A comment on

"The relationship between political ideology and judgements of bias in distributional outcomes"

Clemens Lindner [1] & David F. Urschler [2]*

[1] Friedrich Schiller University Jena, Department of Psychology, Germany
[2] HSD Hochschule Döpfer University of Applied Sciences, Cologne, Germany

*All authors contributed equally to this work. Authors listed in alphabetical order.

Abstract

Kim and Zauberman (2024) examined the relationship between political ideology and judgements of bias in distributional outcomes. They proposed that people's judgements of bias vary with the target group's characteristics and the observer's political ideology. Across 26 studies (two in the main text and 24 in the supplements), they showed that conservatives set a higher threshold for recognizing bias against traditionally non-dominant targets than liberals. Conversely, liberals set a higher threshold for recognizing bias against traditionally dominant targets than conservatives. However, the relationship between political ideology and the recognition of biases reduced when targets had been unknown or irrelevant. Thus, highlighting the impact of context-dependency of bias judgements and the importance of stimulus sampling. These findings were corroborated by an internal meta-analysis. We successfully reproduced the main findings of the article (i.e., the two studies presented in the main text and the internal meta-analysis) by the clean data and the analysis code provided by the authors. However, we noticed some issues including: no access to the raw data, non-reproducible application of exclusion criteria, divergence between versions of the provided data and the analysis script, and multiple measures of conservatism (moderator) with no rationale given why the specific measure was used in the article. Despite these issues, our robustness checks provide further support for the article's findings.

1. Introduction

Kim and Zauberman (2024), henceforth KZ, found that conservatives set a higher threshold for recognizing bias against traditionally non-dominant targets than Liberals. Conversely, Liberals set a higher threshold for recognizing bias against traditionally dominant targets than conservatives. However, the relationship between political ideology and the recognition of biases reduced when targets had been Unknown or irrelevant. Thereby emphasizing the impact of context-dependency of bias judgements and the importance of stimulus sampling. These findings were underpinned by extensive robustness checks.

KZ recruited online workers from two platforms (i.e., MTurk and Prolific) to test their predictions. Additionally, they obtained a nationally representative sample stratified by age, gender, race, and political affiliation through a market research firm (ROI Rocket). In Study 1, participants were randomly assigned to one of five target groups (Men, Women, Black people, White people, or an Unknown target). In Study 2, participants were randomly assigned to evaluate bias against one of three types of college applicants: Liberal, Conservative, or Unknown.

In the present report, which is part of the Göttingen Replication Games workshop in Germany, we investigate whether KZ's results are computationally reproducible based on the code they provided. Beyond KZ's extensive robustness checks, we conducted additional analyses to further validate their findings. Specifically, we (1) developed our own R code using different approaches to obtain the same result; (2) performed quantile regressions and computed estimated marginal means; (3) re-ran the interaction models across all targets and moderator measures (both raw and standardized scores); and (4) calculated the bootstrapped 95% confidence intervals for the floodlight analyses in both Study 1 and Study 2. For the internal meta-analysis, we ran a stress-test on the correlations for each of the target groups by drawing random samples with three sampling proportions (30, 50 and 70 percent) from the original sample. For each sampling proportion, we drew 1000 random samples and subsequently calculated mean and variance of the correlation across random subsamples (within sampling proportion).

2. Computational Reproducibility

KZ provide detailed information to reproduce Study 1 and Study 2, and the internal meta-analysis in the main text, as well as for those studies in the supplements. They provide pre-registrations, additional description about each study's method, information about exclusion criteria, cleaned data sets, R code, and a comprehensive data dictionary. All information can be retrieved from the authors OSF repository (https://osf.io/sa35x/) and from the article's supplementary information (https://shorturl.at/BpgYS). Despite this transparent and comprehensive information, the authors did not provide raw data sets, cleaning codes, and the complete study materials. Furthermore, the authors provided several versions of the studies' data without documenting which changes have been made in the data between ‘older’ and ‘recent’ versions.

We focused on reproducing each study based on the provided data set and R code, and information from the main article. Two versions of the data were tested, (1) the version used in the KZ’s original R code; and (2) a version denoted ‘most recent’. Given that no cleaning codes were provided, we could not reproduce the cleaning process. Overall, we found that findings were computationally reproducible from the provided data (for both versions of the data) with the accompanying R code. However, we found minor discrepancies between the information in the main article and the R code. These discrepancies do not alter KZ findings. In addition, it remained unclear which version of the data set should be used to reproduce the findings in the article. To be clear, we highly appreciate that KZ transparently provide different versions of their data sets. Again, the usage of different versions does not alter their findings. Our codes/programs are available on the Open Science Framework: https://osf.io/9r8w7/

2.1 Study 1

KZ provided cleaned data, data for additional analyses, R code, and a data dictionary. The authors transparently reported exclusions criteria in Supplementary Table 19. However, they did not provide raw data and a cleaning code, nor did the data sets include codings for each exclusion criterion, except for the completion of each study. Therefore, we could not reproduce the cleaning process and the results from raw data.

