

Do Pakistanis Hold Militant Groups in Low Regard?*

Gaurav Sood†

Abstract

Blair et al. (2013) conclude from endorsement experiments that Pakistanis hold militant groups in “low regard.” I reanalyze their survey with the measurement model developed for endorsement experiments (Bullock et al., 2011). In the unweighted analysis sample, the model implies a 43% probability of a positive latent reaction to an endorsement, averaged across four militant groups (95% credible interval 32%–53%). Small negative mean reactions coexist with substantial dispersion. More generally, an average treatment effect does not identify the proportion holding favorable attitudes. Alternative behavioral assumptions yield different implied supporter proportions. The observed response marginals bound the fraction of responses increased by endorsement between 1% and 58%.

1 Introduction

Do Pakistanis hold militant groups in low regard? Blair et al. (2013) use endorsement experiments to measure affect toward four militant groups. They find small negative average effects on stated policy support and interpret them as weakly negative average affect. That interpretation leaves a different question open: how widely are favorable or unfavorable attitudes held?

Using the measurement model of Bullock et al. (2011) in the unweighted analysis sample, I estimate a 43% probability of a positive latent reaction to an endorsement, averaged equally across the four groups (95% credible interval 32%–53%). The mean reactions have magnitudes no greater than 0.08 ideal-point standard deviations, with larger estimated dispersion.

The distinction between an average and a prevalence is general. An average treatment effect (ATE) combines the frequency and size of positive and negative effects. Its sign alone cannot separate them. Under behavioral assumptions linking favorable attitudes to positive effects, the same ATE can imply different supporter proportions. Some such assumptions are contradicted by the observed response distribution. Others produce implications for quantities that should not be equated: an unaffected fraction, a probability of positive latent affect, and a fraction of observed responses increased by an endorsement.

2 Endorsement Experiments

Endorsement experiments randomize whether a policy question includes an endorsement by an actor of interest. Respondents in the control condition evaluate the policy alone; those in the treatment

*<https://github.com/finite-sample/endorsement>.

†gsood07@gmail.com

condition evaluate the policy with the endorsement. With valid assignment and observation, the difference identifies an effect of the cue on stated policy support.

Interpreting that effect as affect toward the endorser requires more. Respondents must understand the endorsement, and its effect must operate through their attitudes toward the group rather than through other information the cue conveys. Even then, the size of a response change can depend on baseline policy preferences and the response scale. The design has been used to study militant groups in Pakistan and combatants in Afghanistan (Bullock et al., 2011; Blair et al., 2013; Lyall et al., 2013; Blair et al., 2014).

A negative ATE can reflect many small negative reactions, a few large ones, or positive and negative reactions that nearly cancel. It therefore does not identify the proportion with either sign of effect. Nor does the sign of an effect, without a measurement assumption, classify a respondent as a supporter or opponent.

3 The Experiment and Data

Blair et al. (2013) surveyed 6,000 adults in Pakistan’s four main provinces in 2009. Their experiment covered Kashmiri tanzeems, the Afghan Taliban, al-Qa’ida, and sectarian tanzeems. Respondents evaluated polio vaccination, Frontier Crimes Regulation reform, the Durand Line, and curriculum standardization on a five-point intensity scale. I rescale responses to $[0, 1]$; a mean on this scale is not a percentage of respondents supporting a policy.

I use the analysis file supplied with the `endorse` R package (Shiraito and Imai, 2022). It contains 5,212 respondents with one observed answer per policy, yielding 10,619 control responses and 10,229 endorsed responses. The pooled control mean is 0.736, the treatment mean is 0.723, and their difference is -0.013 (respondent-clustered SE 0.0052). These are response-weighted means across the available policy items and endorsers.

The analysis is unweighted. Blair et al. use population and policy weights; their survey over-sampled Balochistan and Khyber Pakhtunkhwa. Their unconditional Table 1 has 5,358 respondents and a control mean of 0.796. Differences from the present estimates can reflect sample selection and policy weighting as well as population weighting. The model-based probabilities below therefore describe the fitted analysis sample, not nationally representative prevalence.

The package sample also retains assignment patterns excluded by the original replication code: 35 respondents have control and endorsed items across policies, and 94 receive the same endorser on multiple policies. Excluding these 129 respondents changes the pooled ATE from -0.01283 to -0.01272 . This diagnostic leaves the sign and rounded magnitude intact. The Bayesian models retain the package sample; their sensitivity to these exclusions has not been estimated.

