WHY THE STUDY OF INTERNATIONAL LAW NEEDS EXPERIMENTS

Dustin Tingley & Adam Chilton

ABSTRACT
Scholarship on international law has undergone an empirical revolution. Throughout the revolution, however, studies have been criticized because shortcomings of the observational data that they used posed serious barriers to reliable causal inference. At the same time, both political scientists and legal scholars studying domestic law have increasingly employed experimental methods because they make it easier to make credible causal claims. Yet 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 to 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.

* Dustin Tingley, Assistant Professor of Government, Harvard University. Ph.D. in Politics, Princeton University. Adam Chilton, Bigelow Fellow & Lecturer in Law, University of Chicago School of Law (designee); Ph.D. candidate in Political Science, Harvard University. Our special thanks to research assistance by Alice Miao, and Michael Findley, Tom Ginsburg, Jill Goldenziel, and Michael Tomz for helpful conversations and advice.
WHY THE STUDY OF INTERNATIONAL LAW NEEDS EXPERIMENTS

Table of Contents

Abstract ................................................................................................................................. 1
Table of Contents .................................................................................................................. 2
Introduction .......................................................................................................................... 3
I. The Growth of Experimental Research ................................................................. 9
   A. The Growth of Experiments in Political Science ............................................. 10
   B. The Growth of Experiments in Legal Research ............................................. 13
II. Why Experiments Are Needed for International Law ................................. 16
   A. Lack of Variance in Sources of International Law ..................................... 17
      1. Widespread Ratification of Multinational Treaties ................................ 18
      2. Universal Applicability of Customary International Law .................... 22
   B. Narrow Ranges of Time for Analysis .............................................................. 24
      1. Short Windows Before Widespread Ratification .................................... 25
      2. Laws Regulating Infrequent Events ......................................................... 28
   C. Overlapping Legal Constraints ..................................................................... 29
      1. Overlapping International Treaties ............................................................ 29
      2. Overlapping Domestic Laws .................................................................... 32
   D. Inadequate Dependent Variables .................................................................. 35
      1. Existing Data Sources Do Not Accurately Measure Compliance .......... 36
      2. Data on Compliance Is Not Available ....................................................... 38
   E. Selection Bias ................................................................................................. 40
      1. Selection Bias & International Treaties .................................................... 41
      2. Selection Bias & International Litigation ............................................... 43
III. Conducting Experimental Research ................................................................. 44
   A. Laboratory Experiments ............................................................................... 46
   B. Survey Experiments ....................................................................................... 50
   C. Field Experiments ......................................................................................... 55
Conclusion ....................................................................................................................... 59
INTRODUCTION

Over the last decade, there has been an empirical revolution in the study of international law. 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. This “empirical turn” has produced important new scholarship on a range of substantive areas of international law; including international human rights law, the laws of war, and international economic law.

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

---

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) (proving 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).


7 For a general discussion of this issue, see Shima Baradaran, Michael G. Findley, Daniel L. Nielson, & J. C. Sharma, Does International Law Matter?, MINN. L. REV. 1, 21-24 (forthcoming 2013), available at <http://michael-findley.com/texas/files/does-intl-law-matter_pre-publication-version.pdf> (last visited January 16, 2013). Of course, the struggle to improve causal inference is not unique to international law. Instead, this is a problem that 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
proclaimed that rates of compliance with international law were high as evidence of international law’s importance, the debate has shifted to whether international law actually has a causal impact on changing state behavior. 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 topic—states may simply only select 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, 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. Although researchers have begun to use more complicated empirical methods—like instrumental

---


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).


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 Stien’s criticisms); Shaffer & Ginsburg, supra note 1, at (noting that despite von Stein’s arguments, “we believe that 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).
variable regression and matching—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. 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.

At the same time that international law has been debating how to demonstrate causation, there has been an explosion in the use of experimental research methods by political scientists, and even a growing use by legal scholars studying other areas of law. In the last two decades, experimental methods have become an increasingly common research strategy used to study a range of topics, from what strategies are most likely to turn people out to vote during elections, to whether legislatures are biased against their constituents, to investigating the “democratic peace” theory, and even to study important questions in constitutional law. This

---

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


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).


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).

