

This article was downloaded by: [99.45.168.136]

On: 05 June 2014, At: 09:43

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## International Interactions: Empirical and Theoretical Research in International Relations

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/gini20>

### Survey Research in International Political Economy: Motivations, Designs, Methods

Dustin Tingley<sup>a</sup>

<sup>a</sup> Harvard University

Accepted author version posted online: 14 Mar 2014. Published online: 02 Jun 2014.

To cite this article: Dustin Tingley (2014) Survey Research in International Political Economy: Motivations, Designs, Methods, *International Interactions: Empirical and Theoretical Research in International Relations*, 40:3, 443-451, DOI: [10.1080/03050629.2014.900614](https://doi.org/10.1080/03050629.2014.900614)

To link to this article: <http://dx.doi.org/10.1080/03050629.2014.900614>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## **Survey Research in International Political Economy: Motivations, Designs, Methods**

DUSTIN TINGLEY

*Harvard University*

*The use of surveys and survey experiments by international political economy scholars is increasing, adding to the ability to study a broad array of topics. In doing so, many scholars in international political economy draw on—and are contributing to—insights and arguments from American politics and comparative politics (Milner 1998), substantive fields with a history of using surveys and survey experiments. In this article, I review motivations for using surveys and survey experiments, the research designs, and analysis strategies in light of this issue’s contributions. I contrast these motivations and their accompanying designs and discuss the pros and cons of ways to approach the data generated by these research designs. The goal of this commentary is to situate surveys and survey experiments—especially those within the special issue—within a larger discussion about research motivations, design, and analysis techniques.*

**KEYWORDS** *international political economy, methods, research design, survey research*

The span of these articles and their relevance to international political economy (IPE) is remarkable. So how should we think about the *motivations* for a research design, the benefits and limitations of specific *research design* decisions, and their *analysis techniques*?

### MOTIVATIONS

These articles have very different motivations for their research *designs*. They include concern about confounding variables, demonstrating a manipulation,

---

Address correspondence to Dustin Tingley, Harvard University, Department of Government, 1737 Cambridge Street, Cambridge, MA 02138, USA. E-mail: [dtingley@gov.harvard.edu](mailto:dtingley@gov.harvard.edu)

and measurement. In this section I focus on these motivations to draw out, heuristically, their differences. In the next section I discuss the designs themselves in more detail, and the concerns that arise, given the motivations for selecting a particular design.

Julia Gray and Raymond P. Hicks (2014) study international investment behavior. Extending earlier observational work, they ask whether perceptions of investment prospects change depending on a country's formal associations with other countries. Nathan M. Jensen and Mi Jeong Shin (2014) examine the extent to which framing US agricultural trade policy as less generous, compared to other countries, shapes voter support for farm payments. Respondents show greater support for farm subsidies when the United States is portrayed as less generous. Both studies use experimental designs, given concerns about confounding variables that are unobservable when we look at observational data of the same substantive phenomenon.

Somewhat similarly, K. Amber Curtis (2014) focuses on establishing a causal relationship between two variables by manipulating one variable and observing the effect of this manipulation. One account of why this is necessary for learning about some behavior (here, pocketbook voting) is the concern about confounding variables. But as Curtis does not identify specific concerns, the motivation of the research design in this respect could be taken as a deeper commitment to the mantra (for some) of "no causation without manipulation" (Holland 1986). Part of that mantra deals with concerns about actual confounding, but at a deeper level this mantra posits that we fundamentally cannot talk about the causal effect of a variable if we cannot reasonably define counterfactuals. For some, the ability to define these counterfactuals turns on the ability to manipulate the causal variable. In practice, this motivation can lead to a view that demonstrating the ability to manipulate something is a necessary condition for it to be considered causally relevant. That is, while Curtis does not postulate specific inferential barriers to the topic of pocketbook voting, one *possible* evidentiary standard is that if we can manipulate concerns about pocketbook voting and show that it has an effect, then we have greater confidence that indeed it has a causal effect in the real world.<sup>1</sup>

An additional motivation for the Gray and Hicks research design is the desire to observe decision making at the appropriate level of aggregation. In their case, existing theories turn on microlevel (that is, investor level) mechanisms. However, the observational data used to test these explanations are typically at an aggregate level. Hence the use of surveys is to provide microlevel data where the theoretical mechanisms operate at the

---

<sup>1</sup>Interestingly, the results show that more sophisticated individuals are more susceptible to manipulation. This contradicts classic results in American politics by Zaller (1992) and others. To this end, more theory—theory that identifies the scope conditions of when more sophisticated individuals will resist manipulation—is important.

individual level. This is the goal of measuring phenomena at the same level as the theoretical mechanisms. Jensen and Shin in this sense share such a measurement-oriented motivation.