KZ registered the study and provided a pre-analysis plan (PAP). The PAP specifies the exact design of the study, the data collection process, exclusion criteria, and how the data will be analyzed. While the authors adhered to the PAP, we identified one discrepancy between the article's description and the accompanying R code, which we elaborate on below. Despite this, our results closely match those reported in the original article.

2.1.1 Correlations Between Bias Threshold and Target of the Bias

KZ exactly followed the PAP by running correlations between bias threshold and political ideology for all five target groups. We successfully reproduced all parameters from Table 1 (top half) from the main text based on the provided data set and R code.

2.1.2 Political Ideology by Bias Target Interaction

Again, KZ exactly followed the PAP by conducting six regression analyses in which bias threshold served as the dependent variable, political ideology as the moderator, and various binary predictors representing different target group comparisons: Women vs. Men, Black vs. White, Men vs. Unknown, Women vs. Unknown, White vs. Unknown, and Black vs. Unknown. We successfully reproduced all parameters reported in the bottom half of Table 1 in the main text using the provided dataset and R code. In addition, we were able to fully reproduce all floodlight analyses used to determine the range of political ideology values for which target type significantly affected perceived bias.

However, we identified a discrepancy between the article and the accompanying R code. While the article states that political ideology was mean-centered before being entered into the regression models (along with target group and their interaction), the R code uses the uncentered political ideology variable. This discrepancy affects the regression coefficient estimates but does not alter the *P*-values, R^2, or ΔR^2. Therefore, it has no impact on the main findings.

Table 1. Overview of accessible materials and successful reproduction (fully, partial or not) for Study 1.

	Fully	**Partial**	**No**
Raw data provided			x
Cleaning code provided		x	
Analysis data provided	x		
Analysis code provided	x		
Reproducible from raw data			x
Reproducible from analysis data	x		
Reproducible from pre-registration		x	
Reproducible from Paper		x	
Reproducible with their analysis code	x		

2.2 Study 2

KZ stated that they followed the same analysis procedure as outlined in the (PAP) for Study 1, although they did not provide a separate PAP for Study 2. As with Study 1, KZ made available cleaned data, data for additional analyses, R code, and a data dictionary. The exclusion criteria were transparently reported in Supplementary Table 19. However, raw data and data cleaning scripts were not provided, and the datasets did not include codings for individual exclusion criteria - except for study completion status. As a result, we were unable to reproduce the data cleaning process or verify results using raw data. Nonetheless, we successfully reproduced the findings of Study 2.

2.2.1 Correlation Between Bias Threshold and Political Ideology

KZ conducted three separate correlation analyses between bias threshold and political ideology, corresponding to Liberal applicant targets, Conservative applicant targets, and Unknown targets. We successfully reproduced all reported correlation coefficients from the main text using the provided dataset and R code.

2.2.2 Political Ideology by Bias Target Interaction

KZ conducted three regression analyses with bias threshold as the dependent variable, political ideology as the moderator, and binary predictor variables representing different target group comparisons: Liberal vs. Conservative applicants, Liberal vs. Unknown applicants, and Conservative vs. Unknown applicants. We successfully reproduced all reported parameters using the provided dataset and R code. In addition, we fully reproduced the floodlight analyses used to identify the range of political ideology values for which target type significantly affected perceptions of bias.

2.3. Internal Meta-Analysis of Correlations between Political Ideology and Bias Threshold

We successfully reproduced the results of the internal meta-analysis based on the provided data set and R code.

Table 2. Overview of accessible materials and successful reproduction (fully, partial or not) for Study 2.

	Fully	Partial	No
Raw data provided			x
Cleaning code provided		x	x
Analysis data provided	x		
Analysis code provided	x		x
Reproducible from raw data			x
Reproducible from analysis data	x		
Reproducible from pre-registration		x	
Reproducible from Paper	x		
Reproducible with their analysis code	x		

Table 3. Overview of accessible materials and successful reproduction (fully, partial or not) for the internal Meta-Analysis.

	Fully	Partial	No
Raw data provided			x
Cleaning code provided		x	x
Analysis data provided	x		
Analysis code provided	x		x
Reproducible from raw data			x
Reproducible from analysis data	x		
Reproducible from pre-registration		x	
Reproducible from Paper	x		
Reproducible with their analysis code	x		

3. Robustness Reproduction

We commend KZ for the encompassing robustness checks they conducted to underpin the rigor of the article’s findings. To further examine the robustness of the article's findings, we developed our own R code, performed quantile regressions, computed estimated marginal means, and calculated 95% confidence intervals for the floodlight analyses for Study 1 and Study 2. For the internal meta-analysis, we stress-tested the correlations by drawing 1000 random subsamples (proportions 30, 50, and 70 percent) from the original sample and calculated the distribution of correlations coefficients across the subsamples.

3.1 New R-script

The original authors primarily used the "kim" package (https://rdocumentation.org/packages/kim/versions/0.6.1) for their analyses. Accordingly, we aimed to reproduce the original results with new code based on different packages, approaches and functions. As mentioned above, the first step in reproducing the analyses with our own code was to use the information from the manuscript and supporting information. If the information from the manuscript was not sufficient, we looked at the original authors' code to clarify how the analyses were set up.