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, An Experimental Investigation of the Democratic Peace, AM. POL. SCI. REV. 1, 4-5 (forthcoming 2013), available at
growth in experiments has been fueled in part because it is increasingly easy to conduct experimental research, but also because researchers have increasingly recognized that experimental methods greatly facilitate credible causal claims.

Yet despite the fact that experimental methods have become increasingly popular because of their ability to improve causal inference during the same time that empirical research on international law has struggled to demonstrate causal claims, there have been very few efforts to use experimental methods to study international law. In fact, we are only aware of a handful of experiments—one field experiment, two laboratory experiments, and seven survey experiments—that have been conducted in the field of international law.


21 For a recent examples of legal research that used experimental methods, see, e.g., Dan M. Kahan, David A. Hoffman, Donald Braman, Daniéli 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.


to study international law. Moreover, political scientists, and not legal scholars, conducted all of these experiments.

26 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 emissions 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 2013) (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); 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, Working Paper (2012) (on file with authors) (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 # 16 (2012), available at <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2132948 (last visited January 24, 2013) (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, Does International Law Affect Willingness to Punish Foreign Human Rights Abusers?, Working Paper (2009) (conducting a survey experiment to 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 24 (2008), available at <http://www.stanford.edu/~tomz/working/Tomz-IntlLaw-2008-02-11a.pdf> (last visited November 17, 2012) (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.

27 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, Working Paper (2012) (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 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 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?, Working Paper (2009), available at <http://www.cis.ethz.ch/events/past_events/PEIO2009/Maliank__Tierney> (last visited January 16, 2013).

28 The one exception that we are currently aware of is Shima Baradaran, Associate
This should change. In this paper, we present the first argument 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 credibly 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 a large amount of original data collection 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.

To make this argument, 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 help to 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.

Professor of Law, BYU Law School. Baradaran was a coauthor with political scientists on a forthcoming article using a field experiment examining compliance with international law. See Baradaran et al., supra note 7.
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 to 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: laboratory experiments, survey experiments, and field experiments. 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. Finally, we explain how advances in statistical software and online subject recruitment have made it easier, cheaper, and faster to conduct experimental research.

I. THE GROWTH OF EXPERIMENTAL RESEARCH

The use of experimental methods in political and legal research has increased dramatically in the last two decades. Today, improvements in technology mean that researchers can easily, cheaply, and quickly design and administer experiments to thousands of subjects. Researchers can even design their experiments in ways that directly test theoretical causal mechanisms that have been hypothesized to exist in ways that were previously difficult, if not impossible, with observational studies alone.

---

29 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 & political scientists that present a promising research method in many cases. For more information, see infra text accompanying notes 221 – 224.

30 See generally Druckman et al., supra note 16. See infra text accompanying notes 33 – 47 & 56 – 69.

31 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.

As we will argue, scholars of international law should begin to use experimental methods not only because these developments should make conducting an experiment appealing to any scholars interested in directly testing the plausibility of their theories, but also because there are a number of reasons why observational studies are often particularly poor ways to study international law. Before making that case, however, we will first begin in this part by providing a brief background on the growth of experimental methods generally.

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 then will turn to discussing how legal scholars have recently begun to embrace experimental research, but that this embrace has not extended to international law. As we will discuss, it does not appear that legal scholars studying international law use experimental methods. This discussion will help to 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. Despite this early research on the value of experiments, social scientists generally were slow to adopt experimental methods. In fact, economics and political science were both disciplines where very little experimental research was published in the first fifty years after the first work on the importance of randomization. In fact, the first experiments to test causal mechanisms, see infra PART III.B.

33 See 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 2 ("While scientists have conducted experiments for hundreds of years, modern experimentation made its debut in the 1920s and 1930s.").

34 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.").

35 For evidence on the limited use of experiments by economists before the 1970s, see Falk & Heckman, supra note 34, 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 discussions on the growth of the use of experimental methods in political science, see Kathleen McGraw & Valerie Hoekstra, Experimentation in Political Science: Historical Trends and Future Directions, in RESEARCH IN MICROPOLITICS 3 (M. Delli
experiment was not published in the American Political Science Review (APSR)—the leading political science journal—until 1956. This was not a watershed moment, but instead it took over a decade for the next experiment to appear in the APSR. Moreover, experimental methods were not just used infrequently in the discipline—commentators have suggested that the use of these methods were poorly regarded by the discipline until recently.