Other studies focus much more directly on being able to measure important quantities of interest. Judith L. Goldstein and Margaret E. Peters (2014) study the role of cultural versus material attitudes toward immigration by leveraging changes in the economic conditions of respondents over time. The survey component of their research design is designed to cleanly measure attitudes toward immigrants of different skill levels.

Benjamin E. Bagozzi, Thomas Brawner, Bumba Mukherjee, and Vineeta Yadav (2014) are motivated by something quite fundamental to the survey enterprise. In particular, a respondent's prior knowledge or experience with an issue area might have a large influence on their revealed preferences. Hence their motivation is to think more systematically about the process of survey response rather than a strict focus on the substantive content of the survey topic itself. These motivations are wise and have a long tradition (Zaller 1992). This is very important, as it highlights the fact that there is a data-generating process that we are tapping into. Real individuals are sitting down and giving researchers responses to questions. Assuming that each individual simply has prerecorded responses to our survey questions, and that the act of a survey is just to record them, appears unwise.

## RESEARCH DESIGNS

In light of the motivations for adopting survey and survey experimental research designs, how might we evaluate research designs? In this section I focus on how realistic manipulations are, how manipulations might be perceived by participants, how randomization can be used for measurement purposes, and how we should define and interpret experimental conditions.

The research designs used by Jensen and Shin and Gray and Hicks use a common survey experiment design in the framing and priming literature. Individuals are confronted with a political or economic situation and asked to indicate their support for a policy choice or some other decision. The features of the situation are selected to represent variation that might be observed in the real world. For example, Jensen and Shin portray US farm payments as generous versus ungenerous, and Gray and Hicks vary whether a country has signed a trade agreement with another country with a positive or negative reputation. Given their concerns about confoundingness, they achieve their goal. Curtis' design is similar, with a prime that manipulates the salience of pocketbook considerations in the subject's decision calculus. While Curtis does not need to be concerned about confounds for the prime, the same cannot be said for the operative mechanism of sociotropism.

In each of these designs an important question beyond issues of causal identification is that of the extent to which these manipulations capture things that real voters, investors, etc., or experience, or whether they have no clear connection. As in many experiments, further elaboration of the plausibility of this is valuable. Even if the researchers feel their manipulation has a clear connection to the substantive topic of interest, it is important for experimentalists and readers to ask: How are the experimental conditions interpreted by participants?<sup>2</sup>

The Goldstein and Peters article uses randomization more as a measurement tool rather than as a means to establishment a causal relationship in the face of confounding. They study preferences for immigration by whether immigrants are high or low skill. To do this half of the sample gets a high-skill prompt and half a low-skill one. They create a common dependent variable and then an indicator (which by design is randomly assigned) for whether the respondent was asked about low- or high-skill immigration.<sup>3</sup> To me, the design in Goldstein and Peters is less in the tradition of manipulating a causal variable at the respondent level, but instead a way to measure support for two different reference groups and then backing out the premium individuals assign to one group or the other.

This measurement strategy is of course important because it might be that different individuals have different “types” of immigrants in mind. And asking about both types of immigrant skill levels could prime the respondent. But it is still possible to ask all individuals both conditions and just randomize the order. This, of course, would use up valuable survey space, but it shows how randomization is helpful here for a measurement purpose.

The causal variables in their theory/design remain observational ones in my view, individual skill and changing economic threat perceptions. By utilizing individuals who got the same skill level for the immigration question in the panel, they can look at how within-respondent changes in economic circumstances and perceptions impact the *premium* given to high-skill immigrants. The causal role played by economic circumstances and perceptions are identified through *standard panel analysis assumptions* rather than through random assignment.

