

Why the Study of International Law Needs Experiments

ADAM CHILTON & DUSTIN TINGLEY*

Scholarship on international law has undergone an empirical revolution. Throughout the revolution, however, shortcomings of the observational data that studies have used have posed serious barriers to reliable causal inference. During the same period, political scientists and legal scholars studying domestic law have increasingly employed experimental methods because they make it easier to make credible causal claims. Despite the simultaneous emergence of those trends, there have been relatively few attempts to use experimental methods to study international law. This should change. In this paper we present the first argument that the study of international law could uniquely benefit from the use of experimental research methods. To make this argument, we present data we have collected that illustrates why observational studies will often be unable to provide answers to many of the most important questions of legal scholars. After doing so, we provide guidance on how laboratory, survey, and field experiments can be used by legal scholars to research international law.

* Adam Chilton, Harry A. Bigelow Teaching Fellow & Lecturer in Law, University of Chicago School of Law. Ph.D. in Political Science, Harvard University; J.D., Harvard Law School. Email: adamchilton@uchicago.edu. Dustin Tingley, Paul Sack Associate Professor of Political Economy in the Government Department at Harvard University. Ph.D. in Politics, Princeton University. Email: dtingley@gov.harvard.edu. Website: <http://scholar.harvard.edu/dtingley>. Our special thanks to research assistance by Alice Miao, and Emilie Hafner-Burton, Michael Findley, Tom Ginsburg, Jill Goldenziel, and Michael Tomz for helpful conversations and advice.

INTRODUCTION

Over the last decade, there has been an empirical revolution in the study of international law.¹ During that time, the focus of international legal scholarship has moved away from debating whether international law matters and toward trying to explain when—and why—states commit to and comply with international legal agreements.² This “empirical turn” has produced important new scholarship on a range of substantive areas of international law;³ including international human rights law,⁴ the laws of war,⁵ and international economic law.⁶

Throughout this empirical turn, however, scholars have struggled to provide convincing evidence that international law has a causal impact on state behavior.⁷ Although prominent commentators

1. See generally Gregory Shaffer & Tom Ginsburg, *The Empirical Turn In International Legal Scholarship*, 106 AM. J. INT’L L. 1 (2012) (documenting the growth of empirical research of international law); see also Emilie M. Hafner-Burton, David G. Victor, & Yonatan Lupu, *Political Science Research on International Law: The State of the Field*, 106 AM. J. INT’L L. 47 (2012) (providing a review of political science research that is relevant to the study of international law); Beth Simmons, *Treaty Compliance and Violation*, 13 ANN. REV. POL. SCI. 273 (2010) (reviewing empirical research on treaty commitment and compliance). Of course, it is important to note that during this time there has also been a rise in the use of empirical legal research more generally. See generally Shari Seidman Diamond & Pam Mueller, *Empirical Legal Scholarship in Law Reviews*, 6 ANN. REV. L. SOC. SCI. 581, 589 (2010).

2. See Shaffer & Ginsburg, *supra* note 1, at 1 (“What matters now [in research on international law] is the study of the conditions under which international law is formed and has effects.”).

3. See *id.* at 19–42 (documenting empirical legal scholarship on five substantive areas of international law).

4. For an overview of the current state of the quantitative literature on international human rights, see Emilie M. Hafner-Burton, *International Regimes for Human Rights*, 15 ANN. REV. POL. SCI. 265 (2012). For examples of prominent quantitative scholarship on international human rights, see, e.g., BETH SIMMONS, *MOBILIZING FOR HUMAN RIGHTS: INTERNATIONAL LAW IN DOMESTIC POLITICS* (2009); Oona A. Hathaway, *Do Human Rights Treaties Make a Difference?*, 111 YALE L.J. 1935 (2002).

5. See, e.g., James D. Morrow, *When Do States Follow the Laws of War?*, 101 AM. POL. SCI. REV. 559 (2007); Benjamin A. Valentino, Paul K. Huth, & Sarah Croco, *Covenants without the Sword: International Law and the Protection of Civilians in Times of War*, 58 WORLD POL. 339 (2006).

6. See, e.g., CHRISTINA L. DAVIS, *WHY ADJUDICATE? ENFORCING TRADE RULES IN THE WTO* (2012); Susan D. Franck, *Development and Outcomes of Investment Treaty Arbitration*, 50 Harv. Int’l L. J. 435 (2009).

7. For a general discussion of this issue, see Shima Baradaran, Michael Findley,

initially simply proclaimed that rates of compliance with international law were high as evidence of international law's importance,⁸ the debate has shifted to whether international law actually has a causal impact on changing state behavior.⁹ This shift has been made in part because of the recognition that—perhaps best expressed in Downs, Rocke, and Barsoom's seminal article on the topic¹⁰—states may simply select only into international agreements that codify policies that would have been adopted even in the absence of an international agreement. Although researchers quickly began to respond to that skeptical claim by producing scholarship that tried to demonstrate that international law does in fact change state behavior,¹¹ the empirical methods used by a number of early prominent studies were criticized for their inability to satisfactorily demonstrate that international law has a causal influence on state behavior.¹² Although researchers

Daniel Nielson & J. C. Sharman, *Does International Law Matter?*, 97 MINN. L. REV. 743, 764–67 (2013). Of course, the struggle to improve causal inference is not unique to international law. Efforts to analyze law more broadly have faced the same problems. For a discussion of the history of causal analysis in legal studies, and advice on how to improve research design to aid causal inference, see Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 ANN. REV. L. SOC. SCI. 17 (2011); see also Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHICAGO L. REV. 1 (2002).

8. The most famous statement of this view was Louis Henkin's claim that "almost all nations observe almost all principles of international law and almost all of their obligations almost all of the time." LOUIS HENKIN, *HOW NATIONS BEHAVE* 47 (1979). In addition to Henkin's claim, another prominent expression of this view can be found in Chayes & Chayes theory that has become known as the "managerial school" of international law. See Abram Chayes & Antonia Handler Chayes, *On Compliance*, 47 INT'L ORG. 175 (1993) (discussing explanations for why nations comply with their international legal commitments).

9. See Eric A. Posner, *Some Skeptical Comments on Beth Simmons's "Mobilizing for Human Rights"*, 44 NYU J. INT'L L. & POL. 819 (2012).

10. George W. Downs, David M. Rocke & Peter N. Barsoom, *Is the Good News About Compliance Good News About Cooperation?*, 50 INT'L ORG. 379 (1996) (arguing that evidence that states comply with international law is not evidence that states make meaningful changes in policy as a consequence of international treaties and agreements).

11. See, e.g., Beth A. Simmons, *International Law and State Behavior: Commitment and Compliance in International Monetary Affairs*, 94 AM. POL. SCI. REV. 819 (2000); Hathaway, *supra* note 4.

12. For a critical response to Simmons, *supra* note 11, see Jana Von Stein, *Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance*, 99 AM. POL. SCI. REV. 611 (2005) (arguing that Simmons did not adequately deal with selection effects that influence which states agree to bind themselves with international agreements); *but see* Beth Simmons & Daniel Hopkins, *The Constraining Power of International Treaties: Theory and Methods*, 99 AM. POL. SCI. REV. 623 (2005) (responding to Von Stein's criticisms); Shaffer & Ginsburg, *supra* note 1, at 15 (noting that despite Von Stein's arguments, "we believe that

have begun to use more complicated empirical methods—like instrumental variable regression¹³ and matching¹⁴—to demonstrate causality, these approaches have not entirely silenced critics that are skeptical of states' willingness to change their preferences and actions as a consequence of international law.¹⁵ As a result, improving the empirical methods used to conduct causal analysis is a project that is of great importance for scholars of international law.

While international law scholars have debated how to demonstrate causation, political scientists have embraced experimental research methods,¹⁶ as have, increasingly, legal scholars studying other areas of law.¹⁷ In the last two decades, experimental methods have become a common research strategy used to study a range of topics: from what strategies are most likely to turn people out to vote during elections,¹⁸ to whether legislatures are biased against their constitu-

Simmons' contribution withstands the critique in this particular case . . ."). For a critical response to Hathaway, *supra* note 4, see Ryan Goodman & Derek Jinks, *Measuring the Effect of Human Rights Treaties*, 15 EU. J. INT'L L. 171 (2003) (arguing against a number of empirical decisions and claims made by Hathaway).

13. See, e.g., SIMMONS, *supra* note 4; Morrow, *supra* note 5.

14. See, e.g., Yonatan Lupu, *The Informative Power of Treaty Commitments: Using the Spatial Model to Address Selection Effects*, AM. J. POL. SCI. (forthcoming 2013), available at <http://dss.ucsd.edu/~ylupu/Informative%20Power.pdf>; Rich Nielsen & Beth Simmons, *Rewards for Ratification: Payoffs for Participating in the International Human Rights Regime?* (2011) (working paper) (on file with author); Daniel W. Hill, Jr., *Estimating the Effects of Human Rights Treaties on State Behavior*, 72 J. POL. 1161 (2010); Joseph M. Grieco, Christopher F. Gelpi, & T. Camber Warren, *When Preferences and Commitments Collide: The Effect of Relative Partisan Shifts on International Treaty Compliance*, 63 INT'L ORG. 341 (2009); Simmons & Hopkins, *supra* note 12.

15. See Posner, *supra* note 9 (expressing skepticism at the empirical approach used by Simmons to argue that human rights treaties can change state policies through their potential to alter domestic political calculations).

16. See generally James N. Druckman, Donald P. Green, James H. Kuklinski & Arthur Lupia, *The Growth and Development of Experimental Research in Political Science*, 100 AM. POL. SCI. REV. 627 (2006). For a longer discussion of the growing use of experimental methods in political science, see *infra* PART I.A.

17. See Seidman Diamond, & Mueller, *supra* note 1, at 589 (providing evidence of research using experimental methods published in law reviews). For a longer discussion of the growing use of experimental methods in legal scholarship, see *infra* PART I.B.

18. See generally DONALD P. GREEN & ALAN S. GERBER, *GET OUT THE VOTE! HOW TO INCREASE VOTER TURNOUT* (2008) (explaining experimental research on how political campaigns can increase voter turnout). For a recent example of this line of research, see Daron R. Shaw, Donald P. Green, James G. Gimpel & Alan S. Gerber, *Do Robotic Calls from Credible Sources Influence Voter Turnout or Vote Choice? Evidence from a Randomized Field Experiment*, 11 J. POL. MARKETING 231 (2012).

ents,¹⁹ to the “democratic peace” theory,²⁰ to questions in constitutional law.²¹ This growth in experiments has been fueled by the increasing ease of experimental research,²² and the growing recognition that experimental methods greatly facilitate credible causal claims.²³ Simply put, by using experimental methods, researchers are able to control the data-generation process. That is, they can estimate the influence of a specific intervention on an outcome by randomly assigning treatments.

Yet, despite the fact that empirical research on international law has struggled to demonstrate causal claims, there have been few efforts to use experimental methods to study international law. In fact, we are aware of only a handful of international law experiments²⁴—one field experiment,²⁵ two laboratory experiments,²⁶ and

19. See Daniel M. Butler & David E. Broockman, *Do Politicians Racially Discriminate against Constituents? A Field Experiment on State Legislators*, 55 AM. J. POL. SCI. 463 (2011) (providing evidence that legislators are less likely to help putatively black constituents with requests for help voting).

20. See Michael Tomz & Jessica L. Weeks, *The Democratic Peace: An Experimental Approach*, AM. POL. SCI. REV. 1 (forthcoming 2013) (manuscript at 4–5), available at <https://www.princeton.edu/~pcglobal/conferences/methods/papers/tomz.pdf>.

21. For a recent example of legal research that used experimental methods, see Dan M. Kahan, David A. Hoffman, Donald Braman, Danieli Evans & Jeffrey J. Rachlinski, “*They Saw a Protest*”: *Cognitive Illiberalism and the Speech-Conduct Distinction*, 64 STAN. L. REV. 851 (2012) (conducting a laboratory experiment to show that cultural outlooks effect opinions on speech activities that are relevant to the speech-conduct distinction).

22. See, e.g., Druckman et al., *supra* note 16, at 629.

23. See, e.g., James N. Druckman, Donald P. Green, James H. Kuklinski & Arthur Lupia, *Experimentation in Political Science*, in CAMBRIDGE HANDBOOK OF EXPERIMENTAL POLITICAL SCIENCE 3, 3 (Druckman, Green, Kuklinski & Lupia, eds., 2011) (“The growing interest in experimentation reflects the increasing value that the discipline places on causal inference and empirically-guided theoretical refinement.”); see also D. James Greiner, *Causal Inference in Civil Rights Litigation*, 11 HARV. L. REV. 533, 558 (2008) (“[I]n causal inference, a randomized experiment is the gold standard.”). Cf. Richard A. Berk, *An Introduction to Sample Selection Bias in Sociological Data*, 48 AM. SOC. REV. 386, 392 n.8 (1983) (arguing that randomized experiments are the best way to eliminate measurement and specification errors). For a longer discussion of the benefits and drawbacks of different methods of experimental research, see *infra* PART III.

24. It is important to note that we are referring specifically to papers that explicitly conduct an experiment to test the effect of international laws—that is either treaty or customary international law. There are also a number of papers that use experimental methods to analyze the impact that international institutions have on public opinion. See, e.g., Dustin Tingley & Michael Tomz, *How Does the UN Security Council Influence Public Opinion* (2012) (working paper) (on file with author) (conducting an experiment to test the mechanisms for how the United Nations Security Council may impact public opinion); TERRANCE L. CHAPMAN, *SECURING APPROVAL: DOMESTIC POLITICS AND MULTILATERAL*

eight survey experiments.²⁷ Moreover, political scientists, and not

AUTHORIZATION FOR WAR 121–30 (2011) (testing whether information on bias in the UN Security Council influences the effect that its decisions have on public opinion); Joseph Greico, Christopher Gelpi, Jason Reifler & Peter D. Feaver, *Let's Get a Second Opinion: International Institutions and American Public Support for War*, 55 INT'L S. Q. 563 (2011); Daniel Maliank & Michael J. Tierney, *Do Foreign Public Really Care About IO Approval?* (2009) (working paper), available at http://147.142.190.246/joomla/peio/files/Maliank_Tierney.pdf.

25. Michael G. Findley, Daniel L. Nielson & J. C. Sharman, *Using Field Experiments in International Relations: A Randomized Study of Anonymous Incorporation*, INT'L ORG. (forthcoming 2013), available at http://scholar.byu.edu/danielnielson/files/findley_et_al.2013.experiments_in_ir.25may13.pdf. The results of the same field experiment were published in a law review article. See Baradaran et al., *supra* note 7; see also MICHAEL G. FINDLEY, DANIEL L. NIELSON & JASON SHARMAN, GLOBAL SHELL GAMES: EXPERIMENTS ON ANONYMOUS INCORPORATION IN TRANSNATIONAL RELATIONS (forthcoming 2014).

26. See Emilie M. Hafner-Burton, Brad L. Leveck, David G. Victor & James H. Fowler, *A Behavioral Approach to International Cooperation* (ILAR Working Paper No. 13, 2012), available at <http://ilar.ucsd.edu/assets/001/503027.pdf> (conducting an experiment on students and elites to analyze how individual traits may effect treaty negotiation); Christopher Engel, *The Emergence of a New Rule of Customary Law: An Experimental Contribution*, 7 REV. L. & ECON. 767 (2011) (conducting a laboratory experiment to provide evidence that customary international law develops because practice leads to convergent expectations).

27. See Dustin Tingley & Mike Tomz, *Conditional Cooperation and Climate Change*, COMP. POL. STUD. (forthcoming 2014) (conducting a survey experiment that tests whether respondents are more supportive of retaliatory measures being taken against countries that have promised to reduce fossil fuel omissions compared to countries that said it would do so); Stephen Chaudoin, *Promises or Policies? An Experimental Analysis of International Agreements and Audience Reactions*, INT'L ORG. (forthcoming 2014) (conducting an experiment to illustrate how respondents with expressed preferences have muted reactions when they learn that leaders have broken international agreements on trade policy); Adam Chilton, *The Influence of International Human Rights Agreements on Public Opinion: An Experimental Study*, 15 CHI. J. INT'L L. (forthcoming 2014) (conducting a survey experiment examining the influence of information on international human right agreements on public opinion); Geoffrey P.R. Wallace, *International Law and Public Attitudes Towards Torture: An Experimental Study*, 67 INT'L ORG. 105 (2013) (conducting a survey experiment to determine if the degree of legalization of international law affects public opinion on torture); Adam Chilton, *Public Opinion, the Laws of War, & Saving Civilians: An Experimental Study* (2013) (working paper) (on file with author) (conducting a survey experiment to analyze whether information on the status of international law changes public opinion on the acceptability of targeting civilians during war); Emilie Marie Hafner-Burton, Brad L. Leveck & David G. Victor, *Strategic Enforcement: Results from an Elite Survey Experiment on International Trade Agreements* (ILAR Working Paper No. 16 2012), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2132948 (conducting a survey experiment on elites to show that strategic qualities of individuals explain preferences for the designs of the enforcement provision of trade agreements); Tonya L. Putnam & Jacob N. Shapiro, *International Law and Voter Preferences: The Case of Foreign Human Rights Violations* (2013) (working paper) (on file with author) (conducting a survey experiment to

legal scholars, conducted all of these experiments.²⁸

This trend should change. In this paper, we argue that international law is a field that could uniquely benefit from experimental research. This is not only because international law is a field that is particularly concerned with causal analysis, but also because there are a number of limitations of observational data on international law that severely restrict the ability of observational studies to produce credible causal inferences on many of the topic's most important questions. To make this argument, we articulate a set of general points about observational data in international law and supplement these points with data on international legal agreements. As we will illustrate, using this data and a range of prominent examples from the literature, experimental methods provide a promising way forward to study international law, either in concert with other methods or in places where other methods may be unable to do so.

At the outset, we should note that experimental methods are not a panacea. Our argument is not that observational studies are never valuable, or that experimental methods are always preferable: there are limitations to experimental methods²⁹ and designing valid experiments can be a difficult task requiring careful research on experimental methods and procedures. Our argument is, however, that experimental methods are frequently the most credible way to test causal relationships and that many limitations with the observational

analyze whether views on the use of sanctions to punish foreign human rights abuses change as a result of respondents being provided information on the status of international law); Michael Tomz, *Reputation and the Effect of International Law on Preferences and Beliefs* (Working Paper No. 24, 2008), available at <http://www.stanford.edu/~tomz/working/Tomz-IntlLaw-2008-02-11a.pdf> (conducting experiments to determine if information on international law changes perceptions of elites and individuals through a survey experiment). It is worth noting that Tomz's paper, written in 2008, was the "first-ever experimental analysis of treaty commitments." *Id.* at 1.

28. The one exception that we are currently aware of is Shima Baradaran, Associate Professor of Law, BYU Law School. Baradaran was a coauthor with political scientists on an article using a field experiment examining compliance with international law. See Baradaran et al., *supra* note 7. Another exception worth mentioning is Katerina Linos, Assistant Professor of Law, UC Berkeley School of Law. Linos teaches and writes in international law and has a book and a number of articles using experimental methods. To our knowledge, however, Linos' research using experimental methods has focused on policy diffusion across the OECD. See, e.g., KATERINA LINOS, *THE DEMOCRATIC FOUNDATIONS OF POLICY DIFFUSION: HOW HEALTH, FAMILY, AND EMPLOYMENT LAWS SPREAD ACROSS COUNTRIES* (2013).

29. For example, one difficulty with experimental research is designing an experiment that will produce externally valid results. For more on this point, see *infra* notes 242–44 and accompanying text.

data available to study questions important in international law would benefit greatly from their use.

To make the argument that experiments should be used to study international law, we begin in PART I by offering important background on the use of experimental methods in both political science and legal scholarship. We first explore the recent growth in experimental research in political science, including its increasing use to study international relations. We next document the increasing acceptance and use of experimental methods by legal scholars in a range of fields, and then discuss how this trend has not extended to international law.

