>
Original research	0.20	0.001	(0.08, 0.32)	0.16	0.002
This research	0.20	0.001	(0.08, 0.32)	0.16	0.002
δ	0%	0%	0%	0%	0%
rating	identical	identical	identical	identical	identical

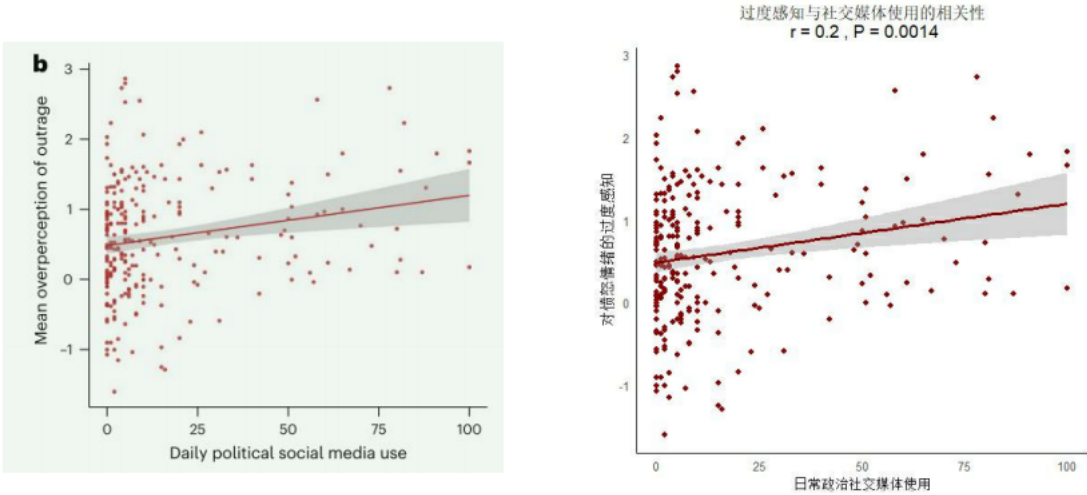


Figure 5 Reproducibility Results of scatter plots in Study 3

Study 4

Descriptive Statistics. Original Study Findings: Recruited 300 Democrats and 300 Republicans. After excluding participants without political party affiliation ($N = 27$) and those who failed comprehension checks ($N = 52$), the final number of participants was 523.

Reproduced Results: The original dataset included a total of 602 participants. After excluding participants without political party affiliation ($N = 27$) and those who failed comprehension checks ($N = 52$), the total number of participants remained 523.

Table 6 Reproducibility test results of study 4 Inferential statistics

	Study4			
	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>
	(initial participants)	(no political)	(failed comprehension check)	(final participants)
Original research	600	27	52	523
This research	602	27	52	523
δ	0.33%	0%	0%	0%
rating	minor deviation	identical	identical	identical

Inferential Statistics.

1. Independent Samples t-test Original Study Findings: Participants' judgments of collective anger on their social networks under conditions of high overperception of news sources ($M = 5.82$) were significantly higher than judgments under conditions of low overperception of news sources ($M = 3.53$), $t(479.50) = 21.56$, $P < 0.001$, Cohen's $d = 1.90$, 95% CI = (2.09, 2.50).

Reproduced Results: $t = 21.56$, $df = 479.5$, $p < 0.001$. The mean for the high overperception of tweets group was 5.82, and for the low overperception of tweets group was 3.53. Cohen's d was 1.91, with a 95% confidence interval for the difference between the means of (2.09, 2.51).

Table 7 Duplicate results of independent sample t test of study 4

	Study4					
	<i>M</i> (collective anger)	<i>M</i> (individual anger)	<i>t</i>	<i>p</i>	Cohen's <i>d</i>	95%CI
Original research	5.82	5.30	8.06	<0.001	0.49	(5.69,5.95)
This research	5.82	5.30	8.06	<0.001	0.49	(5.69,5.95)
δ	0%	0%	0%	0%	0%	0%
rating	identical	identical	identical	identical	identical	identical

Single-Sample t-test.

