Survey Design, Order Effects, and Causal Mediation Analysis

Stephen Chaudoin, Brian J. Gaines, and Avital Livny

September 21, 2020

Abstract

Causal mediation analysis (CMA) requires measurement of an outcome variable (O) with and without treatment, plus a set of mediator variables (M) that constitute possible pathways for the treatment effect. There is no consensus on whether surveys should measure potentially mediating variables before or after the outcome variables – "MO" or "OM." We use a replication exercise to demonstrate how the order of mediator and outcome items can be consequential for the results from CMA. Order can affect mediation conclusions, even if the treatment effect is similar across designs. As such, randomizing order is usually prudent, though best practice depends on the researcher's contextual knowledge about her particular application. (109 words)

keywords: mediation, survey, causality.

Supplementary material for this article is available in the online appendix and at `https://www.stephenchaudoin.com` and at `https://www.alivny.com`.

Replication files are available in the JOP Data Archive on Dataverse (`http://thedata.harvard.edu/dvn/dv/jop`).

This research was conducted after approval by the University of Illinois at Urbana-Champaign Institutional Review Board for human subjects research.

This research was funded by the University of Illinois at Urbana-Champaign.

Causal mediation analysis (CMA) aims to decompose treatment effects by mediating channel. It is rapidly growing in popularity, especially in survey-experimental work. When designing an instrument, an important choice arises: whether to measure outcomes *then* mediators (an "OM" design) or mediators *then* outcomes ("MO"). It is unclear which to prefer, and there is no consensus in the existing literature. MO resembles the posited causal sequence, but researchers often measure the outcome directly after treatment, to prevent intervening questions from moderating treatment effects.

In this short article, we argue that the OM/MO choice can impact: (1) the distribution of outcome variables and/or mediators; (2) the effect of treatment on either; and (3) mediation results, even when it does not moderate the overall treatment effect. The first two occur because ordering can alter how respondents understand later questions, by priming latent thoughts or framing issues, or because survey fatigue induces satisficing or anchoring in later questions. The third effect is more subtle: even if the treatment effect is similar under both designs, the proportions of the effect attributed to a particular mediator can still change. Importantly, this effect can occur even when the first two do not.

After describing these three effects theoretically, we demonstrate them in practice by replicating a prominent survey experiment on democratic-peace theory (Tomz and Weeks, 2013). While OM and MO versions produced similar estimates of the overall treatment effect, survey design impacted distributions of some variables (Effect 1), though it did not affect relationships between variables (Effect 2). Still, design-choice affected mediation results, altering which mediators mattered more for the overall effect (Effect 3). We conclude with guidance on how researchers should proceed, offering some guidelines. In addition to using contextual knowledge to decide whether fatigue and/or priming seem likely in a study, randomly assigning respondents to an OM or MO version is often wise.

1 Why OM vs. MO Can Matter

To assess common practices employed in CMA applications, we searched for existing work using survey-experimental data.[1] We found fifty-five articles published between 2012 and 2019, spanning disciplinary subfields, with ten appearing in the *Journal of Politics*. Across the set of articles, there does not appear to be a modal survey design, and discussion of order is minimal. Sixteen used an OM design, fourteen MO, and twenty-three were not clear about survey structure.[2] Only two – Tomz and Weeks (2013) and Huddleston (2019) – discussed the OM/MO choice, with both conducting experiments using each setup.

Although the extant literature does not often discuss the differences between OM and MO designs, the choice can impact quantities of interest in CMA, including the average treatment effect (ATE), average causal mediation effect (ACME), the average direct effect (ADE), and the proportion mediated (PM), further complicating one's ability to draw clear inferences from CMA (Bullock, Green and Ha, 2010). If $Y_i(t,m)$ is the potential outcome for individual i when she is assigned to treatment ($t = 1$) and when the mediating variable equals m, then the ATE is the difference in outcomes under assignment to treatment versus control.[3] Because treatment can also affect mediators, $M_i(t)$ is the potential value of the mediating variable when i is assigned to treatment, and the ACME is the difference in outcomes under $M_i(1)$ and $M_i(0)$. Meanwhile, the ADE is the effect of treatment on outcome, holding fixed the value of the mediator. Finally, the proportion of the treatment effect that goes "through" the mediating variable – the PM – is usually expressed as a ratio of the ACME to the ATE.