I think it is important to remind readers in light of the excellent Goldstein and Peters article that causal inference need not be the only motivation for randomizing something in an experiment. Indeed, one use of

---

<sup>2</sup>For example, consider the interpretation of the results in the survey experiments reported in Findley, Nielson, and Sharman (2013) that examine willingness of firms that handle the establishment of shell companies to process incorporation without the necessary legal requirements. One finding is that firms in tax-haven countries were especially sensitive to international law. Yet, this could simply be because the response to their email solicitation is to “play it safe,” as such solicitation could be coming from investigative authorities that focus on firms in tax-haven countries.

<sup>3</sup>Contrast this manipulation to those in the contributions to this volume by Jensen and Shin or Gray and Hicks.

experiments, perhaps a bit lost in the current trends/debates, is that survey experiments also enable the interrogation of measurement. In some settings, exposing an individual to two conditions can lead to order effects, so researchers randomly assign a single condition/sequence. When order effects are not a concern, then designs with repeated exposure to different conditions will be more efficient and can be rigorously analyzed (see the following).

Finally, it is important to note a difference between the designs in this volume: the use of a contrast treatments and whether or not to have a pure control condition. For example, the additive design in Curtis compares a condition with no pocketbook prime to one with a pocketbook prime. There is no alternative, or contrasting, prime. This differs, for example, with Jensen and Shin, who contrast generous versus ungenerous framings of US farm policy. In that experiment there was no baseline condition of support for US farm payments sans information about the relative US generosity. Their food stamp experiment, in contrast, uses an additive design. Some research includes a pure control condition, alongside a positive and negative condition. A range of factors can be relevant here, including sample size considerations or reasonableness of the pure control condition (perhaps in the real world individuals only see one of the contrasting conditions). It is important, though, for scholars of IPE to know that there is a debate on this. For example, some have criticized studies lacking a pure control because their interpretation is ambiguous (Gaines, Kuklinski, and Quirk 2007).<sup>4 5</sup>

## ANALYSIS METHODS

Having collected data through an experiment, or some other research design, scholars must face their data. Depending on the design, different methods for

---

<sup>4</sup>Another design consideration in modern surveys deals with mode. We are reaching to new platforms to conduct surveys, and we are increasingly using collaborative Internet-based surveys that have desirable properties compared to traditional modes (Ansolabehere and Rivers 2013). For example, Jensen and Shin used the *Cooperative Congressional Election Study* to field their main US survey. And there are also convenience sample providers for Internet-based surveys, like Mechanical Turk, utilized by both Gray and Hicks and Jensen and Shin. While use of “mTurk” has grown in popularity (Berinsky, Huber, and Lenz 2012; Tingley and Tomz 2014), researchers should continue to think about how properties of mTurk samples (such as being more liberal and younger)—just like any other convenience sample—could interact with treatment conditions. Importantly, researchers and reviewers should avoid assuming that convenience samples *will* pose these problems without specific theoretical justification (Druckman and Kam 2011). Of course, surveys themselves may be used less and less as other data sources like social media become available (For example, Jamal, Keohane, Romney, and Tingley 2013).

<sup>5</sup>Other design issues are important to keep in mind. For example, Goldstein and Peters note the importance of social desirability bias and how their design is unlikely to suffer from it due to their within-subject design. Researchers who are concerned with social desirability issues can also turn to other design-based solutions, such as list and endorsement experiments that now can be analyzed with multivariate methods (Blair and Imai 2012; Bullock, Imai, and Shapiro 2011; Glynn 2013).

analysis will be more or less appropriate. With many types of experimental data, the analysis techniques are relatively straightforward because the key variable of interest has been randomly assigned. In these cases simple difference in means, perhaps with regression adjustment, may be appropriate. In other settings, things are not so simple.

