Bergeron-Boutin, Olivier; Ciobanu, Costin; Cohen, Guila; Erlich, Aaron

Working Paper

Replicating Backfire Effects in Anti-Corruption Messaging: A Comment on Cheeseman and Peiffer (2022)

I4R Discussion Paper Series, No. 94

Provided in Cooperation with:
The Institute for Replication (I4R)

Suggested Citation: Bergeron-Boutin, Olivier; Ciobanu, Costin; Cohen, Guila; Erlich, Aaron (2023) : Replicating Backfire Effects in Anti-Corruption Messaging: A Comment on Cheeseman and Peiffer (2022), I4R Discussion Paper Series, No. 94, Institute for Replication (I4R), s.l.

This Version is available at:
https://hdl.handle.net/10419/280691

INSTITUTE for REPLICATION

No. 94
I4R DISCUSSION PAPER SERIES

Replicating Backfire Effects in Anti-Corruption Messaging: A Comment on Cheeseman and Peiffer (2022)

Olivier Bergeron-Boutin

Costin Ciobanu

Guila Cohen

Aaron Erlich

December 2023

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 94

Replicating Backfire Effects in Anti-Corruption Messaging: A Comment on Cheeseman and Peiffer (2022)

Oliver Bergeron-Boutin[1], Costin Ciobanu[2], Guila Gohen[3], Aaron Erlich[3]

[1]Dartmouth College, Hanover/USA
[2]Royal Holloway (University of London), Egham/Great Britain
[3]McGill University, Montreal/Canada

DECEMBER 2023

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

Replicating Backfire Effects in Anti-Corruption Messaging: A Comment on Cheeseman and Peiffer (2022)*

Olivier Bergeron-Boutin, Costin Ciobanu, Guila Cohen, and Aaron Erlich

November 9, 2023

Abstract

Cheeseman and Peiffer (2022) field a survey experiment in Nigeria to test the effect of five different anti-corruption messages on participants' willingness to bribe public officials. They find that these messages generally fail to reduce bribes and could, in fact, *increase* bribes. They further show that these counterproductive effects of anti-corruption messages are especially pernicious for participants who believe corruption is widespread, whom they call "Pessimistic Perceivers." We find that Cheeseman and Peiffer's findings are computationally reproducible: using the same data and estimation procedures, we arrive at the same output reported in the original article. Furthermore, we find that following Cheeseman and Peiffer's strategy to dichotomize a three-item scale used as a moderating variable, their results are robust to different estimation strategies. However, we draw attention to several shortcomings of the original analysis. First, the distribution of the moderating variable is highly skewed: on a 0-1 scale, the mean value is 0.81. Cheeseman and Peiffer's dichotomization procedure is also sensitive to the cutoff threshold and produces unstable results. Similarly, when we employ more flexible estimation strategies for heterogeneous treatment effects when the moderator is measured on a continuous scale, the results appear less robust.

KEYWORDS: Replication study; Corruption; Nigeria.

*Authors: Bergeron-Boutin: Dartmouth College. Ciobanu: Royal Holloway (University of London). Cohen: McGill University. Erlich: McGill University. E-mail: aaron.erlich@mcgill.ca.

1 Introduction

Cheeseman and Peiffer (2022) investigated the impact of different types of anti-corruption messages on behavior using a bribery game. They fielded an in-person survey experiment in Lagos, Nigeria, over three weeks from December 2019 to January 2020. They chose Lagos due to its corruption issues, evident from its low rank in the Corruption Perception Index. Moreover, the diversity, poverty, and inequality levels made Lagos population representative of those facing development issues. They used Afrobarometer's protocol to recruit a representative sample of individuals. A representative sample of 2,572 individuals was recruited, of whom 1,200 participants engaged in a bribery game.

They designed the treatments to replicate information commonly found in anti-corruption campaigns while avoiding overly emotional language or imagery. They included messages about corruption being widespread (*widespread*), condemnation of corruption by religious leaders (*religious*), government success in fighting corruption (*govsuccess*), local community efforts against corruption (*local*), and the connection between corruption and citizens taxes (*taxes*). Cheeseman and Peiffer describe their main results on p. 1082: "We find that exposure to anti-corruption messages fails to discourage corrupt behaviour and, in some cases, makes individuals more willing to bribe. However, this effect is not universal. Instead, the influence of anti-corruption campaigns is conditioned by an individuals preexisting perceptions regarding the prevalence of corruption."

In the control group, which did not receive an anti-corruption message, 41% of participants elected to bribe during the bribery game. All treatment effects are estimated relative to this control condition. Of the five anti-corruption messages that were randomly assigned to participants, Cheeseman and Peiffer find that two increased propensity to bribe in the ensuing bribery game, on average (*widespread* leads to 0.13 increase $\uparrow$ in probability [$p = 0.013$]; *religious* leads to 0.10 increase $\uparrow$ in probability [$p = 0.046$]). One of the messages (*govsuccess*) showed a positive

but non-significant effect (0.08 increase ↑ in probability [p = 0.114]). On average, the final two messages appeared to have no effect (*local* leads to a 0.03 increase ↑ in probability [p = 0.493]; *taxes* leads to a 0.02 decrease ↓ in probability [p = 0.773]).

The authors also estimated conditional average treatment effects using a dichotomized three-item index of corruption perceptions as moderators. In every treatment condition except for the *local* treatment, Cheeseman and Peiffer find at least some evidence that the anti-corruption messages *increased* the propensity to bribe among participants who already perceived corruption as widespread.

The materials necessary to reproduce the results of this study were publicly available here. We anonymously requested additional data from the authors through the Institute for Replication. The authors gracefully and promptly sent us this data – namely, additional pre-treatment covariates and a binary variable indicating which survey participants were invited to participate in the bribery game.[1] The code that reproduced the findings reported in the paper was written in Stata. We did not find any pre-analysis plan associated with the paper, so we cannot replicate anything written in a pre-analysis plan.

In the present paper, we investigate whether the studys results are reproducible and replicable and assess the robustness of the results using alternative estimation strategies. The results were computationally reproducible and stable in the face of alternative strategies. However, statistical significance appears dependent on the estimation strategies for one of the treatments (i.e., appeal to religion). We also investigate alternative strategies to measure the moderator, and the results appear less robust. In the final section, we explore a replication-related issue: the original study's main effects appear slightly underpowered, and the interaction effects are likely more so.

[1]The authors mention in the paper that due to resource constraints, they could only invite a subset of the full survey sample to participate in the bribery game.

2 Computational Reproducibility

Overall, the results reported by Cheeseman and Peiffer (2022) are computationally reproducible. We translated the code written in Stata into R. Upon inspecting the code, we encountered no coding errors and could reproduce the figures and tables from the article.

3 Robustness Replication

In this section, we proceed with a "robustness replication" of Cheeseman and Peiffer – that is, we use the same data as the original article and explore alternative estimation procedures to detect how sensitive the results are to discretionary analytical choices.

We begin by re-estimating the average treatment effect of the five anti-corruption stimuli (all relative to a control condition). In the original paper, Cheeseman and Peiffer report their experimental results in two ways. First, they use a bar graph to show, for each experimental condition, the share of respondents who choose to bribe the third party in the bribing game (Figure 1 in the original article). Second, they model bribing behavior using a logistic model that regresses a binary bribing variable on the experimental conditions (with the control condition excluded) and a five-item scale that measures respondents' experience with poverty.[2] In addition to Cheeseman and Peiffer's second approach, we tested four additional estimation strategies:

1. A logistic regression model with no pre-treatment covariates.

[2]Cheeseman and Peiffer note on p. 1088 that "DIM [differences-in-means] tests were run on basic demographic indicators. The results revealed that the mean level of poverty in the *local* and *control* groups was significantly higher than that of the religious group." Therefore, they use logistic regression with poverty level as the sole control beyond their treatment indicators. We find this not to be the case. In fact, as Appendix D of the original paper shows, the differences in poverty levels between each pair of experimental conditions (a total of 15 combinations) never reach conventional levels of statistical significance. Using a multinomial logistic regression model, we regressed respondents' experimental conditions on a full set of covariates. Comparing this model to an intercept-only model using a likelihood ratio test, we find that the full model does not fit significantly better ($p = 0.61$). The same is true when comparing an intercept-only model with a model that includes the poverty index as a covariate ($p = 0.28$).

