

Supplemental Material for “Survey Design, Order
Effects, and Causal Mediation Analysis”

August 3, 2021

1 Articles with CMA for Survey-Experimental Data

Following is a chronologically arranged list of recent (post-2012) articles on political-science subjects that report some causal mediation analysis for data obtained from a survey experiment. We combined searches within particular journals with searches on Google Scholar, so this list is non-exhaustive. Following each article, we indicate what survey design we believe was used, OM or MO (or, rarely, both). In many cases, we make inferences about order from description of items in the body of the article, but we did not find the full survey instrument. “Not sure” indicates that we neither obtained the full survey nor felt comfortable guessing about question order based on the article’s description of the items.

1. Arceneaux, Kevin. 2012. Cognitive Biases and the Strength of Political Arguments. *American Journal of Political Science* 56, 2 (April): 271-285. (MO)
2. Jacobs, Alan M., and J. Scott Matthews. 2012. Why Do Citizens Discount the Future? Public Opinion and the Timing of Policy Consequences. *British Journal of Political Science* 42, 4 (October): 903-935. (OM)
3. Lupu, Noam. 2013. Party brands and partisanship: Theory with Evidence from a Survey Experiment in Argentina. *American Journal of Political Science* 57, 1 (January): 49-64. (OM)
4. Tomz, Michael R. and Jessica L. P. Weeks. 2013. Public Opinion and the Democratic Peace. *American Political Science Review* 107, 4 (November): 849-865. (OM, plus MO replication)
5. Egan, Patrick J. 2014. “Do Something” Politics and Double-Peaked Policy Preferences. *Journal of Politics* 726, 2 (February): 333-349. (MO)
6. Hartman, Todd K., Benjamin J. Newman, and C. Scott Bell. 2014. Decoding Prejudice Towards Hispanics: Group Cues and Public Reactions to Threatening Immigrant Behavior. *Political Behavior* 36, 1 (March): 143-164. (not sure)
7. Pedulla, David S. 2014. The Positive Consequences of Negative Stereotypes: Race, Sexual Orientation, and the Job Application Process. *Social Psychology Quarterly* 77, 1 (March): 75-94. (not sure)
8. Gadarian, Shana Kushner, and Bethany Albertson. 2014. Anxiety, Immigration, and the Search for Information. *Political Psychology* 35,2 (April): 133-164. (not sure)
9. Curtis, K. Amber. 2014. In Times of Crisis: the Conditions of Pocketbook Effects. *International Interactions* 40, 3: 402-430. (not sure)
10. Halperin, Eran, Ruthie Pliskin, Tamar Saguy, Varda Liberman, and James J. Gross. 2014. Emotion Regulation and the Cultivation of Political Tolerance: Searching for a New Track for Intervention. *Journal of Conflict Resolution* 58, 6 (September): 1110-1138. (MO)
11. Baker, Andy. 2015. Race, Paternalism, and Foreign Aid: Evidence from US Public Opinion. *American Political Science Review* 109, 1 (February): 93-109. (OM)
12. Naoi, Megumi, and Ikuo Kume. 2015. Workers or Consumers? A Survey Experiment on the Duality of Citizens’ Interests in the Politics of Trade. *Comparative Political Studies* 48, 10 (September): 1293-1317. (not sure)