In PART II, we use examples and data to show that there are aspects of international law that make experimental methods a particularly well-suited approach to studying many important questions in the field. First, we argue that in many cases, the widespread adoption of treaties and universal applicability of customary international law mean that there is often insufficient variance in the “treatment” of international law for observational studies to be appropriate. Experimental methods can address these problems by employing research designs that randomize the information on the applicability of international law. Second, we argue that there is often a short window during which there is sufficient variation in the adoption of international agreements to allow for observational studies, but that experimental research allows for studying phenomena outside of such a narrow time frame. Third, we argue that observational studies are frequently unable to deal with the fact that many states have overlapping legal constraints created by the presence of domestic laws and international treaties covering the same issue areas. This collinearity poses serious inference problems for observational studies, but can easily be overcome by using designs that randomize the laws discussed in experimental vignettes. Fourth, we argue that observational studies have had difficulty finding appropriate dependent variables to measure whether states comply with international agreements. Experimental research can overcome this obstacle by using research designs that directly measure theories of why international law might change a state’s policies and actions. Finally, we conclude this part by arguing that experimental research can help overcome the most widely discussed inferential barrier to the study of international law—selection bias.

After making the case for why experimental research is needed, in PART III of the paper, we turn to specifically discussing how international legal scholars can conduct experimental research. To do so, we discuss each of three experimental methods in turn: la-

laboratory experiments, survey experiments, and field experiments.³⁰ For each, we provide an explanation of the virtues and vices of the method, provide examples of cutting edge research using the approach, and discuss how it might be applied to research on important topics in international law.

I. THE GROWTH OF EXPERIMENTAL RESEARCH

The use of experimental methods in political and legal research has increased dramatically in the last two decades.³¹ Today, improvements in technology mean that researchers can easily, cheaply, and quickly design and administer experiments to thousands of subjects.³² Researchers can now design their experiments in ways that directly test hypothetical theoretical causal mechanisms in ways that were previously difficult, if not impossible, with observational studies alone.³³ We argue scholars of international law should use experimental methods not only because these developments can make possible directly testing the plausibility of their theories, but also because observational methods are often particularly poor ways to study international law. We will begin, in this part, by providing general background on the growth of experimental methods.

This part will proceed in two sections. We will first document the growth of experimental research in political science, as well as its recent popularity as a method of studying international relations. We will then discuss how legal scholars have begun to embrace experimental research, but that this embrace has not extended

30. A fourth category of research that can also be considered experimental is natural experiments. Although we will not consider natural experiments in depth, they are a growing research strategy used by legal scholars and political scientists that present a promising research method in many cases. For more information, *see infra* notes 238–40 and accompanying text.

31. *See generally* Druckman et al., *supra* note 16. *See infra* notes 34–48, 67–80 and accompanying text.

32. *See, e.g.*, Adam J. Berinsky, Gregory A. Huber & Gabriel S. Lenz, *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANAL. 351 (2012). We will provide guidance on how to conduct experiments—including online experiments—in PART III.

33. *See generally* Kosuke Imai, Dustin Tingley & Teppei Yamamoto, *Experimental Designs for Identifying Causal Mechanisms*, 176 J. ROYAL STAT. SOC'Y 5 (2013). *See also* Kosuke Imai, Luke Keele, Dustin Tingley & Teppei Yamamoto, *Unpacking the Black Box of Causality: Learning About Causal Mechanisms from Experiments and Observational Studies*, 105 AM. POL. SCI. REV. 765 (2011). For a longer discussion about designing experiments to test causal mechanisms, *see infra* PART III.B.

to international law. As we will discuss, legal scholars studying international law generally do not use experimental methods. This discussion will help lay the foundation for our argument in PART II that international law could benefit from experimental research.

A. The Growth of Experiments in Political Science

Research on the value of randomization for causal inference, and thus the theoretical underpinnings of modern experimental research, first began to emerge in the 1920s.³⁴ Despite this early research on the value of experiments, social scientists generally were slow to adopt experimental methods.³⁵ In fact, economics and political science were both disciplines where very little experimental research was published in the fifty years after the first work on the importance of randomization.³⁶ The first experiment was not published in the *American Political Science Review* (APSR)—the leading political science journal—until 1956.³⁷ This was not a watershed moment: it took over a decade for the next experiment to appear in the APSR.³⁸ Moreover, commentators have suggested that the use of experimental methods was poorly regarded by the discipline until recently.³⁹

34. Ho & Rubin, *supra* note 7, at 18 (citing R.A. FISHER, *STATISTICAL METHODS FOR RESEARCH WORKERS* (1925); R.A. FISHER, *THE DESIGN OF EXPERIMENTS* (1935)). *See also* Druckman et al., *supra* note 23, at 4 (“While scientists have conducted experiments for hundreds of years, modern experimentation made its debut in the 1920s and 1930s.”).

35. *See* Armin Falk & James J. Heckman, *Lab Experiments Are a Major Source of Knowledge in the Social Sciences*, 326 *SCIENCE* 535, 535 (2009) (“The social sciences have generally been less willing to use laboratory experiments than the natural sciences, and empirical social science has traditionally been considered as largely non-experimental, that is, based on observations collected in naturally occurring situations.”).

36. For evidence on the limited use of experiments by economists before the 1970s, see *id.* at 535 (“Fewer than 10 experimental papers [in economics] per year were published before 1965, which grew to about 30 per year by 1975.”). For a discussion on the growth of the use of experimental methods in political science, see REBECCA B. MORTON & KENNETH C. WILLIAMS, *EXPERIMENTAL POLITICAL SCIENCE AND THE STUDY OF CAUSALITY: FROM NATURE TO THE LAB* 3–10 (2010); Kathleen McGraw & Valerie Hoekstra, *Experimentation in Political Science: Historical Trends and Future Directions*, in *RESEARCH IN MICROPOLITICS* 3 (1994); Druckman et al., *supra* note 16, at 628–30.

37. *See* Druckman et al., *supra* note 16, at 628–29. The first experiment was a study by Samuel Eldersveld that encouraged participants to vote based on personal contact. *See* Samuel J. Eldersveld, *Experimental Propaganda Techniques and Voting Behavior*, 50 *AM. POL. SCI. REV.* 154 (1956).

38. *See* Druckman et al., *supra* note 16, at 629.

39. *See* MORTON & WILLIAMS, *supra* note 36, at 3 (“Until the last decade,

In recent years, however, experimental methods have become widely used in political science. Although there is a debate on the exact number of published articles using experimental research methods,⁴⁰ there is agreement that the number has grown exponentially in the last two decades.⁴¹ In fact, there were over fifty experimental articles published in just three political science journals between 2000 and 2007.⁴² Although perhaps the most prominent experiments published in the last decade have focused on voter turnout,⁴³ experimental methods have been used to study a range of topics: from how individual attitudes towards immigration policy are formed,⁴⁴ to the circumstances under which incorporating private firms comply with international disclosure requirement laws.⁴⁵ Additionally, political scientists have held conferences dedicated solely to experimental research,⁴⁶ formed a section of the American Political

experimentation seemed to have a low standing within the discipline.”). *See also* Druckman et al., *supra* note 16, at 627 (“[P]olitical scientists have long expressed skepticism about the prospects for experimental science.”).

40. *See* MORTON & WILLIAMS, *supra* note 36, at 3 (citing Rose McDermott, *Experimental Methods in Political Science*, 5 ANN. REV. POL. SCI. 32 (2002) (discussing a disagreement over how to correctly classify which articles should be counted as experimental political science)).

41. *See id.* at 4–5. *See also* Druckman et al., *supra* note 16, at 628.

42. *See* MORTON & WILLIAMS, *supra* note 36, at 4.

43. *See, e.g.*, Alan S. Gerber, Donald P. Green & Christopher W. Larimer, *Social Pressure and Voter Turnout: Evidence From a Large-Scale Field Experiment*, 102 AM. POL. SCI. REV. 33 (2008) (conducting a large field experiment to show that voters are more likely to vote after receiving letters that apply social pressure encouraging them to vote). *See also* GREEN & GERBER, *supra* note 18; Shaw et al., *supra* note 18.

44. *See, e.g.*, Dustin Tingley, *Public Finance and Immigration Preferences: A Lost Connection?* 45 POLITY 4 (2013) (using experimental evidence to produce evidence that individual immigration preferences are not driven by states’ public finances environments). *See also* Jens Hainmueller & Michael J. Hiscox, *Attitudes Toward Highly Skilled and Low-Skilled Immigration: Evidence from a Survey Experiment*, 104 AM. POL. SCI. REV. 61 (2010) (conducting an experiment to produce evidence that nonmaterial factors explain individual immigration preferences).

45. *See* Findley et al., *supra* note 25 (conducting a field experiment to test whether information on the status of international law changes private firms’ likelihood to comply with international disclosure requirements in response to requests to incorporate). *See also* Baradaran et al., *supra* note 7 (reporting the results of the same experiments).

46. For example, the New York University Center for Experimental Social Science (NYU – CESS) held its fifth annual conference in March 2012. For more information, see <http://cess.nyu.edu/policon2012/> (last visited Jan. 17, 2013). Additionally, Brigham Young University held a “Field Experiments in International Relations” Conference in September 2012. *See* <https://pedl.byu.edu/Pages/Conference.aspx> (last visited Jan. 24, 2013), for more information.

Science Association dedicated to experimental research,⁴⁷ and even established a regular newsletter on experimental political science.⁴⁸

What should be of interest to scholars of international law, however, is that experimental methods have not only been used to study American politics,⁴⁹ but also have increasingly been used to study international relations.⁵⁰ For example, laboratory experiments have been used to study how personality traits and incentives affect policy on topics ranging from the distribution of foreign aid to the conduct of war,⁵¹ survey experiments have been used to understand how audience costs change political decision making,⁵² and field experiments have helped illuminate how international organizations should improve resource allocation.⁵³ Although experimental research is admittedly far from being the primary method used to study

47. The “Experimental Research” section of the American Political Science Association was organized in 2010. See *Organized Section: 42. Experimental Research*, AM. POL. SCI. ASS’N, <https://www.apsanet.org/sections/sectionDetail.cfm?section=Sec42> (last visited Jan. 17, 2013), for more information.

48. Established in 2010, the “Experimental Political Scientist” is the American Political Science Association’s bi-annual newsletter. It is edited by Dustin Tingley and contains advice on new methods for conducting experimental research and surveys of recent research. For the complete archive, see <http://scholar.harvard.edu/dtingley/pages/expopolisci> (last visited Jan. 18, 2013).

49. Rose McDermott, *New Directions for Experimental Work in International Relations*, 55 INT’L STUD. Q. 503, 503 (2011) (“Experimentation has infiltrated some aspects of political science, most notably in the realms of public opinion and voting behavior and economic gains in comparative perspective. . .”) (citations omitted).

50. For articles reviewing the use of experimental methods to study international relations, see Alex Mintz, Yi Yang & Rose McDermott, *Experimental Approaches to International Relations*, 55 INT’L STUD. Q. 493 (2011); McDermott, *supra* note 49; Natalie Florea Hudson & Michael J. Butler, *The State of Experimental Research in IR: An Analytic Study*, 12 INT’L STUD. REV. 165 (2010).

51. See Dustin Tingley, *The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining*, 55 INT’L STUD. Q. 521 (2011); Joshua Kertzer, *Taking Resolve Seriously: Three Theories of Willpower in International Politics* (Mar. 16, 2011) (unpublished manuscript) (on file with author).

52. See, e.g., Michael Tomz, *Domestic Audience Cost in International Relations: An Experimental Approach*, 61 INT’L ORG. 821 (2007).

53. See Susan D. Hyde, *Experimenting in Democracy Promotion: International Observers and the 2004 Presidential Elections in Indonesia*, 8 PERSPECTIVES ON POL. 511 (2010); Jeremy M. Weinstein & James D. Fearon, *Can Development Aid Contribute to Social Cohesion After Civil War? Evidence from a Field Experiment in Post-Conflict Liberia*, 99 AM. ECON. REV. 287 (2009); Mary Kay Gugerty & Michael Kremer, *Outside Funding and the Dynamics of Participation in Community Associations*, 52 AM. J. POL. SCI. 585 (2008).

international affairs,⁵⁴ this growing body of research shows that conducting experiments is an incredibly valuable way to gain insight into international relations.⁵⁵ Experiments do not just compliment observational research methods; they also provide insight into important questions that have been at a standstill because of inferential problems with observational data.⁵⁶

The increase in the use of experimental methods in both political science generally and international relations specifically has been driven by several factors.⁵⁷ First, as previously noted,⁵⁸ experimental methods make it possible to make credible causal claims when it is frequently difficult to do so with observational data:⁵⁹ by randomly assigning when a specific intervention is given without varying any other factors, it is possible to reliably estimate the influence of that intervention on a given outcome. For example, to test whether a particular drug helps treat a disease, medical researchers randomly assign a group of patients to receive the drug while other patients receive a placebo or no treatment at all. If the group that received the drug sees its condition change relative to the group that did not, researchers can assume that the drug caused the difference. Similarly, political scientists are able to analyze whether international election monitors influence voting results by randomly selecting polling sta-

54. See McDermott, *supra* note 49, at 503.

55. For example, there is a large body of experimental research trying to understand international security. See, e.g., CHRISTOPHER GELPI, PETER FEAVER & JASON A. REIFLER, *PAYING THE HUMAN COSTS OF WAR: AMERICAN PUBLIC OPINION AND CASUALTIES IN MILITARY CONFLICTS* (2009); Adam J. Berinsky, *IN TIME OF WAR: UNDERSTANDING AMERICAN PUBLIC OPINION FROM WORLD WAR II TO IRAQ* (2009); Matthew S. Levendusky & Michael C. Horowitz, *When Backing Down is the Right Decision: Partisanship, New Information, and Audience Costs*, 74 J. Pol. 323 (2012); Michael C. Horowitz & Matthew S. Levendusky, *Drafting Support for War: Conscription and Mass Support for Warfare*, 73 J. Pol. 524 (2011); Robert F. Trager & Lynn Vavreck, *The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party*, 55 Am. J. Pol. Sci. 526 (2011); Matthew A. Baum & Tim Groeling, *Shot by the Messenger: Partisan Cues and Public Opinion Regarding National Security and War*, 31 Pol. Behav. 157 (2009); Scott Sigmund Gartner, *The Multiple Effects of Casualties on Public Support for War: An Experimental Approach*, 102 Am. Pol. Sci. Rev. 95 (2008).

56. See, e.g., Tomz & Weeks, *supra* note 20 (discussing why a number of problems with observational data have made it impossible to provide a satisfactory test of the democratic peace hypothesis).

57. See generally MORTON & WILLIAMS, *supra* note 36, at 12–16.

58. See *supra* note 23.

59. See generally MORTON & WILLIAMS, *supra* note 36, at 12–14. See also Druckman et al., *supra* note 16, at 627.

tions with monitors present during an election.⁶⁰

Secondly, technological improvements have made it easier to conduct experimental research.⁶¹ For example, prominent commentators have in part attributed the rise in political scientists' experimental research to the development of computer-assisted telephone interviews in the 1990s.⁶² The Internet has made it possible to conduct experiments cheaply,⁶³ also giving rise to a new wave of experimental research.⁶⁴ The result of these and other⁶⁵ technological developments means that, as political scientists are increasingly recognizing the value of experiments, it is becoming easier than ever to incorporate them into their research.

B. The Growth of Experiments in Legal Research

The same factors that have led political scientists to increasingly conduct experimental research have also begun to influence legal scholars.⁶⁶ Unsurprisingly, however, legal researchers have been slower to fully adopt experimental research methods.⁶⁷ The first wave of empirical legal studies took shape in the 1920s and 1930s,⁶⁸ which was roughly the same time when the theoretical foundations for modern experimental research were developed.⁶⁹ Despite these early roots, empirical legal research was still rare for much of the

60. See Hyde, *supra* note 53.

61. See also Druckman et al., *supra* note 16, at 629–30. But see MORTON & WILLIAMS, *supra* note 36, at 12 (arguing that although technological advances have made it easier to conduct experimental research, “technology cannot be the primary answer for why experiments have increased in standing”).

62. See Druckman et al., *supra* note 16, at 629–30.

63. Cf. Berinsky et al., *supra* note 32 (demonstrating how Amazon’s “Mechanical Turk” service can be used to conduct reliable social scientific experiments).

64. See, e.g., Tingley & Tomz, *supra* note 27.

65. For example, easy to use computer software that can help analyze survey results.

66. Cf., e.g., Griener, *supra* note 23, at 558.

67. For an excellent primer on the use and misuse of empirical methods by legal scholars generally, see Epstein & King, *supra* note 7.

68. For a fascinating discussion of early legal empirical research, see Ho & Rubin, *supra* note 7, at 18–19 (citing JOHN HENRY SCHLEGEL, AMERICAN LEGAL REALISM AND EMPIRICAL SOCIAL SCIENCE (1995)); see also Herbert M. Kritzer, *The (nearly) Forgotten Early Empirical Legal Research*, in OXFORD HANDBOOK OF EMPIRICAL LEGAL RESEARCH 875 (P. Cane & H.M. Kritzer eds., 2010).

69. See *supra* text accompanying note 34.

twentieth century,⁷⁰ and the use of experimental methods was even rarer still. Perhaps the earliest appreciation for experimental methods came from the law and economics movement.⁷¹ That said, the particular recognition of the value of experiments focused on explaining the insights legal scholars could gain from laboratory experiments conducted by economists; the actual examples of legal researchers conducting experiments on their own were few and far between.⁷²

In the last twenty years, however, this trend has begun to change.⁷³ Leading law journals have published articles involving laboratory experiments,⁷⁴ survey experiments,⁷⁵ field experiments,⁷⁶

70. Cf. Richard H. McAdams & Thomas S. Ulen, *Introduction to the Symposium on Empirical and Experimental Methods in Law*, 2002 U. ILL. L. REV. 791, 791 (2003) (“Empirical methods are still rare in legal scholarship: very few law professors buttress their arguments by appeals to tests of statistical significance or even with descriptive statistics.”). McAdams & Ulen do, however, suggest that empirical and experimental methods are becoming more common in legal scholarship. See *id.* (citing Robert C. Ellickson, *Trends in Legal Scholarship: A Statistical Study*, 29 J. LEGAL STUD. 517 (2000)).

71. See Elizabeth Hoffman & Matthew L. Spitzer, *Experimental Law and Economics: An Introduction*, 85 COLUM. L. REV. 991 (1985). For more recent examples of law and economics scholarship on the value of experiments, see Rachel Croson, *Experimental Law and Economics*, 5 ANN. REV. L. SOC. SCI. 25 (2009), and Rachel Croson, *Why and How to Experiment: Methodologies from Experimental Economics*, 2002 U. ILL. L. REV. 921 (2003).

72. Cf. Hoffman & Spitzer, *supra* note 71, at 1024 n.110 (documenting the exceptions to their observation that “[m]uch of the research discussed [in their article] was neither done by legal scholars nor designed specifically to test or investigate theories of law and economics”).

73. See McAdams & Ulen, *supra* note 70, at 791–92 (discussing the rise of legal scholars using empirical methods generally); cf. Ho & Rubin, *supra* note 7, at 20 (documenting experimental studies relevant to law, with the earliest published in 1990); Seidman, Diamond & Mueller, *supra* note 1, at 590–92 (documenting the use of empirical research in law reviews, including experimental methods).

74. See, e.g., Michael D. Guttentag, Christine L. Porath & Samuel N. Fraidin, *Brandeis’ Policeman: Results From a Laboratory Experiment on How to Prevent Corporate Fraud*, 5 J. EMPIRICAL LEGAL STUD. 239 (2008); see also Jeremy A. Blumenthal, *Group Deliberation and the Endowment Effect: An Experimental Study*, 50 HOUS. L. REV. 41 (2012).