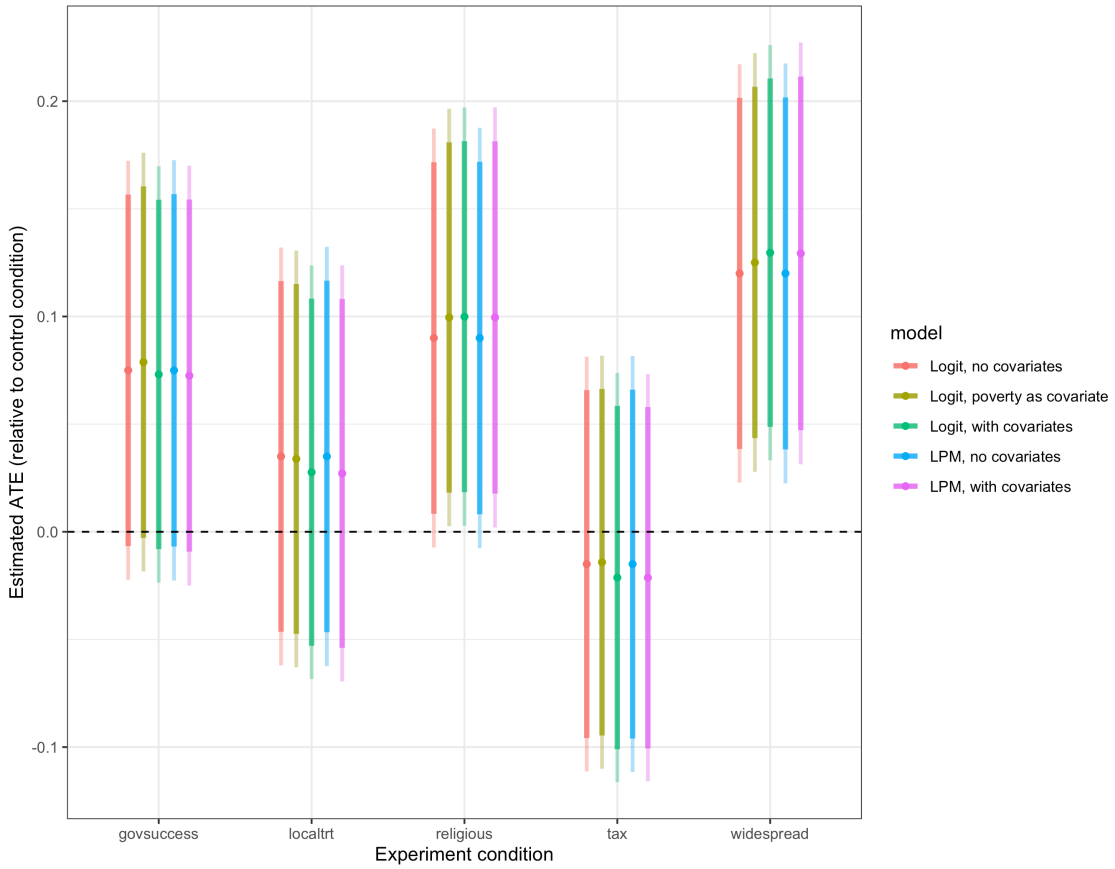
2. A logistic regression model with a more complete battery of pre-treatment covariates.
3. A linear probability model with no pre-treatment covariates.
4. A linear probability model with a more comprehensive battery of pre-treatment covariates.

We find that the results reported by Cheeseman and Peiffer are robust to these different estimation strategies (see Figure 1 and Table A.1 for coefficients and p-values). Across the five estimation strategies (the four above, plus Cheeseman and Peiffer's original approach) and the five experimental conditions, the differences in estimated treatment effects never exceed 0.01 (interpreted as a change of one percentage point in the probability that a respondent engages in bribery). However, since the original result for the "religious" condition borders the conventional but arbitrary threshold for statistical significance (p = 0.046), it is unsurprising that some estimates are no longer significant at $\alpha = 0.05$. Nevertheless, the main idea that these messages do not reduce corruption behavior among the entire sample appears validated by our replication.

Next, we replicate the subgroup effects by Cheeseman and Peiffer. We note that while Cheesman and Peiffer label their section "Testing the interactions," they do not formally test for heterogeneous treatment effects to see if there is a statistical difference between subgroups. Nevertheless, their heterogeneous treatment effect analysis is based on the idea that "Pessimistic Perceivers," defined as individuals who already believe corruption is pervasive, may respond differently to anti-corruption messages. Cheeseman and Peiffer identify Pessimistic Perceivers using a binary variable constructed using factor scores from a three-variable scale. The three variables forming the scale are presented in Table 1.

The authors constructed an index based on these three variables using principal component analysis. The resulting index ranges from -4.11 to 1.11. The authors report that 57% of respondents score above 0; they treat these respondents as Pessimistic Perceivers.

Figure 1: Marginal effects of the treatment in the five treatment conditions, using different estimation strategies



Note: The thick vertical bars show 90% confidence intervals, while thinner vertical bars show 95% confidence intervals.

	Mean	SD	Min	Max	Wording
corr_widespread	3.51	0.68	1.00	4.00	How widespread would you say that corruption is in Nigeria?
common	4.44	0.93	1.00	5.00	Taking into account your own experience or what you have heard, corruption among public officials is...
mostbribe	4.00	1.03	1.00	5.00	How strongly do you agree or disagree with the following statement: Most people I know have paid a bribe. Do you

Table 1: Components of Index

Figure 2 (see A.2 for point estimates and p-values) shows the results of the same four models we presented for the main effects, however, now including interaction terms between treatments and the binary indicator for Pessimistic Perceivers. In the left panel, we show the estimated average marginal effects of the five anti-corruption messages, conditional on Cheeseman and Peiffer's measure of Pessimistic Perceivers. The estimated effects found by Cheeseman and Peiffer are stable across estimation strategies. We note that in all but one of the treatment arms, "Nonpessimistic Perceivers" have an effect in the direction that would be normatively desirable. Still, these effects are only significant and of a substantive magnitude in the case of the Tax treatment. The "Pessimistic Perceivers," on the other hand, have treatment effects that are larger and statistically significant across the same two treatment arms that had some statistical evidence for the main treatment effects and replicate Cheeseman and Peiffer's results, where they find strong evidence for two of the treatment effects and weaker evidence for an additional two among Pessimistic Perceivers.

In the right panel, we estimate the *differences* in treatment effects (Gelman and Stern 2006) between Pessimistic and Nonpessimistic Perceivers. We find that there is only one treatment arm with clear statistical evidence for a *difference* between Pessimistic Perceivers and Nonpessimistic Perceivers, and this is the "Tax treatment." In this case, the difference is largely driven by the large and statistically significant effect in reducing corruption for the Nonpessimistic Perceivers group, a point also highlighted by Cheeseman and Peiffer.