4 Model-Based Positive Latent Reactions

The measurement model of Bullock et al. (2011) relates ordinal policy responses to latent policy preferences and endorsement reactions. Respondent i has a latent position x_i . For question j endorsed by group k , the model adds a reaction s_{ijk} and maps the resulting position to an observed

category through question-specific loadings and cutpoints. In the baseline specification,

$$s_{ijk} \sim \mathcal{N}(\lambda_k, \omega_k^2), \quad P(s_{ijk} > 0) = \Phi(\lambda_k/\omega_k). \quad (1)$$

The mean and variance of the reaction distribution are common across policies. The reaction itself is question-specific; the specification does not assign each respondent a single fixed supporter status across all questions. The positive probability is invariant to the units of the latent scale.

A positive latent reaction need not move a response into a higher observed category. The model also includes response noise, so $P(s_{ijk} > 0)$ is distinct from the probability that the observed treatment response exceeds the control potential response. Its interpretation as favorable affect depends on the measurement model and the meaning of the endorsement cue.

I fit the baseline model without covariates using **endorse**.¹ Table 1 reports posterior means and credible intervals. Averaged equally over groups, the probability of a positive latent reaction is 43% (95% credible interval 32%–53%). This is an average of four marginal probabilities, not the probability of favoring at least one group. The Afghan Taliban have the lowest estimate, 32%, and a wide interval, 2%–49%. Their posterior probability of a negative mean reaction is 0.98; the other groups’ mean-reaction intervals include zero.

Table 1: Positive latent reactions under the baseline measurement model

Endorser	$P(s > 0)$	95% CrI	Mean reaction (ideal-point SDs)	Reaction SD (ideal-point SDs)
Kashmiri tanzeems	45%	[25%, 58%]	−0.03	0.25
Afghan Taliban	32%	[2%, 49%]	−0.08	0.20
Al-Qaeda	47%	[40%, 54%]	−0.03	0.49
Sectarian tanzeems	49%	[38%, 57%]	−0.01	0.36
Average	43%	[32%, 53%]	−0.04	0.32

Unweighted analysis sample, $n = 5,212$. CrI denotes a 95% credible interval. Mean reaction and reaction SD are posterior means in ideal-point SD units. The average row weights the four endorsers equally.

Other analyses of this survey also describe heterogeneity. Blair et al.’s middle-income treatment effects range from −0.006 to 0.001 across specifications, while low-income effects range from −0.026 to −0.019 on the rescaled outcome. Bullock et al. estimate geographic variation in latent affect, with positive estimates in parts of Punjab and negative estimates in Balochistan and NWFP.

Adding gender, rurality, education, and income covariates raises the average positive probability to 44%; a model with division-level means raises it to 49% (Table 2). The increase is in the average, not every group-specific estimate. The Afghan Taliban estimate rises to 45% under the hierarchical specification. The covariate and hierarchical summaries average over the sample’s covariates and division frequencies, respectively.

¹The fitting script uses three chains with starting values dispersed across negative, zero, and positive mean reactions. Each chain runs for 200,000 iterations, discards 20,000, and retains every 40th draw. The largest reported Gelman–Rubin statistic across the mean and variance parameters is 1.016, and the smallest effective sample size across mean reactions is 1355. The fits use the package’s default priors.

Table 2: Positive latent reactions across measurement-model specifications

Endorser	Baseline	Covariates	Hierarchical
Kashmiri tanzeems	45% [25%, 58%]	45% [34%, 54%]	49% [44%, 55%]
Afghan Taliban	32% [2%, 49%]	39% [27%, 48%]	45% [40%, 51%]
Al-Qaeda	47% [40%, 54%]	46% [40%, 52%]	49% [44%, 54%]
Sectarian tanzeems	49% [38%, 57%]	48% [39%, 56%]	51% [46%, 56%]
Average	43% [32%, 53%]	44% [38%, 51%]	49% [44%, 53%]

Posterior means with 95% credible intervals in brackets. All specifications use the unweighted package sample. Baseline reactions have common group means; covariate means depend on individual characteristics; hierarchical means vary by division. The average row weights endorsers equally.