3.1.1 Correlations Between Bias Threshold and Target of the Bias

We ran correlation analyses and were able to reproduce all coefficients as reported. Table 4 depicts all reported (manuscript Table 1) and reproduced correlations, 95% confidence intervals and *P*-values.

Table 4.

	n		*r*		*CI 2.5%*		*CI 97.5%*		*P*	
Target	**KZ**	**LU**	**KZ**	**LU**	**KZ**	**LU**	**KZ**	**LU**	**KZ**	**LU**
Men	235	235	-0.14	-0.1354	-0.26	-0.2589	-0.01	-0.0076	0.038	0.038
Women	221	221	0.24	0.2431	0.11	0.1148	0.36	0.3634	<0.001	0.0003
White people	217	217	-0.19	-0.1864	-0.31	-0.3119	-0.05	-0.0546	0.006	0.0059
Black people	227	227	0.25	0.2454	0.12	0.119	0.36	0.364	<0.001	0.0002
Unknown	208	208	0.05	0.0506	-0.09	-0.086	0.19	0.1854	0.468	0.4676

Note. KZ = Kim and Zauberman (reported); LU = Lindner and Urschler (reproduced)

3.1.2 Political Ideology by Bias Target Interaction - Study 1

Table 5 depicts all reported and reproduced R^2 for the model with target x conservatism interaction, R^2 change for models with and without target x conservatism interaction, as well as *F* and *P*-values of the model comparison.

Table 5.

Contrast		*R² Interaction*		*R² change*		*F*		*P*	
Trgt 1	**Trgt 2**	**KZ**	**LU**	**KZ**	**LU**	**KZ**	**LU**	**KZ**	**LU**
ME	WO	0.066	0.0663	0.034	0.0337	16.3	16.3373	<0.001	0.0001
WP	BP	0.048	0.0481	0.047	0.0466	21.5	21.5311	<0.001	0.000005
ME	UK	0.018	0.0183	0.008	0.0084	3.8	3.7753	0.053	0.0527
WO	UK	0.037	0.0365	0.008	0.0078	3.4	3.4443	0.064	0.0642
WP	UK	0.022	0.0218	0.014	0.0139	6.0	5.9853	0.015	0.0148
BP	UK	0.036	0.0364	0.009	0.0093	4.2	4.1774	0.042	0.0416

Note. KZ = Kim and Zauberman (reported); LU = Lindner and Urschler (reproduced); ME = Men; WO = Women; BL = Black people; WP = White people; UK = Unknown

Table 6.

Contrast	*b*	*P*	*JN lower*	*JN upper*
Liberals - Conservatives				
KZ	-	< 0.001	3.55	5.69
LU	-3.682	0.00004	3.554	5.694
Liberals - Unknown				
KZ	-	0.035	3.28	-
LU	-1.710	0.0351	3.283	26.09
Conservatives - Unknown				
KZ	-	0.031	1.09	-
LU	1.972	0.0313	1.087	9.745

Note. KZ = Kim and Zauberman (reported); LU = Lindner and Urschler (reproduced)

3.1.3 Political Ideology by Bias Target Interaction - Study 2

Table 6 depicts all reported and reproduced betas and *P*-values from the model with target x conservatism interaction, as well as Johnson-Neyman bounds.

3.2 Quantile Regressions

Given that ordinary least squares regression (OLS) only focuses on the average of the dependent variable, such as the effect of political ideology on the average effect on the perception of biases against different target groups. However, such effects can be often quite heterogeneous. For example, people who have a lower threshold of bias perception (e.g., those in 0.25 quantile) might be less affected by their political ideology than those who have a higher threshold (e.g., those in 0.75 quantile). Therefore, we ran quantile regressions (QR) to explore these potentially heterogeneous effects (Koenker et al, 2017). Additionally, QR is more robust to violations of assumptions than OLS regression, especially when the assumption of homoscedasticity is violated (Koenker, 2017).

3.2.1 Correlations Between Bias Threshold and Target of the Bias

We conducted an OLS regression for each bias target and examined assumptions of the models (i.e., outliers, normal distribution of the residuals, and homogeneity of variances) In addition, we ran quantile regressions (QR) at five quantiles (0.10, 0.25, 0.50, 0.75, and 0.90) for each bias target and compared model fit indices, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and root mean square error (RMSE).

For Women as targets, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and heteroscedastic residuals. These violations do not bias the estimates of the regression coefficients, but the standard errors and hence significance tests and confidence intervals will be incorrect (Cohen et al., 2003). To address these issues and to gain a more nuanced understanding of the relationship between political ideology and perceived bias, we employed QRs. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 7), providing additional support for the robustness of the original results.

For Men as targets, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. Again, we employed QRs. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 8), hence providing additional support for the robustness of the original results.

For Black people as targets, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. Again, we employed QRs. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 9), hence providing additional support for the robustness of the original results.