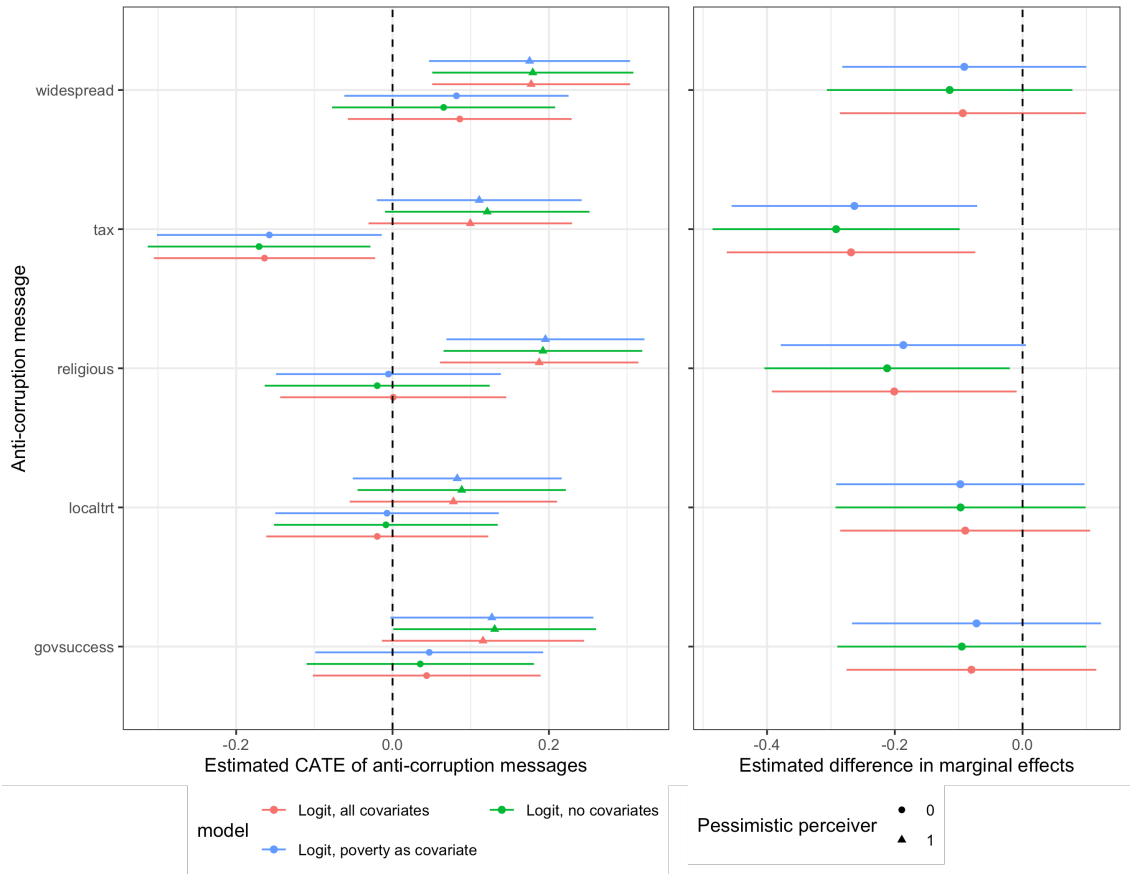


Figure 2: Estimated treatment effects, conditional on different modelling strategies

Before carrying out our robustness replication, we first note that the three variables that form the index appear to have been measured post-treatment. This consideration is not strictly related to replication, so we do not address it here, but it may bias the estimates presented (Sheagley and Clifford 2023, Blackwell et al. 2023, Montgomery et al. 2018).

Regardless of the post-treatment moderator, we suggest that issues may arise from using a binary indicator variable derived from the principal component analysis as employed in the article. The underlying distributions of the three variables composing the Pessimistic Perceiver index are so skewed that creating a dummy variable may lead to problematic classifications. As the authors show, the vast majority of respondents perceive corruption to be widespread in Nigeria. Across the entire sample, nearly 90% of respondents say that "corruption among public officials" in Nigeria is "common" or "very common." The highly skewed nature of this distribution is not well captured by the binary variable that Cheeseman and

Peiffer create. We created a simple additive index composed of the three variables and compared values from that index to values of the binary variable created using principal component analysis. There is a sharp cutoff point on the scale of the additive index that divides respondents into two groups formed by principal component analysis. Respondents who score 0.833 or higher on the simple additive index are classified as "Pessimistic Perceivers" by Cheeseman and Peiffer's procedure. Respondents who score 0.806 or lower on the simple additive index are classified as "Nonpessimistic Perceivers." As an example, a respondent would score 0.806 on the additive index if they answered in the following way:

1. Corruption among public officials is *very common*
2. *Agree* that most people I know have paid a bribe
3. Corruption in Nigeria is *very widespread*

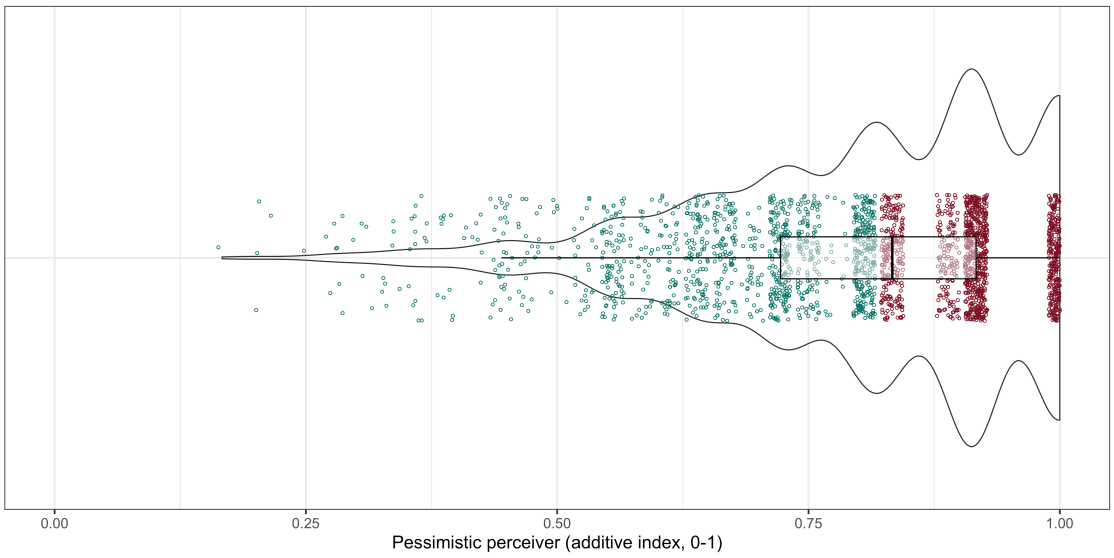
Of course, this is an extreme example, as it is the highest possible value for a respondent categorized by Cheeseman and Peiffer as a Nonpessimistic Perceiver. However, while extreme, it is *not* rare: fully 22.4% of respondents classified as Nonpessimistic Perceivers score 0.806 on the simple additive index. Respondents who score 0.72 and above on the additive index represent just over half of all respondents classified as Nonpessimistic Perceivers by Cheeseman and Peiffer. The full distribution of the "Pessimistic Perceiver" additive index is shown in Figure 3.

Given the potential sensitivity of the moderator, we sought to reproduce the heterogeneous treatment effects analysis using alternative estimation strategies. We pursue two approaches:

1. Modifying the threshold used to classify respondents as Nonpessimistic Perceivers.
2. Using the full scale of the three-variable index as a moderator and allowing non-linearity.