75. See, e.g., David Fontana & Donald Braman, *Judicial Backlash or Just Backlash? Evidence from a National Experiment*, 112 COLUM. L. REV. 731 (2012); David A. Hoffman & Alexander S. Radus, *Instructing Juries on Noneconomic Contract Damages*, 81 FORDHAM L. REV. 1221 (2012).

76. See, e.g., David Schakade, Cass R. Sunstein & Reid Hastie, *What Happened on Deliberation Day?*, 95 CAL. L. REV. 915 (2007); see also D. James Greiner, Cassandra Wolos Pattanayak & Jonathan Hennessy, *How Effective Are Limited Legal Assistance Programs? A Randomized Experiment in a Massachusetts Housing Court* (2012) (working paper), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1880078 (last

and even natural experiments.⁷⁷ In these articles, experimental methods have helped to answer important questions from a range of legal fields, including criminal law,⁷⁸ constitutional law,⁷⁹ and corporate law,⁸⁰ among others. It appears that experimental methods are gaining mainstream acceptance as a credible research method for legal scholars.

While experimental methods have begun to permeate law generally, largely because of their superiority over observational methods in evaluating causality, this has not yet occurred in the field of international law.⁸¹ In fact, as far as we can tell, legal scholars studying international law have essentially not used experimental methods at all.⁸² For example, in Shaffer and Ginsburg's recent extensive review of empirical scholarship on international law, they do not discuss a single experiment.⁸³ Moreover, we have been unable to find any original experimental research published in an international law journal. As far as we can tell, the experimental work on international law that has been conducted has been done entirely by professors in political science departments that are interested in international relations and international legal issues.⁸⁴ Given the growing use of

visited Jan. 15, 2013); Carol Seron, Martin Frankel, Gregg Van Ryzin & Kean Kovath, *The Impact of Legal Counsel on Outcomes for Poor Tenants in New York City's Housing Court: Results of a Randomized Experiment*, 35 LAW & SOC'Y REV. 419 (2001); Ian Ayres, *Fair Driving: Gender and Race Discrimination in Retail Car Negotiations*, 104 HARV. L. REV. 817 (1991).

77. See, e.g., Bert I. Huang, *Lightened Scrutiny*, 124 HARV. L. REV. 1109 (2011); Christopher R. Berry & Jacob E. Gersen, *The Timing of Elections*, 77 U. CHICAGO L. REV. 37 (2010); see also David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145 (2007).

78. See Douglas L. Colbert, Ray Paternoster & Shawn Bushway, *Do Attorneys Really Matter? The Empirical and Legal Case for the Right of Counsel at Bail*, 23 CARDOZO L. REV. 1719 (2002); see also Gary S. Green, *General Deterrence and Television Cable Crime: A Field Experiment in Social Control*, 23 CRIMINOLOGY 629 (1985). See generally CHRISTINE HORNE & MICHAEL LOVAGLIA, *EXPERIMENTAL STUDIES IN LAW AND CRIMINOLOGY* (2008).

79. See, e.g., Kahan et al., *supra* note 21 (using experimental methods to explore the First Amendment speech/conduct distinction).

80. See, e.g., Guttentag et al., *supra* note 74; Lucian A. Bebchuk, Alma Cohen & Charles C.Y. Wang, *Staggered Boards and the Wealth of Shareholders: Evidence from a Natural Experiment*, HARV. OLIN DISCUSSION PAPER NO. 697 (2011).

81. But see sources cited *supra* note 25 (documenting recent studies that use experimental methods to examine questions relevant to the study of international law).

82. But see *supra* note 28.

83. See Shaffer & Ginsburg, *supra* note 1.

empirical methods to study international law⁸⁵ and the problems of inference that have led many to criticize observational studies in the field,⁸⁶ it is time that this changes.

II. WHY EXPERIMENTS ARE NEEDED FOR INTERNATIONAL LAW

As the previous section demonstrated, experimental methods are an increasingly common way for scholars to conduct research on political and legal issues, and the ability of experiments to provide reliable causal estimates is a considerable advantage to using these methods when studying international events. That said, we are unaware of any attempts to argue that experimental methods are uniquely appropriate for the study of international law. In fact, the justifications for using experimental methods in international law run deeper than those for using experimental methods in international relations more generally. The ratification patterns and complexity of international agreements mean that international law is particularly well suited to experimental methods. Accordingly, although there have been considerable advances in the understanding of the effects of international law using observational research designs,⁸⁷ this discussion will demonstrate further why these approaches should be augmented by experimental research.

In this Part, we present five arguments for why experimental methods are particularly appropriate for studying international law. First, we argue that there is frequently insufficient variation in the countries that are bound by sources of international law—specifically multinational treaties and customary law—to assess whether those laws change country behavior. This makes experimental methods an appropriate and cost-efficient way to create that variation by randomizing input information on whether states are bound by international law. Second, we present evidence that there are frequently short windows of time when there is variance in treaty adoption, meaning there may be too few events that occur during that time frame to allow for a reliable study. Given these realities, we explain how exper-

84. See, e.g., Findley, *supra* note 25; Hafner-Burton, *supra* note 26; Putnam & Shapiro, *supra* note 27; Tomz, *supra* note 27; Wallace, *supra* note 27.

85. See *supra* text accompanying notes 1–6.

86. See *supra* text accompanying notes 7–15.

87. For the most extensive effort to use sophisticated statistical methods to test international law, see SIMMONS, *supra* note 4. For a summary of empirical findings from using empirical efforts to study international law, see Shaffer & Ginsburg, *supra* note 1; Hafner-Burton et al., *supra* note 1; Simmons, *supra* note 1.

imental methods present a way to explore hypothetical events outside of the narrow windows available to observational studies. Third, we explain how observational studies frequently do not account for overlapping legal constraints and thus cannot isolate the causal effects of individual agreements on states. We suggest that experimental methods can help address this by testing whether international agreements may have an “additive,” rather than a purely “substitutive,” effect over other sources of law. Fourth, we discuss how observational studies often have difficulty finding appropriate dependent variables to test state compliance. We argue that experiments can be designed in ways to directly test whether state and non-state actors comply, thus overcoming the problems of using large datasets that are poor direct measures of non-compliant behavior. Fifth, we argue that although observational studies have problems accounting for the fact that states do not randomly commit to international treaties or decide to bring international litigation, experimental research can help overcome the selection bias in these situations.

A. Lack of Variance in Sources of International Law

A fundamental requirement of causal analysis is that any explanatory variable must actually vary to assess its causal impact.⁸⁸ If an explanatory variable does not take on multiple values, it is impossible to isolate the effect of that variable when compared to other factors that could influence a particular outcome. This requirement often poses difficulties for qualitative and quantitative researchers who would like to study any fixed feature of an environment.⁸⁹ Experimental research, however, allows researchers to artificially create the variance required to study a given topic.⁹⁰

The need for explanatory variables to vary presents a particular challenge for any observational study analyzing the influence of

88. GARY KING, ROBERT O. KEOHANE & SIDNEY VERBA, *DESIGNING SOCIAL INQUIRY* 146 (1994) (“[T]he causal effect of an explanatory variable that does not vary cannot be assessed.”).

89. *See, e.g.*, KENNETH N. WALTZ, *MAN, THE STATE, AND WAR: A THEORETICAL ANALYSIS* 16–41 (1959) (arguing that it is difficult, if not impossible, to assess the influence of human nature on war because if human nature is fixed, it is not possible to use it as a variable to explain instances of war and peace); *see also* KING, KEOHANE & VERBA, *supra* note 88, at 147 (citing DAVID D. LAITIN, *HEGEMONY AND CULTURE: POLITICS AND RELIGIOUS CHANGE AMONG THE YORUBA* (1986)) (noting Laitin’s difficulty in studying the influence of religion on politics in Somalia because of the religious homogeneity among the population).

90. For example, this can be done by varying information on whether particular countries have ratified a given treaty. *See, e.g.*, Tomz, *supra* note 27.

international law on state behavior: the two primary sources of international law, treaties and customary international law,⁹¹ often do not have sufficient variance to allow for a satisfactory analysis of their causal impact. It is possible, however, to overcome these problems using experimental methods.

1. Widespread Ratification of Multinational Treaties

It is increasingly difficult to assess the influence of many of the most important multinational treaties because they have become so widely adopted. As we will discuss, experimental methods, by varying the information provided to research subjects on whether or not a given country has ratified a treaty, can help with this roadblock in international law study.

One of the most prominent lines of empirical research in international law has been examining whether the ratification of treaties influences the behavior of states.⁹² This line of research has used treaty ratification as the key explanatory variable in a wide range of substantive issue areas in international law.⁹³ This includes research on compliance with international monetary rules,⁹⁴ environmental standards,⁹⁵ human rights,⁹⁶ and the laws of war,⁹⁷ among others.⁹⁸ The difficulty, however, is that many of the most important treaties have become so ubiquitous that there is very little variance in their

91. See Curtis A. Bradley & Mitu Gulati, *Withdrawing from International Custom*, 120 YALE L.J. 202, 204 (2010) (“There are two basic types of international law—treaties and customary international law (CIL).”).

92. See, e.g., Simmons, *supra* note 1, at 274; see also Hafner-Burton et al., *supra* note 1, at 89 (“Research on whether an international legal agreement has had an effect on state behavior often probes whether states parties comply with the terms of agreements more often than nonparties.”).

93. See generally Simmons, *supra* note 1, at 280–92 (reviewing empirical research on treaty compliance by issue area).

94. See Simmons, *supra* note 11; see also von Stein, *supra* note 12; Simmons & Hopkins, *supra* note 12.

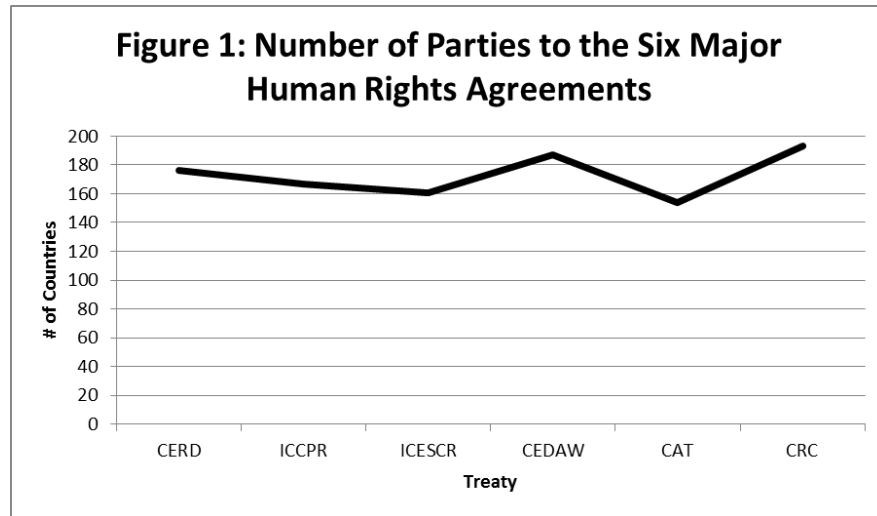
95. See, e.g., Oran A. Young & Michael Zürn, *The International Regimes Database: Designing and Using a Sophisticated Tool for Institutional Analysis*, 6 GLOBAL ENVTL. POL. 121 (2006).

96. See, e.g., SIMMONS, *supra* note 4; Hathaway, *supra* note 4.

97. See, e.g., Morrow, *supra* note 5; Valentino et al., *supra* note 5.

98. For example, there is a growing literature on the effect of ratification of Bilateral Investment Treaties. See, e.g., Zachary Elkins, Andrew Guzman & Beth Simmons, *Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000*, 60 INT'L ORG. 811 (2006).

ratification. Moreover, since many studies have argued that treaties likely have the largest influence on certain types of countries,⁹⁹ there may be even less variance among this subset. Given the insufficient variance in states that have ratified major agreements, using observational studies is a less viable option for studying the influence of treaty ratification on state behavior.¹⁰⁰



To help illustrate this point, Figure 1 and Figure 2 present data on the number of states party to two categories of multilateral treaties.¹⁰¹ Figure 1 presents the number of states party to the six “core” human rights treaties:¹⁰² the International Covenant on Civil and Po-

99. See, e.g., SIMMONS, *supra* note 4, at 150–54 (arguing that human rights treaties have the largest impact on “Transnational/Partly Democratic Countries”); Morrow, *supra* note 5, at 561 (arguing that democracies are more likely to comply with treaties on the laws of war).

100. To be clear, there are sources of variation that can be exploited for observational studies, such as the examination of the time period before treaties have been widely ratified and the comparison of individual countries’ behaviors before and after ratification. Our broader point, however, is that experimental methods can introduce variation that is otherwise not present in observational data.

101. The source for the data presented in Figure 1 is from the United Nations Treaty Collection, available at <http://treaties.un.org/Pages/ParticipationStatus.aspx> (last visited Nov. 17, 2012). Figure 2 additionally contains data from the International Committee of the Red Cross, available at <http://www.icrc.org/eng/war-and-law/treaties-customary-law/geneva-conventions/index.jsp> (last visited Nov. 17, 2012). The data presented in Figures 1 & 2 are based on the number of state parties to these agreements as of Nov. 16, 2012.

102. See SIMMONS, *supra* note 4, at 59–64 (discussing the six “core” human rights

litical Rights (ICCPR);¹⁰³ the International Covenant on Economic, Social and Cultural Rights (ICESCR);¹⁰⁴ the International Convention on the Elimination of All Forms of Racial Discrimination (CERD);¹⁰⁵ the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW);¹⁰⁶ the Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT);¹⁰⁷ and the Convention on the Rights of the Child (CRC).¹⁰⁸ Figure 2 presents the number of states party to the major multinational conventions that have sought to regulate crimes against humanity and armed conflicts post WWII: the Convention on the Prevention and Punishment of the Crime of Genocide;¹⁰⁹ the Geneva Conventions of 1949 (Geneva Conventions I–IV);¹¹⁰ Additional Pro-

treaties being the ICCPR, ICESCR, CERD, CEDAW, CRC, and CAT). *See also International Instruments*, OFFICE OF THE HIGH COMM’R FOR H.R., available at <http://www.ohchr.org/EN/Issues/Disability/Pages/Instruments.aspx> (last visited Nov. 17, 2012) (providing information on the “six core human rights conventions”).

103. International Covenant on Civil and Political Rights, G.A. Res. 2200A (XXI), U.N. GAOR, 21st Sess., Supp. No. 16, U.N. Doc. A/6316, at 52 (Dec. 16, 1966), 999 U.N.T.S. 171, *entered into force* Mar. 23, 1976.

104. International Covenant on Economic, Social and Cultural Rights, G.A. Res. 2200A (XXI), U.N. GAOR, 21st Sess., Supp. No. 16, U.N. Doc. A/6316, at 49 (Dec. 16, 1966), 993 U.N.T.S. 3, *entered into force* Jan. 3, 1976.

105. International Convention on the Elimination of All Forms of Racial Discrimination, G.A. Res. 2106 (XX), U.N. GAOR, 20th Sess., Supp. No. 14, U.N. Doc. A/6014, at 47 (Dec. 21, 1965), 660 U.N.T.S. 195, *entered into force* Jan. 4, 1969.

106. Convention on the Elimination of All Forms of Discrimination Against Women, G.A. Res. 34/180, U.N. GAOR, 34th Sess., Supp. No. 46, U.N. Doc. A/34/46, at 193 (Dec. 18, 1979), 1249 U.N.T.S. 13, *entered into force* Sept. 3, 1981.

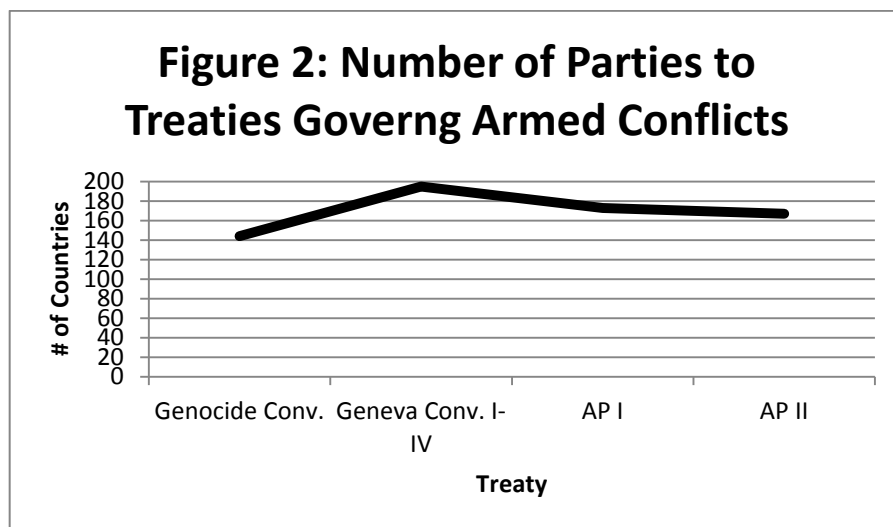
107. Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, G.A. Res. 39/46, U.N. GAOR, 39th Sess., Supp. No. 51, U.N. Doc. A/39/51, at 197 (Dec. 10, 1984), 1465 U.N.T.S. 85, *entered into force* June 26, 1987.

108. Convention on the Rights of the Child, G.A. Res. 44/25, U.N. GAOR, 44th Sess., Supp. No. 49, U.N. Doc. A/44/49 at 167 (Nov. 20, 1989), 1577 U.N.T.S. 3, *entered into force* Sept. 2, 1990.

109. Convention on the Prevention and Punishment of the Crime of Genocide, G.A. Res. 260 (III), U.N. Doc. A/RES/260(III) (Dec. 9, 1948), 78 U.N.T.S. 277, *entered into force* Jan. 12, 1951.

110. Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Aug. 12, 1949, 75 U.N.T.S. 31, *entered into force* Oct. 21, 1950; Geneva Convention for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, Aug. 12, 1949, 75 U.N.T.S. 85, *entered into force* Oct. 21, 1950; Geneva Convention Relative to the Treatment of Prisoners of War, Aug. 12, 1949, 75 U.N.T.S. 135, *entered into force* Oct. 21, 1950; Geneva Convention Relative to the Protection of Civilian Persons in Time of War, Aug. 12, 1949, 75 U.N.T.S. 287, *entered into force* Oct. 21, 1950.

Protocol I to the Geneva Conventions of 1949 (AP I);¹¹¹ and Additional Protocol II to the Geneva Conventions of 1949 (AP II).¹¹² As Figures 1 & 2 show, the number of parties to these agreements ranges from 154 to 195 states. To put this in perspective, with the addition of South Sudan in 2011, there were 193 members of the United Nations.¹¹³ Thus, an overwhelming number of states are party to the most important treaties regulating human rights and armed conflicts. This leaves little variance in ratification for scholars seeking to assess the impact of these agreements using traditional observational studies.¹¹⁴



Although the lack of variance in the ratification of major multilateral treaties poses a serious obstacle for observational studies,

111. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the protection of victims of international armed conflicts (Protocol I), Jun. 8, 1977, 1125 U.N.T.S. 3, *entered into force* Dec. 7, 1978.

112. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the protection of victims of non-international armed conflicts (Protocol II), Jun. 8, 1977, 1125 U.N.T.S. 609, *entered into force* Dec. 7, 1978.

113. *United Nations Member States—Growth in United Nations Membership, 1945–present*, UN.ORG, <http://www.un.org/en/members/growth.shtml> (last visited November 17, 2012). The Cook Islands are the only country that has signed the Geneva Conventions but is not a member of the United Nations.