One example is the analysis of causal mechanisms, where even if the treatment variable is randomized, the role of a causal mechanism is not identified. Furthermore, some existing techniques for analyzing causal mechanisms are ill founded. For example, Curtis offers sociotropic evaluations as a mechanism linking pocketbook evaluations to political approval. To answer this question Curtis begins with a three-step procedure that checks the statistical significance of (1) the main explanatory variable on the outcome, (2) the main explanatory variable on the hypothesized mediating variable, and then (3) whether the impact of the main explanatory variable diminishes once controlling for the mediating variable (Baron and Kenny 1986). This procedure is potentially problematic. We learn very little from it because it distracts researchers from thinking about their research design, what assumptions are necessary to establish causal relationships, and what substantive quantities should be estimated.<sup>6</sup>

Appropriately, Curtis raises a discussion of the assumptions behind mediation analysis, as well as ways to directly calculate mediating relationships rather than simply conducting significance tests. These approaches have received substantial discussion elsewhere (Imai, Keele, Tingley, and Yamamoto 2011). Curtis also, appropriately, calculates how sensitive the mediation results are to the assumption required by the research design. Future survey experimental work that is interested in causal mechanisms might also adopt alternative designs that require different, perhaps more defensible, assumptions (Imai, Tingley, and Yamamoto 2013). A recent example of this is the application of the Imai et al. 2013 cross-over design to the study trade policy attitudes (Strezhnev 2013).

Another example of the need to use more sophisticated analysis strategies is discussed in the Bagozzi et al. article, which refocuses our attention on our model of survey *response*. What we observe in survey data might be a mixture of different modes of survey response, and this can be modeled using finite mixture methods. These methods are quite powerful and can be used in a variety of ways beyond uncovering modes of survey responses, including testing competing nonnested theories (Imai and Tingley 2012). The importance of using information about different information levels is likely important in many areas of IPE where survey respondents might be relatively, or massively, unfamiliar with particular topic areas.

---

<sup>6</sup>While Curtis claims that the procedure can lead to overstated relationships, this is not necessarily the case (for example, in the case of inconsistent mediation this procedure leads to false negatives).



Other recent methodological advances will also be appropriate for IPE scholarship studying public opinion. I conclude by providing three examples. First, testing for heterogeneities in treatment effects is another common way to add more nuance to a study. However, just like with any design that conducts multiple hypothesis tests, researchers need to take extra care. This might go beyond simple adjustments to  $p$  values and instead to more principled investigation of effect heterogeneities (Green and Kern 2012; Imai and Ratkovic 2013). Ideally theoretical considerations should identify where we might expect heterogeneities, but data-mining approaches like these that try to prevent false positives are potentially powerful. Second, new methods are available for analyzing open-ended responses in surveys. One such method, the structural topic model, permits the incorporation of covariates, such as treatment assignments, directly into the model, yielding interesting quantities of interest about the topics individuals express in an open-ended format (Roberts, Stewart, Tingley, Lucas, Leder-Luis, Albertson, Gadarian, and Rand forthcoming). Third, many policies are extremely multifaceted, containing many different dimensions that respondents consider. While standard experimental randomizations manipulating one dimension at a time are possible in principle, this quickly becomes unrealistic without massive sample sizes. An alternative approach is conjoint analysis (Hainmueller, Hopkins, Yamamoto 2014), which has been largely used in marketing research. Conjoint analysis has individuals choose between multiple sets of contrasting options. Each option is varied along a range of dimensions, and by repeatedly asking a respondent multiple questions, across multiple respondents, the relative effect of many different attributes can be ascertained. In my view, it is simply a means to collect more information, more efficiently, than would be available by only asking each respondent one contrast.<sup>7</sup>

## CONCLUSION

Surveys and survey experiments help researchers address a number of barriers to making clear inferences about important substantive questions. They can assist in controlling for confounding variables, demonstrate manipulability, and improve measurement. Their success depends on how well they capture real-world behavior and beliefs and whether respondents interpret interventions in the way researchers claim. I suspect we will continue to see research in IPE using these methods, and the contributions to this volume help point in some fascinating directions.

---

<sup>7</sup>Just as conjoint analysis involves reinterviewing the same subject multiple times, I see great promise in the class of cross-over designs, which involve treatment assignment in multiple stages. For example, it would be interesting to think about what we could learn from the members of the Goldstein and Peters panel who received different skill treatments over time.