In Figure 4, we show the subgroup analyses when we vary the threshold used to separate Pessimistic and Nonpessimistic Perceivers. When we vary the cutoff for

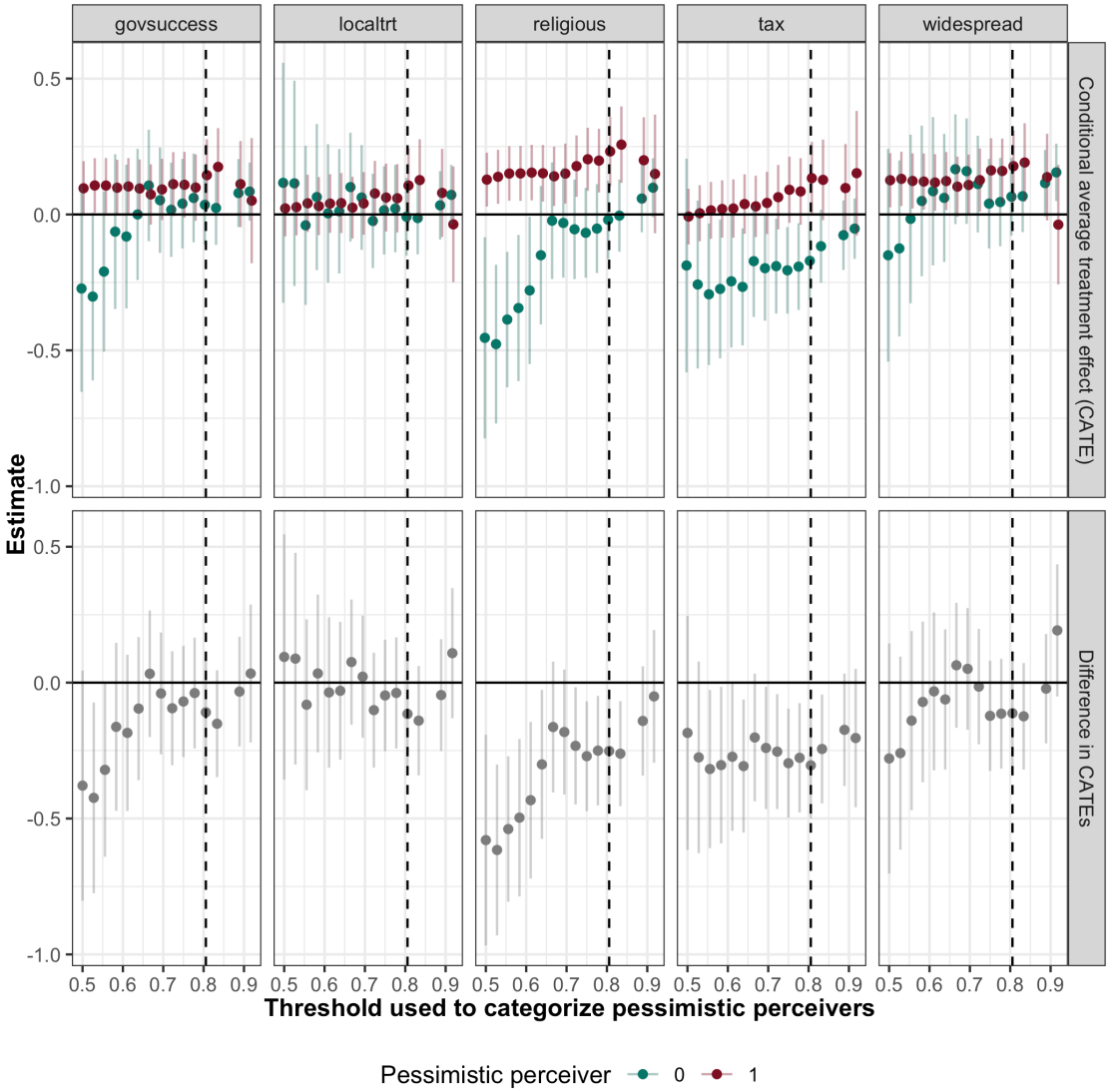
Figure 3: Distribution of the "Pessimistic Perceiver" additive index



Note: Respondents categorized as Pessimistic Perceivers by Cheeseman and Peiffer are in red.

the threshold, we see that the estimated effects of the anti-corruption messages for Nonpessimistic Perceivers at lower values of the threshold are more reliably negative and, in many cases, statistically significant across four of the five treatment arms, as would be predicted from theory. Of course, these effects are much less precisely estimated since the number of Nonpessimistic Perceivers in each experimental treatment arm is small when the threshold is low. We also see that, while generally signed positively, the coefficients estimated for Pessimistic Perceivers are not reliably statistically significant, and several show signs of non-linearity as the cutoff varies. Therefore, this variation of the thresholds raises concerns about the statistical reliability of Cheeseman and Peiffer's findings concerning Pessimistic Perceivers. It also highlights a present but underemphasized point in the manuscript — that those who genuinely do not believe corruption is a significant problem may be more amenable to anti-corruption messaging.

Figure 4: Estimated conditional average treatment effects for Pessimistic Perceivers and Nonpessimistic Perceivers, using different dichotomization thresholds



Note: The dashed horizontal line shows the threshold used by Cheeseman and Peiffer.

Given how vital the definition and coding of Pessimistic Perceivers is for the results, we conducted additional checks by employing a continuous measure of the moderator. Specifically, we use (1) the additive index mentioned above ("Pessimistic Perceivers (additive index)") and (2) the initial factor scores that, in the paper, are employed to dichotomize the moderator and conduct the main analysis ("Pessimistic perceivers (continuous factor)"). In addition to examining marginal effects over the common support of the moderator, the continuous moderator allows us to test for

two assumptions important for multiplicative interaction models but not directly addressed in empirical work (see discussion in Hainmueller et al. 2019): 1) a linear interaction effect and 2) the existence of common support of the moderator.

We employ the `interflex` R package developed by Hainmueller et al. (2019) for these tests. Below, in Figure 5 (additive index), we show the marginal effects of the linear moderator and those effects when splitting the moderator at low, medium, and high values. However, we reach a similar conclusion using a different operationalization of the moderator on the raw factor score (Appendix B.1) and using a non-linear kernel estimator for both operationalizations (additive index and raw factor score) (Appendix B.2).

Figure 5 and those present in Appendix B reveal three main findings. First, as the plot shows (see Table B.4 for estimates of differences and p-values), the linearity assumption does not hold for four comparisons with the control condition. We only have some indication that linearity is present for the "taxes" and "religious" experimental conditions. Formally, we conduct a Wald test to determine if we can reject the linear multiplicative interaction model by comparing it with a more flexible model of multiple bins. The Wald test yields a value of 0.073, which suggests that, for a threshold value of 0.1, we can reject the null hypothesis that the linear interaction model and the three-bin model are statistically equivalent. Thus, there is some evidence that the relationship between the key explanatory variable and the moderator is not linear.

Second, we find little reliable statistical evidence for backfire effects, with the linear and non-linear estimators providing contradictory evidence. The non-linear models suggest a backfire effect only at median levels of the moderator. In contrast, the linear model suggests that there are backfire effects only at the highest values of the moderator.

Third, we again see a lack of common support at the low levels of the moderator, which seems to impact the robustness of the finding about the "taxes" experimental condition. While the tax treatment at low levels of the moderator is the *only* finding

reliably statistically significant across models estimated and has a properly signed and negative effect for those respondents at lower levels of moderator, the large negative effects found when dichotomizing the moderator are likely driven by only a few observations.

Is sum, although our analysis of the continuous moderator is not direct critique of the paper (as the authors opt for a binary moderator), these findings reveal the importance of paying attention to the concerns raised by Hainmueller et al. (2019). Moreover our analysis suggests that, in empirical work, more research is needed to thoroughly theorize and empirically identify who constitutes Pessimistic Perceivers.