13. Claassen, Christopher. 2016. Group Entitlement, Anger and Participation in Intergroup Violence. *British Journal of Political Science* 46, 1 (January): 127-148. (MO)
14. Ha, Shang E., Soo Jin Cho, and Jeong-Han Kang. 2016. Group Cues and Public Opposition to Immigration: Evidence from a Survey Experiment in South Korea. *Journal of Ethnic and Migration Studies* 42, 1: 136-149. (not sure)
15. Clifford, Scott, and Ben Gaskins. 2016. Trust me, I Believe in God: Candidate Religiousness as a Signal of Trustworthiness. *American Politics Research* 44, 6 (November): 1066-1097. (not sure)
16. Jacobs, Alan M., and J. Scott Matthews. 2017. Policy Attitudes in Institutional Context: Rules, Uncertainty, and the Mass Politics of Public Investment. *American Journal of Political Science* 61, 1 (January): 194-207. (MO)
17. Shelef, Nadav G., and Yael Zeira. 2017. Recognition Matters! UN State Status and Attitudes toward Territorial Compromise. *Journal of Conflict Resolution* 61, 3 (March): 537-563. (not sure)
18. Tingley, Dustin. 2017. Rising Power on the Mind. *International Organization* 71,S1 (April): S165-S188. (OM)
19. Valentino, Nicholas A., and Fabian G. Neuner. 2017. Why the Sky Didn't Fall: Mobilizing Anger in Reaction to Voter ID Laws. *Political Psychology* 38, 2 (April): 331-350. (not sure)
20. Mossler, Max V., et al. 2017. How Does Framing Affect Policy Support for Emissions Mitigation? Testing the Effects of Ocean Acidification and Other Carbon Emissions Frames. *Global Environmental Change* 45 (July): 63-78. (not sure)
21. Caverley, Jonathan D., and Yanna Krupnikov. 2017. Aiming at Doves: Experimental Evidence of Military Images' Political Effects. *Journal of Conflict Resolution* 61, 7 (August): 1482-1509. (OM)
22. Falco-Gimeno, Albert, and Jordi Munoz. 2017. Show Me Your Friends: A Survey Experiment on the Effect of Coalition Signals. *Journal of Politics* 79, 4 (November): 1454-1459. (not sure)
23. Koch, Thomas, and Christina Peter. 2017. Effects of Equivalence Framing on the Perceived Truth of Political Messages and the Trustworthiness of Politicians. *Public Opinion Quarterly* 81, 4 (winter): 847-865. (MO)
24. Huddleston, R. Joseph, and Nicholas Weller. 2017. Unintended Causal Pathways: Probing Experimental Mechanisms Through Mediation Analysis. SSRN 2964336. (both OM and MO, with discussion)
25. Alkon, Meir, and Erik H. Wang. 2018. Pollution Lowers Support for China's Regime: Quasi-experimental Evidence from Beijing. *The Journal of Politics* 80, 1 (January): 327-331. (not sure)
26. Piston, Spencer, Yanna Krupnikov, Kerri Milita, and John Barry Ryan. 2018. Clear as Black and White: the Effects of Ambiguous Rhetoric Depend on Candidate Race. *Journal of Politics* 80, 2 (April): 662-674. (OM)
27. Suhay, Elizabeth, and Jeremiah Garretson. 2018. Science, Sexuality, and Civil Rights: Does Information on the Causes of Sexual Orientation Change Attitudes?. *Journal of*

