Empathy for the Devil: Pre-Analysis Plan

October 27, 2023

Contents

1 Design

This is the pre-analysis plan for a survey experiment investigating how AI-illustrated vignettes emphasizing tolerance and counter-stereotypes can foster empathy toward political adversaries. I will use a two-condition (control vs. treatment) between-subjects design. Participants will indicate their political preferences and be categorized into three groups: Bolsonaristas, Lulistas, or neutral. Within each group, participants will be randomly assigned to either the treatment or control groups. Participants in the neutral group assigned to the treatment condition will randomly receive either the Bolsonaristas' treatment or the Lulistas' treatment.

2 Research Questions

- RQ1: Do the vignettes reduce affective polarization in specific?

- RQ2: Do the vignettes increase empathy toward out-group members in general?

- RQ3: Do the vignettes reduce support for measures that could impede the fairness and integrity of elections?

- RQ4: Do the vignettes reduce support for state actions aimed at combatting corruption but that jeopardize individual rights?

- RQ5: Do the vignettes reduce support for political violence?

3 Sample and Exclusion Criteria

The participants will be drawn from a non-probabilistic, nationally diverse sample of distinct Brazilian residents aged 18 or older, recruited by QuestionPro. The survey vendor will be instructed to apply quotas in alignment with the demographic composition of the 2022 Brazilian Electoral Study (*Estudo Eleitoral Brasileiro* or ESEB), which is a nationally

representative survey. These quotas will be matched for age, gender, geographic region, education, and household income. Participants who do not pass both pretreatment attention checks will be excluded from the sample. Please refer to the accompanying questionnaire for details about the attention checks.

4 Main Model

All models examining the primary and secondary dependent variables will be specified as follows:

$$y = \alpha + \beta treatment + \gamma pref\_candidate + \delta X + \varepsilon, \tag{1}$$

where y will be one of the primary and secondary dependent variables described below, $treatment$ will be a dummy variable that takes a value of 1 when the participant was assigned to the treatment group, $pref\_candidate$ will be a categorical variable that indicates if the participant is Bolsonarista, Lulista, or neutral, X will be a vector of demographic covariates to increase efficiency (Lin 2013), and ε represents Huber-White heteroscedasticity-robust standard errors (White 1980). Demographic covariates will be all categorical and include gender ($D1$), age ($D2$), geographic region ($D3$), household income ($D4$), education ($D5$), and race ($D6$). Specific categories for these covariates can be found in the accompanying questionnaire.

5 Dependent Variables

5.1 Primary Dependent Variables

- Affective Polarization ($affpol$): measuring dislike toward out-group members.

- Social Distance Index (SD_index): index aggregating $SD1$, $SD2$, and $SD3$.

- Free and Fair Elections Index (FFE_index): index aggregating $FFE1$, $FFE2$, and $FFE3$.

- Corruption and Individual Rights Index (CIR_index): index aggregating $CIR1$, $CIR2$, and $CIR3$.

- Political Violence Index (PV_index): index aggregating $PV1$, $PV2$, $PV3$, and $PV4$.

All these variables will be measured using a 101-degree feeling thermometer. Please refer to the accompanying questionnaire for the precise wording of the questions. The summary indices will be calculated using the standardized inverse-covariance weighted average of the mentioned indicators, following the approach developed by Anderson (2008) and the software implementation by Schwab et al. (2020). The control group will serve as the reference for standardizing the variables. The primary dependent variable $affpol$ will consistently be assessed in the first dependent-variable question in the survey.

To test each of the five research questions, I will examine the coefficient β from the model described above for each of the five primary dependent variables (or outcome domains). This procedure will provide confirmatory evidence about intervention effects for each domain separately and will use unadjusted p-values. Additionally, I will conduct a confirmatory test for the effects of the intervention across the five outcome domains. As a multiplicity correction to the latter, I will compute q-values to control for the false discovery rate (FDR) using the two-stage Benjamini-Hochberg (TSBH) step-up procedure (Benjamini, Krieger and Yekutieli 2006) with a nominal type I error rate (α) of 0.05.

5.2 Secondary Dependent Variables

- Social Distance Variables: Discuss Politics ($SD1$), Employee Hiring ($SD2$), and Daughter's Marriage ($SD3$).

- Free and Fair Elections variables: Reduce Public Transportation ($FFE1$), Contest Election Results ($FFE2$), and Increase Traffic Stops ($FFE3$).

- Corruption and Individual Rights Variables: Disclose Classified Information ($CIR1$), Extend Pretrial Detention ($CIR2$), and Illegal Wiretapping ($CIR3$).

- Political Violence Variables: Threaten Political Leader ($PV1$), Harass Out-Group Supporter ($PV2$), Violence for Political Goals ($PV3$), and Violence After Electoral Defeat ($PV4$).

All of these variables will be measured using a 101-degree feeling thermometer. Please refer to the accompanying questionnaire for the precise wording of the questions. The order of the questions examining all these dependent variables (with the exception of $affpol$, as noted above) will be determined randomly without consideration for their respective outcome domains.

To analyze each of these secondary dependent variables, I will evaluate the coefficient β using the model described above. Similar to the analysis conducted for the primary dependent variables, I will present results using both unadjusted p-values and FDR-sharpened q-values.

6 Heterogeneity Analyses

In addition to the analyses that consider all participants, I will divide the sample into three groups based on political preferences (Bolsonaristas, Lulistas, and neutrals) to investigate both primary and secondary variables once more. I will use a model specified as follows:

$$y = \alpha + \beta treatment + \gamma X + \varepsilon, \tag{2}$$

where the the coefficient of interest is β, and all terms are defined as in the main model. Similar to the previous analyses, I will present results using both unadjusted p-values and FDR-sharpened q-values.

References

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103(484):1481–1495. ↑ p.4

Benjamini, Yoav, Abba M. Krieger and Daniel Yekutieli. 2006. "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate." *Biometrika* 93(3):491–507. ↑ p.4

Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique." *The Annals of Applied Statistics* 7(1):295–318. ↑ p.3

Schwab, Benjamin, Sarah Janzen, Nicholas P. Magnan and William M. Thompson. 2020. "Constructing a summary index using the standardized inverse-covariance weighted average of indicators." *The Stata Journal* 20(4):952–964. ↑ p.4

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48(4):817–838. ↑ p.3