Existing research suggests numerous ways that the OM/MO choice could impact these quantities and, therefore, the results of CMA. In the context of surveys with multiple experiments, Transue, Lee and Aldrich (2009) identify how treatment in one experiment can affect outcome

[1] We focused on survey-experimental studies because they have delineated treatments manipulated by the researcher and they often consider mediators pertaining to psychological constructs. In other settings, where theory implies a clearer MO sequencing, researchers should use the design that fits their causal model.

[2] We list the full set in the Appendix.

[3] With randomized treatment assignment, this quantity can be calculated in straightforward ways, such as simple comparisons of means in outcomes across treatment conditions.

variables in later experiments within a given survey, and a similar phenomenon is possible in single-experiment studies, where earlier questions can prime respondents to frame issues in particular ways (Tourangeau, Rips and Rasinski, 2000). In his study of audience costs, Huddleston (2019) finds that the presence or absence of a question about the costs of military action has both a direct effect on approval of a military intervention (ATE) and an indirect effect, as it alters the time horizon respondents use to assess interventions (ACME).[4] For this reason, MO designs (compared to OM) can even induce a mediation effect where there would otherwise be none.

Just as question order can impact responses, so too can response order within questions (Sudman and Bradburn, 1974). The two can even interact, with later question inducing more "satisficing" wherein respondents are more likely to anchor on a first or last response option (Holbrook et al., 2007).[5] For studies that are part of lengthy surveys, or for those that involve cognitively demanding tasks, simple respondent fatigue can generate substantial differences between OM and MO designs, especially when outcome and mediator(s) are separated by one or more questions.

Whether due to priming and/or fatigue, OM and MO designs can affect outcome variable (Y_i) and/or the mediator(s) (M_i) in their distributions (Effect 1), or in the treatment's effect on either (Effect 2). Any of these can change the results of CMA by altering the relationship between treatment and mediator, impacting either $M_i(1)$, $M_i(0)$, or the relationship between them, thereby affecting ACME.[6] In addition, if OM/MO affects the relationship between treatment and outcome (ATE), then it can change the ADE. Finally, even if a design choice affects the ADE, but not the ATE, then the estimated proportion of the treatment effect that goes "through" the mediating variable (PM) will also be impacted (Effect 3).

It would take strong theoretical expectations to specify, *a priori*, whether an MO or OM result is less bias-prone. The possibility that certain items frame or cue later items is extremely general: the

[4] Of course, question ordering may not always affect responses, as has been shown in the case of more benign items, such as factual manipulation checks (Kane and Barabas, 2019). But when they involve sensitive topics or respondent perceptions, framing effects are more likely. In cases where there are multiple such mediators, prudence suggests that their ordering ought to be randomized.

[5] Sometimes, question order is itself the experimental treatment, to explore framing effects directly.

[6] As Imai et al. (2011) note, "If the treatment has no effect on the mediator, i.e., $M_i(1) = M_i(0)$, then the causal mediation effects are zero" (p. 769).

effects might be on any moment of the later distribution, and can alter associations across variables in any direction. Which items are most difficult to measure, and most prone to fatigue-based error can also be hard to anticipate.

2 Replication

To illustrate the significance of OM/MO choice, we replicated a well-known study by Tomz and Weeks (2013), which uses a survey experiment to study micro-foundations of democratic-peace theory. We chose the Tomz-Weeks ("T-W") study because it was among the *most* diligent in its discussion of potential survey-design issues. The authors used an OM design but, recognizing that this choice matters, also took the rare step of replicating the survey, under an MO design, which produced a similar ATE.

The study examines the role of public opinion in explaining rarity of militarized conflict between democracies. To assess the effect of an adversary's regime-type on citizen preferences, the authors gave subjects a hypothetical vignette in which their government was confronted by a foreign country developing nuclear weapons. Randomly, this country was a democracy or not. Of interest was subjects' approval of attacking the adversary. The study investigates the direct effect of democracy plus several theoretically motivated channels through which it might impact citizen preferences, perceptions of: *costs* associated with attack; *threats* from not attacking; likelihood of attack *success*; and the *(im)morality* of a preemptive strike.