Table 7. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	1814.6	1824.8	14.5
QR 10	1734.7	1741.5	20.6
QR 25	1732.1	1738.9	17.9
QR 50	1812.9	1819.7	16.0
QR 75	1927.7	1934.5	18.0
QR 90	1992.7	1999.5	29.3

Table 8. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	1980.2	1990.6	16.1
QR 10	1963.0	1969.9	24.4
QR 25	1965.1	1972.0	21.5
QR 50	2047.4	2054.3	16.6
QR 75	2073.6	2080.5	22.7
QR 90	2035.0	2041.9	29.8

Table 9. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	1910.0	1920.3	16.0
QR 10	1911.1	1918.0	26.5
QR 25	1915.2	1922.1	21.2
QR 50	1968.8	1975.6	16.2
QR 75	1988.0	1994.8	21.5
QR 90	1981.3	1988.1	30.6

For White people as targets, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. Again, we employed QRs. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 10), hence providing additional support for the robustness of the original results.

Table 10. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	1821.9	1832.1	16.0
QR 10	1816.5	1823.2	24.7
QR 25	1810.5	1817.3	20.9
QR 50	1864.7	1871.4	16.4
QR 75	1913.2	1919.9	21.2
QR 90	1896.9	1903.7	30.7

For Unknown targets, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. Again, we employed QRs. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 11), hence providing additional support for the robustness of the original results.

Table 11. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	1759.6	1769.6	16.5
QR 10	1710.0	1716.7	24.7
QR 25	1711.9	1718.6	20.0
QR 50	1801.2	1807.9	19.1
QR 75	1866.4	1873.1	21.0
QR 90	1846.6	1853.3	32.4

3.2.2 Political Ideology by Bias Target Interaction - Study 1

As for the correlations between bias target and bias threshold, we conducted an OLS regression for each bias target comparison, used the mean centered political ideology scaly, and examined the same model assumptions. Again, we ran quantile regressions at five quantiles (0.10, 0.25, 0.50, 0.75, and 0.90) for each bias target comparison and compared model fit indices.

For the Women vs. Men comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and heteroscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 12), hence providing additional support for the robustness of the original results.

Table 12. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	3795.5	3816.1	15.4
QR 10	3704.9	3721.4	22.6
QR 25	3705.0	3721.5	19.8
QR 50	3867.6	3884.1	16.3
QR 75	4001.6	4018.1	20.6
QR 90	4031.3	4047.8	29.5

For the Black vs. White people comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 13), hence providing additional support for the robustness of the original results.

Table 13. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	3730.0	3750.5	16.0
QR 10	3727.6	3744.0	25.6
QR 25	3726.0	3742.3	21.0
QR 50	3833.6	3850.0	16.3
QR 75	3901.2	3917.6	21.3
QR 90	3878.2	3894.6	30.6

For the Women vs. Unknown target comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and heteroscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 14), hence providing additional support for the robustness of the original results.

For the Men vs. Unknown target comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 15), hence providing additional support for the robustness of the original results.

Table 14. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	3575.4	3595.7	15.4
QR 10	3448.4	3464.6	22.6
QR 25	3448.1	3464.3	18.9
QR 50	3619.7	3635.9	17.5
QR 75	3795.8	3812.0	19.5
QR 90	3839.8	3856.1	30.8

Table 15. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	3737.9	3758.3	16.3
QR 10	3673.5	3689.8	24.5
QR 25	3677.5	3693.9	20.8
QR 50	3848.7	3865.1	17.8
QR 75	3940.6	3956.9	21.9
QR 90	3882.9	3899.3	31.0

For the White people vs. Unknown target comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 16), hence providing additional support for the robustness of the original results.

Table 16. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	3579.7	3600.0	16.1
QR 10	3527.0	3543.2	24.7
QR 25	3522.7	3538.9	20.4
QR 50	3666.0	3682.2	17.8
QR 75	3780.2	3796.4	21.1
QR 90	3744.0	3760.2	31.6

For the Black people vs. Unknown target comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 17), hence providing additional support for the robustness of the original results.

Table 17. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	3667.7	3688.1	16.3
QR 10	3622.2	3638.5	25.7
QR 25	3628.3	3644.6	20.7
QR 50	3770.0	3786.3	17.7
QR 75	3855.6	3871.9	21.3
QR 90	3828.4	3844.7	31.6

3.2.3 Political Ideology by Bias Target Interaction - Study 2

We followed the same procedure as for the bias target comparison as for those in Study 1. For the Liberal vs. Conservative target comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and heteroscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 18), hence providing additional support for the robustness of the original results.

Table 18. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	2101.3	2119.2	12.8
QR 10	2023.5	2037.8	18.0
QR 25	1992.4	2006.7	15.3
QR 50	2069.2	2083.5	13.9
QR 75	2219.8	2234.1	14.4
QR 90	2332.6	2346.9	25.6

For the Liberal vs. Unknown target comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and heteroscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 19), hence providing additional support for the robustness of the original results.