2. High Overperception of Tweets Group Original Study Findings: Under conditions of high overperception, participants' perception of collective anger ($M = 5.82$) was significantly greater than the perceived anger mean ($M = 5.30$) when viewing each news source individually, $t(264) = 8.06$, $P < 0.001$, $d = 0.49$, 95% CI = (5.69, 5.95). Reproduced Results: $t = 8.01$, $df = 264$, $p < 0.001$, 95% CI = (5.69, 5.95), Cohen's $d = 0.49$. The mean perception of collective anger was 5.82, while the mean perception of anger when viewing each news source individually was 5.30.

Table 8 Single-Sample t-test (High Overperception of Tweets)

	Study4					
	M (collective anger)	M (individual anger)	t	p	Cohen's d	95%CI
Original research	5.82	5.30	8.06	<0.001	0.49	(5.69,5.95)
This research	5.82	5.30	8.06	<0.001	0.49	(5.69,5.95)
δ	0%	0%	0%	0%	0%	0%
rating	identical	identical	identical	identical	identical	identical

3. Low Overperception of Tweets Group Original Study Findings: Under conditions of low overperception, participants' perception of collective anger did not differ significantly from the perceived anger mean when viewing each news source individually, $t(252) = 1.38$, $P = 0.169$, $d = 0.09$, 95% CI = (3.36, 3.69).

Reproduced Results: $t = 1.38$, $df = 252$, $p = 0.169$, 95% CI = (3.36, 3.69), Cohen's $d = 0.09$. The mean perception of collective anger was 3.53, while the mean perception of anger when viewing each news source individually was 3.41.

Table 9 Single-Sample t-test (low Overperception of Tweets)

Study4				
	<i>t</i>	<i>p</i>	Cohen's <i>d</i>	95%CI
Original research	1.38	0.169	0.09	(3.36,3.69)
This research	1.38	0.169	0.09	(3.36,3.69)
δ	0%	0%	0%	0%
rating	identical	identical	identical	identical

Finally, visualize observers’ perceptions of collective anger under two different conditions (high overperception and low overperception). Use a dashed line to represent the mean perception of individual anger by observers, points for observers’ perceptions of collective anger, and bars for the mean perception of collective anger. Ensure that the shape of the generated plot matches that of the original literature (left: original plot,right: generated plot).

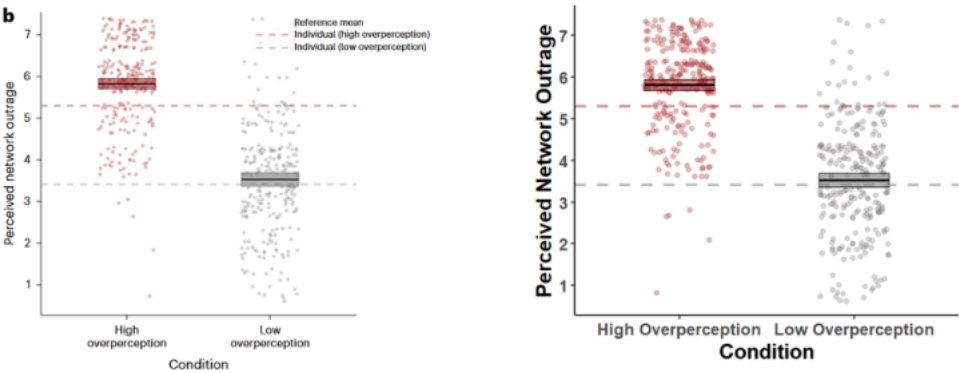


Figure 6 Visualized repeatable test results in Study 4

Study 5

Descriptive Statistics. Researchers recruited 600 Democrats and 600 Republicans to participate jointly in a preregistered study on making social judgments [https://osf.io/mjftk]. On the OSF platform, researchers reported how they determined

sample size, data exclusion criteria, and all measurement methods. In this study, participants without political party affiliation ($N = 100$) and those who failed comprehension checks ($N = 87$) were excluded, resulting in a final sample size of $N = 1013$. Replication results were consistent with the original study findings.