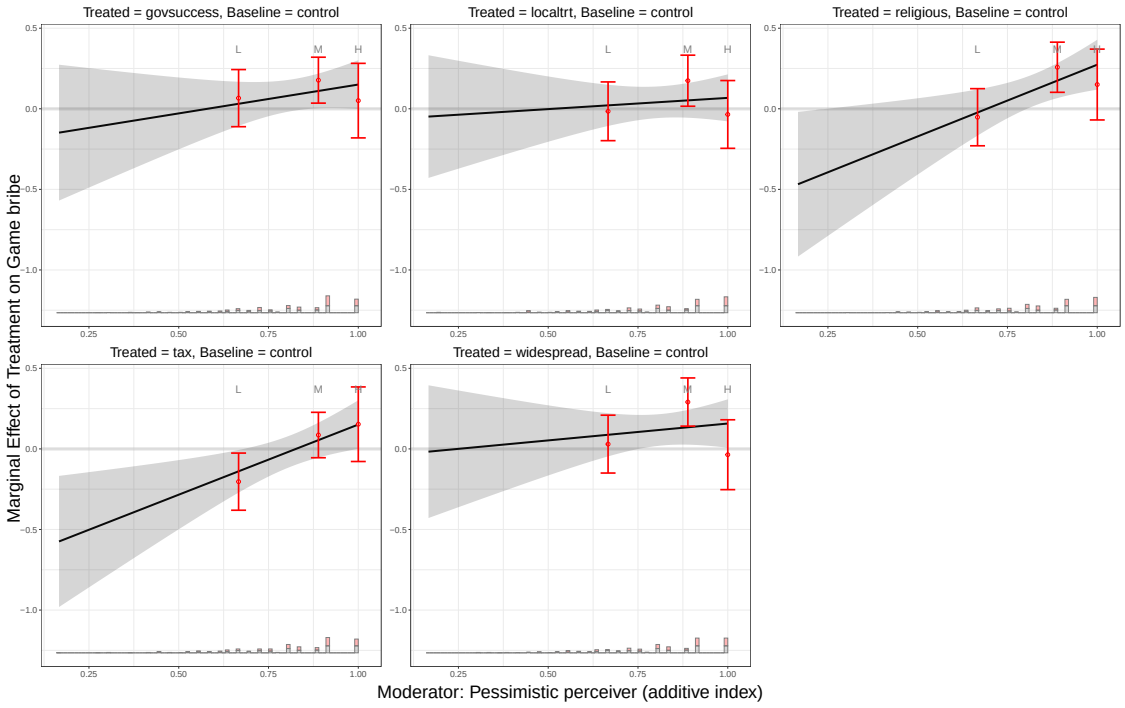


Figure 5: Pessimistic perceivers (additive index). Testing the interaction assumptions. The black bar and grey lines show marginal effects and 95% confidence intervals from a linear interactive model. The red dots vertical lines and show the marginal effects and 95% confidence intervals from the binning estimator.

4 Replication related critiques: Power calculations

A key concern in quantitative political science, particularly germane experimental research, is the lack of statistical power (Arel-Bundock et al. 2022). For the Cheeseman and Peiffer (2022) article, we could not identify a discussion on power

calculations conducted before the fielding of the experimental study in the article's main body or the online appendix. Thus, we sought to calculate the Minimal Detectable Effect (MDE) — the smallest effect size detectable in a study with this sample size. For a two-sample test with 200 subjects per experimental condition, 0.05 statistical significance level and 0.8 power, we obtain an effect size (Cohen's d) of 0.28. As this effect is calculated as the difference between the means divided by the pooled standard deviation, we can conclude that the MDE is 0.14. Based on the effect sizes shown in the Cheeseman and Peiffer paper's Table 1 (i.e., 0.13 for the "Widespread" condition, and 0.1 for the "Religious" condition), we conclude that the study is potentially slightly underpowered for the main effects reported in the article.

A simulation approach further supports this conclusion. For the direct effect, we consider 200 respondents per experimental condition and assume an effect that goes from 0 to 0.3 by an increment of 0.01. The alpha level is 0.05. Given this setup, we found that an effect size of 0.14 can be detected with a power of 0.81. The simulation results are shown in Figure 6.

We also conduct simulations related to conditional effect based on the effect sizes seen in Figure 2 of Cheeseman and Peiffer's article. Hence, we assume an effect size ranging from 0 to 0.2 with an increment of 0.01 for the Pessimistic Perceivers and an effect size ranging from -0.15 - 0.1 with an increment of 0.05 for the Nonpessimistic Perceivers. The alpha level is 0.05. Here, given the simulated effect sizes for Nonpessimistic and Pessimistic Perceivers, we seek to calculate the power of detecting such effects. The findings (see Figure C.4 in Appendix C) reveal that the study could be underpowered in uncovering the conditional effects, further highlighting the need for additional research to address the concerns related to the reliability of the results based on subgroup analysis.

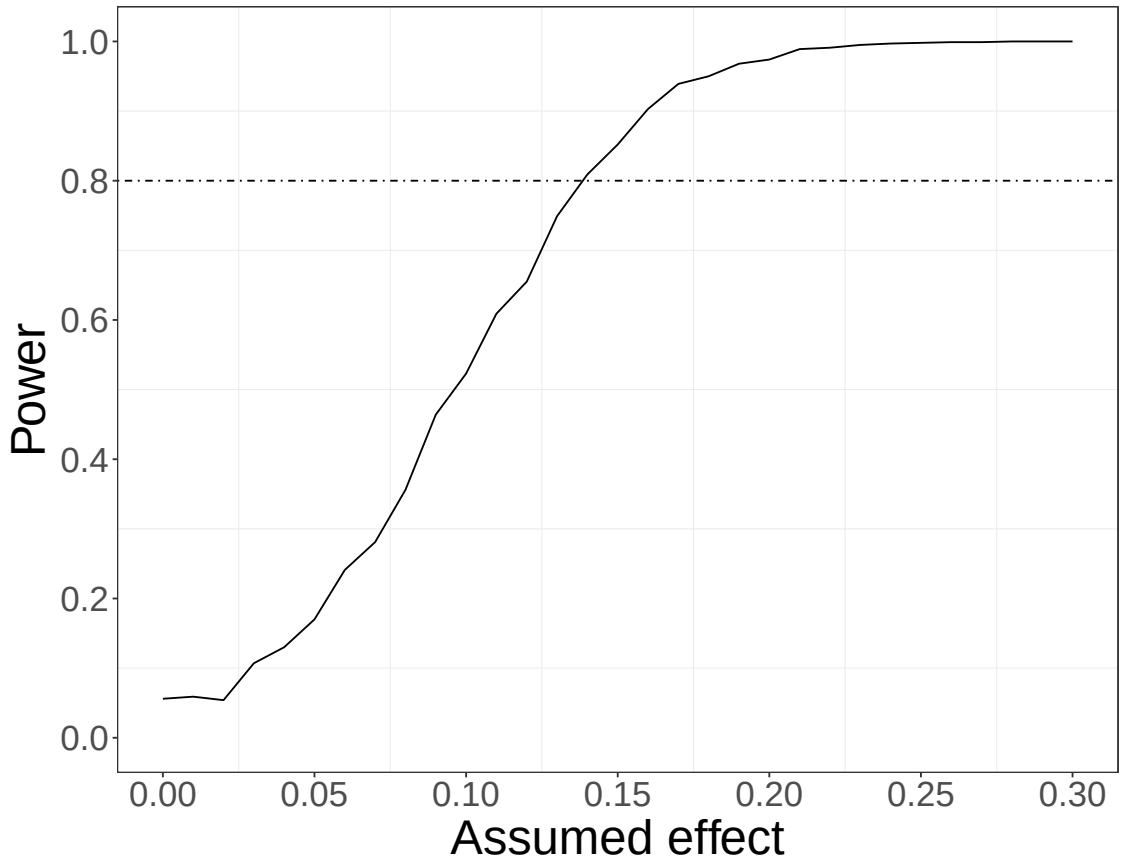


Figure 6: The power to detect the direct effect (see Table 1 in the C&P article)

5 Conclusion

In conclusion, Cheeseman and Peiffer's study explored the impact of various anti-corruption messages on behavior using a bribery game conducted in Lagos, Nigeria. The research addressed significant issues related to corruption in the region, aiming to shed light on the effectiveness of anti-corruption campaigns. For our replication study, we sought to replicate and assess the robustness of their findings while raising important critiques and considerations.

Our findings demonstrate that the study's results were computationally reproducible, and we successfully reproduced the figures and tables in the original article. Robustness checks using alternative estimation strategies revealed substantial consistency in the average treatment effects across various analytical approaches. However, the statistical significance of the "appeal to religion" treatment showed dependency on the estimation strategy.

However, with respect to the subgroup analysis of Pessimistic Perceivers, our robustness checks demonstrated that the estimated size and statistical significance of the subgroup effects were sensitive to the choice of classification criteria. Indeed, different reasonable thresholds for dichotomizing the binary indicator lead to different substantive conclusions about subgroups. Additionally, we identified common support issues and non-linearity as potential concerns.