Table 19. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	2048.0	2065.9	11.8
QR 10	2000.7	2015.0	16.9
QR 25	1942.0	1956.2	14.0
QR 50	2025.0	2039.3	12.7
QR 75	2154.6	2168.9	13.3
QR 90	2298.7	2313.0	21.5

Table 20. Comparison of model fit indices

	AIC	BIC	RMSE
OLS	2108.0	2125.8	13.3
QR 10	2047.8	2062.0	18.6
QR 25	2011.5	2025.8	15.9
QR 50	2100.9	2115.2	14.0
QR 75	2237.5	2251.8	14.9
QR 90	2346.8	2361.1	26.0

For the Conservative vs. Unknown target comparison, the OLS regression's results matched with those reported in the original article. The tests for assumptions revealed no outliers, non-normally distributed residuals, and homoscedastic residuals. The QR models yielded results consistent with the original findings, while also highlighting variation across the distribution of bias perception. Furthermore, comparison of model fit indices indicated that the OLS model demonstrated good overall fit (Table 20), hence providing additional support for the robustness of the original results.

3.3 Estimated Marginal Means

To further examine the robustness of the original floodlight analyses, we computed pairwise comparisons for each target comparison by applying estimated marginal means (emmeans) for the full range of the political ideology scale in steps of 0.01. We used Šidák correction for multiple testing. This procedure further tests the region of the moderating effect of political ideology.

3.3.1 Political Ideology by Bias Target Interaction - Study 1

For the Women vs. Men comparison, the emmeans analysis revealed that participants with a score of 4.79 on the 7-point political conservatism scale and below showed significantly higher bias thresholds for Men than for Women, $t(452) = 1.98$, $P = 0.049$. Those who scored higher than 4.79 did not show significantly different bias thresholds, $t(452) = 1.95$, $P = 0.052$.

For the Black vs. White people comparison, the emmeans analysis revealed that participants with a score of 3.42 on the conservatism scale and below showed significantly higher bias thresholds for White people than for Black people, $t(440) = 1.96$, $P = 0.049$. Whereas participants with a score of 4.92 on the conservatism scale and above showed significantly higher bias thresholds for Black people than for White people, $t(440) = -1.97$, $P = 0.049$. Those who had conservatism scores higher than 3.42 ($t(440) = 1.95$, $P = 0.052$) and lower than 4.92 ($t(440) = -1.95$, $P = 0.052$) did not show significantly different levels of bias thresholds.

For the Women vs. Unknown target comparison, the emmeans analysis revealed that participants with a score of 4.03 on conservatism scale and below showed significantly lower bias thresholds for Women than for Unknown targets, $t(425) = -1.97$, $P = 0.049$. Those who scored higher than 4.03 did not show significantly different bias thresholds, $t(425) = -1.96$, $P = 0.050$.

For the Men vs. Unknown target comparison, the emmeans analysis revealed that participants with a score of 3.93 on conservatism scale and below showed significantly higher bias thresholds for Men than for Unknown targets, $t(439) = 1.97$, $P = 0.050$. Those who scored higher than 3.93 did not show significantly different bias thresholds, $t(439) = 1.96$, $P = 0.051$.

For the Black vs. Unknown target comparison, the emmeans analysis revealed that participants with a score of 4.83 on conservatism scale and above showed significantly higher bias thresholds for Black people than for Unknown targets, $t(431) = 1.97$, $P = 0.050$. Those who scored lower than 4.83 did not show significantly different bias thresholds, $t(431) = 1.96$, $P = 0.051$.

For the White vs. Unknown target comparison, the emmeans analysis revealed that participants with a score of 3.62 on conservatism scale and below showed significantly higher bias thresholds for White people than for Unknown targets, $t(421) = 1.97$, $P = 0.049$. Those who scored higher than 3.62 did not show significantly different bias thresholds, $t(421) = 1.96$, $P = 0.051$.

In sum, our results underpin the robustness of the floodlight analyses' results in Study 1.

3.3.2 Political Ideology by Bias Target Interaction - Study 2

For the Liberal vs. Conservatism comparison, the emmeans analysis revealed that participants with a score of 3.55 on the conservatism scale and below showed significantly higher bias thresholds for Conservative than for Liberal targets, $t(260) = -1.98$, $P = 0.049$. Whereas participants with a score of 5.69 on the conservatism scale and above showed significantly higher bias thresholds for Liberal than for Conservative targets, $t(260) = 1.96$, $P = 0.050$. Those who had conservatism scores higher than 3.55 ($t(260) = -1.96$, $P = 0.051$) and lower than 5.69 ($t(260) = 1.96$, $P = 0.052$) did not show significantly different levels of bias thresholds.

For the Liberal vs. Unknown target comparison, the emmeans analysis revealed that participants with a score of 3.28 on the conservatism scale and below showed significantly lower bias thresholds for Liberal than for Unknown targets, $t(258) = -1.97$, $P = 0.050$. Those who scored higher than 3.28 did not show significantly different bias thresholds, $t(258) = -1.96$, $P = 0.051$.

For the Conservative vs. Unknown target comparison, the emmeans analysis revealed that participants with a score of 1.12 on the conservatism scale and below showed significantly higher bias thresholds for conservative than for Unknown targets, $t(258) = 1.96$, $P = 0.050$. Those who scored higher than 1.12 did not show significantly different bias thresholds, $t(258) = 1.97$, $P = 0.051$. These results diverge slightly from the original results. Nevertheless, we consider the results of the floodlight analyses' results in Study 2 as robust.