Table 10 Descriptive Statistics of Participant Replication Results

	Study5	
	$N(\text{original quantity})$	$N(\text{filtered quantity})$
Original research	1200	1013
This research	1200	1013
δ	0%	0%
rating	identical	identical

In Study 5, the researchers computed frequencies for gender, party affiliation, and familiarity, and created histograms for age distribution and political ideology. The original study did not provide mean and standard deviation for age, only histograms of age distribution. In the replication, the mean age was calculated as $M = 36.96$, with a standard deviation of $SD = 13.95$.

According to the author’s provided code, the frequency for Democrats was 533, accounting for 52.67% of the total participants; the frequency for Republicans was 479, accounting for 47.33% of the total participants. Ideological histograms were generated based on the provided code.

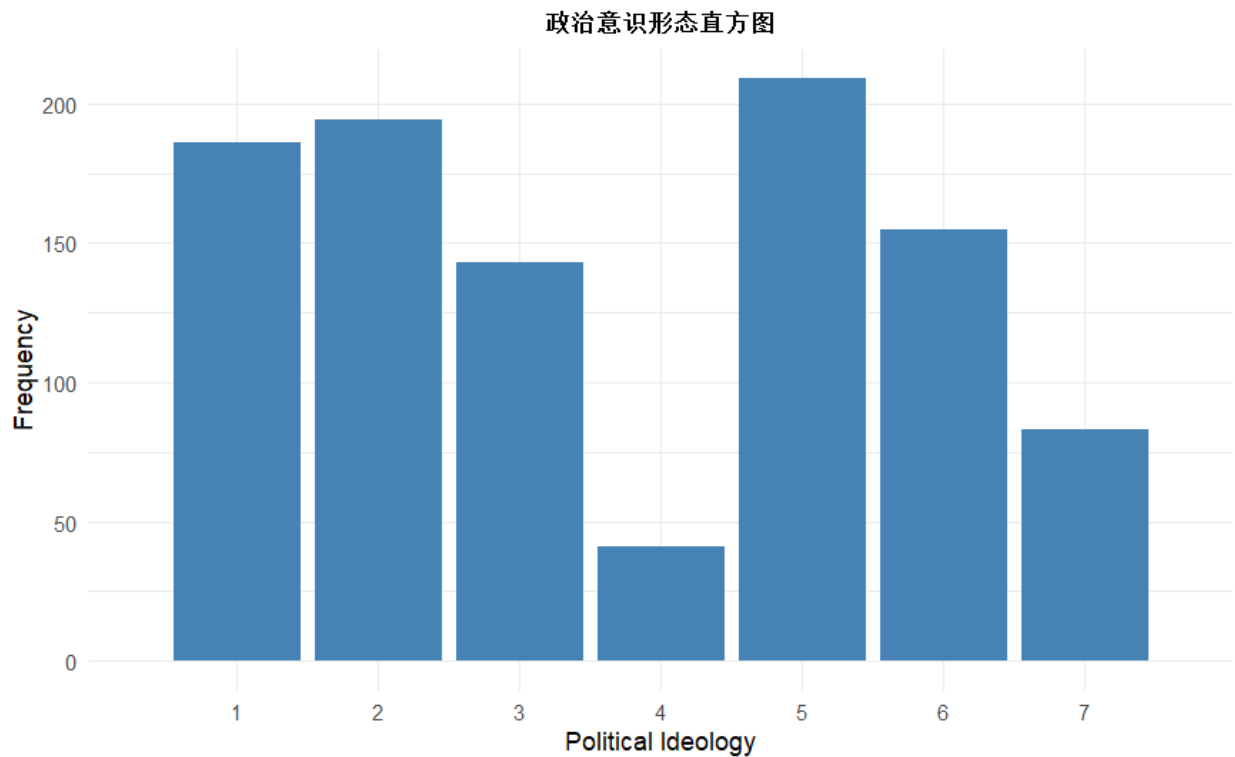


Figure 7 Political Ideology Histogram in Study 5

Inferential Statistics.