In sum, this replication study underscores the importance of reproducibility and robustness in social science research. While Cheeseman and Peiffer's study provides valuable insights into the complex relationship between anti-corruption messages and behavior, our analysis emphasizes the need for critical evaluation of measurement and statistical approaches to ensure the validity and reliability of findings, particularly concerning subgroup analysis. Replication studies are crucial in advancing scientific knowledge by confirming or challenging existing results and enhancing the overall research rigor in the social sciences.

References

Arel-Bundock, V., Briggs, R. C., Doucouliagos, H., Mendoza Aviña, M. and Stanley, T. D.: 2022, Quantitative political science research is greatly underpowered, *Technical report*, I4R Discussion Paper Series.

Blackwell, M., Brown, J. R., Hill, S., Imai, K. and Yamamoto, T.: 2023, Priming bias versus post-treatment bias in experimental designs, *arXiv preprint arXiv:2306.01211* .

Cheeseman, N. and Peiffer, C.: 2022, The Curse of Good Intentions: Why Anticorruption Messaging Can Encourage Bribery, *American Political Science Review* **116**(3), 1081–1095.

Gelman, A. and Stern, H.: 2006, The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant, *The American Statistician* **60**(4), 328–331.

Hainmueller, J., Mummolo, J. and Xu, Y.: 2019, How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice, *Political Analysis* **27**(2), 163–192.

Montgomery, J., Nyhan, B. and Torres, M.: 2018, How conditioning on posttreatment variables can ruin your experiment and what to do about it, *American Journal of Political Science* **62**(3), 760–775.

Sheagley, G. and Clifford, S.: 2023, No evidence that measuring moderators alters treatment effects, *American Journal of Political Science* .

Appendices

A Tables and Coefficients form Regression Models

A.1 Main Effects Models

Table A.1: Replicated ATEs for all five experimental conditions, using different estimation strategies

	(1)	(2)	(3)	(4)	(5)
govsuccess	0.075	0.073	0.075	0.079	0.073
	(0.132)	(0.144)	(0.130)	(0.111)	(0.138)
localtrt	0.035	0.027	0.035	0.034	0.028
	(0.481)	(0.582)	(0.480)	(0.493)	(0.573)
religious	0.090	0.100	0.090	0.099	0.100
	(0.071)	(0.045)	(0.069)	(0.043)	(0.043)
tax	−0.015	−0.021	−0.015	−0.014	−0.022
	(0.761)	(0.658)	(0.760)	(0.773)	(0.661)
widespread	0.120	0.129	0.120	0.125	0.129
	(0.016)	(0.010)	(0.015)	(0.011)	(0.008)
Controls	-	All	-	Poverty	All
Num.Obs.	1200	1176	1200	1188	1176
Model	Linear	Linear	Logit	Logit	Logit

Note:
Models 1 and 2 show results from a linear regression model with HC2 standard errors. Models 3-5 show results from a logistic regression model. P-values are shown in parentheses below each estimate.

A.2 Models with Interaction Terms

Table A.2: Replicated CATEs for all five experimental conditions, using different estimation strategies

	(1)	(2)	(3)	(4)	(5)
govsuccess 0	0.047	0.035	0.044	0.036	0.044
	(0.528)	(0.633)	(0.557)	(0.631)	(0.565)
govsuccess 1	0.127	0.131	0.116	0.126	0.110
	(0.055)	(0.048)	(0.079)	(0.057)	(0.093)
localtrt 0	0.007	0.008	0.020	0.009	0.020
	(0.923)	(0.907)	(0.787)	(0.907)	(0.784)
localtrt 1	0.083	0.088	0.078	0.083	0.072
	(0.225)	(0.193)	(0.250)	(0.212)	(0.277)
religious 0	0.005	0.020	0.001	0.020	0.000
	(0.943)	(0.790)	(0.990)	(0.788)	(0.996)
religious 1	0.196	0.192	0.188	0.189	0.184
	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)
tax 0	0.158	0.171	0.164	0.170	0.164
	(0.032)	(0.019)	(0.023)	(0.022)	(0.028)
tax 1	0.111	0.121	0.099	0.116	0.095
	(0.097)	(0.069)	(0.134)	(0.080)	(0.149)
widespread 0	0.082	0.065	0.086	0.066	0.086
	(0.264)	(0.369)	(0.239)	(0.372)	(0.250)
widespread 1	0.175	0.179	0.177	0.175	0.174
	(0.008)	(0.006)	(0.006)	(0.009)	(0.009)
Controls	Poverty	-	All	-	All
Model	Logit	Logit	Logit	Linear	Linear
Num.Obs.	1188	1200	1176	1200	1176

Note:
Model 1 shows results from a logistic regression model with the 5-item poverty index as the sole covariate (as originally presented by Cheeseman and Peiffer). Model 2 shows the same model without this covariate. Model 3 includes a more complete battery of covariates. Models 4 and 5 are linear regression models with no covariates and all covariates, respectively. p-values are shown in parentheses under each estimate.

B Models with a Continuous Moderator

B.1 Linearity Assumptions of Models with Interaction Terms

	Estimate	Std. Error	t value	Pr(>|t|)
Pessimistic perceiver (additive index)	-0.6590	0.2342	-2.81	0.0050
govsuccess	-0.2076	0.2693	-0.77	0.4408
Pessimistic perceiver (additive index) x govsuccess	0.3580	0.3251	1.10	0.2710
localtrt	-0.0715	0.2541	-0.28	0.7783
Pessimistic perceiver (additive index) x localtrt	0.1393	0.3080	0.45	0.6512
religious	-0.6164	0.2884	-2.14	0.0328
Pessimistic perceiver (additive index) x religious	0.8901	0.3462	2.57	0.0103
tax	-0.7194	0.2639	-2.73	0.0065
Pessimistic perceiver (additive index) x tax	0.8705	0.3192	2.73	0.0065
widespread	-0.0520	0.2639	-0.20	0.8437
Pessimistic perceiver (additive index) x widespread	0.2095	0.3191	0.66	0.5118
(Intercept)	0.9485	0.1931	4.91	0.0000
Num. obs.		1154		

Table B.3: Linear model - Pessimistic perceivers (additive index) as a moderator, game bribe as outcome (see Figure 5)

	diff. estimate	**SD**	**z-value**	**p-value**	**lower CI (95%)**	**upper CI (95%)**
govsuccess						
75% vs 50%	0.07	0.07	1.06	0.29	-0.06	0.21
100% vs 75%	0.04	0.07	0.59	0.56	-0.10	0.18
100% vs 50%	0.12	0.14	0.83	0.41	-0.16	0.39
localtrt						
75% vs 50%	0.03	0.06	0.55	0.58	-0.09	0.15
100% vs 75%	0.04	0.07	0.59	0.56	-0.09	0.17
100% vs 50%	0.07	0.13	0.58	0.56	-0.17	0.32
religious						
75% vs 50%	0.19	0.07	2.59	0.01	0.05	0.34
100% vs 75%	0.16	0.08	2.02	0.04	0.01	0.31
100% vs 50%	0.35	0.15	2.32	0.02	0.05	0.65
tax						
75% vs 50%	0.17	0.06	2.71	0.01	0.05	0.30
100% vs 75%	0.18	0.07	2.67	0.01	0.05	0.31
100% vs 50%	0.35	0.13	2.74	0.01	0.10	0.60
widespread						
75% vs 50%	0.05	0.07	0.71	0.48	-0.09	0.18
100% vs 75%	0.01	0.07	0.14	0.89	-0.13	0.15
100% vs 50%	0.06	0.14	0.43	0.67	-0.21	0.33

Table B.4: Kernel estimations: Difference in treatment effects at the 50, 75 and 100 percentiles of the moderator (Pessimistic perceiver as an additive index). The control condition is the baseline.