- Politics* 80,2 (April): 692-696. (OM)
28. Utych, Stephen M. 2018. How Dehumanization Influences Attitudes Toward Immigrants. *Political Research Quarterly* 71, 2 (June): 440-452. (OM)
 29. Aytaç, S. Erdem, Luis Schiumerini, and Susan Stokes. 2018. Why Do People Join Backlash Protests? Lessons from Turkey. *Journal of Conflict Resolution* 62,6 (July): 1205-1228. (OM)
 30. Delton, Andrew W., Michael Bang Petersen, and Theresa E. Robertson. 2018. Partisan Goals, Emotions, and Political Mobilization: The Role of Motivated Reasoning in Pressuring Others to Vote. *Journal of Politics* 80, 3 (July): 890-902. (not sure)
 31. Brutger, Ryan, and Joshua D. Kertzer. 2018. A Dispositional Theory of Reputation Costs. *International Organization* 72, 3 (summer): 693-724. (OM)
 32. Hameleers, Michael, Linda Bos, and Claes de Vreese. 2018. Framing Blame: Toward a Better Understanding of the Effects of Populist Communication on Populist Party Preferences. *Journal of Elections, Public Opinion and Parties* 28, 3 (August): 380-398. (MO)
 33. Kreps, Sarah, and Sarah Maxey. 2018. Mechanisms of Morality: Sources of Support for Humanitarian Intervention. *Journal of Conflict Resolution* 62, 8 (September): 1814-1842. (not sure)
 34. Arnesen, Sveinung, Mikael P. Johannesson, Jonas Linde, and Stefan Dahlberg. 2018. Do Polls Influence Opinions? Investigating Poll Feedback Loops Using the Novel Dynamic Response Feedback Experimental Procedure. *Social Science Computer Review* 36, 6 (December): 735-743. (not sure)
 35. Hassell, Hans J.G., and Emily E. Wyler. 2019. Negative Descriptive Social Norms and Political Action: People Aren't Acting, So You Should. *Political Behavior* 41, 1 (March): 231-256. (not sure)
 36. Sulitzeanu-Kenan, Raanan, and Reimut Zohlnhöfer. 2019. Policy and Blame Attribution: Citizens' Preferences, Policy Reputations, and Policy Surprises. *Political Behavior* 41, 1 (March): 53-77. (not sure)
 37. Andrews-Lee, Caitlin. 2019. The Revival of Charisma: Experimental Evidence From Argentina and Venezuela. *Comparative Political Studies* 52, 5 (April): 687-719. (OM)
 38. Fisk, Kerstin, Jennifer L. Merolla, and Jennifer M. Ramos. 2019. Emotions, Terrorist Threat, and Drones: Anger Drives Support for Drone Strikes. *Journal of Conflict Resolution* 63, 4 (April). (MO)
 39. Jones, Philip Edward, and Paul R. Brewer. 2019. Gender Identity as a Political Cue: Voter Responses to Transgender Candidates. *Journal of Politics* 81, 2 (April): 697-701. (not sure)
 40. Lupu, Yonatan, and Geoffrey P.R. Wallace. 2019. Violence, Nonviolence, and the Effects of International Human Rights Law. *American Journal of Political Science* 63, 2 (April): 411-426. (OM)
 41. Arora, Maneesh, and Christopher T. Stout. 2019. Letters for Black Lives: Co-ethnic Mobilization and Support for the Black Lives Matter Movement. *Political Research Quarterly* 72, 2 (June): 389-402. (not sure)

42. Borau, Sylvie, and Jean-François Bonnefon. 2019. The imaginary Intrasexual Competition: Advertisements Featuring Provocative Female Models Trigger Women to Engage in Indirect Aggression. *Journal of Business Ethics* 157, 1 (June): 45-63. (MO)
43. Gervais, Bryan T. Rousing the Partisan Combatant: Elite Incivility, Anger, and Antide-liberative Attitudes. *Political Psychology* 40, 3 (June): 637-655. (not sure)
44. Kane, John V., and Benjamin J. Newman. 2019. Organized Labor as the New Undeserving Rich?: Mass Media, Class-Based Anti-Union Rhetoric and Public Support for Unions in the United States. *British Journal of Political Science* 49, 3 (July): 997-1026. (MO)
45. Silverman, Daniel. 2019. What Shapes Civilian Beliefs about Violent Events? Experimental Evidence from Pakistan. *Journal of Conflict Resolution* 63, 6 (July): 1460-1487. (MO)
46. Huddleston, R. Joseph. 2019. Think Ahead: Cost Discounting and External Validity in Foreign Policy Survey Experiments. *Journal of Experimental Political Science* 6, 2 (summer): 108-119. (both OM and MO, with discussion)
47. Johnson, Austin P., Nehemia Geva, and Kenneth J. Meier. 2019. Can Hierarchy Dodge Bullets? Examining Blame Attribution in Military Contracting. *Journal of Conflict Resolution* 63, 8 (September): 1965-1985. (OM)
48. Perez, Efren O., Maggie Deichert, and Andrew M. Engelhardt. 2019. E Pluribus Unum? How Ethnic and National Identity Motivate Individual Reactions to a Political Ideal. *Journal of Politics* 81, 4 (October): 1420-1433. (MO)
49. Sanaei, Ali. 2019. Time is of the Essence: The Causal Effect of Duration on Support for War. *Journal of Peace Research* 56, 6 (November). (not sure)
50. Tolochko, Petro, Hyunjin Song, and Hajo Boomgaarden. 2019. That Looks Hard!: Effects of Objective and Perceived Textual Complexity on Factual and Structural Political Knowledge. *Political Communication* 36, 4 (October-December): 609-628. (MO)
51. Fang, Songying, and Xiaojun Li. 2020. Historical Ownership and Territorial Disputes. *Journal of Politics* 82, 1 (January): 345-360. (OM)
52. Tomz, Michael R., and Jessica L. P. Weeks. 2020. Human Rights and Public Support for War. *Journal of Politics* 82, 1 (January): 182-194. (MO)
53. Kapatadze, A., and T. Zeitzoff. Forthcoming. In the Shadow of Conflict: How Emotions, Threat Perceptions and Victimization Influence Foreign Policy Attitudes. *British Journal of Political Science*, published online, July 2019): 1-22. (not sure)