1. Perception of Anger across Different News Input Groups Researchers analyzed the distribution, homogeneity of variance, significance, and effect size of data under conditions of high and low perception of anger. The sign of the t-value indicates the direction of the mean difference between the two groups. The original results showed: $t(964.58) = -6.89, p < 0.001, d = 0.43, 95\% \text{ CI} = (0.47, 0.85)$; our replication results showed: $t(964.58) = 6.89, p < 0.001, d = 0.43, 95\% \text{ CI} = (0.47, 0.85)$. Except for the sign of the t-value, our results are consistent with the original study. A positive t-value indicates that the appropriateness scores of the high perception group are higher than those of the low perception group, suggesting that the high perception group perceives anger as more suitable for society.

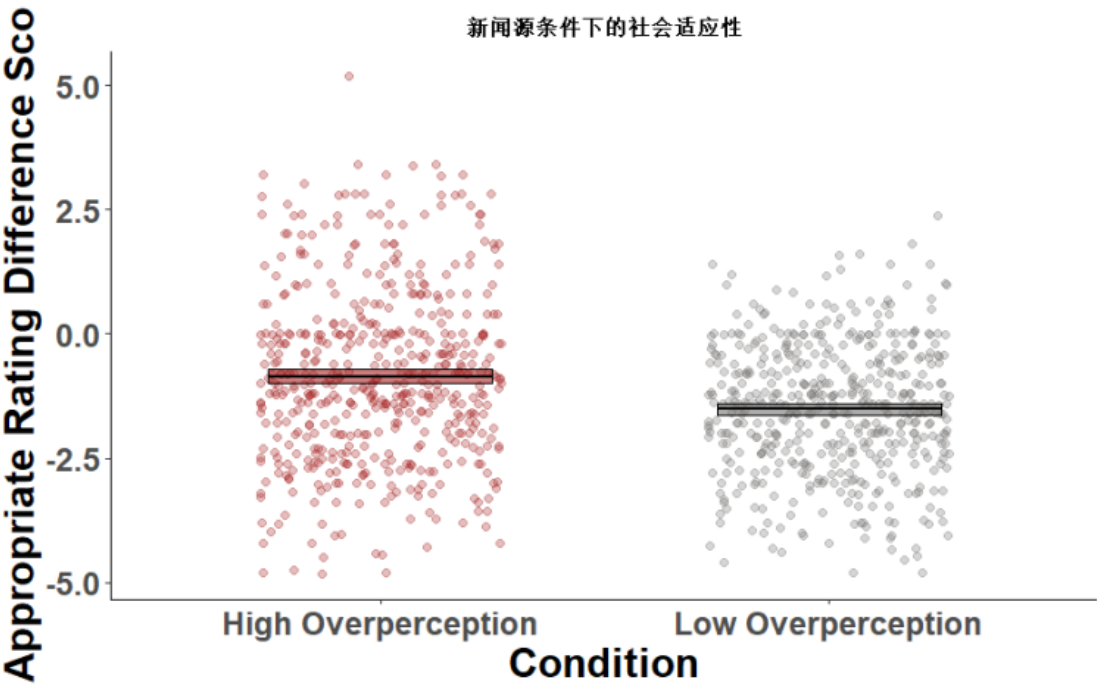


Figure 8 Social Adaptability under the Conditions of News Sources in Study 5

Table 11 Description of the Number of Participants with Repetition Results Statistics in Study 5

t test(Anger perception)				
	<i>t</i>	<i>p</i>	<i>Cohen's d</i>	95%CI
Original research	-6.89	<0.001	0.43	(0.47, 0.85)
This research	6.89	<0.001	0.43	(0.47, 0.85)
δ	200%	0%	0%	0%
rating	significant deviation	identical	identical	identical