5 Behavioral Assumptions and Supporter Proportions

The measurement model estimates a latent distribution under parametric assumptions. A simpler decomposition shows why the ATE alone cannot identify a supporter proportion. Suppose a proportion p holds favorable attitudes and the remainder holds unfavorable attitudes. Let δ^+ and δ^- be their average effects on the observed response. Then

$$\beta = p\delta^+ + (1 - p)\delta^-. \quad (2)$$

Interpreting the two classes as supporters and opponents requires behavioral assumptions, such as $\delta^+ > 0$ and $\delta^- < 0$. One observed mean effect cannot determine the three unknowns. This is a partial-identification problem (Manski, 1990). The control mean α supplies information about baseline responses, but does not resolve the decomposition without restrictions on how baseline preferences relate to attitudes and effects.

5.1 Binary Switching and Baseline Information

Suppose every supporter ends at 1 and every opponent at 0 under treatment. Then the treatment mean directly identifies p :

$$p = E[Y(1)] = \alpha + \beta. \quad (3)$$

This endpoint restriction identifies p without separately identifying both class-specific baseline means. Applied to the pooled data it gives 72.3%. But the restriction also predicts treatment variance $p(1 - p)$ and excludes all interior responses. The observed interior categories contradict this model.

The baseline still matters. Reporting α and β reveals both treatment and control means, whereas reporting only β discards information about the scale’s available room for movement. The control mean of 0.736 is a mean intensity, not a supporter fraction. For comparison, 70.7% of control responses occupy the top two categories.

Partial awareness requires another assumption. If a fraction q has an opinion and follows binary

switching, while the rest is unaffected with baseline mean a_U , then

$$p_{\text{aware}} = \frac{\alpha + \beta - (1 - q)a_U}{q}. \quad (4)$$

The expression $\alpha + \beta/q$ follows only if the unaffected subgroup has baseline mean α . With $q = 0.7$ as an illustrative assumption, allowing feasible subgroup means gives a range from 60% to 100%, rather than a unique estimate near the full-sample treatment mean. This calculation remains conditional on binary switching among opinion-holders.

5.2 Alternative Moment Restrictions

To illustrate sensitivity to assumptions, I hold the pooled ATE fixed and vary the restrictions on endorsement effects.

Symmetric class-average effects. If $\delta^+ = d$ and $\delta^- = -d$, then

$$p = \frac{\beta}{2d} + \frac{1}{2}, \quad |\beta| \leq d. \quad (5)$$

Using the pooled ATE, $d = 0.05$ implies 37%, whereas $d = 0.02$ implies 18%. These choices constrain class-average effects; they do not require every individual to move by a noninteger number of response categories. Neither value of d is identified by the pooled ATE and baseline alone.

Extreme opponent reactions. If opponents move from 1 to 0 and everyone else is unaffected, then $\beta = -(1 - p)$, so $p = 1 + \beta$, or 99%. Here p is the unaffected fraction. Calling it a supporter fraction requires an additional classification assumption; the calculation supplies no positive reactions. The restriction also implies $\beta \leq 0$ and specific changes in the endpoint masses.

A continuous normal approximation. If one approximates individual effects by $\delta_i \sim \mathcal{N}(\beta, \sigma^2)$, then $P(\delta_i > 0) = \Phi(\beta/\sigma)$. Setting $\sigma = 0.05$ gives 40%. An exact normal distribution is incompatible with differences on a bounded five-point scale. This calculation is a continuous approximation, not a fitted ordinal model. It differs from Equation (1), where normality applies to a latent reaction rather than an observed response difference.

6 Observed Distributions, Bounds, and Sensitivity

6.1 Endpoint Restrictions and Response Variation

Table 3 applies the moment calculations separately by endorser. Each ATE compares the group's pooled endorsed responses with the pooled control responses. The Afghan Taliban have the most negative difference and the lowest implied fraction under each formula. Each formula is an increasing function of the ATE, so the group ordering is the same under all of them.

Binary switching predicts a pooled treatment SD of 0.45, compared with the observed 0.29. Across policy-endorser cells, the average observed variance divided by the average predicted variance is 0.42 (Table 4). Interior responses directly violate the endpoint restriction; the variance comparison summarizes the discrepancy. Given a treatment mean m and variance v , the diagnostic is $v/[m(1 - m)]$.