In recent years, however, experimental methods have become widely used in political science. Although there has been a debate on the exact number of articles using experimental research methods that have been published to date, there is agreement that the number has grown exponentially in the last two decades. In fact, there were over fifty experimental articles published in just three political science journals between 2000 and 2007. Although perhaps the most prominent experiments published in the last decade have focused on voter turnout, experimental methods have been used to study a range of topics from how individual attitudes towards immigration policy are formed to the

Carpini, Leoni Huddy, & Robert Y. Shapiro, eds., 1994); REBECCA B. MORTON & KENNETH C. WILLIAMS, FROM NATURE TO THE LAB: EXPERIMENTAL POLITICAL SCIENCE AND THE STUDY OF CAUSALITY 1-6 (2010); Druckman et al., supra note 16, 628-30.  
36 See Druckman et al., supra note 16, at 628-9. 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).  
37 See Druckman et al., supra note 16, at 629.  
38 See MORTON & WILLIAMS, supra note 35, at 1 (“Until the last decade, 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 . . . .”).  
39 See MORTON & WILLIAMS, supra note 35, at 1-2 (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).  
40 See id. at 2. See also Druckman et al., supra note 16, at 628.  
41 See MORTON & WILLIAMS, supra note 35, at 2.  
42 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 to encouraging them to vote). See also Green & Gerber, supra note 18; Shaw et al., supra note 18.  
43 See, e.g., Dusting Tingley, Public Finance and Immigration Preferences: A Lost Connection?, POLITY (forthcoming 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
circumstances under which private firms comply with international law on disclosure requirements during incorporation. It is worth noting that the growing popularity of experimental methods is not limited to simply the number of articles that have been published. In recent years, political scientists have held conferences dedicated solely to experimental research, formed a section of the American Political Science Association dedicated to experimental research, and even established a regular newsletter on experimental political science.

What should be of interest to scholars of international law, however, is that experimental methods have not only been used to study American politics, but also have increasingly been used to study international relations. 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, survey experiments factors explain individual immigration preferences).

---

44 See Findley et al., supra note 24 (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).

45 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 January 17, 2013). Additionally, Brigham Young University held a “Field Experiments in International Relations” Conference in September 2012. For more information, see <https://pedl.byu.edu/Pages/Conference.aspx> (last visited January 24, 2013).

46 The “Experimental Research” section of the American Political Science Association was organized in 2010. For more information, see <https://www.apsanet.org/sections/sectionDetail.cfm?section=Sec42> (last visited January 17, 2013).

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

48 Rose McDermott, New Directions for Experimental Work in International Relations, 55 INT’L STUD. Q. 503 (2011) (“Experimentalism 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).

49 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 48; Natalie Florea Hudson & Michael J. Butler, The State of Experimental Research in IR: An Analytic Study, 12 INT’L STUD. REV. 165 (2010).

have been used to understand how audience costs change political decision making, and field experiments have been used to help understand how international organizations should improve their allocation of resources. Although experimental research is admittedly still far from being the primary method used to study international affairs, this growing body of research has shown that conducting experiments is an incredibly valuable way to gain insight into international relations. Moreover, experiments are not simply being used to compliment observational research methods, but also to provide insight into important questions that have previously been at a standstill because of inferential problems with observational data.

B. The Growth of Experiments in Legal Research

Unsurprisingly, legal researchers have been slower to fully adopt experimental research methods. The first wave of empirical legal studies began to take shape in the 1920s and 1930s, which, as previously noted, See, e.g., Michael Tomz, Domestic Audience Cost in International Relations: An Experimental Approach, 61 INT’L ORG. 821 (2007).


53 See McDermott, supra note 48, at 503.


55 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).