2. Does perceived anger amplify group polarization? The original results found that the analysis of variance showed a significant interaction between conditions and online networks, $F(1, 1010) = 369.58, p < 0.001, \eta^2_p = 0.27$. Participants who viewed news from highly anger-inducing sources expressed greater liking for their political in-group compared to those who viewed news from sources inducing lower

levels of anger ($p < 0.001$). Participants who viewed news from highly anger-inducing sources also reported higher dislike for political out-groups on their social networks, whereas those who viewed news from sources inducing lower anger reported lower dislike for political out-groups, with significant differences between the two groups ($p < 0.001$). Post hoc tests indicated that the effect of news source condition on increasing dislike for political out-groups was nearly twice as strong as its effect on increasing liking for political in-groups. These findings suggest that when individuals perceive excessive anger in news sources, they tend to perceive a significant emotional polarization in their social networks. Our test results align with the original findings, where the analysis of variance showed $F(1, 1010) = 369.58$, $p < 0.001$, $\eta^2_p = 0.27$. The source code did not successfully produce results or statistical figures, and did not include the post hoc test section, which was also absent from the original manuscript. After modifying the code, the following figure was obtained.

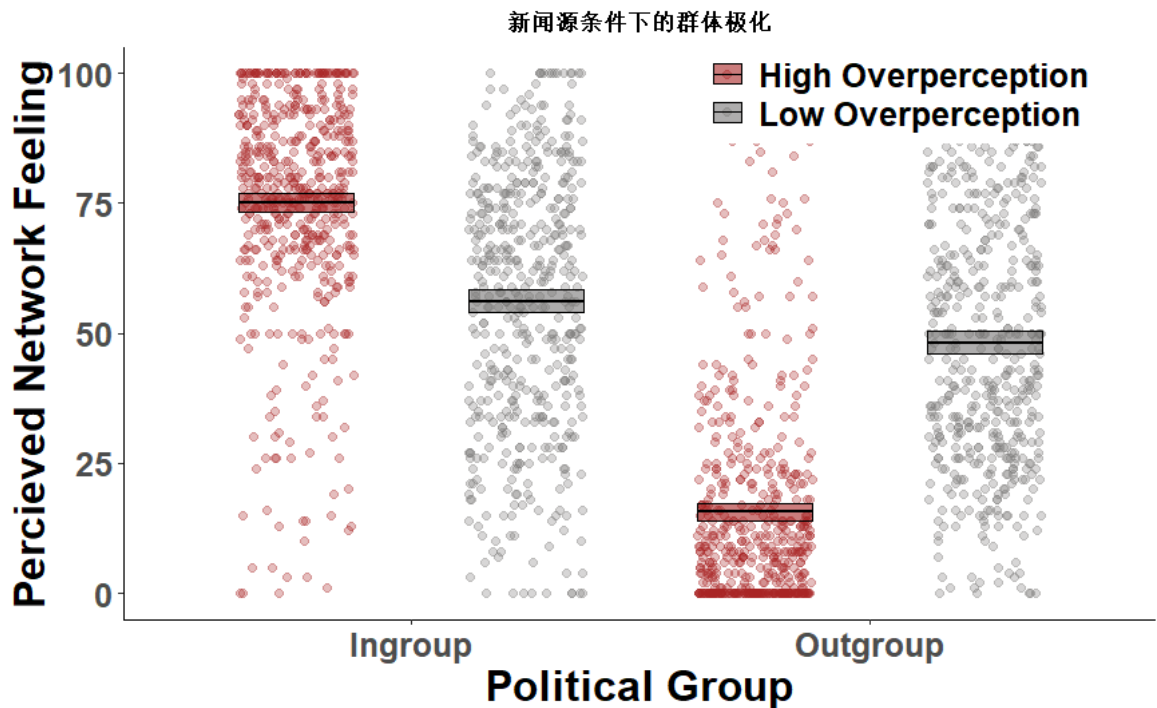


Figure 9 Group Polarization under News Source Conditions in Study 5

376

Table 12 Results of F-test (Group Polarization) in Study 5

	F-test (Anger Perception)		
	<i>F</i>	<i>p</i>	<i>n</i> ² _{<i>p</i>}
Original research	369.58	<0.001	0.27
This research	369.58	<0.001	0.27
<i>δ</i>	0%	0%	0%
rating	identical	identical	identical