Table 3: Moment-based implications by endorser

Endorser	ATE	Binary	Symmetric	Normal	Extreme
Kashmiri tanzeems	−0.0113	72.5%	38.7%	41.1%	98.9%
Afghan Taliban	−0.0174	71.8%	32.6%	36.4%	98.3%
Al-Qaeda	−0.0142	72.2%	35.8%	38.8%	98.6%
Sectarian tanzeems	−0.0085	72.7%	41.5%	43.3%	99.2%

ATEs are differences on the 0–1 response scale in the unweighted sample. Binary denotes the endpoint model; symmetric uses $d = 0.05$; normal uses the continuous approximation with $\sigma = 0.05$; extreme denotes the unaffected fraction under maximal opponent reactions.

Table 4: Observed variance versus the binary-switching prediction

Policy	Endorser	Var(T) Obs	Var(T) Pred	Ratio
Polio	Kashmiri tanzeems	0.057	0.151	0.38
Polio	Afghan Taliban	0.064	0.159	0.40
Polio	Al-Qaeda	0.060	0.144	0.41
Polio	Sectarian tanzeems	0.061	0.154	0.40
FCR	Kashmiri tanzeems	0.074	0.215	0.35
FCR	Afghan Taliban	0.085	0.225	0.38
FCR	Al-Qaeda	0.088	0.220	0.40
FCR	Sectarian tanzeems	0.081	0.214	0.38
Durand	Kashmiri tanzeems	0.101	0.223	0.45
Durand	Afghan Taliban	0.094	0.219	0.43
Durand	Al-Qaeda	0.106	0.225	0.47
Durand	Sectarian tanzeems	0.099	0.216	0.46
Curriculum	Kashmiri tanzeems	0.085	0.196	0.43
Curriculum	Afghan Taliban	0.090	0.194	0.47
Curriculum	Al-Qaeda	0.085	0.196	0.43
Curriculum	Sectarian tanzeems	0.089	0.199	0.45
Average		0.082	0.197	0.42

Treatment responses are rescaled to 0–1. Predicted variance is $m(1 - m)$ for each cell’s treatment mean m ; observed variance uses the sample variance. The final ratio divides the two column averages. Interior response categories contradict the endpoint model.

6.2 Ceiling Effects on the Discrete Scale

The response ceiling can attenuate upward movement. In the control group, 42% of responses are at 1. A respondent at that endpoint cannot report greater policy support, although a respondent with an unfavorable reaction can move down. This is a measurement constraint, not bias in an ATE defined on the observed response scale.

For an illustration, suppose supporters and opponents share the observed control distribution, and each moves up or down by the same number of categories, subject to the endpoints. Let $\bar{\delta}^+$ and $\bar{\delta}^-$ denote the resulting average shifts. Solving Equation (2) gives 59% for a one-category shift, 66% for two, and 70% for three (Table 5). The baseline distribution within attitude classes is unobserved; compatibility of the resulting treatment distribution with the data would require a separate check. The calculations show how truncation can produce a negative mean despite a majority assigned to the favorable class.

Table 5: Fixed category shifts with truncation at the scale endpoints

Shift	Eff. δ^+	Eff. δ^-	Ceiling Loss	Implied p
1 category	0.145	-0.236	0.105	58.6%
2 categories	0.218	-0.454	0.282	65.6%
3 categories	0.251	-0.631	0.499	70.1%

Both attitude classes are assigned the pooled control response distribution. Effects and ceiling loss are in 0–1 scale units; ceiling loss is the nominal upward shift minus its truncated average. Implied p matches the pooled ATE under each restriction.

6.3 Comparing the Illustrations

Table 6 collects the moment implications. The range from 18% to 99% demonstrates how different restrictions and definitions can produce different answers from the same means. It is not a confidence interval, an identification region, or evidence that the data fit every restriction.

Table 6: Illustrative implications of the pooled outcome means

Restriction	Assumption	Implied fraction
Extreme opponent reaction	$\delta^- = -1, \delta^+ = 0$	98.7%
Binary switching	All supporters $\rightarrow 1$, opponents $\rightarrow 0$	72.3%
5-point scale, 2-step shift	Shift ± 2 categories, truncated	65.6%
5-point scale, 1-step shift	Shift ± 1 category, truncated	58.6%
Normal heterogeneity ($\sigma = 0.05$)	$\delta_i \sim \mathcal{N}(\beta, 0.05^2)$	39.9%
Symmetric effects ($d = 0.05$)	$\delta^+ = +0.05, \delta^- = -0.05$	37.2%
Symmetric effects ($d = 0.02$)	$\delta^+ = +0.02, \delta^- = -0.02$	17.9%

All rows use the unweighted pooled ATE -0.013 and, where required, control mean 0.736 . Fractions refer to the class defined by each restriction. Binary switching is contradicted by the observed interior categories. The normal row is a continuous approximation; the extreme row is an unaffected fraction. No sampling intervals are shown.