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

57 For a fascinating discussion of early legal empirical research, see Ho & Rubin, supra
was roughly the same time that the theoretical foundations for modern experimental research began to develop.\(^{58}\) Despite these early roots, empirical legal research was still rare for much of the twentieth century.\(^{59}\) Even more rare, however, were the use of experimental methods by legal scholars. Perhaps the earliest appreciation for experimental methods came from the law and economics movement.\(^ {60}\) That said, this early recognition of the value of experiments focused on explaining the insights that legal scholars could gain from laboratory experiments that had been conducted by economists, but actual examples of legal researchers conducting experiments of their own were few and far between.\(^{61}\)

In the last twenty years, however, this trend has begun to change.\(^{62}\) During this time, leading law journals have published articles involving laboratory experiments,\(^{63}\) survey experiments,\(^{64}\) field experiments,\(^{65}\) and

\[\text{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).}

\[\text{See supra text accompanying note 33.}

\[\text{See Richard H. McAdams & Thomas S. Ulen, Introduction to the Symposium on Empirical and Experimental Methods in Law, 2002 U. Ill. L. Rev. 1, 1 (“Empirical methods are still rare in legal scholarship: very few law professors buttress their arguments by appeals to tests of statistical significance of even with descriptive statistics.”). McAdams & Ulen do, however, suggest that this trend is has been changing. Id. at 1 (citing Robert C. Ellickson, Trends in Legal Scholarship: A Statistical Study, 29 J. Legal Stud. 517 (2000)).}


\[\text{See Hoffman & Spitzer, supra note 60, 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”).}

\[\text{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 591-92 (documenting the use of empirical research in law reviews, including experimental methods). See also McAdams & Thomson, supra note 59, at 1-2 (discussing the rise of legal scholars using empirical methods generally).}


even natural experiments. In these articles, experimental methods have been used to explore important questions from a range of legal fields. These include criminal law, constitutional law, and corporate law, among others. As a result, it appears that experimental methods are gaining mainstream acceptance as a credible research method for legal scholars to use.

While experimental methods have begun to permeate law generally, this has not occurred in the field of international law. In fact, as far as we can tell, legal scholars studying international law have essentially not used experimental methods at all. For example, in Shaffer and Ginsburg’s recent extensive review of empirical scholarship on international law, they do not discuss a single experiment. 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. Given the growing use of empirical methods to

---


68 See, e.g., Kahan et al., supra note 17 (using experimental methods to explore the speech/conduct distinction).

69 See, e.g., Guttenrag et al., supra note 63; Bebchuk, supra note 29.

70 But see supra notes 24 – 27 (documenting recent studies to use experimental methods to examine questions relevant to the study of international law).

71 But see supra note 28.

72 See Shaffer & Ginsburg, supra note 1.

73 See, e.g., Findley, supra note 24; Hafner-Burton, supra note 25; Putnam & Shapiro, supra note 26; Tomz, supra note 26; Wallace, supra note 26.
study international law,\textsuperscript{74} and the problems of inference that have led many to criticize observational studies in the field,\textsuperscript{75} it is time that this changes.

II. WHY EXPERIMENTS ARE NEEDED FOR INTERNATIONAL LAW

As the previous section demonstrated, experimental methods have become an increasingly common way that scholars conduct research on political and legal issues, and there are considerable advantages to using these methods when studying international events. Until now, however, we are unaware of any attempts to argue that there are reasons unique to the study of international law that make experimental methods particularly appropriate. As we will argue, however, the justifications for using experimental methods run deeper than the justifications for using experimental methods for the study of international relations more generally. Instead, the ratification patterns and complexity of international agreements mean that international law is a field that is particularly well suited to experimental methods. As a result, although there have been considerable advances in our understanding of the effects of international law using observational research designs,\textsuperscript{76} we hope that this discussion will help to demonstrate 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 not sufficient variation in which countries are bound by sources of international law—specifically multinational treaties and customary international law—to assess whether those laws are able to change behavior. As a result, experimental methods are an appropriate and cost efficient way to create that variation by randomizing 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, and then additionally explain that there may be few events that occur during that time frame to allow for a reliable study. Given these realities, we explain how experimental methods present a way to explore hypothetical events outside of the narrow windows available to

\textsuperscript{74} See supra text accompanying notes 1 - 6
\textsuperscript{75} See supra text accompanying notes 7 - 15.
\textsuperscript{76} 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.
observational studies. Third, we explain how observational studies are frequently unable to take into account the overlapping legal constraints that states face, and are thus unable to isolate the causal effects of individual agreements. We suggest that experimental methods can help address this problem by testing whether international agreements may have an “additive”, and not purely “substitute”, 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 states and non-state actors comply, and as a consequence, overcome the problems posed by the use of large dataset sets 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 to overcome the selection bias that exists in these situations.