2 Supplemental Tables

2.1 Replication of Treatment Effect

Table A1: Democracy’s Effect on Favorability to Attack

	Tomz-Weeks				Replication	
	UK.OM	US.OM (b)	US.OM (w)	US.MO	OM	MO
Democracy	20.9	41.9	38.5		17.7	19.3
Not Dem.	34.2	53.3	50.0		26.5	28.6
Difference	-13.3	-11.4	-11.5	-11.7	-8.8	-9.3
95% CI	(-19.6, -7.0)	(-16.8, -5.9)	(-14.7, -8.3)		(-16.0,-1.7)	(-16.7, -2.0)
N.Dem	398	634	972		260	254
N.Not	364	639	972		260	262
N	762	1273	972	797	520	516

In Table A1, we show estimated democracy effects on support for attacking the hypothetical nuclearizing nation. The first four columns correspond to published results from Tomz and Weeks (2013) (“T-W” for brevity). Following their precedent, we dichotomized support for attack, separating strong and not-strong support from uncertainty, opposition, and strong opposition. On that scheme, we show the percentages favorable to attack for a democratic foe and for a not-democratic foe, with the difference estimating the direct treatment effect (of democracy). Column 1 shows their results from UK respondents. Column 2 shows more between-subject results, from an OM design and US sample. Column 3 shows their within-subject results, based on American respondents asked about both kinds of foe, in separate sessions. (In tables that follow, all results from the T-W study relate to these data.) Our calculated confidence intervals differ trivially from those reported in their Table 1 (p. 854). Column 4 describes their replication under an MO design, employing Amazon Mechanical Turk workers. We could not compute a confidence interval because the article provides only the point estimate and total N (footnote 13, page 855). Columns 5 and 6 show results from our replication using OM and MO designs, respectively.

2.2 Distributions of Outcome and Mediators

Table A2 shows descriptive statistics of the outcome variable and mediators for our replication, according to design choice. Only immorality differed significantly in means across the modules.

These numbers reflect the same method that T-W used for aggregating each mediator. They measured *threats* by combining respondent choices given a list of six threatening events:

1. the country would build nuclear weapons (*Build*);

Table A2: Moments of Key Variables in OM and MO Modules

	OM		MO		Diff. Means
	Mean	(Std.Dev.)	Mean	(Std.Dev.)	<i>p</i> -value
Attack Favorability	0.22	(0.42)	0.24	(0.43)	0.46
Threats	1.89	(1.39)	1.85	(1.41)	0.64
Costs	2.14	(1.34)	2.21	(1.35)	0.40
Success	1.05	(0.72)	1.11	(0.73)	0.15
Immorality	0.55	(0.50)	0.61	(0.49)	0.04

2. the country would build nuclear weapons and threaten to use them against another country (*Threaten*);
3. the country would build nuclear weapons and threaten to use them against the U.S. or a U.S. ally (*Threaten US*);
4. the country would build nuclear weapons and launch a nuclear attack against another country (*Attack*);
5. the country would build nuclear weapons and launch a nuclear attack against the U.S. or a U.S. ally (*Attack US*);
6. the credibility, prestige, or reputation of the U.S. would suffer (*Lose Prestige*).