3.4 95% CIs for Floodlight Analyses

To assess the robustness of the Johnson-Neyman (JN) significance regions from the Floodlight Analyses, we conducted a nonparametric bootstrap procedure (1,000 iterations) for each interaction model. For each bootstrap sample, we re-estimated the linear model and computed the JN bounds to identify the range of the moderator (conservatism) at which the predictor had a statistically significant effect on the outcome (bias threshold). From the distribution of the bootstrapped bounds, we derived 95% confidence intervals for both the lower and upper JN thresholds. This approach provides an estimate of the sampling variability of the JN bounds and helps evaluate the robustness of the original results (Hayes, 2022).

3.4.1 Political Ideology by Bias Target Interaction - Study 1

Table 21 shows all the results from the nonparametric bootstrap. The original estimates of the JN bounds are all within the bootstrapped 95% confidence intervals. For some predictors, the 95% confidence interval for the lower (or upper) JN bound could not be determined, as the bound approached infinity in the bootstrap distribution. An infinite JN bound indicates that no meaningful lower (or upper) threshold could be estimated within the observed range of the moderator. Infinite values of the 95% confidence intervals were only observed for the contrasts with the Unknown target groups, suggesting caution when interpreting the significance regions for these specific contrasts.

For the Women vs. Men contrast, only the lower JN bound is within the range on the seven-point conservatism scale (1-7). The upper bound falls outside the scale's range (7.482). Bootstrapping of the upper bound therefore yields a much broader confidence interval (5.554, 20.688).

For the Black vs. White people both JN bounds are within the range on the seven-point conservatism scale (1-7) and within the bootstrapped 95% confidence interval.

For the Men vs. Unknown target contrast, only the upper JN bound (3.931) falls within the range of the seven-point conservatism scale. However, the 95% confidence interval for this bound was unbounded on the upper end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate. The lower JN bound (–119.3) falls far outside the range of the seven-point conservatism scale. Similarly, the 95% confidence interval for this bound was unbounded on the lower end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate.

For the Women vs. Unknown target contrast, only the upper JN bound (4.028) falls within the range of the seven-point conservatism scale. However, the 95% confidence interval for this bound was unbounded on the upper end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate. The lower JN bound (-20.241) falls outside the range of the seven-point conservatism scale. Similarly, the 95% confidence interval for this bound was unbounded on the lower end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate.

For the White people vs. Unknown target contrast, only the lower JN bound (3.615) falls within the range of the seven-point conservatism scale. However, the 95% confidence interval for this bound was unbounded on the lower end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate. The upper JN bound (9.046) falls outside the range of the seven-point conservatism scale. Similarly, the 95% confidence interval for this bound was unbounded on the upper end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate.

For the Black people vs. Unknown target contrast, only the upper JN bound (4.826) falls within the range of the seven-point conservatism scale. However, the 95% confidence interval for this bound was unbounded on the upper end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate. The lower JN bound (-24.469) falls outside the range of the seven-point conservatism scale. Similarly, the 95% confidence interval for this bound was unbounded on the lower end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate.

In sum, our results support the robustness of the floodlight analysis findings in Study 1. We observe stable estimates of the 95% confidence intervals for the JN bounds in contrasts involving known target groups, where the interaction effects are highly significant. In contrast, for comparisons involving Unknown target groups - where interaction effects are non-significant or only marginally significant - we do not obtain robust estimates of the 95% confidence intervals, suggesting instability in the underlying moderation effects. Nonetheless, all observed JN-bounds are within the bootstrapped 95% confidence interval.

Table 21.

Contrast	JN lower	JN upper	Valid Bootstraps
Men vs Women			
Original	4.7940	7.482	
95% CI	(3.957, 5.639)	(5.554, 20.688)	1000
Black vs White people			
Original	3.42	4.917	
95% CI	(1.589, 4.171)	(4.153, 6.691)	1000
Men vs Unknown			
Original	-119.3	3.931	
95% CI	(-Inf, 4.876)	(3.303, Inf)	1000
Women vs Unknown			
Original	-20.241	4.028	
95% CI	(-Inf, 4.941)	(3.205, Inf)	1000
White people vs Unknown			
Original	3.615	9.046	
95% CI	(-Inf, 4.805)	(3.155, Inf)	1000
Black people vs Unknown			
Original	-24.469	4.826	
95% CI	(-Inf, 5.293)	(3.618, Inf)	1000

Table 22.

Contrast	JN lower	JN upper	Valid Bootstraps
Liberals vs Conservatives			
Original	3.555	5.695	
95% CI	(1.219, 4.357)	(4.237, 17.404)	1000
Liberals vs Unknown			
Original	3.284	26.095	
95% CI	(-Inf, 4.233)	(2.622, Inf)	1000
Conservatives vs Unknown			
Original	1.087	3.844	
95% CI	(-Inf, 3.844)	(2.541, Inf)	1000

3.4.2 Political Ideology by Bias Target Interaction - Study 2

We conducted the same analysis as for Study 1 to check the robustness of the JN bounds of the original study. Table 22 depicts all results from the nonparametric bootstrapping.