A. Lack of Variance in Sources of International Law

A fundamental requirement of causal analysis is that any explanatory variable must vary to be able to assess its causal impact. The reason is that 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 basic requirement often poses difficulties for both qualitative and quantitative researchers that would like to study any fixed feature of an environment. The need for explanatory variables to vary presents a particularly difficult challenge for any observational study that attempts to analyze the influence that international law has on state behavior. The reason is that the two primary sources of international law that often do not have sufficient variance to allow for a satisfactory analysis of their causal impact: treaties and

77 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.”).

78 See, e.g., Kenneth N. Waltz, Man, The State, and War: A Theoretical Analysis 16-41 (1959). In a famous example for students of international relations, Waltz argues that it is difficult, if not impossible, to assess the influence of human nature on war. This is because if human nature is fixed, it is not possible to use it as a variable to explain instances of war and peace. Id. See also King, Keohane, & Verba, supra note 77, at 147 (citing David D. Laitin, Hegemony and Culture: Politics and Religious Change Among the Yoruba (1986)) (noting Laitin’s difficulty studying the influence of religion on politics in Somalia because of the religious homogeneity among the population).
customary international law.\footnote{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).”).}

1. Widespread Ratification of Multinational Treaties

As we have suggested, it is becoming increasingly difficult to assess the influence of many of the most important multinational treaties because they have become so widely adopted. 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.\footnote{See 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.”).} This line of research has used treaty ratification as the key explanatory variable while studying a wide range of substantive issue areas in international law.\footnote{See generally Simmons, supra note 1, at 280 – 292 (reviewing empirical research on treaty compliance by issue area).} This includes research on compliance with international monetary rules,\footnote{See Simmons, supra note 11. See also von Stein, supra note 12; Simmons & Hopkins, supra note 12.} environmental standards,\footnote{See, e.g., O.R. Young & M. Zurn, The International Regimes Database: Designing and Using a Sophisticated Tool for Institutional Analysis, 6 Global Environmental Pol. 121 (2006).} human rights,\footnote{See, e.g., Simmons, supra note 4; Hathaway, supra note 4.} and the laws of war,\footnote{See, e.g., Morrow, supra note 5; Valentino et al., supra note 5.} among others.\footnote{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).} The difficulty, however, is that many of the most important treaties have become so ubiquitous that there is very little variance in their ratification. Moreover, since many studies have argued that treaties have the largest influence on certain types of countries,\footnote{See, e.g., Simmons, supra note 4, at 150 - 154 (arguing that human rights treaties have the largest impact on “Transnational / Partly Democratic Countries”); Morrow, supra note 85, at 561 (arguing that democracies are more likely to comply with treaties on the laws of war).} there may be even less variance among the subset of countries that treaties are likely to influence. Taken together, the import of these facts is that, given the insufficient variance in states that have ratified major agreements, using observational

\footnote{See, e.g., Simmons, supra note 4, at 150 - 154 (arguing that human rights treaties have the largest impact on “Transnational / Partly Democratic Countries”); Morrow, supra note 85, at 561 (arguing that democracies are more likely to comply with treaties on the laws of war).}
studies may be an increasingly less viable options to try and study the influence that treaty ratification has on state behavior.⁸⁸

To help illustrate this point, Figure 1 and Figure 2 present data on the number of states that are currently a party to two categories of multilateral treaties.⁸⁹ Figure 1 presents the number of states that are currently a party to the six “core” human rights treaties.⁹⁰ Those

---

⁸⁸ To be clear, there are sources of variation that can be exploited for observational studies. For example, the time before treaties have been widely ratified, or comparing individual countries behavior before and after ratification. Our broader point, however, is they experimental methods can introduce variation that is otherwise not present in observational data.