T-W detected a negative overall treatment effect: respondents were significantly less supportive of strikes against foreign democracies, compared to non-democracies. Further, they found that this effect operates through some of the four mediating channels. Approximately 34% of the effect could be attributed to changing threat-perceptions, and 15% stemmed from beliefs about the morality of an attack. There was less evidence of mediation through costs (4%) or likelihood of success (6%), leaving the balance of the democracy effect (approximately 40%) unmediated. They concluded, "democracies view other democracies as less threatening, which in turn reduces

support for using force." They also suggested that morality's part in the democratic peace become "a major topic of future research" (p. 862).

Although T-W obtained very similar ATE estimates with OM and MO designs, we wondered about design's impact on other quantities of interest in the CMA. So we reconstructed their full survey instrument, altering question order to create OM and MO versions. We fielded these to 1,041 respondents recruited though Amazon's Mechanical Turk (AMT) from June 5–7, 2018. Respondents were randomly assigned to either a democracy or non-democracy scenario and, independently, to either OM or MO setup.

Our survey differed from the original in two ways. First, we held fixed the value of other treatments, which randomized features of the foreign country's alliances and military strength. Second, T-W analyzed within-subject data, as respondents read two scenarios about a foreign country, one a democracy and the other a non-democracy, several days apart. Our design followed the majority of survey experiments by relying on between-subject variation only. We do not expect these differences, alone or jointly, to magnify differences between OM and MO designs.

2.1 Outcome and Mediator Distributions

Our estimated treatment effect, difference in support for attacking a democracy versus a non-democracy, was about 9%, slightly smaller than estimates from distinct samples in the T-W article, albeit at lower levels of support. Tables in the Appendix show these comparisons and that the distributions of our *threats*, *costs*, and *success* mediators were also similar to those in T-W.

Our respondents were more likely to view attack as immoral, with 65% and 52% saying that attacking a democracy or a non-democracy, respectively, would be immoral (compared to 38% and 31% in T-W). Survey design had a significant effect on levels of this mediator (Effect 1), as the percentages declaring attack immoral were 61% for MO versus 55% for OM ($p = 0.04$). Further, design may have moderated the relationship between treatment and the *morality* mediator (ACME) (Effect 2), though it did not affect this relationship for other mediators. In the OM design, democracy increased the perceived immorality of an attack by 10%, from 50% to 60%. For MO, the

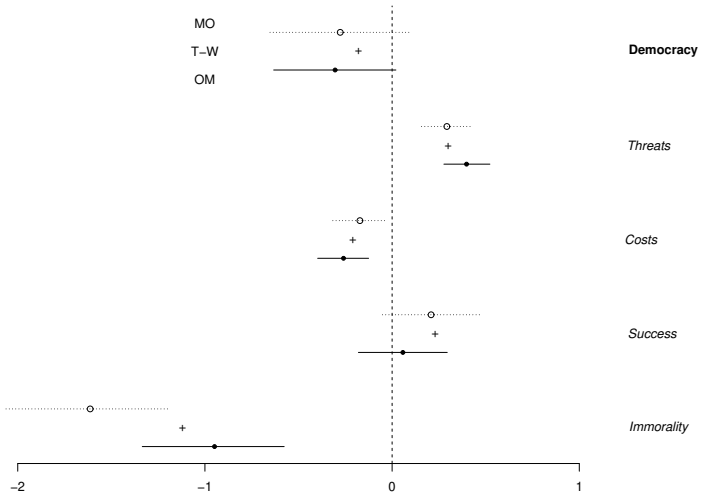

Figure 1: Probit Coefficients with 95% Confidence Intervals

Notes: Results of probit models of binary support of attack on democratic country. Control variables omitted. Full results are in the Appendix. Point estimates only (without intervals) for T-W.

levels were 54% and 69%, so the democracy effect was stronger, though this 10-*vs*.-15 difference did not itself reach statistical significance.

Following the original study, we estimated probit models of attack attitudes, dichotomized into support (strong or not) *vs*. opposition (strong or not) or neutrality. Figure 1 shows that our results resemble those of T-W. Our pooled results (see Appendix) match closely, though the smaller *N*s conspire against statistical significance of the democracy treatment in the distinct modules. Still, both designs yield substantively similar estimates: democracy lowers the probability of approval by 4.5% in the MO design and by 5.9% in the OM design.[7]

2.2 Mediation Analysis

We employed the popular R package `mediation` (Tingley et al., 2014) to estimate the percentage of the democracy treatment effect that travels through the relevant channel for each mediator (PM),

[7]These quantities were calculated holding other variables at their sample means.