For the Liberals vs. Conservatives contrast, both JN bounds fall within the range of the seven-point conservatism scale (1–7) and lie within the bootstrapped 95% confidence intervals. Interestingly, the confidence interval for the upper bound is substantially wider (4.237, 17.404), which may reflect a skewed distribution of conservatism in the original sample (i.e., a larger proportion of Liberal participants).

For the Liberals vs. Unknown target contrast, only the lower JN bound (3.284) falls within the range of the seven-point conservatism scale. However, the 95% confidence interval for this bound was unbounded on the lower end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate. The upper JN bound (26.095) falls far outside the range of the seven-point conservatism scale. Similarly, the 95% confidence interval for this bound was unbounded on the upper end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate.

For the Conservatives vs. Unknown target contrast, both upper JN bounds (1.087, 3.844) fall within the range of the seven-point conservatism scale. However, the 95% confidence interval for both bounds was unbounded. For the lower bound, the lower end of the confidence interval was unbounded (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate. Similarly, for the upper JN-bound the 95% confidence interval was unbounded on the upper end (i.e., infinite), indicating substantial uncertainty and a lack of robustness in the estimate.

In sum, our results support the robustness of the floodlight analysis findings in Study 2, since all original JN-bounds fall within the bootstrapped confidence interval. We observe stable estimates of the 95% confidence intervals for the JN bounds for the contrast involving only known target groups (Liberals, Conservatives), where the interaction effects are significant. However, for comparisons involving Unknown target groups - where interaction effects are only marginally significant - we do not obtain robust estimates of the 95% confidence intervals.

3.5 Distribution of correlations for the Internal Meta-Analysis

To assess the robustness of the observed correlations between *conservatism* and *bias threshold*, we conducted a subsampling analysis using three proportions (30%, 50%, and 70% of the original sample) within each bias target. Specifically, we repeatedly (1,000 iterations) drew random subsamples comprising 30%, 50% and 70% of the data (without replacement) and computed the correlation coefficient within each subsample. This allowed us to evaluate how stable the correlations were across different random subsets of the data. By analyzing the distribution of the resulting correlation coefficients - particularly their variability and consistency in direction - we were able to assess the sensitivity of the observed relationships to sampling fluctuations.

Table 23 depicts all mean correlations with standard deviation across the 1000 random subsamples for the three sampling proportions. Our results show that for all three sampling proportions, the mean correlation is very close to the original (unweighted) correlation in the whole sample, further supporting the robustness of the meta-analytical findings coefficients.

Table 23. Raw correlation (unweighted) of the whole sample, mean correlation, and variance across random subsamples within the three sampling proportions for each target group.

		M (SD)		
		Random Subsample (n = 1000)		
Target	**Original r (unweighted)**	**30%**	**50%**	**70%**
Blacks	0.222	0.222 (0.026)	0.221 (0.018)	0.222 (0.012)
Conservatives	-0.205	-0.204 (0.088)	-0.206 (0.057)	-0.204 (0.036)
Immigrants	0.247	0.242 (0.075)	0.247 (0.047)	0.247 (0.032)
Irrelevant	0.067	0.075 (0.129)	0.069 (0.084)	0.069 (0.054)
Liberals	0.132	0.132 (0.093)	0.133 (0.062)	0.133 (0.04)
Men	-0.089	-0.089 (0.033)	-0.089 (0.022)	-0.09 (0.014)
Natives	-0.237	-0.236 (0.093)	-0.238 (0.062)	-0.239 (0.04)
Unknown	0.084	0.083 (0.022)	0.083 (0.015)	0.084 (0.009)
Whites	-0.177	-0.177 (0.031)	-0.179 (0.019)	-0.177 (0.013)
Women	0.184	0.186 (0.028)	0.185 (0.018)	0.184 (0.012)

Figure 1 illustrates the results from the random subsampling technique.