377

378 3. The impact of anger perception on group polarization In the original study,
379 researchers found that compared to participants who viewed news from sources
380 inducing lower levels of anger, those who viewed news from sources inducing higher
381 levels of anger perceived their networks to be more ideologically extreme, $t(1003.60)$
382 $= -11.39$, $p < 0.001$, $d = 0.72$, 95% CI = (0.46, 0.64). Replication results closely
383 aligned with the original findings, except for a discrepancy in the sign of the t-value,
384 which was $t(1003.60) = 11.39$, $p < 0.001$, $d = 0.72$, 95% CI = (0.46, 0.64).
385 Integrating findings from Study 5 suggests that excessive perception of anger in news
386 sources directly amplifies perceptions of norms for expressing anger, emotional
387 polarization, and ideological extremism in social networks.

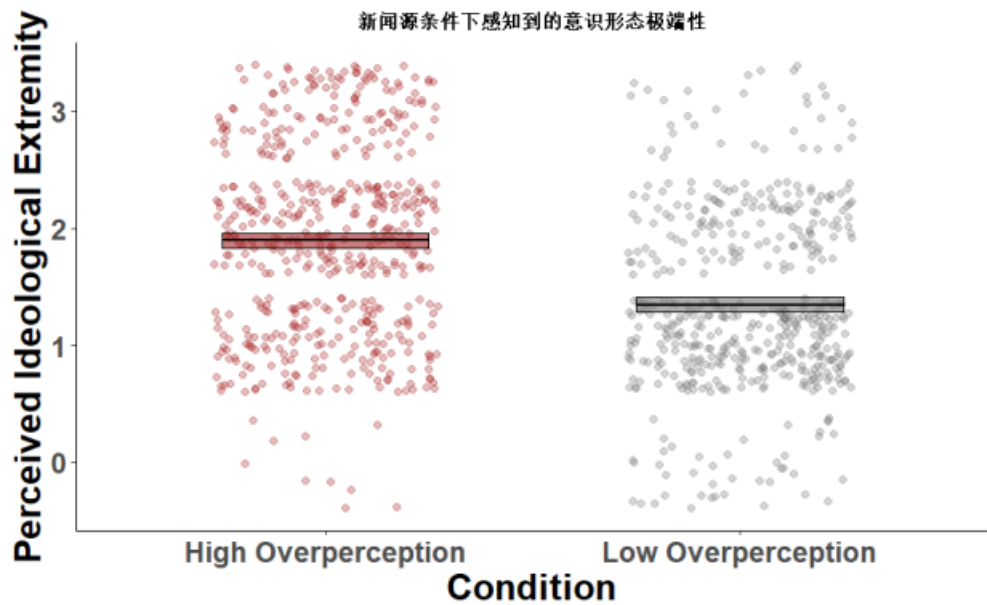


Figure 10 Perceived Ideological Extremity under News Source Conditions in Study 5

Table 13 Test Results (Ideological Extremity) in Study 5

t-test (Ideological Extremity)				
	<i>t</i>	<i>p</i>	<i>Cohen's d</i>	95%CI
Original research	-11.39	<0.001	0.27	0.72
This research	11.39	<0.001	0.27	0.72
δ	200%	0%	0%	0%
rating	significant deviation	identical	identical	identical