114. Of course it is still possible to gain some leverage by studying the effects of these treaties during windows before ratification was so prevalent. For a discussion of this approach, *see infra* PART II.B.1.

experimental research is able to help solve this problem. It is possible to employ an experimental design that randomizes whether countries are subject to a given treaty to try to estimate the impact that agreement might have on individual preferences or predictions. For example, Michael Tomz of Stanford University has conducted an experiment on members of the British House of Commons to determine “the impact of treaties on expectations.”¹¹⁵ During interviews with these policy makers, Tomz asked for their opinions on whether a country in a hypothetical scenario was pursuing the development of nuclear weapons. Half of the respondents were presented with a series of facts that included that the country in question *had* signed the Nuclear Non-Proliferation Treaty (NPT). The other respondents received an identical set of facts, but were told that the country *had not* signed the NPT. 35% of the NPT-signed respondents thought it was likely that the country was pursuing a nuclear weapon, but of those told the country had *not* signed the NPT, 61% thought this pursuit was likely.¹¹⁶ Although this is not proof that the NPT is having an impact, it is at least evidence that high-level policy makers believe it is an important signal. Moreover, since there are 190 countries currently party to the NPT,¹¹⁷ it is nearly impossible to try and estimate whether the NPT has any influence without taking this kind of creative approach, made possible by experimental research design.

2. Universal Applicability of Customary International Law

Experimental methods also provide a way to deal with a related and perhaps more difficult to account for problem: there is essentially no variance in the applicability of Customary International Law (CIL) because it applies to (almost) every country. Experiments help deal with this by comparing reactions to specific treaty obligations to the reactions to similar obligations simply grounded in general principles of international law.

CIL is universally applicable international law, regardless of whether a country has signed a formal treaty.¹¹⁸ The only way for a country to not be bound by CIL is to be a persistent objector while a

115. Tomz, *supra* note 27.

116. *Id.* at 24–28.

117. *Status of the Treaty on the Non-Proliferation of Nuclear Weapons*, United Nations Office for Disarmament Affairs, U.N. OFFICE FOR DISARMAMENT AFFAIRS, <http://disarmament.un.org/treaties/t/npt> (last visited Nov. 17, 2012).

118. See generally Bradley & Gulati, *supra* note 91. See also JACK L. GOLDSMITH & ERIC A. POSNER, *THE LIMIT OF INTERNATIONAL LAW* (2005).

norm of CIL is being formed.¹¹⁹ In practice, there are very few examples of countries being successful objectors.¹²⁰ Even when a state has persistently objected, CIL is often considered a “peremptory norm” where objections are not valid.¹²¹ Moreover, these peremptory norms include many of the strongest norms of international law.¹²²

For our purposes, the relevant takeaway is that CIL is a nearly universal source of international law prohibiting actions on a range of topics. Although treaties may arguably be more important than CIL in modern international affairs,¹²³ CIL is still the subject of a great deal of international legal scholarship.¹²⁴ Qualitative or quantitative observational studies are, however, largely powerless in assessing the influence of CIL on state behavior, due to lack of variance in its applicability. In other words, since this source of law applies to every country, traditional research designs cannot determine whether CIL has played a role in many important changes in state behavior.

Although observational studies are generally unable to adequately deal with the lack of variance of CIL, experimental research

119. See Bradley & Gulati, *supra* note 91, at 205 (discussing Int’l Law Ass’n, Comm. on the Formation of Customary (Gen.) Int’l Law, Statement of Principles Applicable to the Formation of General Customary International Law 27 (2000)). There was a symposium on this issue in 21 DUKE J. COMP. & INT’L L. 1 (2010).

120. See Bradley & Gulati, *supra* note 91, at 211 (“Persistent objection must involve affirmative international communications, not mere silence or adherence to contrary laws or practices, and there are few examples of agreed-upon successful persistent objection.”). For a discussion of the theory behind the persistent objector doctrine, see Joel P. Trachtman, *Persistent Objectors, Cooperation, and the Utility of Customary International Law*, 21 DUKE J. COM. & INT’L L. 221 (2010).

121. See Bradley & Gulati, *supra* note 91, at 212–13 (discussing the “modern” view that there are a small subset of norms with special status); *id.* at 212–13, n.37 (quoting Vienna Convention on the Law of Treaties art. 53, May 23, 1969, 1155 U.N.T.S. 331, 344) (“For the purposes of the present Convention, a peremptory norm of general international law is a norm accepted and recognized by the international community of States as a whole as a norm from which no derogation is permitted and which can be modified only by a subsequent norm of general international law having the same character.”).

122. See Bradley & Gulati, *supra* note 91, at 212–13 (listing genocide, torture, and slavery as practices banned by CIL) (citing RESTATEMENT (THIRD) OF THE FOREIGN RELATIONS LAW OF THE UNITED STATES § 702 cmt. n (1987)).

123. Cf. Andrew T. Guzman, *Saving Customary International Law*, 27 MICH. J. INT’L L. 115, 119 (2005) (“modern international relations have made the treaty a more important tool, relative to CIL, than it has been in the past”). *But cf.* John J. Chung, *Customary International Law as Explained by Status Instead of Contract*, 37 N.C. J. INT’L L. & COM. REG. 609, 609 (2012) (“[CIL] forms the foundation of international law.”).

124. See Shaffer & Ginsburg, *supra* note 1, at 11–15 (discussing lines of research in CIL while arguing that there should be more empirical research on the topic).

can help establish whether it is at least plausible that customary international law is having an impact on state behavior. A number of experiments have sought to assess the impact of international law without references to specific treaties.¹²⁵ For example, Tonya Putnam and Jacob Shapiro conducted an experiment to evaluate whether individuals are more likely to support actions being taken against a state that violates human rights when they are informed those harms are violating international law.¹²⁶ As part of their experiment, they told some respondents nothing about international law, others that a country was violating a treaty, and still another group just that they were violating international law generally.¹²⁷ The results indicate that although information about international law changes public opinion, general references to international law have roughly the same effect on public opinion as references to signed treaties.¹²⁸ The result is particularly interesting, because it suggests that international law without a treaty—customary international law—may still have an influence on policy preferences and outcomes. Thus, experimental research can estimate whether CIL may matter, while observational studies cannot, given a lack of variance in the applicability of customary norms.

B. Narrow Ranges of Time for Analysis

To draw reliable inferences from the data in a large-*n* empirical study, it is necessary to analyze a sufficiently long time period to have a big enough sample of observations.¹²⁹ This can pose a problem for observational studies of international events because often only relatively short time frames can be analyzed. By giving the researcher greater control, for example by allowing the generation of new data, experimental methods can help overcome this obstacle.

Narrow time frames of analysis have posed a considerable problem for observational studies. Mindful of this concern, scholars of international phenomena (wars, alliances, trading patterns, etc.) of-

125. See, e.g., Putnam & Shapiro, *supra* note 27; Wallace, *supra* note 27.

126. See Putnam & Shapiro, *supra* note 27.

127. *Id.* at 15. It is worth noting that a fourth treatment group was told that there was a relevant treaty, but that the country had not signed it. *Id.*

128. *Id.* at 19.

129. Obviously there is great variation in the time frame that should be analyzed depending on the question being studied. There are excellent studies that are able to focus on a relatively few number of years and others that examine hundreds of years.

ten construct datasets that cover the entire post-war period,¹³⁰ the entire twentieth century,¹³¹ or even over 200 years.¹³² The approach of historically extending datasets, however, is often simply not available to scholars of international law. There are often very short windows when there is variation in the number of countries that have signed treaties. A related problem results if an insufficient number of events to study occurs during these windows. Experimental methods can help overcome both of these problems by exploring hypothetical scenarios or interning in the field to generate new data.

1. Short Windows Before Widespread Ratification

When researching the effect of international law, scholars are typically only able to analyze the period during which the relevant treaty, court, or institution was in effect. However, although it has ancient roots,¹³³ the corpus of international law has mostly formed relatively recently.¹³⁴ As a consequence, even in the best cases, scholars of international law are left studying the short time periods

130. See, e.g., James D. Fearon & David D. Laitin, *Ethnicity, Insurgency, and Civil War*, 97 AM. POL. SCI. REV. 75 (2003) (analyzing a dataset of civil war onset from 1945 to 1999).

131. See, e.g., Alexander B. Downes, *Restraint or Propellant? Democracy and Civilian Fatalities in Interstate Wars*, 51 J. CONFLICT RESOL. 872 (2007) (analyzing a dataset of interstate wars between 1900 and 2003 to assess whether democracies are more likely to kill civilians during wars).

132. See, e.g., Jason Lyall, *Do Democracies Make Inferior Counterinsurgents? Reassessing Democracy's Impact on War Outcomes and Duration*, 64 INT'L ORG. 167 (2010) (analyzing a dataset from 1800 to 2005 to study the effectiveness of democracies in fighting counterinsurgencies).

133. See, e.g., Gabriella Blum, *The Laws of War and the "Lesser Evil"*, 35 YALE J. INT'L L. 1, 8 (2010) ("'International Humanitarian Law' is a term of the past century alone even though the notion of regulating and limiting warfare is almost as ancient as wars themselves").

134. For example, much of human rights law emerged after World War II. See Jacob Katz Cogan, *The Regulatory Turn in International Law*, 52 HARV. INT'L L.J. 321, 322–23 (2011) ("While strands of such international human rights law date back hundreds of years in the protection of certain foreign nationals, such as diplomats, from state action or inaction deemed unlawful, its flowering would only begin to occur in the mid-twentieth century when, in the wake of World War II, international law's shelter extended fundamentally beyond that limited population to encompass a state's control of its own people.") (citations omitted). For documentation of the growth of international courts and tribunals during this period, see Cesare P.R. Romano, *A Taxonomy of International Rule of Law Institutions*, 2 J. INT'L DISP. SETTLEMENT 241 (2011).

during which international law has gone into effect.¹³⁵ This problem is exacerbated by the fact that it is often not prudent, or possible, to study the entire time that an international treaty or institution has been in effect. As was explained in PART III.A.,¹³⁶ analyzing causation requires the presence of variation, and many international treaties have been so widely adopted that there is not sufficient variation in their treatment.¹³⁷

A window during which there was sufficient variation in the adoption of the treaty could ameliorate this causation problem. For example, if there were several decades during which only half the world had ratified the CEDAW, it might be possible to compare the treatment of women during that period to understand whether the presence of the treaty changed behavior.¹³⁸ However, many recent international agreements are being quickly ratified by many states. To illustrate this point, we have collected data on when states ratified the human rights agreements and armed conflict treaties discussed in PART II.B.¹³⁹ This information is presented in Figure 3 & Figure 4.¹⁴⁰

135. See, e.g., Rachel Brewster & Adam Chilton, *Supplying Compliance: Domestic Sources of Trade Law & Policy* (Nov. 4, 2012) (working paper) (on file with authors) (analyzing United States compliance with WTO decisions from 1996 to 2010). *But see* Morrow, *supra* note 5 (looking at compliance with the laws of war for the entire twentieth century by using a series of different treaties as the “treatment”).

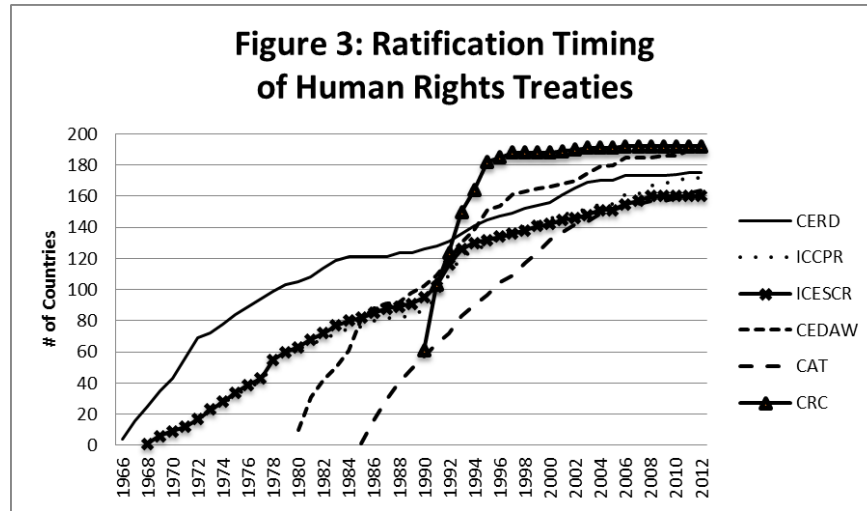
136. See *supra* text accompanying notes 88–128.

137. See *supra* Figure 1 & Figure 2.

138. This is the approach implicitly taken by Beth Simmons in her book analyzing the success of human rights agreements. See SIMMONS, *supra* note 4.

139. See *supra* notes 103–12.

140. For information on the sources used to collect data for these Figures, see *supra* note 101.



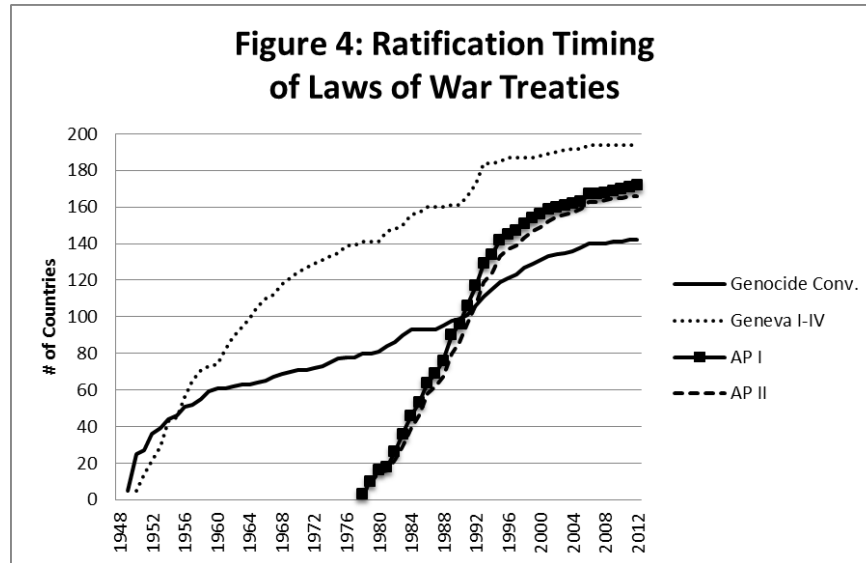
As the figures show, not only do these major agreements have many signatories, but several of the more recent treaties reached these signatory numbers in just a few decades. In the most extreme case, there were 150 countries that were parties to the Convention on the Rights of the Child within three years of when it went into effect in 1990. Similarly, there were 150 parties to Additional Protocol I to the Geneva Conventions twenty years after it opened for signature, and 142 parties to Additional Protocol II during that same time. This same trend emerged with a number of other major international agreements. For example, the Kyoto Protocol went into force on February 16, 2005.¹⁴¹ Today, there are 192 parties to the agreement.¹⁴² Similarly, the Rome Statute for the International Criminal Court (establishing the ICC) went into force on July 1, 2002.¹⁴³ Today, there are 122 members to the agreement.¹⁴⁴

141. See Kyoto Protocol to the United Nations Framework Convention on Climate Change, Dec. 11, 1997, 37 I.L.M. 22.

142. *Id.*

143. See Rome Statute of the International Criminal Court, July 17, 1998, U.N. Doc. A/CONF.183/9.

144. *Id.*



Although these issues present serious hurdles to research designs based on observational data, experimental methods can provide one way to overcome them. First, because experiments can use hypothetical vignettes that vary relevant facts about countries, researchers are not restricted to analyzing events that occurred in the time frame before widespread adoption of treaties. For example, as discussed in PART II.A., Mike Tomz conducted an experiment on the impact that ratification of the NPT has on threat perception, even though he did not conduct the experiment when there was still variance in which countries had signed the NPT.¹⁴⁵ Even if the amount of time before treaties are widely adopted grows shorter in the future, it can still be possible to see if the treaty regime can theoretically alter policy preferences and beliefs.

2. Laws Regulating Infrequent Events

The problem created by the short time frames between when treaties are negotiated and when they are widely adopted is exacerbated if the treaty regulates events that happen infrequently. An example is the Additional Protocol I to the Geneva Conventions of 1949, which went into effect in 1978.¹⁴⁶ This treaty is the most com-

145. See Tomz, *supra* note 27; see also *supra* text accompanying notes 115–17.

146. Protocol I, *supra* note 111.

prehensive attempt to regulate armed interstate conflicts.¹⁴⁷ By at least one count, there were only fourteen interstate wars before the treaty had over 160 signatures in 2003.¹⁴⁸ It thus is incredibly difficult to conduct large-*n* observational research on whether the treaty has helped change the behavior of states engaged in armed conflicts.

An advantage of experimental designs is that they can explore topics that have occurred relatively infrequently. For example, the aforementioned wide adoption of Additional Protocol I of the Geneva Conventions has certainly made it difficult to conclusively determine whether the law has influenced state behavior during conflict.¹⁴⁹ As a way to address this problem, author Chilton presented respondents with a hypothetical future conflict in which violations of the laws of war were occurring.¹⁵⁰ This experiment produced evidence that information on the status of the laws of war changes public opinion¹⁵¹ but suggested that the information does not have an additive effect over other similar arguments.¹⁵² Although the experiment does not conclude the debate on whether states change their behavior as a consequence of ratifying treaties on the laws of war, it does bring to the discussion new evidence simply unavailable through reexamining the observational data on the same set of conflicts. As these examples illustrate, experimental designs can help avoid the problems created by the short windows of time between when a treaty is ratified and when it is widely adopted.

147. See George H. Aldrich, *Prospects for United States Ratification of Additional Protocol I to the 1949 Geneva Conventions*, 85 AM. J. INT'L L. 1, 1 (1991) (“[Additional Protocol I] is the most important treaty codifying and developing international humanitarian law since the adoption of the four [Geneva] Conventions themselves; and it is the first such treaty since 1907 to deal with methods and means of warfare and the protection of the civilian population from the effects of warfare.”).

148. See Alexander B. Downes, Web Appendix for *How Smart and Tough Are Democracies? Reassessing Theories of Democratic Victory in War*, 33 INT'L SECURITY 9 (on file with author) (providing documentation for the wars used in his large empirical project to analyze the number of civilians killed during interstate wars).

149. Compare Morrow, *supra* note 5 (finding that democracies are likely to follow the laws of war when reciprocation is likely), with Valentino et al., *supra* note 5, (finding no evidence that states, including democracies, change their behavior as a result of the laws of war).

150. See Chilton, *supra* note 27.

151. *Id.* at 16–21.

152. *Id.* at 21–24. For an explanation of how experiments can be designed to test causal mechanisms, see Imai, Tingley & Yamamoto, *supra* note 33; Imai, Keele, Tingley & Yamamoto, *supra* note 33. For other examples of research using related approaches to test causal mechanisms, see Tingley & Tomz, *supra* note 24, and Tomz & Weeks, *supra* note 20.

C. Overlapping Legal Constraints

To conduct reliable causal analysis, researchers must also isolate the effects of competing causal claims. This requirement, however basic, is often very difficult to satisfy. For example, the difficulty may arise because two or more factors hypothesized to cause a given phenomenon may occur simultaneously: democracies may be less likely to fight each other because they have more transparent governments or because they are more likely to have intertwined economic relationships.¹⁵³ Since both of these factors are observed in the real world, it is difficult to know which (if either) makes democracies less likely to fight. Experimental methods overcome this problem by isolating which causal claims are tested.

Qualitative or quantitative observational studies assessing the impact of international law on state behavior often have difficulty untangling the effects of overlapping legal constraints. As the study of international law moves from analyzing whether states comply with the treaties and agreements they sign toward assessing whether those agreements make a causal impact on state behavior,¹⁵⁴ it is increasingly important for scholars to perform credible causal analysis by isolating effects of specific agreements.¹⁵⁵ To do so, scholars have focused on analyzing the impact of the ratification of an individual treaty¹⁵⁶—or participation in an international institution¹⁵⁷—on changes in policy.