⁸⁹ 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 November 17, 2012). Figure 2 additional 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 November 17, 2012). The data presented in Figures 1 & 2 is based on the number of state parties to these agreements as of November 16, 2012.

agreements are: the International Covenant on Civil and Political Rights (ICCPR);\(^{91}\) the International Covenant on Economic, Social and Cultural Rights (ICESCR);\(^{92}\) the International Convention on the Elimination of All Forms of Racial Discrimination (CERD);\(^{93}\) the Convention on the Elimination of All Forms of Discrimination Against Woman (CEDAW);\(^{94}\) the Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT);\(^{95}\) and the Convention on the Rights of the Child (CRC).\(^ {96}\) Figure 2 presents the number of states that are party to the major multinational conventions that have sought to regulate crimes against humanity and armed conflicts in the post WWII period. Those agreements are: the Convention on the Prevention and Punishment of the Crime of Genocide;\(^ {97}\) the Geneva Conventions of 1949 (Geneva Conventions I-IV”);\(^ {98}\) Additional Protocol I to the Geneva Conventions of 1949 (AP I);\(^ {99}\) and Additional Protocol II to the Geneva Conventions of 1949 (AP II).\(^ {100}\) As Figures 1 & 2 show, the number of parties to these agreements ranges from 153 to 194 states. To put this in perspective, with the addition of South Sudan in 2011, there were 193 members of the United Nations.\(^ {101}\)

---


The result is that an overwhelming number of states are currently parties to the most important treaties regulating human rights and armed conflicts, leaving very little variance in ratification for scholars seeking to assess the impact of these important agreements using traditional observational studies.\textsuperscript{102}

Although the lack of variance in the ratification of major multilateral treaties poses a serious obstacle for observational studies, experimental research is able to help solve this problem. The reason is that it is possible to employ an experimental design that randomizes whether countries are subject to a given treaty to try and 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.”\textsuperscript{103} During interviews with these policy makers, Tomz asked their opinion on whether a country in a hypothetical scenario is 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 half of respondents were given an identical set of facts, but were told that the country HAD NOT signed the NPT. The result was that 35% of respondents that were

\textsuperscript{102} 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.

\textsuperscript{103} Tomz, supra note 26.
told the country had signed the NPT thought it was likely that the country was pursuing a nuclear weapon, but 61% of the respondents that were told the country had not signed the NPT thought that it was likely that the country was pursuing a weapon.\textsuperscript{104} Although this is not proof that the NPT is having an impact, it is at least evidence that high-level policy makers believe that it is an important signal. Moreover, since there are 190 countries that are currently party to the NPT,\textsuperscript{105} it is nearly impossible to try and estimate whether the NPT has any influence without taking this kind of creative approach that is possible while using an experimental research design.

2. Universal Applicability of Customary International Law

A related, and perhaps more difficult to account for, problem is that there is essentially no variance in the applicability of Customary International Law (CIL) because it applies to (almost) every country. CIL is international law that applies universally, regardless of whether a country has signed a formal treaty.\textsuperscript{106} The only way for a country to not be bound by CIL is to be a persistent objector while a norm of CIL is being formed.\textsuperscript{107} In practice, there are very few examples of countries being successful objectors from the formation of CIL.\textsuperscript{108} Even in cases where a state has persistently objected, customary international law is often considered a “preemptory norm” where objections are not valid.\textsuperscript{109} Moreover, these

\textsuperscript{104}Id. at 24 – 28.
\textsuperscript{106}See generally Bradley & Gulati, supra note 79. See also JACK L. GOLDSMITH & ERIC A. POSNER, THE LIMIT OF INTERNATIONAL LAW (2005).
\textsuperscript{107}See Bradley & Gulati, supra note 79, 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).
\textsuperscript{108}See id. 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. COMP. & INT’L L. 221 (2010).
\textsuperscript{109}See Bradley & Gulati, supra note 79, at 212-13 (discussing the “modern” view that there are a small subset of norms with special status) (citing Vienna Convention on the Law of Treaties, art. 53, May 23, 1969, 1155 U.N.T.S. 331, 8 I.L.M. 679) (“For the purposes of the present Convention, a preemptory norm of general international law is a norm accepted and recognized by the international community of States as a whole as a
preemptory norms include many of the strongest norms of international law.  