Respondents were asked which of the events “would have more than a 50 percent chance of happening if the U.S. did not attack the country’s nuclear sites,” with multiple selections permitted and a “None” option as well. The first five items have a nested quality insofar as each of (2) through (5) implies (1), (3) is arguably a subset of (2), and so on. We followed the authors in ignoring that feature, and merely counting the total number chosen. We did, however, discard 69 respondents who picked at least one item *plus* the “None” option and 7 respondents who skipped the question completely.

Table A3 shows that some of the threats were viewed as less likely by our replication respondents than by the original subjects, while others were seen as more likely. In general, our respondents perceived not-attacked democracies as less threatening than non-democracies, as in the original study. For individual items, the effect of democracy on perceived threat varied across OM versus MO designs. However, when summing the number of threats, the total-threats counts for the two groups were very similar. A chi-squared test supports pooling ($p = 0.89$), with about 10 percent seeing no threats as more likely than not, then 37, 23, 15, 5, 7 and 3 percent judging as plausible 1 through 6 of these threats, respectively.

T-W ignored the sixth threat from their list, constructing their threats variable from only the first five events. (This is why two rows are blank in Table A3). We follow their precedent

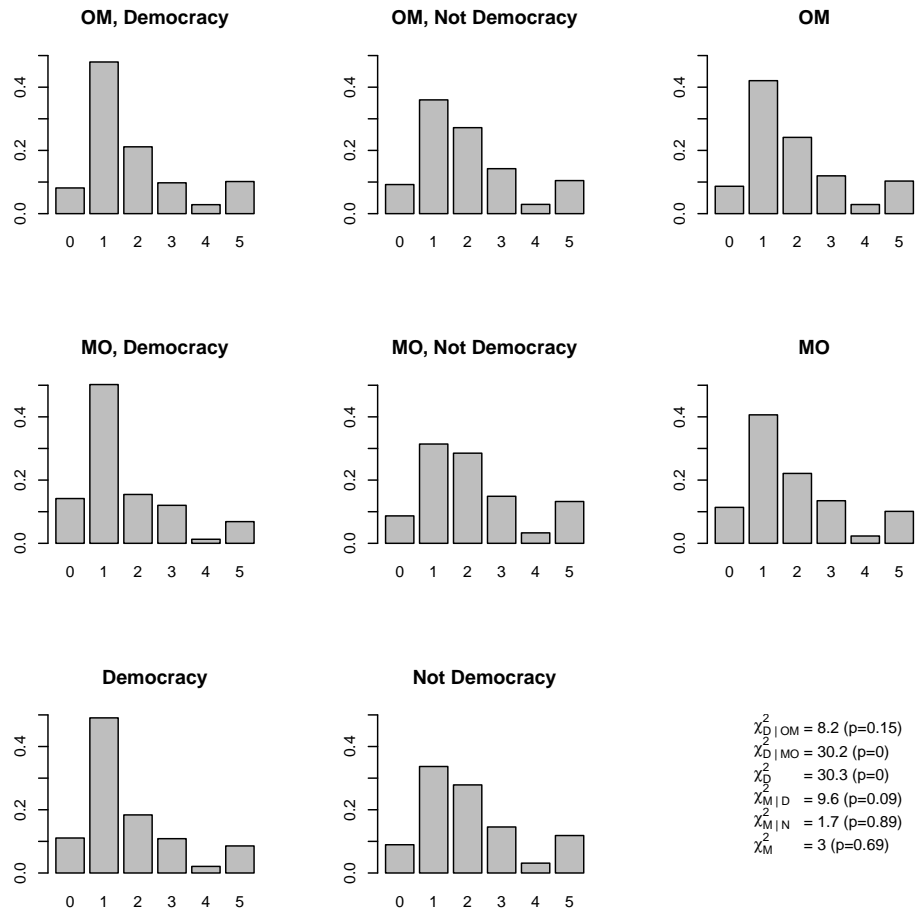


Figure A1: Distributions of Threat Perceptions, by Treatment and Design

Table A3: Distribution of Threat Perceptions, by Treatment and Design

	Tomz-Weeks		OM Replication		MO Replication	
	Dem.	Not	Dem.	Not	Dem.	Not
Build	72	75	88	85	78	87
Threaten	38	52	41	54	35	55
Threaten US	34	45	19	24	15	28
Attack	26	34	22	20	18	25
Attack US	24	30	12	14	10	17
Lose Prestige			15	17	15	18
None			8	9	14	8
<i>N</i>	972	972	246	239	233	242

hereafter. Figure A2 shows the distributions for our OM and MO modules for the five-count version of the *threats* variable. To the naked eye, the effect of democracy (or not) on threats is more dramatic than the effect of design module, and the chi-squared statistics confirm as much. But the interactions are more subtle. The chi-squared statistics for conditional distributions reveal that the contrast between the democracy and not-democracy distributions is stronger in the MO than OM module, and that the module effect is marginally significant conditional on democracy, but not on not-democracy. So there is some evidence here that the relationship between the main treatment, democracy, and the *threats* mediator is sensitive to questionnaire design.