Figure 1 varies the ATE while holding each restriction fixed. Lines stop where the implied fraction would leave $[0, 1]$. For example, symmetric effects with magnitude d require $|\beta| \leq d$;

positive ATEs are incompatible with the extreme-opponent restriction.

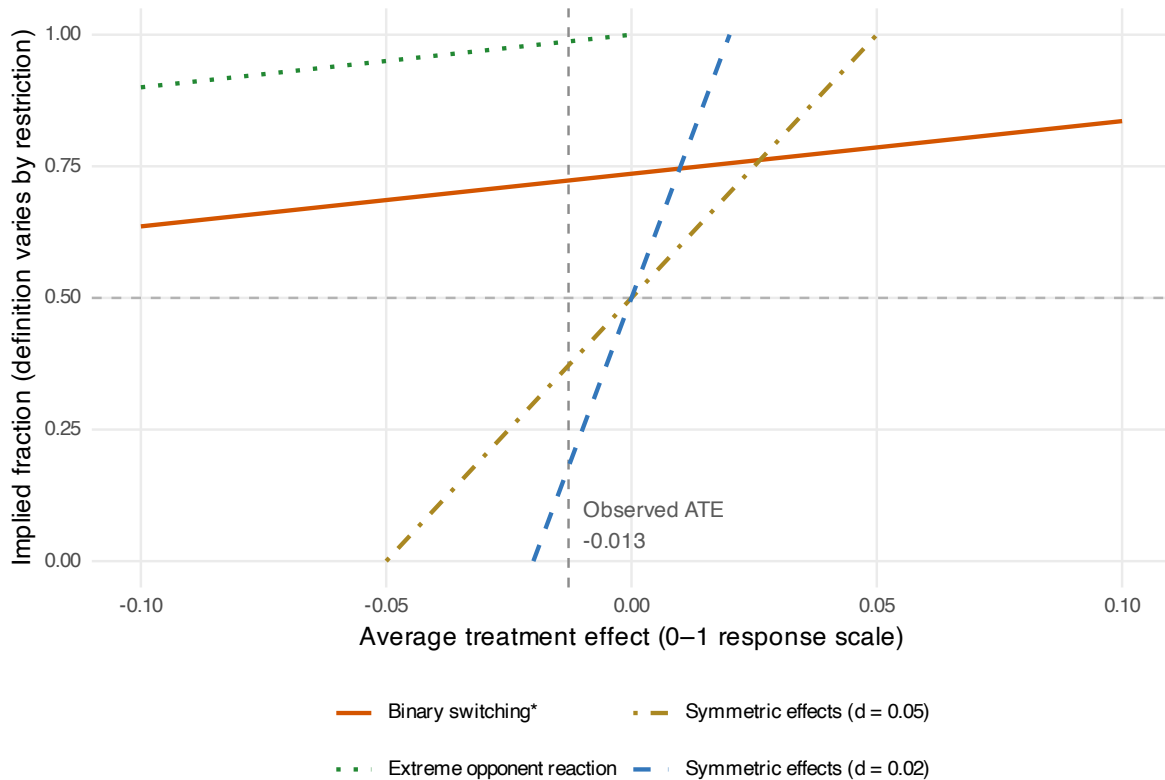


Figure 1: Sensitivity of moment implications to the ATE on the 0–1 scale. The baseline mean for binary switching is held at 0.736. Curves omit infeasible probabilities; they do not flatten at zero or one. The vertical line marks the unweighted pooled ATE, and the horizontal line marks one half. The asterisk identifies the contradicted endpoint restriction. Curves describe deterministic formulas, not sampling uncertainty.

6.4 Bounds on Observed Response Increases

A different calculation bounds observed response changes without specifying a latent reaction distribution. Under valid randomization, observation, and consistency assumptions, the experiment identifies treatment and control marginal outcome distributions, but not their joint distribution. Allowing every coupling of the estimated pooled marginals gives bounds of 1% to 58% on the fraction of person–policy responses for which $Y(1) > Y(0)$ (Table 7). These are plug-in bounds conditional on the empirical marginals, not confidence limits. Pooling also discards policy-specific restrictions on the joint distribution.