For our purposes, the relevant takeaway is that CIL is an important source of international law that prohibits actions on a range of topics that applies nearly universally. Although treaties may arguably be more important than CIL in modern international affairs, CIL is still the subject of a great deal of international legal scholarship. Assessing the influence that CIL may have on state behavior, however, maybe nearly impossible to do using either qualitative or quantitative observational studies because there is not any variance in its applicability. In other words, since this source of law applies to every country, using many traditional research designs it is impossible to determine if CIL has played either a major role, or no role at all, in many important changes in state behavior.

Although observational studies are unable to adequately deal with the lack of variance of CIL in most cases, experimental research can help establish whether it is at least plausible that customary international law is having an impact on state behavior. There have been a number of experiments that have sought to assess the impact of international law without references to specific treaties. For example, Tonya Putnam and Jacob Shapiro have conducted an experiment trying 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. As part of their experiment, they told some respondents nothing about international law, while telling others that a country was violating a treaty, and another group just that they were violating international law generally. Interestingly the results of their experiment suggest that although information about international law changes public opinion, general references to international law have roughly

---

110 See Bradley & Gulati, supra note 79, at 212-213 (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)).


112 See Shaffer & Ginsburg, supra note 1, at 11-15 (2012) (discussing lines of research in CIL while arguing that there should be more empirical research on the topic).

113 See, e.g., Putnam & Shapiro, supra note 26; Wallace, supra note 26.

114 See Putnam & Shapiro, supra note 26.

115 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.
the same effect on public opinion as references to specific treaties that have been signed. Why this result is interesting, however, is that it suggests that international law without a treaty—customary international law—may still have an influence on policy preferences and outcomes. The impact is that it suggests a way that experimental research can estimate whether CIL may matter while observational studies are unable to do so given a lack of variance in the applicability of customary norms.

B. Narrow Ranges of Time for Analysis

Another problem that observational studies of international law face is that it is often the case that only relatively short-time frames can be analyzed. When conducting a large-n empirical study, it is necessary to analyze a sufficiently long time period to have a large enough sample of observations to draw reliable inferences from the data. Mindful of this concern, scholars of international relations study international phenomena (wars, alliances, trading patterns, etc.), often construct datasets that cover the entire post-war period, the entire twentieth century, or even over 200 years. The approach of extending datasets further back into history, however, is often simply not available to scholars of international law. This is because there are often very short windows when there is variation in the number of countries that have signed treaties. In some cases, a related problem results because an insufficient number of events to study may have occurred during those windows.

---

116 Id. at 19.
117 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.
119 See, e.g., Alexander B. Downes, Restraint or Propellant? Democracy and Civilian Fatalities in Interstate Wars, 51 J. CONFLICT RES. 872 (2007) (analyzing a dataset of interstate wars between 1900 and 2003 to assess whether democracies are more likely to kill civilians during wars).
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. The obvious difficulty is that although international law has ancient roots, much of the corpus of international law has been formed relatively recently. The consequence is that even in the best cases, scholars of international law are often left studying relatively short time periods during which international law has gone into effect. This problem is exacerbated, however by the fact that it is often not prudent, or possible, to study the entire time period that an international treaty (or institution) has been in effect. As was explained previously in PART III.A., 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.

The fact that nearly all countries have now adopted many major treaties would not pose such a serious problem, however, if there was a window during which there was sufficient variation in the adoption of the treaty. For example, if there were several decades during which only half the world at 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. The difficulty, however, is that it appears

---

121 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”).

122 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. 322, 322-23 (2011) (“While strands of such international human rights law date back hundreds of years in the protecting 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 See Cesare P.R. Romano, A Taxonomy of International Rule of Law Institutions, 2 J. INT’L DISP. SETTLEMENT 241 (2011).

123 See, e.g., Rachel Brewster & Adam Chilton, Supplying Compliance: Domestic Sources of Trade Law & Policy, Working Paper (2012) (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”).