When measuring perceived *costs* associated with an attack, T-W asked respondents which among the following set of four possible effects had a greater than 50% chance of happening, in the event of an attack:

1. The country would respond by attacking the U.S. or a U.S. ally (*Retaliate*);
2. The U.S. military would suffer many casualties (*Casualties*);
3. The U.S. economy would suffer (*Economy*);
4. U.S. relations with other countries would suffer (*Relations*).

The likelihood of *success* was measured with two items asking whether, following an attack, there would be a greater than 50% chance that this attack “would prevent the country from making nuclear weapons in the near future” (*Deterrence (Short)*) and “... in the long run” (*Deterrence (Long)*). The cost and success events, all conditional on attack, were combined in a single survey item, affecting the meaning of the “None” response (which T-W omitted from their Table 4).

Table A4 shows the percentage of respondents in a particular study and treatment condition that indicated that a particular cost or success result was likely. For both the costs

Table A4: Distribution of Perceived Costs and Success, by Treatment and Design

	Tomz-Weeks		OM Replication		MO Replication	
	Dem.	Not	Dem.	Not	Dem.	Not
Retaliate	39	39	61	53	59	59
Casualties	33	32	53	47	52	51
Economy	31	31	45	39	48	45
Relations	53	49	66	63	65	64
Deterrence (Short)	61	66	72	72	73	79
Deterrence (Long)	25	30	30	35	34	37
None			6	4	5	2
<i>N</i>	972	972	256	247	244	249

and success mediators, our OM and MO respondents differed very little from one another. For the aggregate number of costs, a chi-squared test favors pooling the modules ($p = 0.82$). Percentages seeing 0 through 4 costs as likely were roughly: 14, 20, 22, 23, and 21, respectively. (Note that the response option “None” reflected in Table A4 applies to costs and successes, whereas we coded zeros for the costs variable according to the four cost outcomes only.)

Figure A2 shows that, unlike *threats*, *costs* shows no sensitivity in its distribution to module or democracy status. From these patterns, we might not expect to see much variation in the proportion mediated attributable to this mediator, across modules.

2.3 Full Replication Results

Figure 1 in the paper illustrates the main coefficients in our replication model in which we follow the T-W article in collapsing attack favorability into a dichotomous measure. That figure includes only select effects from a probit model. Full results are shown in Table A5.