Summary and Discussion

Reproducibility Assessment in the Original Study

Table 14 Analysis of Reasons for Non-reproducibility

Reproducibility Status	Number and Proportion	
	<i>N</i> *	%
Completely Consistent ($\delta = 0\%$)	54	88.52
Slight Deviation ($0\% < \delta < 10\%$)	1	1.64
Significant Deviation ($\delta > 10\%$)	4	6.56
Deviation due to rounding	2	3.28
Unable to conduct reproducibility test	0	0

395

396 In this reproducibility study, we identified several instances where results deviated or
397 could not be replicated. These issues primarily centered around the following areas. Firstly,
398 some results could not be replicated due to incomplete data or insufficient information
399 provided in the original paper. For example, age data of participants were not disclosed in
400 the original paper, making comparisons impossible. Secondly, discrepancies arose from
401 inconsistencies between the original data and code files. This included operations in the
402 code that lacked corresponding data entries, complicating the replication process.
403 Additionally, ambiguities were found in the procedural reporting, such as unclear or
404 incorrect reporting of variables used in regression models and criteria for subgroup
405 inclusion. Descriptive statistics also exhibited deviations; for instance, the reported original
406 participant count of 600 individuals in the paper differed from the actual count of 602
407 individuals in the dataset, impacting the analysis process. In the statistical inference part,
408 rounding discrepancies were noted; for example, the confidence interval for happiness index
409 in Study 2 shifted from 0.14 as reported in the paper to 0.15 due to rounding, affecting the
410 interpretation of results. Some confidence intervals or effect sizes also differed slightly from
411 the original paper; for instance, Study 3 showed variations in confidence intervals between
412 observers and authors in the perception of excessive happiness. Furthermore, discrepancies
413 in the sign of statistical test values were observed; for example, in Study 4, although the
414 numerical value of the t-statistic was consistent, the opposite sign indicated a different

direction in the mean differences between groups. While most results aligned with the original paper during replication, these minor deviations in critical findings underscore the need for rigorous validation. These findings emphasize the importance of transparent data and code sharing practices, as well as detailed reporting in academic research to enhance reproducibility and credibility.

Analysis of Reproducibility Test Results

For Study 3, the inconsistent confidence interval results can be attributed to differences in sample sizes and data collection procedures compared to Studies 1 and 2. Additionally, the absence of a random seed setting in Study 3's data collection process may have contributed to these discrepancies. In the descriptive statistics section of Study 4, regarding participant numbers, the original paper reported 600 participants, while the dataset actually contained 602 participants, and all analyses were conducted based on the latter number. This discrepancy is suspected to be an authorial error, impacting reproducibility. In Study 5, the original code encountered issues and could not run smoothly. We made modifications to the "mutate" function within "descriptives," specifically changing the levels from "Over perception" and "Accurate Perception" to "High Overperception" and "Low Overperception." This modification was necessary because the data did not contain columns labeled "Over perception" and "Accurate Perception." These issues may also stem from problems with code modifications, authorial errors, software package or version issues, or delayed updates by the original authors.

Further Reflections

In terms of course learning, this reproducibility study deepened our understanding of R programming and statistical methods, particularly in handling complex datasets and conducting multilevel model analyses practically. Moreover, in scientific research, we gained insights into research methodologies and recognized the importance of transparent

data and detailed reporting in ensuring the reliability and reproducibility of research results. Regarding teamwork, this project enhanced our collaborative and communication skills, ensuring smooth project progress through clear division of labor and mutual support. Finally, concerning coursework, the integration of practical tasks with theoretical learning equipped us not only with technical skills but also fostered critical thinking, which is crucial for future research endeavors. We suggest that courses include more practical components to better assist students in understanding and applying the knowledge they acquire.

References

Brady, W. J., McLoughlin, K. L., Torres, M. P., Luo, K. F., Gendron, M., &
Crockett, M. J. (2023). Overperception of moral outrage in online social networks inflates
beliefs about intergroup hostility. *Nature Human Behaviour*.
<https://doi.org/10.1038/s41562-023-01582-0>