1. Overlapping International Treaties

Isolating the causal effect of any individual treaty or institution, however, is complicated because countries are often bound by several international legal agreements with overlapping legal obligations and constraints. Given the rapid growth of international law in the last sixty years, most states are subject to overlapping international treaty obligations in many areas. For example, there were over 50,000 treaties on file with the UN Treaty System as of 2005.¹⁵⁸ The

153. See Tomz & Weeks, *supra* note 27 at 21–24.

154. See Posner, *supra* note 9.

155. Cf. Ho & Rubin, *supra* note 7, at 21–22 (explaining the importance of focusing on a single “treatment” when performing causal analysis).

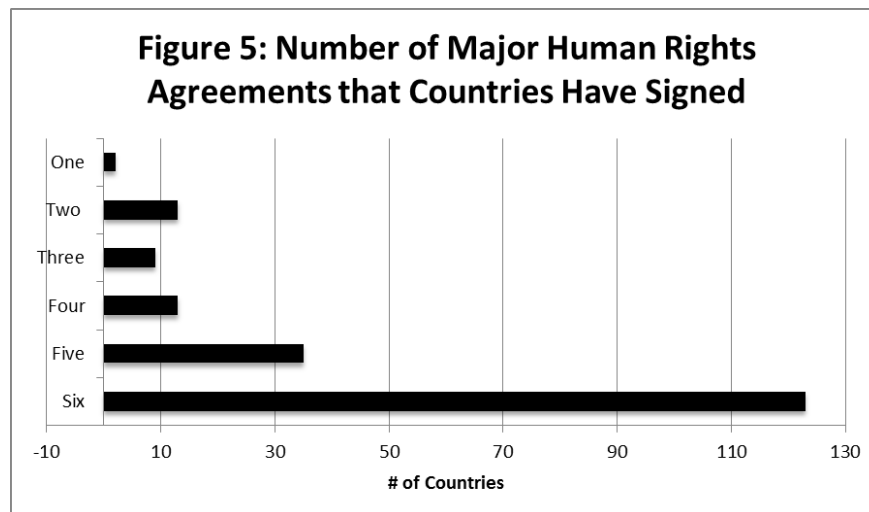
156. See, e.g., Hathaway, *supra* note 4.

157. See, e.g., Simmons, *supra* note 11.

158. Christopher J. Borgen, *Resolving Treaty Conflicts*, 37 GEO. WASH. INT’L L. REV. 573, 574 (2005).

result of this massive proliferation in international legal agreements has been that these agreements increasingly create conflicting and overlapping obligations.¹⁵⁹ As a result, it can be difficult to know whether any specific international law is having an effect; states may be complying with it because of a different international agreement, or failing to comply or remain compliant because of other conflicting international legal agreements.

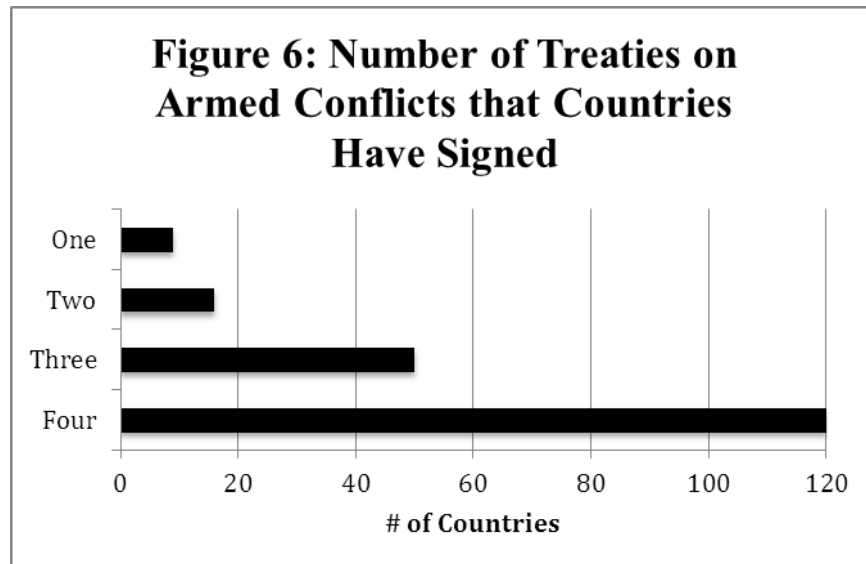
To illustrate, we present Figures 5 & Figure 6, showing how many of the treaties discussed in PART II.A.1 & PART II.B.1 states are party to. As Figure 5 shows, 123 states are party to all six major human rights treaties. Figure 6 shows that 125 states are party to the four treaties governing armed conflict. Significantly, these treaties contain common elements, and efforts that states take in response to one agreement may fall under the scope of another agreement.¹⁶⁰



159. See *id.* at 574 (“The very success of treaties as a policy tool has caused a new dilemma: a surfeit of treaties that often overlap and, with increasing frequency, conflict with one another.”). The increase in conflicting and overlapping international legal agreements has led the international legal community to study the impact of this “fragmentation.” See generally Study Group of the Int’l Law Comm’n, *Report on the Fragmentation of International Law: Difficulties Arising from the Diversification and Expansion of International Law*, Int’l Law Comm’n, U.N. Doc. A/CN.4/L.682 (Apr. 13, 2006) (by Martti Koskeniemi); see also Roger P. Alford, *The Proliferation of International Courts and Tribunals: International Adjudication in Ascendance*, 94 AM. SOC’Y INT’L L. PROC. 160 (2000).

160. For example, Article I of both the ICCPR and the ICESCR protect the “right of self-determination.” See ICCPR, *supra* note 103, at art. I; ICESCR, *supra* note 104, at art. I.

These constraints are often collinear, making it particularly difficult for observational studies to isolate the effect of a single treaty. Collinearity occurs when a variable of interest strongly correlates with another variable (or set of variables) that offer a potential explanation for a phenomenon.¹⁶¹ When this occurs, it is difficult—if not impossible—to tease out the causal effects of these strongly related factors.¹⁶² For example, when there are multiple treaties that seek to protect minority rights with similar ratification patterns, it might be impossible to tell which, if any, treaty is doing anything to protect minority rights. That is to say, if the set of countries that have signed agreement X is the same as that of countries that have signed agreement Y, the ratification of those two agreements is collinear, making it impossible to know the causal effect of either treaty on an outcome of interest.



Using experimental research designs can help researchers directly address the problems posed by the presence of overlapping legal constraints that confound observational studies. For one, experimental designs vary information that would be collinear in

161. As an informal definition, two variables can be said to be collinear when one variable can perfectly predict the other. See KING, KEOHANE & VERBA, *supra* note 88, at 122–24, 213–15, for a discussion on this problem.

162. See Tomz & Weeks, *supra* note 20, at 4–5, for an excellent discussion of the problem that collinearity poses for testing whether there is a democratic peace.

observational data. That is, even if nearly all countries that have ratified treaty X have also ratified treaty Y—making observational studies futile—that information can be varied during experimental research. For example, there has been a great deal of debate among scholars of international relations on whether there is a “democratic peace” (that is, are democracies less likely to fight wars with each other).¹⁶³ However, the presence of democracy is often collinear with other variables that might explain peace (i.e. shared political interests), hindering scholars using observational data.¹⁶⁴ To solve this problem, Mike Tomz and Jessica Weeks recently conducted a survey experiment that varied the information provided to respondents on a number of variables that had been hypothesized to both cause peace and be collinear by democracy.¹⁶⁵ This way, Tomz and Weeks provided new evidence that individuals are less supportive of war with countries that are democratic, even compared to autocracies similar in all relevant respects. Importantly, this example illustrates how it would be possible to use an experimental design that varies whether a state is party to different treaties, even if this distribution of ratification is not common in the actual world.¹⁶⁶

2. Overlapping Domestic Laws

A related problem is that states that have signed international legal agreements on a given topic may also have domestic laws on the same subjects. Although there has been a continual academic debate on degrees of difference between international and domestic law,¹⁶⁷ scholars have long recognized that countries often have inter-

163. See generally MICHAEL E. BROWN, SEAN M. LYNN-JONES & STEVEN E. MILLER, *DEBATING THE DEMOCRATIC PEACE* (1996).

164. See Henry S. Farber & Joanne Gowa, *Common Interests or Common Politics? Reinterpreting the Democratic Peace*, 59 J. POL. 393 (1997); see also JOANNE GOWA, *BALLOTS AND BULLETS: THE ELUSIVE DEMOCRATIC PEACE* (1999).

165. See Tomz & Weeks, *supra* note 20.

166. If the research design used hypothetical countries, then this would be possible to do without deceiving subjects. If, however, the researcher hopes to ask questions about specific countries, it might only be possible to utilize this approach if subjects are deceived about whether a country has ratified a specific treaty (or set of treaties). This introduces an ethical question that is central to experimental research. For a discussion on the issue of deceiving subjects, see *infra* text accompanying notes 266–68.

167. There has obviously been an incredible amount of scholarship over the years analyzing the relationship between international law and domestic law. For just a few recent examples, see, Oona Hathaway & Scott J. Shapiro, *Outcasting: Enforcement in Domestic and International Law*, 121 YALE L.J. 252 (2011) (arguing that international law is enforced

national laws and domestic laws that cover the same subjects.¹⁶⁸ This could either be because countries select into international legal agreements that do not impose onerous new requirements,¹⁶⁹ or because countries quickly change their domestic laws to mirror the international legal agreements they have signed.¹⁷⁰ In either scenario, the presence of two laws covering the same topic makes it difficult to causally assess the degree to which the international law is responsible for changes in state behavior.¹⁷¹ This phenomenon will increase if international law continues regulating more areas previously solely in the sphere of national regulation.¹⁷²

through the process of denying violators of law from the benefits of membership in a community, while domestic law is primarily—although not exclusively—enforced through a monopoly on the use of force within a territory); Jack Goldsmith & Daryl Levinson, *Laws for States: International Law, Constitutional Law, Public Law*, 122 HARV. L. REV. 1791 (2009) (analyzing the similarities and differences between domestic constitutional law and international law); Christopher A. Whytock, *Thinking Beyond the Domestic-International Divide: Toward a Unified Concept of Public Law*, 36 GEO. J. INT'L L. 155 (2004) (arguing that we should move away from the “structural/functional” divide in the study of international and domestic law).

168. Cf. Karen Knop, Ralf Michaels & Annelise Riles, *International Law in Domestic Courts: A Conflict of Laws Approach*, 103 AM. SOC'Y INT'L L. PROC. 269 (2009) (arguing that the relationship between international law and domestic law should be understood through a conflict of laws approach).

169. See *supra* PART II.E; see generally Downs, Locke & Barsoom, *supra* note 10, at 382–87 (arguing that countries may only sign international legal agreements when they do not constitute a major change from what the country would have done in the absence of the agreement).

170. For a number of excellent case studies documenting countries that have changed their laws to incorporate international legal agreements they have signed, see SIMMONS, *supra* note 4. For example, Simmons documents how Japan passed a number of laws to expand the rights of women in the workplace following the country becoming a party to the CEDAW. *Id.* at 237–45.

171. Of course, if signing an international legal agreement were to result in changes in domestic law it would still be evidence of the causal effect of international law. Cf. Shaffer & Ginsburg, *supra* note 1, at 15 (reviewing a previous debate between Simmons and Von Stein and concluding that anticipatory changes to currency regimes to join the IMF still can be viewed as evidence of the effect of the IMF's legal rules on state policy). The difficulty, however, is that with large n-observational studies, it can be difficult to be confident of this causal link without more in depth research.

172. See generally Jacob Katz Cogan, *The Regulatory Turn in International Law*, 52 HARV. INT'L L.J. 321 (2011) (arguing that international law has moved from just placing obligations on states to direct regulation of individuals, corporations, and other actors); see also Anne-Marie Slaughter & William Burke-White, *The Future of International Law is Domestic (or, The European Way of Law)*, 47 HARV. INT'L L.J. 327 (2006) (arguing that international law must increasingly move from regulation of nations to direct engagement with domestic institutions).

In addition to the problem that exists when these agreements are collinear—as discussed in the last section—observational studies have difficulty accounting for all of a state’s legal obligations. This requires a huge amount of country-level research, or expertise, that is impracticable for most large-*n* studies. Although it may be possible to account for all of a country’s treaty commitments in a statistical model in some cases,¹⁷³ or even for a country’s constitutional laws,¹⁷⁴ having a model that incorporates data on countries’ domestic laws on specific topics may be an all but impossible task. Instead, scholars have included variables that account for characteristics of countries’ legal regimes—like the degree to which a country is democratic—without including terms to account for all relevant legal obligations.¹⁷⁵

Experimental designs can also help test whether international law has an effect beyond domestic law without needing huge amounts of country-specific data. This is because experiments can test whether international law has an “additive” effect—as opposed to a “substitute” effect—on mass or elite opinion beyond the presence of domestic law or other international agreements.¹⁷⁶ Imagine, for instance, a survey measuring public support for the war on terror, with one group of respondents told that torture violates domestic laws, a second one told that torture violates a signed international treaty, and a third told that torture violates both. If public opinion on

173. See Yonatan Lupu, *The Informative Power of Treaty Commitments: Using the Spatial Model to Address Selection Effects*, AM. J. POL. SCI. (forthcoming 2013), available at <http://dss.ucsd.edu/~ylupu/Informative%20Power.pdf> (last visited Jan. 13, 2013) (using countries’ history of ratifying treaties to perform statistical matching as a means of controlling for selection effects).

174. See ZACHARY ELKINS, TOM GINSBURG, & JAMES MELTON, *THE ENDURANCE OF NATIONAL CONSTITUTIONS* (2009) (exploring a huge amount of data collected on countries’ current and historical national constitutions). See also David S. Law & Mila Versteeg, *The Declining Influence of the United States Constitution*, 87 N.Y.U. L. REV. 762 (2012) (analyzing data on the world’s constitutions to assess the influence of the United States Constitution on elements of constitutions adopted by other countries).

175. For example, in perhaps the most careful empirical analysis of the effect that human rights agreements have on domestic laws, Simmons analyzed whether countries that had ratified specific treaties were more likely to respect the human rights covered by those agreements. See SIMMONS, *supra* note 4. Although her models include a huge number of variables, they do not include terms of other related treaties or the presence of other relevant domestic laws.

176. For a discussion of the difference between “additive” and “substitute” effects as it relates to international law, see Tomz, *supra* note 27, at 19–21. See also Chilton, *supra* note 27 (using an experimental research design that tests if the laws of war have an additive effect beyond moral arguments against targeting civilians).

torture were to be the same in all groups (Group 1 = Group 2 = Group 3), then international law would here have a “substitute” effect.¹⁷⁷ However, if information on international law changed opinion beyond information on domestic law alone (Group 3 > Group 1), then international law could have an “additive” effect.¹⁷⁸ This kind of experimental design makes it possible to test whether international law might influence policy outcomes, even when domestic law already exists on a given topic.¹⁷⁹

D. Inadequate Dependent Variables

Being able to accurately measure a dependent variable, or outcome, is a basic requirement of efficient causal analysis.¹⁸⁰ For example, a scholar studying whether states comply with the laws of war might use civilian deaths during a given conflict as a dependent variable.¹⁸¹ This often poses problems for observational studies, because relevant dependent variables may be difficult, if not impossible, to measure. Experimental methods, on the other hand, often have the advantage of generating the dependent variable of interest itself.

As we have discussed, as scholarship on international law has become more empirical,¹⁸² scholars have shifted their attention to trying to understand whether international law causes states to change their behavior.¹⁸³ To quantitatively analyze whether states change their behavior to become compliant with a given treaty as a consequence of committing to that agreement, however, researchers must have a dependent variable that measures compliance.¹⁸⁴ This founda-

177. That is to say that the information on international law merely substituted the effect that was created by simply being told about domestic law.

178. That is to say that the information on international law added to the change in public opinion beyond the change in public opinion created by domestic law alone.

179. It is worth noting that the same approach can be used to test whether two international legal agreements might have a larger effect than a single international legal agreement.

180. See Chilton, *supra* note 27, at 77. For a longer discussion of dependent variables, see *id.* at 107–09.

181. See, e.g., Valentino et al., *supra* note 5.

182. See generally Shaffer & Ginsburg, *supra* note 1.

183. See Posner, *supra* note 9; see also Shaffer & Ginsburg, *supra* note 1, at 1.

184. Cf. KING, KEOHANE & VERBA, *supra* note 88, at 109 (“[I]n social science, we must be careful to ensure that we are really interested in understanding our dependent variable, rather than the background factors that our research design holds constant”).

tional requirement poses two significant problems for researchers studying international law compliance: first, existing data sources might not be directly interpretable as a measure of compliance; and second, in many cases, observational data of compliance might not exist, or even be possible to collect.

1. Existing Data Sources Do Not Accurately Measure Compliance

In some cases, scholars have been able to identify research questions where it is possible to directly observe and measure non-compliance. For example, in a groundbreaking article on compliance with international law, Beth Simmons analyzed whether states comply with commitments they have made under the IMF Articles of Agreement by leaving their current accounts free from restriction.¹⁸⁵ In this case, Simmons was able to use a dependent variable that was a direct and clear measure of non-compliance.¹⁸⁶

Things are not always so straightforward. Instead, available sources of data may not provide easily interpretable measures of compliance or non-compliance. One criticism that has been raised along these lines is that scholars have often used dependent variables that may not perfectly map to whether a state is compliant with an international legal obligation. For example, in her book on compliance with human rights, Beth Simmons uses a binary variable for states that have free practice of religion.¹⁸⁷ Eric Posner has criticized this measure, however, for not necessarily measuring compliance with the ICCPR.¹⁸⁸

A related problem occurs when the dependent variables that are used are not binary—as in the two examples just discussed from Simmons’ research—but instead are on some kind of scale.¹⁸⁹ The difficulty in these cases is that even if a state changes its behavior after ratifying an international treaty, and there is movement on the dependent variable, it is unclear when “compliance” has occurred. As

185. Simmons, *supra* note 11.

186. *Id.* at 833.

187. See SIMMONS, *supra* note 4, at 386.

188. Posner, *supra* note 9, at 6–7.

189. See, e.g., Yonatan Lupu, *Limited Constraints: Veto Players and the Effects of International Human Rights Agreements* (2012) (working paper) (on file with author) (using the measure of human rights protection from the CIRI project as a dependent variable for compliance with treaty obligations); see also Moonhawk Kim, Yvonne M. Dutton, & Cody Eldredge, *Why Ratify? Reservations, Institutional Changes, and Commitment to Human Rights Treaties* (2012) (working paper) (on file with author).

a result, existing empirical studies of compliance with international law may open themselves to criticism for using dependent variables that fail to fully capture compliance.¹⁹⁰

A third problem is that the best available dependent variables for use might not measure compliance directly at all, but instead be proxies for consequences of compliance. For example, when analyzing whether states change their policies after ratifying the Convention on Political Rights of Women, Oona Hathaway uses the percentage of men in a country's legislature as the dependent variable.¹⁹¹ As Hathaway admits, however, equal representation is not required by the convention. The dependent variable is thus a proxy for compliance, making it debatable whether it adequately captures compliance.¹⁹²

Experimental methods offer a way to directly solve all three of these problems. That is, experiments can directly generate evidence of whether a state complies with international law. For example, Baradaran et al. recently conducted possibly the first such field experiment on compliance with international law.¹⁹³ The researchers sent requests for information on incorporating a shell company to 1,015 firms offering incorporation services in 182 countries.¹⁹⁴ As the experimental treatment, they randomized whether firms were given any facts on the status of international law.¹⁹⁵ To comply with international financial transparency laws, firms must demand proof of identity before providing incorporation services.¹⁹⁶ By carefully wording the requests, the researchers were able to directly measure whether firms were willing to violate international law in their responses.¹⁹⁷ This experimental approach bypassed complaints of existing data sources inadequately measuring international law compliance, because the responses to the experiment itself were evidence of

190. See, e.g., Goodman & Jinks, *supra* note 12, at 173–78 (criticizing Hathaway, *supra* note 4, for using dependent variables that do not fully capture compliance with human rights agreements).