2.4 Assessing Differences in Proportion Mediated

The proportion mediated is a ratio of two estimates, both of which often have large confidence intervals (quasi-Bayesian intervals as the defaults for the `mediation` package, for example). In practice, the confidence intervals around the proportion mediated quantity tend to be large and are not often reported in applied research.

To provide some sense for the role of sample size in generating large amounts of uncertainty about point estimates, we re-estimated these quantities by duplicating, triplicating, etc. each of our observations, to generate larger and larger data sets with otherwise identical properties.

Figure A2 illustrates the proportion mediated through the *threats* channel in a sample of N , $2 \times N$, $4 \times N$, and $6 \times N$. By tripling our sample, we were able to achieve smaller confi-

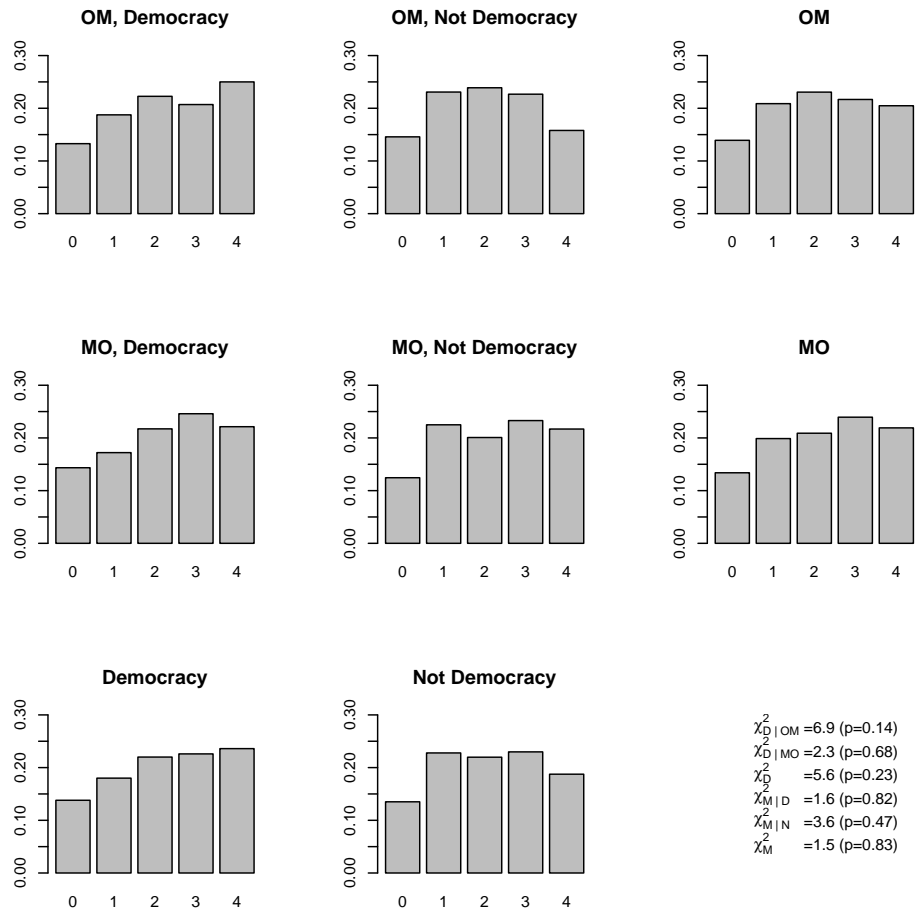


Figure A2: Distributions of Cost Perceptions, by Treatment and Design

Table A5: Replication Results: Probit Models

	Tomz-Weeks	Replication		
	(OM)	Pooled	OM	MO
Treatments				
Democracy	-0.18	-0.27	-0.30	-0.28
Ally	-0.06			
Trade	-0.05			
Mediators				
Threats	0.30	0.34	0.40	0.29
Costs	-0.21	-0.20	-0.26	-0.17
Success	0.23	0.15	0.06	0.21
Immorality	-1.12	-1.18	-0.95	-1.61
Controls				
Militarism	-0.02	0.13	0.12	0.06
Internationalism	0.02	0.18	-0.00	0.43
Republicanism	0.10	-0.03	0.07	-0.02
Ethnocentrism	0.09	0.11	-0.04	0.25
Religiosity	-0.03	0.29	0.40	0.11
Male	0.05	0.07	-0.00	0.20
White	-0.18	0.05	0.09	-0.02
Age	0.01	-0.01	-0.01	-0.00
Education	-0.06	0.00	-0.12	0.13
Intercept	-0.30	-0.48	-0.03	-0.96
<i>N</i>	972 × 2	939	475	464