	diff. estimate	SD	z-value	p-value	lower CI (95%)	upper CI (95%)
govsuccess						
75% vs 50%	0.124	0.102	1.213	0.225	-0.076	0.323
100% vs 75%	-0.053	0.124	-0.427	0.669	-0.296	0.190
100% vs 50%	0.071	0.163	0.433	0.665	-0.249	0.390
localtrt						
75% vs 50%	0.065	0.091	0.719	0.472	-0.112	0.243
100% vs 75%	-0.019	0.129	-0.146	0.884	-0.273	0.235
100% vs 50%	0.046	0.150	0.307	0.759	-0.248	0.341
religious						
75% vs 50%	0.259	0.100	2.588	0.010	0.063	0.455
100% vs 75%	0.051	0.132	0.391	0.696	-0.206	0.309
100% vs 50%	0.310	0.167	1.854	0.064	-0.018	0.638
tax						
75% vs 50%	0.235	0.097	2.429	0.015	0.045	0.424
100% vs 75%	0.148	0.128	1.160	0.246	-0.102	0.399
100% vs 50%	0.383	0.160	2.390	0.017	0.069	0.697
widespread						
75% vs 50%	0.065	0.093	0.694	0.487	-0.118	0.248
100% vs 75%	-0.137	0.121	-1.132	0.258	-0.374	0.100
100% vs 50%	-0.072	0.140	-0.512	0.609	-0.347	0.203

Table B.5: Kernel estimations: Difference in treatment effects at the 50, 75 and 100 percentiles of the moderator (Pessimistic perceiver as a continuous factor). The control condition is the baseline.

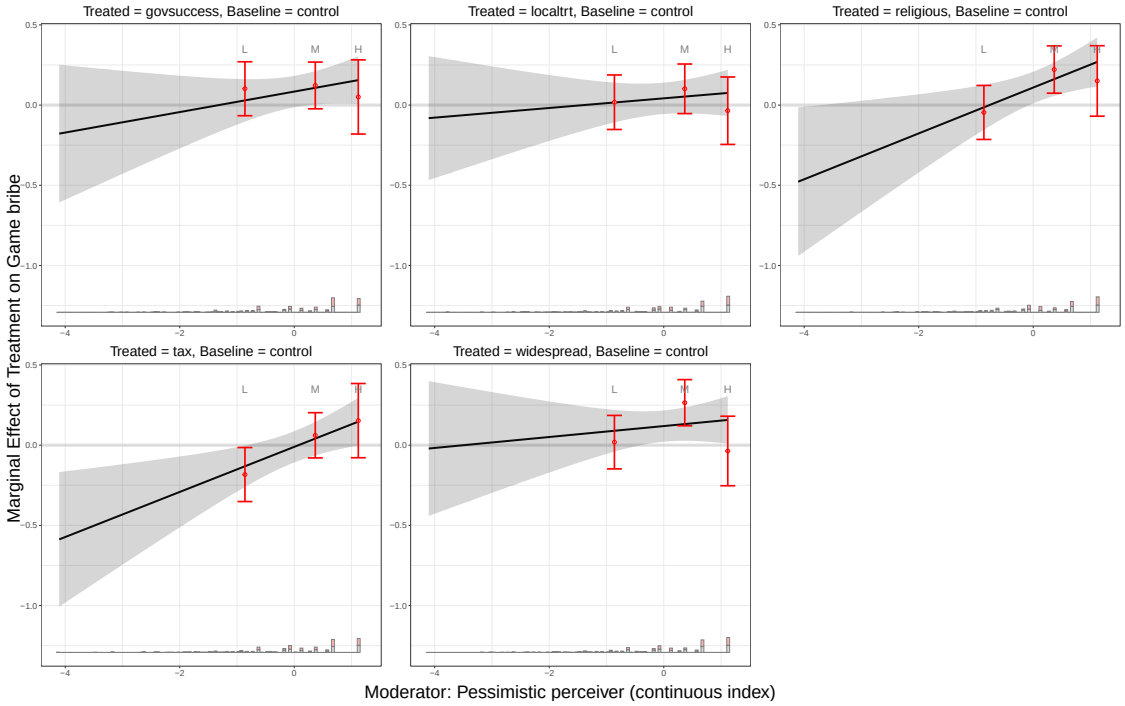


Figure B.1: Pessimistic Perceivers (continuous factor). Testing the interaction assumptions. The black bar and grey lines show marginal effects and 95% confidence intervals from a linear interactive model. The red dots vertical lines and show the marginal effects and 95% confidence intervals from the binning estimator.

B.2 Non-linearity and common support

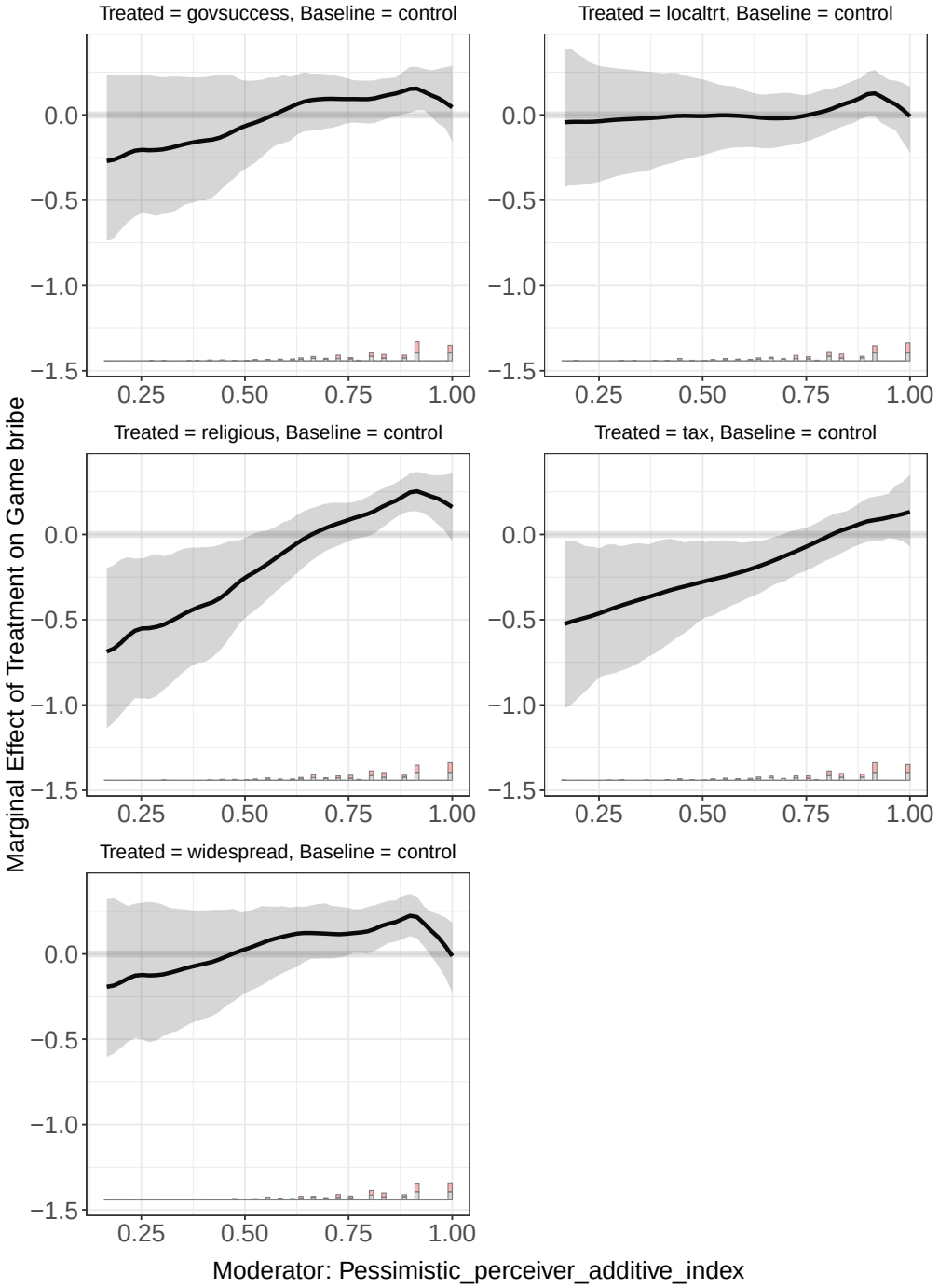


Figure B.2: Pessimistic Perceivers (additive index). Black lines show the marginal effects from a non-linear kernel estimator, and the gray shaded areas display the associated 95% confidence intervals