Table 1: Mediation Analysis by Survey Design

	Tomz-Weeks		Replication			
	OM		OM		MO	
Threats	**-4.0**	(34%)	-1.4	(12%)	**-6.5**	(52%)
Costs	-0.4	(4%)	**-1.7**	(16%)	-0.3	(2%)
Success	**-0.7**	(6%)	-0.7	(7%)	-1.0	(8%)
Immorality	**-1.7**	(15%)	**-3.9**	(34%)	**-6.4**	(52%)

Notes: Average effect of each mediator on outcome. The estimated percent mediated (PM) quantities are in parentheses (these can exceed 100%). Bold ACME estimates are individually significant at the 0.05 level.

taken one at a time, always with the same set of covariates.[8] Table 1 shows notable differences between the OM and MO estimates (Effect 3). The proportions of the democracy effect attributed to three of the four mediators vary a good deal by design, even though the overall treatment effects are similar. Moreover, neither module matches the T-W study in rankings of PM.

Even with no sharp threshold for what constitutes an "important" mediator, researchers using different designs would likely draw different conclusions from otherwise identical studies. Our MO module points to two strong channels (the same two identified by T-W), while the OM version suggests one medium-strength channel and two weaker channels. T-W found that immorality accounted for 15% of the treatment effect, which led them to conclude that it was an important, under-studied channel. In these results, the costs channel looks comparably important under OM, but not under MO.

Figure 2 shows one approach to assessing the likelihood that differences in PM estimates across designs occurred by chance. We draw two ternary plots, displaying sets of PM estimates for OM and MO designs. Each point is based on calculations from 1,000 simulations, with the simulation algorithm repeated 250 times. For the left pane, we randomly drew 500 of these PM quantities for the immorality and threats mediators and plotted the pair, leaving the effects operating through morality and success plus any unmediated direct effect summed in the residual (distance from the

[8]The Appendix includes estimates from an alternative method allowing causal relationships between mediators, plus some discussion of sample-size importance for this stage.

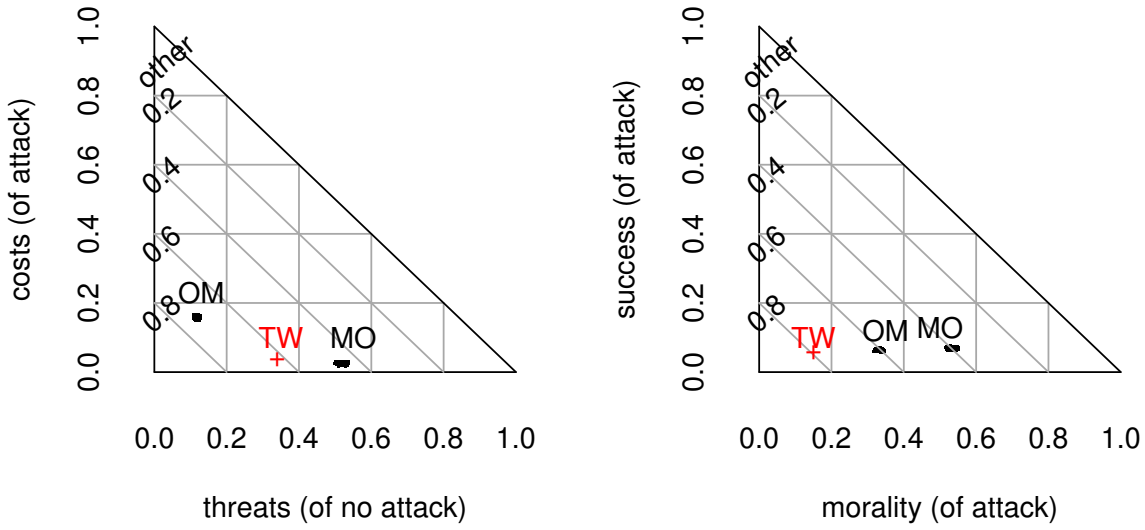

Figure 2: Estimated Proportions Mediated

Notes: Estimated proportions of democracy effect mediated through each mediator, across simulations.

hypotenuse). The right panel shows draws of the PM estimates for the costs and success mediators.