191. Hathaway, *supra* note 4, at 1975–76.

192. For other criticisms about Hathaway's selection of dependent variables, see Goodman & Jinks, *supra* note 12, at 174 for an argument that Hathaway's selection of dependent variables do not "account for strategies governments often adopt in response to improved enforcement of a norm."

193. Baradaran et al., *supra* note 7, at 7–8.

194. *Id.* at 37.

195. *Id.* at 38–39.

196. *Id.* at 43.

197. *Id.* at 42–44.

non-compliance. In other words, the dependent variable was a good measure of compliance, because it was itself the original, direct measure.

2. Data on Compliance Is Not Available

In addition to existing data sources not being ideal measures of non-compliance, sometimes data on whether compliance is occurring is altogether unavailable. When dependent variables do not exist, scholars must collect original data on state treaty compliance. Notably, despite the effort it involves, original data collection is an incredibly valuable service. It is a worthwhile step; the data should be made public more often.¹⁹⁸

That said, original data collection on compliance with international law is not always possible. First, it may require extensive research into the practices of many countries. For example, if field research is required to determine whether a state complies with an international agreement, it is challenging for a researcher to visit enough countries to conduct a large-*n* study. Second, data on compliance with a particular international agreement may not be publicly available. For example, it may be impossible to collect accurate data on whether states comply with the Convention Against Torture.¹⁹⁹ Moreover, even if it is possible to find some national-level data, there might be considerable unaccounted-for subnational variation.

In addition to these obstacles, scholars using empirical methods to research compliance with international law overlook the fact that international treaties often guarantee the protection of a large number of rights.²⁰⁰ As a result, even if a dependent variable is available to measure one commitment contained within the treaty, it might not be plausible to find observational data that measures all of the steps required for a country to be compliant.

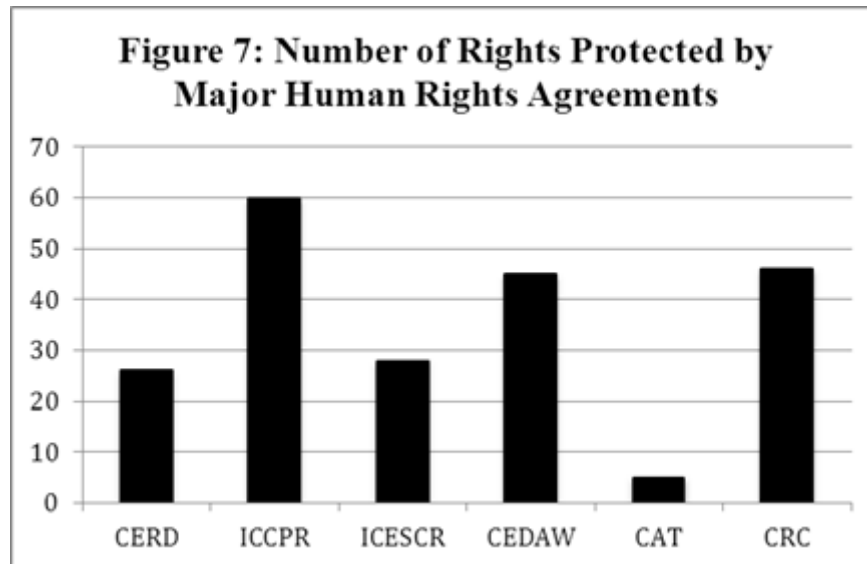
To illustrate, Figure 7 presents the number of individual rights

198. See Gary King, *Ensuring the Data-Rich Future of the Social Sciences*, 331 *SCI. MAG.* 719, 720 (2011) (discussing the value of posting and sharing data). For an example of a recent international law article that engaged in original data collection for the project's dependent variable, see Brewster & Chilton, *supra* note 135, where the authors collect data on the dates on which the United States took steps to comply with adverse rulings in the WTO.

199. See CAT, *supra* note 107.

200. See Posner, *supra* note 9, at 7 for a criticism of the use of dependent variables in Simmons, *supra* note 4, because human rights treaties "collectively contain dozens or maybe even hundreds of provisions."

protected in the six major human rights treaties.²⁰¹ As it shows, the average number of rights protected in the agreements is 35. The ICCPR protects 60 different rights, ranging from the right to not be imprisoned due to debt²⁰² to the right to marry and found a family.²⁰³ As a result, when dependent variable data is only available for a few of the rights within such a treaty, it may be difficult to reliably test whether states comply.



Experiments provide an appealing alternative to observational studies when it is not possible to collect data on compliance. In addition to using the above-discussed approach of Baradaran et al. to collect direct evidence of non-compliance,²⁰⁴ experimental methods can also test the theoretical mechanisms presented for how a specific treaty might influence behavior. For example, in a recent study, Geoffrey Wallace set out to examine whether the Convention Against Torture²⁰⁵ might have an influence on America's use of torture in the war on terror.²⁰⁶ Wallace studied whether international law exerts an independent influence on state behavior by directly examining one theoretical mechanism by which this might occur: altering domestic

201. See sources cited *supra* notes 103–08.

202. See ICCPR, *supra* note 103, at art. 11.

203. *Id.* at art. 23.

204. See *supra* text accompanying notes 193–97.

205. CAT, *supra* note 107.

206. Wallace, *supra* note 27.

mass political views.²⁰⁷ Wallace conducted two survey experiments on national samples of American adults, and the experimental treatment was that he varied the information about international law before gauging respondent views on the acceptability of torture.²⁰⁸ Wallace's experiment showed that international law changes support for the use of torture by roughly six percentage points, meaning international law might change policies on torture without having to have a dependent variable that directly measures torture itself.²⁰⁹

E. Selection Bias

Another barrier to reliable causal inference—and one that has received the most attention in the international law context—is selection bias.²¹⁰ Selection bias occurs when the observations that have received a particular treatment are systematically related to a potential outcome.²¹¹ For example, if a researcher were attempting to estimate the effect of taking a test preparation course on LSAT scores, producing a reliable causal estimate may be complicated by the fact that the students who elect to take such a course (receive the treatment) may be more conscientious students, who were likely to score higher (the potential outcome) than other students, regardless of enrollment. Experimental research helps solve this problem because subjects are randomly assigned to treatments, instead of self-selecting into them.

Although at least one other scholar has explained how experiments can help overcome the problems caused by selection bias in international law,²¹² we will also explain briefly. In the study of international law, producing reliable causal estimates of international agreements and institutions has been frustrated, because countries

207. *Id.* at 106–07.

208. *Id.* at 117–19.

209. *Id.* at 119.

210. *See generally* Hafner-Burton et al., *supra* note 1, at 89–90 (summarizing research on the influence of selection effects on compliance with international law). For an early and prominent example of scholarship on international law drawing attention to selection effects, see Downs, Rocke & Barsoom, *supra* note 10, which argues that previous claims that international law is generally complied with ignore a variety of barriers to inference, including selection effects.

211. For a clear statement of the definition of selection bias, see ALAN S. GERBER & DONALD P. GREEN, *FIELD EXPERIMENTS: DESIGN, ANALYSIS, AND INTERPRETATION* 37–39 (2012). *See also* King, Keohane & Verba, *supra* note 88, at 135–37.

212. *See* Tomz, *supra* note 27, at 11.

choosing to participate in international legal regimes may be systematically different than those that are not.²¹³ This reality is a major problem for observational studies on international law²¹⁴ that experimental research can help solve.

1. Selection Bias & International Treaties

The issue that has received perhaps the most attention from scholars of international law is whether states comply with international treaties that they have consented to.²¹⁵ Although scholars were initially primarily concerned with examining whether states comply with agreements, in recent years research has evolved to focus more on whether states alter their behavior as a consequence of committing to international legal agreements.²¹⁶ The difficulty, however, is that whether a given state chooses to commit to an international treaty is not random.²¹⁷ Instead, there are a number of strategic calculations that determine whether states choose to agree to international treaties.²¹⁸ As a result, it is reasonable to think that the states that select into international treaties are systematically different from those states that do not. This selection bias thus presents an inferential problem making it difficult for international law scholars to study the effects of international agreements.²¹⁹

213. See, e.g., von Stein, *supra* note 12 (discussing how selection bias called the causal effects of previous research into question). But see Simmons & Hopkins, *supra* note 12 (responding to many of von Stein's criticisms).

214. But cf. Hafner-Burton et al., *supra* note 1, at 89 n.231 (citing studies that have used "sophisticated methods" like instrumental regression and matching to obtain valid inferences on the effects of international law).

215. Cf., e.g., Simmons, *supra* note 1.

216. See Posner, *supra* note 9, at 5 (distinguishing between "compliance" and "causation" approaches). Compare Chayes & Chayes, *supra* note 8 (discussing how states have generally high levels of compliance with the international agreements they consent to), with SIMMONS, *supra* note 4 (empirically testing whether committing to international human rights agreements alters those states' human rights practices).

217. For a discussion of this issue, i.e., the inferential problem caused by the fact that states do not randomly choose to sign treaties, see Tomz, *supra* note 27, at 7–8.

218. See, e.g., Geoffrey P.R. Wallace, *Regulating Conflict: Historical Legacies and State Commitment to the Laws of War*, 8 FOREIGN POL'Y ANALYSIS 151 (2012) (examining why states commit to treaties on the laws of war); see also Oona A. Hathaway, *Why Do Nations Join Human Rights Treaties?*, 51 J. CONFLICT RES. 588 (2007) (examining why states consent to international human rights treaties); SIMMONS, *supra* note 4, at 57–111 (developing a theory of why states commit to international human rights agreements).

219. See, e.g., von Stein, *supra* note 12 (discussing the problems posed by selection

Experimental methods can help to solve this problem because they allow researchers to randomize whether participants are told that a country has ratified a particular treaty.²²⁰ For example, political scientists Tingley & Tomz recently tested Americans' willingness to take retaliatory actions against a country that has increased its consumption of fossil fuels.²²¹ Half of the respondents were told that the country had said it would not increase its use of fossil fuels; the other half were told that the country had signed a treaty promising that it would not do so.²²² The respondents that were told that the country had signed a treaty were 14% more likely to support economic sanctions against the country than those that were simply told the country said it would not do so (51% compared to 37%).²²³ Since the assignment of the experimental treatment (being told a treaty had been signed) was randomized, the survey experiment suggests that leaders may have incentives to treat countries that have signed international agreements differently than those that have not. If the authors had instead simply observed US responses to polluting countries that had signed treaties, it would be difficult to draw reliable inferences on the influence of the treaty because countries that signed it may have been systematically different from those that did not. These differences could be directly related to the dependent variable—for example, the states that ratified the treaty may be the same states that US citizens are more willing to take actions against in general—and thus induce bias. Experimental methods thus help to overcome selection bias problems and directly measure the causal inference of treatments.

2. Selection Bias & International Litigation

Relatedly, litigation in international courts suffers from selection bias.²²⁴ Simply put, disputes that parties choose to litigate until a

bias for Simmons, *supra* note 11). It is important to note that scholars have begun to try to address these problems using a number of sophisticated statistical methods. See *supra* text accompanying notes 13–14; see also Hafner-Burton et al., *supra* note 1, at 89 n.231.

220. See, e.g., Tomz, *supra* note 27 (using this approach in what we believe to be the first experimental test of whether information on the status of international law changes policy views).

221. Tingley & Tomz, *supra* note 27.

222. *Id.* at 26–27.

223. *Id.* at 28.

224. See, e.g., Eric A. Posner & Miguel F. P. de Figueiredo, *Is the International Court of Justice Biased?*, 34 J. LEGAL STUD. 599, 614 (2005) (discussing how selection bias causes a problem for studying litigation in the ICJ because the cases that make it to the ICJ may not be representative of all possible disputes); see also Geoffrey Garrett, R. Daniel Kelemen &

judicial decision is reached are likely to be systematically different from other cases. This is not just true of international law, but of litigation generally.²²⁵ For example, disputes that reach a judgment may be systematically different from the universe of cases because easy cases are settled and hard cases are not, or because it may not be economically feasible to pursue cases when small amounts of money are at stake. Regardless of the reason, the implication is that it is difficult to draw inferences about the effects of litigation on the settlement of disputes.

Once again, experimental research can help overcome these selection biases. For example, even if observational research of scholars interested in whether countries change their policies after WTO disputes²²⁶ showed that countries changed their policies as a result of adverse WTO decisions, this would not necessarily be a result of any adverse WTO decision. Instead, states may only bring complaints in situations when they believed a country would be responsive to an adverse ruling, as when the issue is not too politically sensitive. Experimental methods could be used to test this by, for instance, presenting a hypothetical trade dispute and randomizing whether respondents are told a foreign country has strongly alleged violations of international trade law, or whether the WTO has issued a decision supporting those allegations. This would allow a researcher to directly test whether the litigation itself could influence opinions—and in turn policy responses by democratically accountable officials—or whether the nature of the dispute and allegation itself has the same effect. As a result, experimental methods also present one way to overcome the inferential problem posed by selection bias in international litigation.

3. Selection Bias & Institutional Design

An additional selection problem common in the study of international law, along with treaty ratification and international litiga-

Heiner Schulz, *The European Court of Justice, National Governments, and Legal Integration in the European Union*, 52 INT'L ORG. 149, 151–52 (1998) (discussing their efforts to develop a case identification strategy that will help minimize selection bias).

225. Cf. Anna Harvey & Barry Friedman, *Ducking Trouble: Congressionally Induced Selection Bias in the Supreme Court's Agenda*, 71 J. POL. 574 (2009) (empirically testing how the selection bias introduced by the fact that the Supreme Court has discretionary jurisdiction influences the Court's agenda).

226. See, e.g., Brewster & Chilton, *supra* note 135 (analyzing United States' compliance with adverse WTO decisions).

tion, is institutional design. The influence that the designs of international institutions have on compliance rates and policy outcomes has been a major line of inquiry for international law and international relations scholars.²²⁷ These design features, however, are not necessarily exogenous to joining international institutions. For example, the dispute settlement mechanism in a particular institution may be designed to encourage or discourage participation based on ex-ante compliance rates. In other words, features of international institutions are not randomly selected, making it difficult to study their efficacy using observational data. Still, many international relations and law scholars are interested in the consequences of institutional design choices.

This is yet another problem that experimental methods can help to address. Experiments can be designed in a way that makes the particular features of international institutions exogenous. For example, in the same way that Tomz tested whether being told that a country had ratified the NPT changed the views of British MPs,²²⁸ it would be possible to design an experiment where the willingness of decision makers to utilize a particular international institution is influenced by being randomly told information on a relevant feature of the design of that institution.²²⁹

III. CONDUCTING EXPERIMENTAL RESEARCH

As we have argued, scholars of international law should begin to embrace experimental methods. Although experimental research methods are not appropriate ways to study every question, they are underutilized. Increasing their use, however, requires that international legal scholars and political scientists interested in international law understand in concrete terms how to design, field, and interpret experiments.

227. For a selection of research on this topic, see *DELEGATION AND AGENCY IN INTERNATIONAL ORGANIZATIONS* (Darren G. Hawkins, et al. eds., 2006). For a prominent article on the institutional design of international organizations, see Barbara Koremenos, Charles Lipton & Duncan Snidal, *The Rational Design of International Institutions*, 55 INT'L ORG. 761 (2001).

228. See *infra* text accompanying notes 282–83.

229. As a hypothetical example, a scholar may be interested to know whether the presence of amicus briefs in a given dispute settlement mechanism affects the likelihood that countries will pursue litigation through that process. To answer this question, this feature of an institution could be randomly manipulated in a survey given to government officials that forced them to consider whether they should pursue litigation in a particular case.

Although formal training in experimental methods may be the best way to gain the knowledge and skills required to conduct experimental research, it may not always be possible. Given this reality, we provide here a brief overview of some issues that researchers hoping to conduct experimental research should consider. Importantly, researchers interested in conducting experimental research should first consult one of the many excellent textbooks on experimental methodology.²³⁰ Second, it might be fruitful for legal scholars to collaborate with researchers trained in experimental methodology.²³¹ Third, any scholar considering conducting experimental research should make sure to obtain the required research permits from their university's Institutional Review Board (IRB).²³² Finally, researchers should carefully consider what type of experimental design would be most appropriate for the question that they are hoping to study.

A common schema of experiment types breaks experiments into three categories: laboratory, survey, and field.²³³ Laboratory experiments range considerably in what they entail, but share a focus on experimentation in highly controlled laboratory settings.²³⁴ Survey experiments embed manipulations and randomized frames into standard public opinion surveys.²³⁵ Field experiments turn to naturally occurring behavior, but manipulate features of the environment in order to make inferences.²³⁶ A fourth category of experiments is natural experiments. Natural experiments take advantage of naturally occurring randomizations to make inferences,²³⁷ such as arbitrarily

230. For good textbook introductions to experimental methods, see MORTON & WILLIAMS, *supra* note 36 (providing an overview on how to conduct experimental research); GERBER & GREEN, *supra* note 211 (providing an overview specifically on field experiments).

231. For one example of collaboration between political scientists and a legal scholar, see Baradaran et al., *supra* note 7.

232. Like many types of research, experimental research requires approval by University Institutional Review Boards (IRB). IRBs protect both subjects and researchers. Typically this process entails submitting an application describing the research and identifying any potential harms. Specific procedures vary somewhat across institutions and so we direct researchers to their own IRB for more information.

233. See James N. Druckman, Donald P. Green, James H. Kuklinski & Arthur Lupia, *Experimentation in Political Science*, in CAMBRIDGE HANDBOOK OF EXPERIMENTAL POLITICAL SCIENCE 3, 6 (James N. Druckman et al. eds., 2011) (“[M]ost experiments have been implemented in one of three contexts: laboratories, surveys, and the field.”).

234. *Id.* at 6-7.

235. *Id.* at 7.

236. *Id.*

237. For guidance on finding natural experiment, see Gregory Robinson, John E. McNulty & Jonathan S. Krasno, *Observing the Counterfactual? The Search for Political*

drawn legal borders,²³⁸ or the distribution of international election monitors during elections.²³⁹ Although natural experiments can be used in legal research,²⁴⁰ we will not consider them in depth in this article, focusing instead on experimental methods where the researcher controls the randomization and other experimental processes. We suspect that survey and field experiments will be used more likely in studies of international law than laboratory experiments,²⁴¹ but we discuss all for completeness. We also discuss how issues of randomization, as well as challenges to internal,²⁴² external,²⁴³ and construct validity²⁴⁴ affect each type of experimentation.

In this part, we provide a minimal foundation in experimental methods and point scholars to further resources. We specifically discuss the virtues and vices of different experimental techniques with respect to international law. Depending on the research questions or

Experiments in Nature, 17 POL. ANAL. 341 (2009). For additional advice on establishing whether an event can be considered a natural experiment, see Jasjeet S. Sekhon & Rocio Titiunik, *When Natural Experiments Are Neither Natural Nor Experiments*, 106 AM. POL. SCI. REV. 35 (2012) (providing a framework for thinking through assignment and other concerns that arise when considering if an event is a natural experiment).

238. See, e.g., Daniel N. Posner, *The Political Salience of Cultural Differences: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi*, 98 AM. POL. SCI. REV. 530 (2004) (exploiting a natural experiment created by the border between Zambia and Malawi to argue that political salience of cultural cleavages does not depend on the nature of the cleavages, but instead on the relative size of the group).

239. See Susan D. Hyde, *The Observer Effect in International Politics: Evidence from a Natural Experiment*, 60 WORLD POL. 37 (2007) (exploiting a natural experiment created during the 2003 Armenian presidential election to show that election monitors can help reduce fraud at the polling stations they visit).

240. See, e.g., Huang, *supra* note 77 (using a natural experiment to study the impact of increased workload on reversal rates in the federal judiciary); Berry & Gersen, *supra* note 77 (using a natural experiment to study the impact of election timing on public policy outcomes); see also Benjamin A. Lindy, Note, *The Impact of Teacher Collective Bargaining Laws of Student Achievement: Evidence from a New Mexico Natural Experiment*, 120 YALE L.J. 1130 (2011); Bebachuk et al., *supra* note 80.