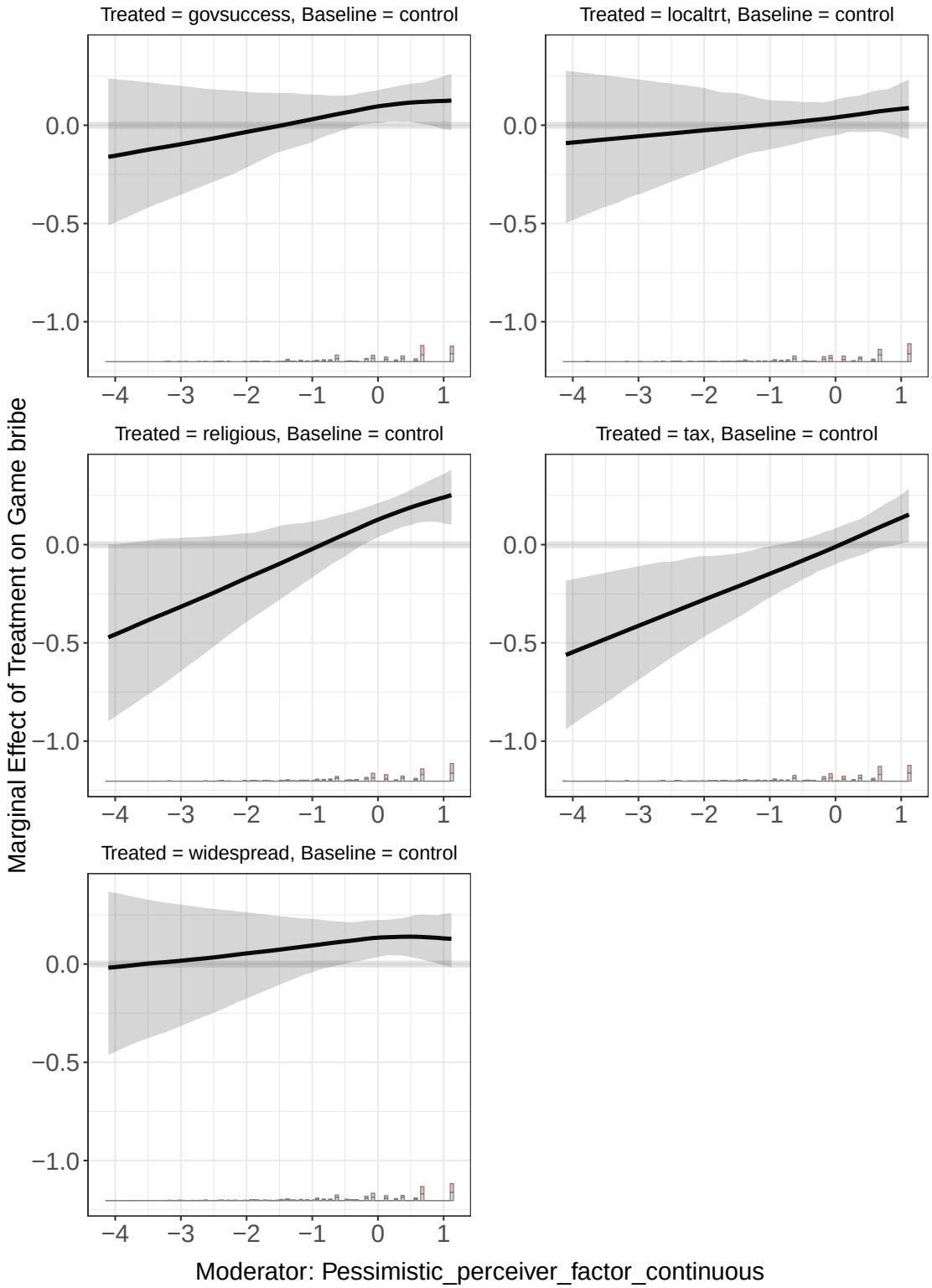


Figure B.3: Pessimistic perceivers (continuous factor). Black lines show the marginal effects from a non-linear kernel estimator, and the gray shaded areas display the associated 95% confidence intervals

C Power calculations

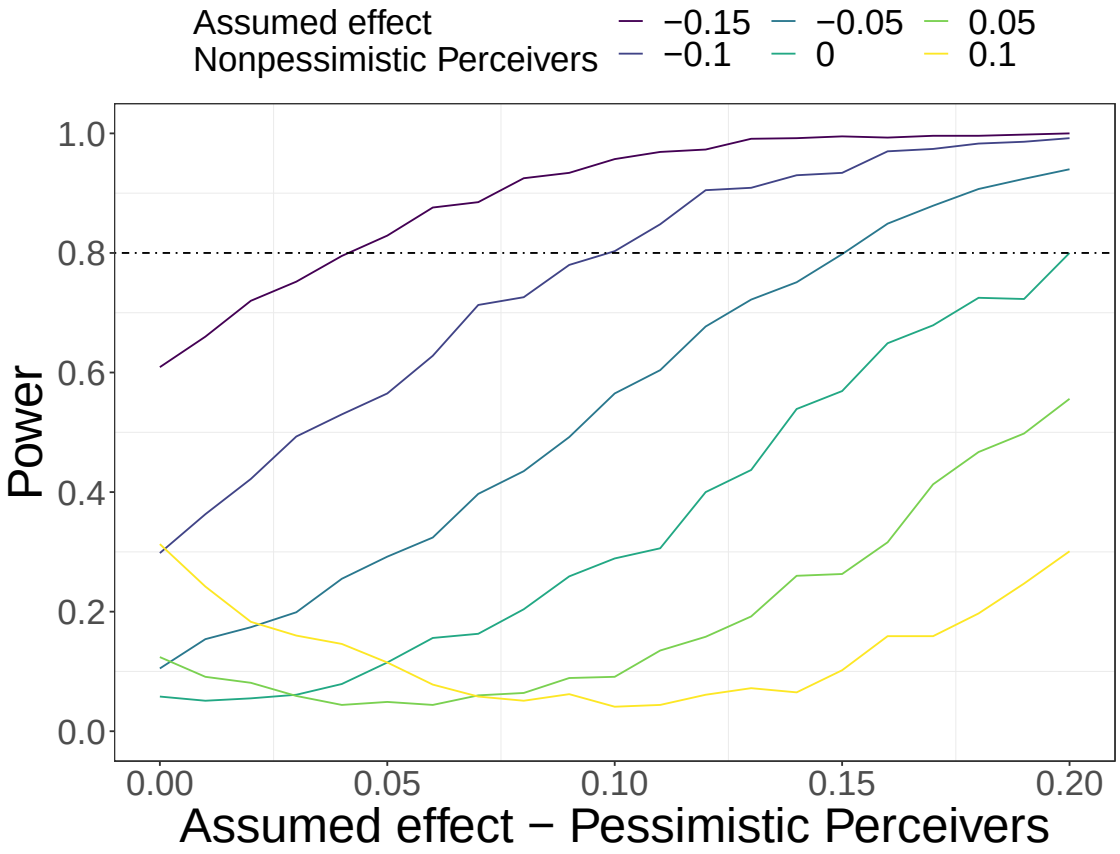


Figure C.4: The power to detect the conditional effect (see Figure 2 in the C&P paper)