In both cases, PM estimates are tightly clustered by survey design, and distant from those in T-W. These plots increase our confidence that design-choice mattered, and that the different "punchlines" in the OM and MO modules were not an anomalous draw or an accident of the simulation approach.

One might prefer a more formal test for which mediators are sufficiently important to warrant attention. Not surprisingly, the confidence intervals of the PM estimates, which are ratios of estimated quantities, tend to be quite wide. The appendix shows simulations to confirm that these intervals shrink as sample sizes increase, but the number of respondents needed to consistently reject equivalence in proportion mediated can be larger than – at least twice the size of – the samples collected in typical applied research. In general, researchers should also be cautious when the ACME and direct effects have different signs, sometimes called "inconsistent mediation," which significantly complicates interpretation of the PM.

Table 2: Suggestions for OM/MO Design Choice

Effect of OM/MO	**Potential Source(s)**	**Suggested Diagnostic(s)**
Changes value or shifts distribution of Y_i or M_i	Survey fatigue; Anchoring	Compare distributions of outcomes and mediators across designs.
Moderates ATE, ACME, or other direct effect	Framing; Desirability bias; Fatigue; Anchoring	Split-sample regressions; Models effect of design on T-M, T-O, or M-O relationships.
Affects PMs, individually or collectively	Any of the sources above	Split-sample calculations of PMs; ternary plots with multiple simulations.

3 Discussion

In our discussion and in our replication, we have identified three ways in which the order of instruments in CMA impacts the conclusions drawn (Table 2). Variables that mediate between treatments and outcomes can be affected by response biases, including priming, satisficing, and anchoring. Our advice to researchers conducting CMA is as follows. At a minimum, design type should always be specified, ideally with discussion of possible order effects.[9] Whenever possible, researchers should directly assess whether their results are conditioned on design choice, by randomly assigning respondents to OM or MO versions. Since Effect 3 can affect PMs even in the absence of the other two effects, and given that CMA is used to draw substantive inferences from these estimates, the effect of design on PMs is likely to be important.

There is some cost to this approach. Agreement between OM and MO modules validates pooling data from both designs. Disagreement, however, means reduced statistical power. But it also gives the researcher valuable information about the complexity of the associations between variables, that can help inform interpretation of results and direct further inquiry.

[9]For example, an MO design may be justified because the mediator questions are unlikely to influence responses to the outcome item(s), or that an OM design is preferable because of a large number of mediator items, some of which appear to be plausible cues.

Acknowledgements: We appreciate the helpful research assistance of Ekrem Baser and Audrey Christopher. Jessica Weeks and Michael Tomz graciously provided advice on replications. We also appreciate feedback from Andy Baker, Ryan Brutger, Joseph Huddleston, Joshua Kertzer, Dustin Tingley, the editors and anonymous reviewers at the *Journal of Politics*.

References

Bullock, John G., Donald P. Green and Shang E. Ha. 2010. "Yes, but What's the Mechanism? (Don't Expect an Easy Answer)." *Journal of Personality and Social Psychology* 98(4):550–558.

Holbrook, Allyson L., Jon A. Krosnick, David Moore and Roger Tourangeau. 2007. "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes." *Public Opinion Quarterly* 71(3):325–348.

Huddleston, R. Joseph. 2019. "Think Ahead: Cost Discounting and External Validity in Foreign Policy Survey Experiments." *Journal of Experimental Political Science* 6(2):108–119.

Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review* 105(4):765–789.

Kane, John V. and Jason Barabas. 2019. "No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments." *American Journal of Political Science* 63(1):234–249.

Sudman, Seymour and Norman M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago, IL: Aldine/NORC.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele and Kosuke Imai. 2014. "Mediation: R package for Causal Mediation Analysis." *Journal of Statistical Software* 59(5).

Tomz, Michael R. and Jessica L.P. Weeks. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107(4):849–865.

Tourangeau, Roger, Lance J. Rips and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Transue, John E., Daniel J. Lee and John H. Aldrich. 2009. "Treatment Spillover Effects Across Survey Experiments." *Political Analysis* 17(2):143–161.

Biographical Information:

Stephen Chaudoin in an assistant professor at Harvard University, Cambridge, MA 02138.

Brian J. Gaines is a professor at the University of Illinois at Urbana-Champaign, Urbana, IL 61801.

Avital Livny is an assistant professor at the University of Illinois at Urbana-Champaign, Urbana, IL 61801.