241. See, e.g., Hafner-Burton et al., *supra* note 26 (conducting an experiment in a laboratory setting to test the behavioral influences of international cooperation).

242. Internal validity is defined as: “[t]he approximate truth of the inference or knowledge claim within a target population studied.” MORTON & WILLIAMS, *supra* note 36, at 188.

243. External validity is defined as: “[t]he approximate truth of the inference or knowledge claim for observations beyond the target population studied.” *Id.*

244. Construct validity is defined as: “[w]hether the inferences from the data are valid for the theory (or constructs) the researcher is evaluating in a theory testing experiment.” *Id.* at 189.

context, some types of experiments will be more applicable or easier to apply. For example, it may be easier to use survey experiments to study how international law affects public opinion,²⁴⁵ but it may be more appropriate to use a field experiment to study how private firms and public officials respond to information on the status of international law.²⁴⁶ We also update scholars of international law on new developments within the experimental social sciences. One exciting area of growth, not surprisingly, is conducting experiments via the Internet.²⁴⁷ We also discuss pertinent methodological innovations in the design and analysis of experiments, and how these relate to the study of international law.

A. Laboratory Experiments

Laboratory experiments use a physical space where experimental subjects respond to carefully controlled stimuli and/or interact with other subjects.²⁴⁸ In the social sciences, psychology uses laboratory experiments the most.²⁴⁹ In recent years, laboratory experiments have become more common in economics and political science. Most laboratory experiments use subjects from either a pre-existing pool or by targeted recruiting of subjects with specific characteristics (like gender or work experience).²⁵⁰ Many psychology departments have laboratories with subject pools, and some universities have broader social science laboratories.²⁵¹ Experimental ses-

245. See e.g., Wallace, *supra* note 27 (conducting a survey experiment to evaluate how information on the legalization of international law affects public opinion on the use of torture in the war on terror).

246. See e.g., Findley et al., *supra* note 25 (conducting a field experiment to evaluate whether private firms violate international law when responding to requests to incorporate shell companies).

247. See *infra* text accompanying notes 278–79.

248. For general background on the use of laboratory experiments in political science, see Shanto Iyenger, *Laboratory Experiments in Political Science*, in CAMBRIDGE HANDBOOK OF EXPERIMENTAL POLITICAL SCIENCE 126–55 (James N. Druckman et al. eds., 2011). For an overview of the use of laboratory experiments to study political economy, see Thomas R. Palfrey, *Laboratory Experiments in Political Economy*, 12 ANN. REV. POL. SCI. 379 (2009).

249. Cf. Falk & Heckman, *supra* note 35, at 535 (“With the exception of psychology, the adoption of laboratory experiments has been much slower in the social sciences, although during the past two decades use of lab experiments has accelerated.”).

250. For an extended discussion on subject recruitment, focusing on the use of students as subjects, see MORTON & WILLIAMS, *supra* note 36, at 237–58.

251. For example, Harvard University runs the Harvard Decision Science Laboratory (HDSL) (<http://decisionlab.harvard.edu/>), Princeton University runs the Princeton

sions begin with a consent and instructional process, record data through computer interfaces, paper and pencil, or experimenter observation, and end with some form of compensation and debriefing. Compensation often takes the form of monetary compensation or class credit.²⁵²

Laboratory experiments are useful for a number of reasons.²⁵³ The ability to precisely control the experimental environment means that effects are most likely due to the experimental intervention.²⁵⁴ Confounding variables or uncontrolled sources of stimuli are minimized. For example, let us imagine one were interested in whether international law had any impact on perceptions by members of the public about the legality of their own government's policies. A laboratory experiment could randomly expose individuals to scripts or videos about either the entailed international law or some neutral control condition, and then ask questions about views on domestic policies. Were this intervention not randomized, and we just correlated awareness of international law with preferences, then any number of variables might commonly cause both.

The controls afforded in laboratory experiments let researchers create more variation than what would exist naturally in the real world. This is helpful for precisely some of the reasons discussed in PART II.A (i.e., observational studies have difficulty accounting for certain treaties being essentially universally adopted).²⁵⁵ In the real world of international law, we believe that X has an impact on Y. But the variation in X is quite small, making it difficult, perhaps impossible, to detect such a relationship. By manipulating X in a hypothetical scenario, we can begin to explore this relationship.

Laboratory experiments also enable clear and direct measurement of subject decision-making or responses to stimuli. Ambi-

Laboratory for Experimental Social Science (PLESS) (<http://pless.princeton.edu/>), and New York University runs the Center for Experimental Social Science (CESS) (<http://cess.nyu.edu/>). Additionally, a group of social scientists at the University of California San Diego have recently established the Laboratory on International Law and Regulation (ILAR) (<http://ilar.ucsd.edu/about/>).

252. For a discussion on compensating subjects, and implications that it may have on experimental validity, see MORTON & WILLIAMS, *supra* note 36, at 259–91.

253. For a general discussion on the advantages of laboratory experiments, see MORTON & WILLIAMS, *supra* note 36, at 225–26, 305–07.

254. For a discussion on the advantage that laboratory experiments afford for controlling variation and the experimental environment, see Falk & Heckman, *supra* note 35, at 535.

255. *See supra* PART II.A.1.

guity about what is being measured is minimized. Part of this comes from the sterility of laboratory environments. We can remove as many potential confounding stimuli as possible. Some experimenters in American politics have even sought to replicate the consumption environments their subjects are used to. For example, Ansolabhere and Iyengar, who were interested in the effects of television on political attitudes, recreated a prototypical living room replete with couch.²⁵⁶ In some social science experiments, researchers have begun to draw on measurement techniques used in the medical sciences²⁵⁷ such as galvanic skin response²⁵⁸ and functional magnetic resonance imaging,²⁵⁹ or measuring subject-specific physical attributes, like genetic profiles.²⁶⁰ Such physiological recordings would be difficult outside of the laboratory.

The decision to use a laboratory experiment depends on the type of stimulus/situation under investigation. As a general rule, when greater control is necessary, laboratory experiments are necessary. These needs can arise in different ways. Some studies require extremely fine-grained measurement, both in terms of content and timing. For example, the implicit association test (IAT) measures prejudices and biases by comparing reaction times to different stimuli at the millisecond level.²⁶¹ Greater control is also required when re-

256. STEPHEN ANSOLABEHERE & SHANTO IYENGAR, *GOING NEGATIVE: HOW POLITICAL ADVERTISEMENTS SHRINK & POLARIZE THE ELECTORATE* (1997).

257. For an early justification of this approach, see Albert F. Ax, *Goals and Methods of Psychophysiology*, 1 *PSYCHOPHYSIOLOGY* 8 (1964).

258. See, e.g., Dustin Tingley, Jooa Julia Lee & Jonathan Renshon, *Physiological Responses to Shifting Bargaining Power: Micro-Foundations of Commitment Problems in International Politics* (2012) (working paper), available at <http://scholar.harvard.edu/files/dtingley/files/bargaininganxietyfinal.pdf> (last visited January 23, 2013) (using skin conductance reactivity as a measure of arousal in the automatic nervous system to test emotional responses to changes in bargaining power). See also Claudia Civai, Corrado Corradi-Dell'Acqua, Matthias Gamer, & Raffaella Rumiati, *Are Irrational Reactions to Unfairness Truly Emotionally-Driven? Dissociated Behavioral and Emotional Responses in the Ultimatum Game Task*, 114 *COGNITION* 89 (2010).

259. See, e.g., Dustin Tingley, *Neurological Imaging as Evidence in Political Science: A Review, Critique, and Guiding Assessment*, 45 *SOC. SCI. INFO.* 5 (2006) (reviewing the use of neurological imaging in political science research).

260. See, e.g., Rose McDermott, Dustin Tingley, Jonathan Cowden, Giovanni Frazzetto & Dominic D. Johnson, *Monoamine Oxidase a Gene (MAOA) Predicts Behavioral Aggression Following Provocation*, 106 *PROC. NAT'L ACAD. SCI.* 2118 (2009); James H. Fowler & Christopher T. Dawes, *In Defense of Genopolitics*, 107 *AM. POL. SCI. REV.* 1 (2013); James H. Fowler & Darren Schreiber, *Biology, Politics, and the Emerging Science of Human Nature*, 322 *SCIENCE* 912 (2008).

261. Anthony G. Greenwald, Debbie E. McGhee & Jordan L. K. Schwartz, *Measuring*

searchers want to make sure that subjects are focused on their task and not distracted by other events that could compromise the manipulation. Decreases in focus on the experimental task can lead to decreases in experimental effects, and even bias estimates, if the level of focus differs across experimental conditions.

Laboratory experiments also let researchers to allow for interactions between subjects. For example, nearly all experiments testing game theoretic models are implemented in the laboratory,²⁶² though this is changing with the advent of web-based platforms.²⁶³ In these experiments, the crucial quantities that are controlled by the experimenter correspond to components of game theoretic models, including the roles/positions of subjects, the information subjects have, the strategies available to subjects, and the payoffs for reaching different outcomes. By carefully manipulating one or more of these parameters, experimenters can compare predictions from formal game theoretic models to actual human behavior.²⁶⁴ Experiments of this nature could be important in testing views on international law emphasizing strategic interaction between decision-makers.²⁶⁵

While experiments with deception remain common in psychology, most economic labs eschew or formally prohibit deception on the grounds that it could contaminate future experiments that rely on subjects not suspecting any deception.²⁶⁶ There exists considerable debate on this subject.²⁶⁷ In his classes on experimental political science, the author Tingley suggests avoiding deception if possible,

Individual Differences in Implicit Cognition: The Implicit Association Test, 74 J. PERSONALITY & SOC. PSYCHOL. 1464 (1998).

262. DOUGLAS D. DAVIS & CHARLES A. HOLT, EXPERIMENTAL ECONOMICS (1992).

263. John J. Horton, David G. Rand & Richard J. Zeckhauser, *The Online Laboratory: Conducting Experiments in a Real Labor Market*, 14 EXPERIMENTAL ECON. 399 (2011). For an early account of the emerging use of the Internet as a platform of laboratory experiments, see Alison I. Piper, *Conducting Social Science Laboratory Experiments on the World Wide Web*, 20 LIB. & INFO. SCI. RES. 5 (1998).

264. Andrew Shotton, Strong and Wrong: *The Use of Rational Choice Theory in Experimental Economics*, 18 J. THEORETICAL POL. 498 (2006); Tingley, *supra* note 51; Dustin Tingley & Barbara F. Walter, *The Effect of Repeated Play on Reputation Building: An Experimental Approach*, 65 INT'L ORG. 343 (2011).

265. See, e.g., Engel, *supra* note 26 (using a laboratory experiment to show how expectations help shape how customary international law develops).

266. See DAVIS & HOLT, *supra* note 262, at 24 n.28.

267. Shane Bonetti, *Experimental Economics and Deception*, 19 J. ECON. PSYCHOL. 377 (1998); Eric Dickson, *Economics versus Psychology Experiments: Stylization, Incentives, and Deception*, in CAMBRIDGE HANDBOOK OF EXPERIMENTAL POLITICAL SCIENCE 58 (Druckman, Green, Kuklinski & Lupia, eds., 2011).

but not letting it get in the way of interesting research questions if one can identify a laboratory that will allow its use.²⁶⁸

Laboratory experiments face a series of challenges.²⁶⁹ Perhaps the biggest are construct and external validity, since most laboratory experiments involve convenience samples of college students in highly artificial settings (though this is changing).²⁷⁰ Simply put, a laboratory setting may not aptly capture the relevant real world situation. Another challenge is that experimenters may induce bias by creating demand effects.²⁷¹ Given the highly controlled setting, experimenters may inadvertently encourage subjects to make choices that conform to the hypotheses under investigation: not because the subjects actually want to make such a choice, but because they want to conform to what the experimenter wants. Ultimately, these limitations are the flip side of advantages.²⁷² Careful control of the experimental setting helps establish causal relationships and explore relationships impossible to examine in the real world, but which are important.

Although laboratory experiments certainly could be a viable method of studying questions of international law, the validity challenges that laboratory experiments present may make them less popular compared to other methods. In the international law context, much of the appeal of experimental methods is that they present a way to rigorously test existing theories of how individuals and organizations respond to international legal requirements. The further removed that tests are from the actual operation of international law, however, the less likely they are to produce evidence convincing skeptical scholars or policymakers. Of course, such concerns should always be balanced against the advantages of laboratory-based experiments, including their often superior ability in controlling the data generating process.

B. Survey Experiments

A survey experiment involves administering surveys to indi-

268. For an extended discussion on the ethical implications of conducting experiments that involve deception, see MORTON & WILLIAMS, *supra* note 36, at 500–21. See also McDermott, *supra* note 40, at 41.

269. See McDermott, *supra* note 40, at 40–41 (summarizing the challenges and disadvantages of experimental research).

270. *Id.*

271. *Id.* at 33–34.

272. *Id.* at 38–39.

viduals, with various types of randomizations.²⁷³ These can include randomized textual prompts/frames, question wordings, response options, or other stimuli like video feeds or audio clips.²⁷⁴ These manipulations let researchers explore how survey responses differ across different settings. Additional, non-experimental, questions allow researchers to have additional subject-specific covariates.²⁷⁵ Survey experiments can be delivered via professional polling firms²⁷⁶ or researcher-created surveys using online platforms.²⁷⁷ Survey experiments can utilize nationally representative samples via the use of probability weighting or other techniques, or convenience samples using subject recruitment platforms like Amazon's Mechanical Turk.²⁷⁸ Survey experiments are an efficient means for researchers to

273. See generally Brian J. Gaines, James H. Kuklinski & Paul J. Quirk, *The Logic of the Survey Experiment Reexamined*, 15 POL. ANALYSIS 1 (2007); Martin Gilens, *An Anatomy of Survey-Based Experiments*, in NAVIGATING PUBLIC OPINION: POLLS, POLICY, AND THE FUTURE OF AMERICAN DEMOCRACY 232 (Manza et al., eds., 2002). For advice on designing survey experiments, see Yusaku Horiuchi, Kosuke Imai & Naoko Taniguchi, *Designing and Analyzing Randomized Experiments: Application to a Japanese Election Survey Experiment*, 51 AM. J. POL. SCI. 669 (2007).

274. For an example of using video clips as an experimental treatment, see Jonathan Renshon, Joa Julia Lee, & Dustin Tingley, *Physiological Arousal and Political Beliefs*, POLITICAL PSYCHOLOGY (forthcoming 2014).

275. For example, researchers are able to ask all respondents basic biographical questions about themselves. This may include: age, gender, ethnic background, educational level, political ideology, partisan political affiliations, or income level. This information can then later be used to either subset the sample to examine differences between groups, or it can alternatively be used as covariates in a multivariate regression to control for any remaining imbalance that might exist within the control and treated experimental groups.

276. Two examples of professional polling firms that are popular with social scientists are Knowledge Networks (<http://www.knowledgenetworks.com/>) and Polimetrix/YouGov (<http://research.yougov.com/>). Both of these firms can provide help in creating surveys, and then fielding them to respondents through a variety of means for a fee.

277. One example of an easy to use online platform that researchers can use to create survey experiments is Qualtrics (<https://www.qualtrics.com/>). Researchers are able to use Qualtrics to write their own survey experiments. The platform allows users to select from a range of question types when developing their experiments. Moreover, the platform allows for a range of randomization options that are essential for experimental research. This includes randomizing elements of individual questions, which questions respondents receive, or the order in which questions are presented. After experiments have been drafted, researchers are provided with a link that they can use to direct respondents to take the experiments through various means of subject recruitment (e.g. Amazon's Mechanical Turk).

278. For information on conducting experiments using Amazon's Mechanical Turk service, see generally Winter Mason & Siddharth Suri, *Conducting Behavioral Research on Amazon's Mechanical Turk*, 44 BEHAV. RES. METHODS 1 (2012); Gabriele Paolacci, Jesse Chandler & Panagiotis G. Ipeirotis, *Running Experiments on Amazon Mechanical Turk*, 5

understand how individuals respond to information and stimuli, often with relatively larger sample sizes than laboratory experiments.²⁷⁹ Many experiments conducted in a laboratory can also be surveys, though their success or feasibility will depend on the experiment's relative level of complication as well as whether interactivity between subjects, or dynamic responses from the experimenter, are required.

An example of a survey experiment on international law comes from the discussed research by Tomz,²⁸⁰ which we briefly review here. In the experiment, respondents took a survey that asked a question about support for a foreign policy of prohibiting trade with Burma.²⁸¹ All respondents were randomly assigned various pro and con arguments about the policy. One of these randomly assigned pieces of information stated that the United States had signed an international treaty prohibiting it from breaking off trade with Burma under international law. Tomz found that individuals receiving information about international law were 17% more likely to oppose

JUDGMENT & DECISION MAKING 411 (2010).

279. There is a growing body of evidence showing that the experimental results produced by using Amazon's Mechanical Turk are the same as other experimental methods. See generally Adam J. Berinsky, Gregory A. Huber & Gabriel S. Lenz, *Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk*, 20 POL. ANALYSIS 351 (2012). For similar research evaluating the use of web-based experiments for psychological research, see Laura Germine, Ken Nakayama, Bradley C. Duchaine, Christopher F. Chabris, Garga Chatterjee & Jeremy B. Wilmer, *Is the Web as Good as the Lab? Comparable Performance from Web and Lab in Cognitive/Perceptual Experiments*, 19 PSYCHONOMIC BULL. REV. 847 (2012). For research documenting the performance of Mechanical Turk for economic research, see Horton et al., *supra* note 263 (providing a review of Mechanical Turk for economic experiments and successfully replicating previous laboratory experiments). Additionally, experiments conducted using Mechanical Turk have been published, or are forthcoming, in leading political science journals. See, e.g., Gregory A. Huber, Seth J. Hill & Gabriel S. Lenz, *Sources of Bias in Retrospective Decision Making: Experimental Evidence on Voters' Limitations in Controlling Incumbents*, 106 AM. POL. SCI. REV. 720 (2012) (using Mechanical Turk to conduct experimental games to show that participants were susceptible to biases when retroactively assessing overall incumbent performance); Kevin Arceneaux, *Cognitive Biases and the Strength of Political Arguments*, 56 AM. J. POL. SCI. 271 (2012) (using Mechanical Turk to conduct an experiment testing whether cognitive biases influence participants' views of the strength of political arguments). See also Tomz & Weeks, *supra* note 20 (using Mechanical Turk to test hypothesis on the democratic peace); Tingley & Tomz, *supra* note 27 (using Mechanical Turk to conduct an experiment testing how cooperation influences participants' views on actions to address climate change).

280. See Tomz, *supra* note 27; see also *supra* text accompanying notes 115–16.

281. Tomz, *supra* note 27, at 13–21.

restricting trade with Burma.²⁸² Tomz also reports a follow-up study using British members of Parliament.²⁸³ MP's were provided with information about a country that may or may not be pursuing nuclear weapons. The various bits of information were all akin to what one would see in a standard intelligence report. In the treatment group, the country was listed as having signed the Nuclear Non-Proliferation Treaty (NPT) while in the control group the country had not signed the NPT. The MP's were asked whether they thought the country was pursuing nuclear weapons. The key finding was that MP's in the treatment condition reported lower expectations that the country was building weapons than those in the control group.²⁸⁴ Hence the effect of international law on its own—not factors that might change the propensity of signing on to international law—had an effect on elite decision-maker expectations. These expectations could, in principle, drive subsequent behavior.