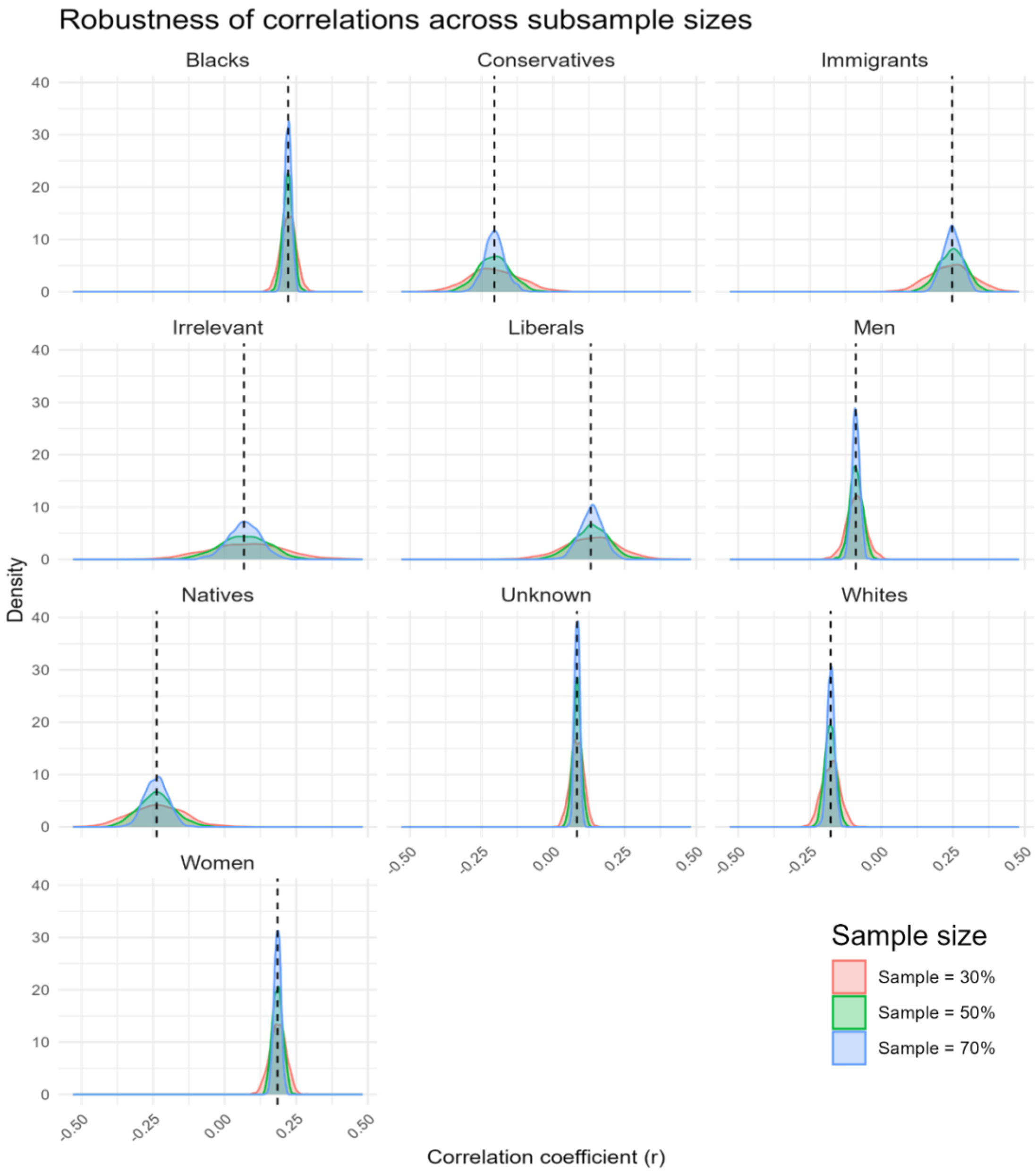

Figure 1. Distribution of correlation coefficients across the 1000 random subsamples for the three sampling proportions (30%, 50%, and 70%) each target group. The dashed line is the original unweighted correlation in the whole sample.

4. Conclusion

Overall, the main results of the paper *"The Relationship Between Political Ideology and Judgements of Bias in Distributional Outcomes"* by Kim and Zauberman (2024) were successfully reproduced using the provided analysis code and cleaned data. In addition, our independent robustness checks further support the results reported in the original article. Despite this successful reproduction and the additional support from our robustness checks, we encountered a few issues.

First, despite the transparent and comprehensive availability of information, no raw data sets were provided. Therefore, we were unable to reproduce the data cleaning process. We argue that the availability of raw data would have further enhanced the transparency of the original findings.

Second, the exclusion criteria were clearly described for all studies, and the dataset included a variable indicating whether participants met the inclusion criteria. However, individual exclusion criteria, aside from successful study completion, were not coded in the dataset. Including filter variables corresponding to each exclusion criterion would ensure that the data cleaning process can be fully traced from raw to cleaned data, thereby improving reproducibility.

Third, while a detailed description of the study materials was provided in the supplementary materials, the complete study materials themselves (as well as the rationale for the selected conservatism measure from other assessed conservatism measures) were not included. We suggest that sharing the full materials, or representative screenshots, would further strengthen the replicability of the studies.

Fourth, we found that the version of the dataset used in the R script was not the most recent version available. Although this discrepancy did not alter the results, clearly indicating which version of the dataset was used would help readers to better track changes and ensure consistency.

Finally, as already outlined above, we identified a discrepancy between the article and the accompanying R code. While the article states that political ideology was mean-centered, the R code uses the uncentered political ideology scale. We would recommend to update the R code to resolve this issue.

To conclude, our reproduction and robustness checks suggest few ways to further enhance transparency, reproducibility, and replicability. Nevertheless, we commend the authors for their strong commitment to open science practices and for findings that are both computationally reproducible and robust.

References

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences, 3rd ed*. Lawrence Erlbaum Associates Publishers.

Hayes, A. F. (2022). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach* (3rd ed.).

Kim, J., & Zauberman, G. (2024). The relationship between political ideology and judgements of bias in distributional outcomes. *Nature Human Behaviour*, *8*(2), 228–242. https://doi.org/10.1038/s41562-023-01779-3

Koenker, R. (2017). Quantile Regression: 40 Years On. *Annual Review of Economics*, *9*, 155–176. https://doi.org/10.1146/annurev-economics-063016-103651

Koenker, R., Chernozhukov, V., He, X., & Peng, L. (Eds.). (2017). *Handbook of Quantile Regression*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315120256