Notes: Probit model of supporting military strike, rather than opposing or declining to support or oppose. T-W estimates are based on within-variance for 972 respondents (two responses each, one for democracy and one for non-democracy), replication estimates are from between-variance (one response each). Bolded coefficients indicate p values less than 0.05.

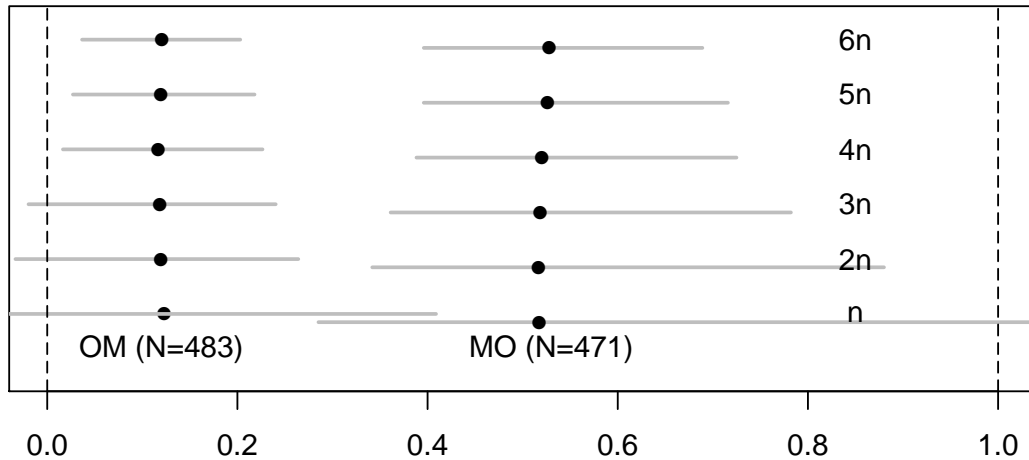


Figure A2: Estimated Proportion Mediated by Sample Size, Threats Mediator

dence intervals that stayed within the logical bounds of 0 and 1, and the precision improved somewhat as we increased the sample size further. It would seem, therefore, that to find statistically distinct results by survey design, quite large samples—larger than those collected in most applied research—can be needed.

2.5 Handling Multiple Mediators

The `mediation` packages allows for estimate of mediation results when multiple mediators are causally dependent on one another. We opted not to use this option, because the study we replicated did not make theoretical claims about causal dependence among mediators.

Still, to ensure that results still differed across survey designs, we replicated the original estimates using the `multimed` option for multiple mediators. Results are qualitatively similar, though the exact numbers differ.

Table A6: Mediation Analysis by Survey Design, Multiple-Mediators

	Tomz-Weeks		Replication			
	OM		OM		MO	
Threats	-4.0	(34%)	-1.3	(12%)	-3.6	(28%)
Costs	-0.4	(4%)	-1.4	(12%)	9.9	(1%)
Success	-0.7	(6%)	-0.1	(1%)	-0.3	(2%)
Immorality	-1.7	(15%)	-2.4	(33%)	-6.1	(47%)

Notes: These estimates, ACMEs and percentage mediated, were obtained from the `multimed` function, wherein individual mediators are chosen as “main” and others as “alternatives.”