Beyond the content of what is being manipulated, survey experiments come in a variety of different forms. A common form is a framing experiment. Here, a particular issue is described in several different ways and a respondent answers questions about the issue.²⁸⁵ Another form of survey experiment is a “conjoint” survey, popular in marketing research.²⁸⁶ In this type of survey, subjects are asked to evaluate a product, or some other quantity like an international agreement,²⁸⁷ but attributes of the product are randomized with the subject evaluating the same quantity but under multiple different profiles of attributes. Reactions to different attribute profiles enable collecting much more information than when individuals only evaluate one attribute profile. Finally, survey experiments can test the mechanisms hypothetically linking the experimental treatment/frame with a respondent's stated preferences. Here, researchers ask additional questions that measure changes in these intermediate variables.²⁸⁸

282. *Id.* at 18.

283. *Id.* at 24–28.

284. *Id.* at 27.

285. See, e.g., Dennis Chong & James N. Druckman, *Framing Public Opinion in Competitive Democracies*, 101 AM. POL. SCI. REV. 637 (2007).

286. See generally Paul E. Green & Vithala R. Rao, *Conjoint Measurement for Quantifying Judgmental Data*, 8 J. MARKETING RES. 355 (1971).

287. See, e.g., Michael Bechtel & Kenneth Scheve, *Public Support for Global Climate Cooperation* (2012) (working paper), available at http://iicas.ucsd.edu/_files/papers/pia/Bechtel-2012.pdf (last visited Jan. 21, 2013).

288. For guidance on how to design experiments in a way that allows mediation analysis to be conducted, see generally Kosuke Imai, Luke Keele, Dustin Tingley & Teppei Yamamoto, *Unpacking the Black Box of Causality: Learning About Causal Mechanisms*

For example, Tomz and Weeks use a survey experiment to explore the connection between political regime type and willingness to go to war.²⁸⁹ After providing information about the potential adversary's regime type they also ask questions about the perceived costs of the conflict, its morality, and other mediating variables, which are hypothesized to then have an impact on support for military intervention.²⁹⁰

As discussed before, survey experiments provide a cost effective way to collect a substantial amount of data. Such experiments, though, face many of the same problems that standard surveys face: creating representative samples can be costly, creating panels to track individuals over time is difficult, and subjects might give socially desirable answers rather than their true opinions.²⁹¹

In addition to these general problems, survey experiments face some specific problems.²⁹² If there are multiple experimental manipulations in a survey, then there can be spillover effects when

from *Experiments and Observational Studies*, 105 AM. POL. SCI. REV. 765 (2011). See also Kosuke Imai, Dustin Tingley & Teppei Yamamoto, *Experimental Designs for Identifying Causal Mechanisms*, 176 J. ROYAL STAT. SOC'Y 5 (2013). For examples of experiments using these methods, see Tingley & Tomz, *supra* note 27; Tomz & Weeks, *supra* note 20.

289. Tomz & Weeks, *supra* note 20.

290. *Id.* at 19–28.

291. It is worth noting that there is an area of research concerned with survey experiments on sensitive subjects. This line of research is specifically focused on developing ways to survey respondents so that they are willing to reveal preferences that they may be unwilling to share with researchers when directly asked (i.e., when there is a stigma associated with the view). See generally Adam N. Glynn, *What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment*, 77 PUB. OPINION Q. 159 (2013), for advice on constructing surveys that use list experiments to allow for the research of sensitive topics. See also Graeme Blair & Kosuke Imai, *Statistical Analysis of List Experiments*, 20 POL. ANALYSIS 47 (2012). For an example of research using these methods, see Jason Lyall, Graeme Blair, & Kosuke Imai, *Explaining Support for Combatants During Wartime: A Survey Experiment in Afghanistan*, AM. POL. SCI. REV. (forthcoming). For resources and more information on conducting survey experiments on sensitive topics, see Graeme Blair, *Methods and Software for Sensitive Survey Questions*, <http://www.princeton.edu/~gblair/sensitive.html> (last visited Jan. 24, 2013).

292. Many of the problems that we discuss can be at least mitigated by careful survey design. For advice on survey design, see, for example, Josh Pasek & Jon A. Krosnick, *Optimizing Survey Questionnaire Design in Political Science*, in THE OXFORD HANDBOOK OF AMERICAN ELECTIONS AND POLITICAL BEHAVIOR 27 (Jan E. Leighley ed., 2010). See also Chris Tausanovitch & Christopher Warshaw, *How Should We Choose Survey Questions to Measure Citizens' Policy Preferences?*, (Feb. 2012) (working paper), available at http://cwarshaw.scripts.mit.edu/papers/MeasuringPreferences_Feb142012.pdf (last visited Jan. 23, 2013).

treatment status in one part of the survey impacts responses to a different experiment.²⁹³ Construct validity concerns can also be an issue. Online surveys that ask questions about politics can only approximately simulate considerations about politics that interest the researcher. Here, use of imbedded media, such as video clips, can be of great help. Finally, there is the question of external validity. We might detect relationships in our survey experiments that, were the same things to happen in the real world, there would be no change in behavior.²⁹⁴ This might be the case for a variety of reasons, including the fact that, in a survey experiment, the respondent's attention is (ideally) exclusively on the survey. The complexities of the real world might not allow this, or there might be competing considerations in the real world that the survey experiment has abstracted away.²⁹⁵

Survey experiments are proving to be a promising way to study international law.²⁹⁶ Many scholars have previously theorized that states may comply with international legal commitments when there are no threats of external enforcement (like in the human rights context) because the ratification of international legal agreements gives domestic political actors a new tool.²⁹⁷ One of those tools is that public opinion may change as a result of prior legal commitments. Thus, using survey experiments to test whether information on the status of international law changes public opinion provides a direct way to test the plausibility of important theories of international law.

C. Field Experiments

Laboratory and survey experiments typically investigate causal relationships in environments where the participants are not actively engaged in behavior that represents what the researcher is studying. This can pose problems for the types of inferences that are

293. See generally John E. Transue, Daniel J. Lee, & John H. Aldrich, *Treatment Spillover Effects Across Survey Experiments*, 17 POL. ANALYSIS 143 (2009).

294. See generally Jason Barabas & Jennifer Jerit, *Are Survey Experiments Externally Valid?*, 104 AM. POL. SCI. REV. 226 (2010).

295. See generally Paul M. Sniderman, *The Logic and Design of the Survey Experiment: An Autobiography of a Methodological Innovation*, in CAMBRIDGE HANDBOOK OF EXPERIMENTAL POLITICAL SCIENCE 102 (Druckman, Green, Kuklinski, & Lupia eds., 2011).

296. See, e.g., Wallace, *supra* note 27.

297. See, e.g., Simmons, *supra* note 4.

drawn from them. Field experiments minimize some of these concerns by examining behavior in the setting in which it would naturally take place.²⁹⁸ We suspect that most scholars of international law will gravitate towards field experiments because of the realism they afford and connectedness to behavior directly relevant to international law.²⁹⁹

While field experiments are done “in the field,” they nevertheless attempt to retain many of the redeeming qualities of other forms of experiments. The assignment of treatment conditions is randomized and dependent variables are clearly defined using previously identified behavior. What can be more difficult is what is known as compliance:³⁰⁰ while one can randomize the treatment conditions, it is sometimes difficult to ensure that the condition an individual is placed in is the one in which they actually participate. This might arise from conscious subterfuge or something more innocuous. In these settings, researchers can leverage econometric techniques like instrumental variable estimation³⁰¹ or report alternative quantities of interest, such as the intention to treat (ITT).³⁰²

Field experiments have been remarkably successful in the study of American politics, especially the study of voter turnout.³⁰³ In recent years there also have been field experiments that directly speak to topics of interest to international law scholars.³⁰⁴ One prom-

298. For a comprehensive guide on how to conduct field experiments, see GERBER & GREEN, *supra* note 211.

299. As we have previously noted, to date there has only been one field experiment conducted that has directly studied international law. See Findley et al., *supra* note 25. That said, field experiments are becoming increasingly popular in international relations generally, and the study of political economy and development specifically, for exactly this reason. For a survey of how field experiments are being used to research the political economy of development, see Macartan Humphreys & Jeremy M. Weinstein, *Field Experiments and the Political Economy of Development*, 12 ANN. REV. POL. SCI. 367 (2009); see also Hyde, *supra* note 53.

300. See generally GERBER & GREEN, *supra* note 211, at 131–210. See also MORTON & WILLIAMS, *supra* note 36, at 116.

301. See MORTON & WILLIAMS, *supra* note 36, at 106–08.

302. For more information on using ITT as an alternative quantity of interest, see GERBER & GREEN, *supra* note 211, at 141–43.

303. See generally GERBER & GREEN, *supra* note 18, for a summary of experimental research on voter turnout. For what is perhaps the most prominent individual example of a field experiment on voter turnout, see Gerber et al., *supra* note 43, which conducts a large-scale field experiment on voters in Michigan to show that social pressure can increase voter turnout.

304. For a general overview of how field experiments can be used to study international

inent example is the influence of international observers on election fraud.³⁰⁵ In the first field experiment on democracy promotion, political scientist Susan Hyde conducted a large-scale experiment to determine how international observers influenced voting during the 2004 Indonesian presidential election.³⁰⁶ Hyde worked with the Carter Center to gain permission to randomly assign to which villages their teams of election monitors would be sent.³⁰⁷ These teams then monitored the voting conduct at 147 individual polling stations.³⁰⁸ After the election, Hyde compared the results of randomly monitored polling places with the overall election results, and counterintuitively showed that the incumbent candidate actually did better at polling stations that received election monitoring.³⁰⁹ Hyde's experiment provides concrete evidence that election monitors can change voting patterns; that large international organizations may be willing to work with researchers to introduce scientific experimentation into their work; and that well reasoned hypotheses (like the belief that monitors would help the challenger) may not bear out after being tested in the field.

Another example clearly engaging with international law issues is the forthcoming work by Findley et al.³¹⁰ As previously mentioned, the researchers conducted a field experiment by sending emails to 1,264 firms providing incorporation services, expressing an interest in anonymously incorporating a shell corporation.³¹¹ The experiment randomly varied whether information on the status of international law was included. Firms that responded but failed to request information on the senders' identity were in violation of international law. The experiment thus not only directly measured

relations, see Susan D. Hyde, *The Future of Field Experiments in International Relations*, 628 ANNALS AM. ACAD. POL. & SOC. SCI. 72 (2010).

305. Another area where field experiments have been used to study international relations is in the study of public goods provision. See, e.g., James Habyarimana, Macartan Humphreys, Daniel N. Posner, & Jeremy M. Weinstein, *Why Does Ethnic Diversity Undermine Public Goods Provision?*, 101 AM. POL. SCI. REV. 709 (2007); James D. Fearon, Macartan Humphreys, & Jeremy M. Weinstein, *Can Development Aid Contribute to Social Cohesion After Civil War? Evidence from a Field Experiment in Post-Conflict Liberia*, 99 AM. ECON. REV. 287 (2009).

306. See Hyde, *supra* note 53.

307. *Id.* at 516–17.

308. *Id.* at 517.

309. *Id.* at 517–20.

310. See *supra* text accompanying notes 193–97.

311. Baradaran et al., *supra* note 7, at 7.

compliance, but also was a rare example of studying compliance with international law by private actors.

From a practical standpoint, the work by Findley et al. showcases how to study the reaction of commercial entities via a simple intervention: email. Similar designs that solicit a response from “real” decision makers, using email or mail,³¹² offer a practical means

312. See, e.g., David E. Broockman, *Black Politicians Are More Intrinsically Motivated to Advance Blacks' Interests: A Field Experiment Manipulating Political Incentives*, 57 AM. J. POL. SCI. 521 (2013), for a finding that black state legislators were more likely to respond to a request for assistance from a putatively black citizen that did not reside in their district than white state legislators; see also Daniel M. Butler, Christopher F. Karpowitz, & Jeremy C. Pope, *A Field Experiment on Legislators' Home Styles: Service Versus Policy*, 74 J. POL. SCI. 474 (2012), for a finding that legislatures are more likely to respond to letters with service concerns than policy concerns; Gwyenth McClendon, *Co-ethnicity and Democratic Governance: A Field Experiment with South African Politicians*, (2012) (working paper) (on file with authors), for a finding that South African politicians were more likely to respond to requests from a fictional constituent that shares their ethnic background; Daniel M. Butler & David E. Broockman, *Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators*, 55 AM. J. POL. SCI. 463 (2011), for evidence that legislators are less likely to respond to putatively black constituents' letters with requests for help voting; Daniel M. Butler & David W. Nickerson, *Can Learning Constituency Opinion Affect How Legislators Vote? Results from a Field Experiment*, 6 Q. J. POL. SCI. 55 (2011) (for a finding that sending letters providing state legislators with information from a public opinion survey on their constituents' opinions affected their voting); Daniel Butler, *Monitoring Bureaucratic Compliance: Using Field Experiments to Improve Governance*, PUBLIC SECTOR DIGEST 41 (2010) (for a finding that high school principals that were sent letters informing them of their obligation to provide students with information on voter registration were more likely to do so. Sending letters or emails can also be an effective strategy for performing experiments on actors that are not officials, but still may influence policy.); see, e.g., Daniel M. Butler & Emily Schofield, *Were Newspapers More Interested in Pro-Obama Letters to the Editor in 2008? Evidence from a Field Experiment*, 38 AM. POL. RES. 356 (2010) (for a finding that letter submissions to the editor at 100 newspapers yielded a determination that pro-McCain letters received more interest. It is worth noting that it is possible to conduct field experiments on decision makers without using mail or email); see, e.g., Edmund Malesky, Paul Schuler, & Anh Tran, *The Adverse Effects of Sunshine: A Field Experiment on Legislative Transparency in an Authoritarian Assembly*, 106 AM. POL. SCI. REV. 762 (2012) (for a finding of no evidence that increased transparency impacts delegate performance in authoritarian Vietnam after working with a newspaper to randomly create websites devoted to the activities of a randomly selected group of delegates). For an example of non-experimental international law research that used email to gather data by conducting human rights monitoring groups from around the world, see Cosette Creamer & Beth Simmons, *Transparency at Home: How Well Do Governments Share Human Rights Information with Citizens?*, in *TRANSPARENCY IN INTERNATIONAL LAW* (Andrea Bianchi & Anne Peters eds., 2013). It is also worth mentioning that experimenting on public officials raises additional ethical considerations beyond those already present in experimental research. For a thorough treatment of this issue, see Gwyenth H. McClendon, *Ethics of Using Public Officials as Field Experiment Subjects*, 3 EXPERIMENTAL POL. SCIENTIST 13 (2012).

of engagement without the high expenses of direct face-to-face contact necessary in other field experiment designs. Of course, one must be careful in interpreting any results because in many settings there might be no response. People might not write back, either because an email is intercepted in a spam box or because they are alienated by the content of the correspondence. However, even a non-response might be thought of as a quantity of interest.³¹³ Needless to say, practical considerations of this sort loom large in the design, conduct, and analysis of field experiments. No experimental methodology is a perfect solution.

These examples show that field experiments are a promising but under-utilized method of studying international law. Well-designed field experiments directly test existing theories in a way that has the potential to generate evidence convincing to both scholars and policymakers. This is not to say that designing and conducting field experiments on international law is not a difficult task, but instead that the potential payoff for doing so may be quite high.

With that fact in mind, when deciding between experimental methods, researchers must consider whether they are interested in studying mass or elite opinion and decision-making. Obviously, survey experiments present a particularly easy way to gauge mass opinion.³¹⁴ Researchers of international law might be particularly interested in measuring mass opinion: First, there are strong theoretical reasons to believe that compliance with international law may be driven by changes in domestic political opinions that result from making international commitments.³¹⁵ Second, there is research suggesting that elite opinions mirror mass opinions on questions of international affairs, and thus that mass surveys can be an excellent

313. See, e.g., Baradaran et al., *supra* note 7, at 45–48, for an analysis of the response rate for emails sent to incorporation firms with request to create shell corporations as a quantity of interest. Of course, non-response issues can also arise during survey research. For a discussion of this issue, see Adam J. Berinsky, *Survey Non-Response*, in *HANDBOOK OF PUBLIC OPINION RESEARCH* 309 (Wolfgang Donsbach & Michael W. Traugott eds., 2008).

314. For more information on how to cheaply conduct survey experiments, see *supra* notes 278–79.

315. See generally SIMMONS, *supra* note 4, at 125–55, (developing a theory of compliance with international human rights law based on domestic politics). See also Raustiala & Slaughter, *International Law, International Relations, and Compliance*, in *HANDBOOK OF INTERNATIONAL RELATIONS*, 538, 547 (Thomas Risse and Beth Simmons eds., 2002) (arguing that the relationship between international law and state behavior may depend in part on domestic institutions); Xinyuan Dai, *Why Comply? The Domestic Constituency Mechanism*, 59 *INT'L ORG.* 363 (2005).

way to predict elite views.³¹⁶ If scholars wish to directly study elite behavior, they will likely either have to conduct a survey of a sample of elites,³¹⁷ or conduct further field experiments. As we have mentioned, field experiments have recently been used in exciting ways to directly test the views and responses of elites, including public officials,³¹⁸ and may be a promising way forward for scholars of international law.

CONCLUSION

Scholars of international law often devote their time to studying some of the world's most important problems: How can we reduce indiscriminate violence against civilians? How can we promote economic development? How should we try to combat global climate change? Despite the fact that international legal scholars ask these important questions, they frequently use research methods that are incapable of answering them. Observational research methods are frequently unable to tell us if, and how, international law changes the behavior of states and non-state actors.

During the same period that scholars of international law have grappled with the limitations of observational research methods, social scientists and legal researchers have increasingly conducted experiments. Experiments are allowing these researchers to directly test existing theories in ways that produce reliable causal estimates while avoiding barriers to inference posed by the use of observational data. International law shares many of the same issues that have forced other disciplines to turn to experimental research methods. But, international law faces many of these same issues to the extreme, as well as problems that are unique to the field.

316. See, e.g., Gregory G. Holyk, *Individual-level Predictors of Leader and Public Support for the Use of Force*, 23 INT'L. J. PUB. OPINION RES. 214 (2011); Kerry G. Herron & Hank C. Jenkins-Smith, *U.S. Perceptions of Nuclear Security in the Wake of the Cold War: Comparing Public and Elite Belief Systems*, 46 INT'L STUD. Q. 451 (2002). But see Emilie M. Hafner-Burton, D. Alex Hughes, & David G. Victor, *The Cognitive Revolution and the Political Psychology of Elite Decision Making*, 11 PERSP. ON POL. 368 (2013) (arguing that experienced policy elites differ from inexperienced subjects in how they make decisions because of their 'sophistication,' which is a learned skill that is derived from experience and tends to be greater in elite than non-elite populations).

317. See, e.g., Tomz, *supra* note 27, for a survey experiment on members of the British Parliament; see also, Hafner-Burton et al., *supra* note 27, for a survey on elites over trade preferences.

318. See *supra* note 312.

In this paper, we have identified five problems that are extremely salient to the study of international law and indicate the field could particularly benefit from experimental methods. First, there is frequently insufficient variance in the sources of international law—treaties and customs—for observational studies. Second, the window of time that scholars of international law can study may not have produced enough relevant observations for large- n empirical analysis. Third, it is difficult, if not impossible, to disentangle the causal influence of different sources of international law using observational methods. Fourth, data on international law outcomes is often either incomplete or impossible to collect. Finally, international law is rife with sources of selection bias.

Experimental research presents a solution to all of these problems, and scholars of international law should thus begin to incorporate experimental methods into their research agendas. This may require international law scholars to either seek additional methodological training, undertake significant self-study, or partner with social scientists. Although all of these experimental based methods pose their own difficulties, difficulties that are not easily dismissed, it is our hope that the importance of the topics international law scholars study will lead them to be willing to find a way to engage in experimental-based research. With the increasing array of experimental techniques and research strategies already available, we are optimistic. After all, although the empirical revolution in international law has been a welcome development, if we truly hope to work towards answers to many of the most important questions facing the world, it is time that international law undergoes an experimental revolution as well.