## REFERENCES

- Ansolabehere, Stephen, and Douglas Rivers. (2013) Cooperative Survey Research. *Annual Review of Political Science* 16(1):307–329.
- Bagozzi, Benjamin E., Thomas Brawner, Bumba Mukherjee, and Vineeta Yadav. (2014) Regional International Organizations and Individual Immigration Attitudes: Results from Finite Mixture Models. *International Interactions* 40(3): 350–375.
- Baron, Reuben M., and David A. Kenny. (1986) The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology* 51(6): 1173–1182.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. (2012) Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk. *Political Analysis* 20(3):351–368.
- Blair, Graeme, and Kosuke Imai. (2012) Statistical Analysis of List Experiments. *Political Analysis* 20(1):47–77.
- Bullock, Will, Kosuke Imai, and Jacob N. Shapiro. (2011) Statistical Analysis of Endorsement Experiments: Measuring Support for Militant Groups in Pakistan. *Political Analysis* 19(4):363–384.
- Curtis, K. Amber. (2014) In Times of Crisis: The Causes, Conditions, and Consequences of Pocketbook Voting. *International Interactions* 40(3):402–430.
- Druckman, Jamie, and Cindy D. Kam. (2011) Students as Experimental Participants: A Defense of the “Narrow Data Base.” In *Cambridge Handbook of Experimental Political Science*, edited by James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. New York: Cambridge University Press.
- Findley, Michael, Daniel Nielson, and Jason Sharman. (2013) Using Field Experiments in International Relations: A Randomized Study on Anonymous Incorporation. *International Organization* 67(4):657–693.
- Gaines, Brian J., James H. Kuklinski, and Paul J. Quirk. (2007) The Logic of the Survey Experiment Reexamined. *Political Analysis* 15(1):1–20.
- Glynn, Adam N. (2013) What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment. *Public Opinion Quarterly* 77(Suppl. 1):159–172.
- Goldstein, Judith L., and Margaret E. Peters. (2014) Nativism or Economic Threat: Attitudes Toward Immigrants During the Great Recession. *International Interactions* 40(3):376–401.
- Gray, Julia, and Raymond P. Hicks. (2014) Reputations, Perceptions, and International Economic Agreements. *International Interactions* 40(3):325–349.
- Green, Donald P., and Holger L. Kern. (2012) Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees. *Public Opinion Quarterly* 76(3):491–511.
- Hainmueller, Jens, Daniel Hopkins, and Teppei Yamamoto. (2014) Causal Inference in Conjoint Analysis: Understanding Multi-Dimensional Choices Via Stated Preference Experiments. *Political Analysis* 22(1):1–30.
- Holland, Paul W. (1986) Statistics and Causal Inference. *Journal of the American Statistical Association* 81(396):945–960.

- 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.
- Imai, Kosuke, and Marc Ratkovic. (2013) Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation. *Annals of Applied Statistics* 7(1):443–470.
- Imai, Kosuke, and Dustin Tingley. (2012) A Statistical Method for Empirical Testing of Competing Theories. *American Journal of Political Science* 56(1):218–236.
- Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. (2013) Experimental Designs for Identifying Causal Mechanisms. *Journal of the Royal Statistical Society, Series A* 176(1):5–51.
- Jamal, Amaney, Robert Keohane, David Romney, and Dustin Tingley. (2013) America in the Eyes of Arabic Tweeters. Working Paper.
- Jensen, Nathan M., and Mi Jeong Shin. (2014) Globalization and Domestic Trade Policy Preferences: Foreign Frames and Mass Support for Agriculture Subsidies. *International Interactions* 40(3):305–324.
- Milner, Helen V. (1998) Rationalizing Politics: The Emerging Synthesis of International, American, and Comparative Politics. *International Organization* 52(4):759–786.
- Roberts, Margaret, Brandon Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Bethany Albertson, Shana Gadarian, and David Rand. (Forthcoming) Structural Topic Models for Open Ended Survey Responses. *American Journal of Political Science*.
- Strezhnev, Anton. (2013) The Effect of Trading Partner Democracy on Public Attitudes Toward Preferential Trade Agreements. Working Paper.
- Tingley, Dustin, and Michael Tomz. (2014) Conditional Cooperation and Climate Change. *Comparative Political Studies* 47(3):344–368.
- Zaller, John. (1992) *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.