Independent random pairing yields a higher treatment response 34% of the time, a lower one 37%, and a tie 29%. Independence is an additional assumption about the coupling, not a feature identified by the experiment. The bounds permit either a minority or a majority of observed responses to increase. They do not bound the fraction holding favorable attitudes or the latent probability in Section 4.

Table 7: Random pairing and bounds on observed response increases

Endorser	$P(T < C)$	$P(T = C)$	$P(T > C)$	Min. Positive	Max. Positive
Kashmiri tanzeems	37%	29%	34%	2%	58%
Afghan Taliban	38%	29%	34%	1%	58%
Al-Qaeda	37%	29%	34%	1%	58%
Sectarian tanzeems	36%	29%	35%	1%	58%
All groups	37%	29%	34%	1%	58%

T and C denote treatment and control responses on the five-point scale. The first three columns pair them independently; the last two optimize $P(T > C)$ over all couplings of the empirical pooled marginals. The unit is a person-policy response. Bounds omit sampling uncertainty and do not measure positive latent affect.

6.5 Comparison with Direct Polling

Pew’s 2012 survey reports 14% favorable views of the Afghan Taliban, 45% unfavorable views, and 41% with no opinion (Pew Research Center, 2012). If stated opinions are classified correctly, assigning all nonrespondents to the favorable category gives an upper bound of 55%. That bound does not allow respondents who report unfavorable opinions to conceal favorable ones.

The corresponding Afghan Taliban estimate from the 2009 endorsement data is 32% (95% credible interval 2%–49%). It falls within the range obtained by varying the classification of Pew nonrespondents. But the surveys differ in year, sample composition, and measurement, and the latent probability is not the same response as direct favorability. Social desirability and fear may contribute to discrepancies; these data do not identify their contribution.

7 Discussion

Researchers should distinguish the average effect on an observed response, the fraction of responses with positive effects, and the prevalence of favorable attitudes. Random assignment identifies the first under the usual design assumptions. Marginal response distributions can bound the second. Estimating the third requires a model connecting responses to attitudes, and its credibility depends on that connection.

Reporting treatment and control distributions alongside the ATE makes such assumptions easier to assess. Endpoint restrictions can be checked against interior responses; ceiling constraints are visible in the baseline distribution. Model-based positive probabilities should be reported with credible intervals, the population over which they are averaged, and evidence about model fit and sensitivity to priors.

The present reanalysis leaves several substantive checks open: population adjustment, Bayesian sensitivity to the assignment exclusions, prior sensitivity, and posterior predictive fit. Endorsement credibility is another measurement question. If respondents disbelieve a cue or infer information beyond the endorser’s identity, the mapping to attitudes can change; the direction of that change is not established by the ATE alone.

8 Conclusion

A negative average endorsement effect does not establish how widely unfavorable attitudes are held. In the unweighted analysis sample, the fitted measurement model implies a 43% probability of a positive latent reaction, averaged across four endorsers, with a 95% credible interval of 32%–53%. Moving from an average response effect to a distribution of attitudes requires stating and defending the additional assumptions.

References

- Blair, G., Fair, C. C., Malhotra, N., and Shapiro, J. N. (2013). Poverty and support for militant politics: Evidence from Pakistan. *American Journal of Political Science*, 57(1):30–48.
- Blair, G., Imai, K., and Lyall, J. (2014). Comparing and combining list and endorsement experiments: Evidence from Afghanistan. *American Journal of Political Science*, 58(4):1043–1063.
- Bullock, W., Imai, K., and Shapiro, J. N. (2011). Statistical analysis of endorsement experiments: Measuring support for militant groups in Pakistan. *Political Analysis*, 19(4):363–384.
- Lyall, J., Blair, G., and Imai, K. (2013). Explaining support for combatants during wartime: A survey experiment in Afghanistan. *American Political Science Review*, 107(4):679–705.
- Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review*, 80(2):319–323.
- Pew Research Center (2012). Pakistani public opinion ever more critical of U.S. Technical report, Pew Research Center.
- Shiraito, Y. and Imai, K. (2022). *endorse: Bayesian Measurement Models for Analyzing Endorsement Experiments*. R package version 1.6.2.