124 See supra text accompanying notes 77 - 116.

125 See supra Figure 1 & Figure 2.

126 This is the approach implicitly taken by Beth Simmons in her book analyzing the success of human rights agreements. See SIMMONS, supra note 4.
that many recent international agreements are being ratified by a huge number of states quickly. To illustrate this point, we have collected data on when states ratified each of the six major human rights agreements discussed in PART II.B.,¹²⁷ and each of the four treaties regulating armed conflicts as well.¹²⁸ This information is presented in Figure 3 & Figure 4.¹²⁹

As the figures clearly show, not only do these major agreements have a large number of signatories, but several of the more recent treaties achieved these huge numbers of signatories 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. It is also worth noting that this same trend has presented itself with a number of other major international agreements. For example, the Kyoto Protocol went into force on February 16, 2005.¹³⁰ Today, there are 191 parties to the agreement.¹³¹

---

¹²⁷ See supra notes 91 - 96.
¹²⁸ See supra notes 97 - 100.
¹²⁹ For information on the sources used to collect data for these Figures, see supra note 89.
¹³¹ Id.
Similarly, the Rome Statute for the International Criminal Court (establishing the ICC) went into force on July 1, 2002.\textsuperscript{132} Today, there are 121 members to the agreement.\textsuperscript{133}

![Figure 4: Ratification Timing of Laws of War Treaties](image)

Although these issues present serious hurdles to research designs based on observational data, experimental methods can provide one way to overcome them. First, because experimental research methods 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 previously discussed in PART II.A., Mike Tomz was able to conduct an experiment on the impact that ratification of NPT has on threat perception, even though he did not conduct the experiment during the window of time when there was variance in which countries had signed the NPT.\textsuperscript{134} The implication is that 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.


\textsuperscript{133} Id.

\textsuperscript{134} See Tomz, supra note 26. See supra text accompanying notes 103 - 105.
2. Laws Regulating Infrequent Events

Moreover, the problem created by the short-time frames between when treaties are negotiated and when they are widely adopted is exacerbated if the treaty attempts to regulate events that happen infrequently. As an example, Additional Protocol I to the Geneva Conventions of 1949 went into effect in 1978. This treaty constitutes the most comprehensive attempt to provide laws to regulate interstate-armed conflicts. Between when the treaty was open for signature and when the treaty had over 160 signatures in 2003, by at least one count there were only 14 interstate wars. The result is that, it thus becomes incredibly difficult to conduct large-n observational research on whether this treaty has helped to change the behavior of states engaged in armed conflicts.

Another advantage of experimental designs is that they can explore topics that have occurred relatively infrequently. For example, the relatively small number of interstate wars since the adoption of Additional Protocol I of the Geneva Conventions was adopted has made it difficult to conclusively determine whether the international humanitarian law influence state behavior during conflict. As a way to address this problem, an experiment by author Chilton posed respondents with a hypothetical future conflict in which violations of the laws of war were occurring. This experiment produced evidence that information on the status of the laws of war change public opinion, and moreover, was able to suggest that this was the case because this information changed respondents’ perceptions of the risks associated with targeting civilians.

---

135 AP I, supra note 99.
136 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) (“[AP 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.”).
137 See Alexander B. Downes, Web Appendix for “Restraint or Repellant?” Democracy and Civilian Fatalities in Interstate Wars (2007) (on file with authors) (providing documentation for the wars used in his large empirical project to analyze the number of civilians killed during interstate wars).
138 Compare Morrow, supra note 5 (finding that democracies are likely to follow the laws of war when there is 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).
139 See Chilton, supra note 26.
140 Id. at 16 – 21.
141 Id. at 21 – 24. For an explanation of how experiments can be designed to test causal mechanisms, see Imai, Tingley & Yamamoto, supra note 32; Imai, Keele, Tingley, Yamaoto, supra note 32. See infra note 272. For other examples of research using related
Although the experiment obviously 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 new evidence to the discussion that simply reexamining the observational data on the same set of conflicts would not. As these examples should thus illustrate, experimental designs can help to avoid the problems created by the short windows of time that exist between when a treaty is ratified and when it is widely adopted.

C. Overlapping Legal Constraints

Another limitation of using exclusively qualitative or quantitative observational studies to assess the impact of international law on state behavior is that they often have difficulty untangling the effects of overlapping legal constraints. As the study of international law has moved from analyzing whether states comply with the treaties and agreements that they have signed towards assessing whether those agreements make a causal impact on state behavior,\(^\text{142}\) it has become increasingly important for scholars to be able to isolate the effects of specific agreements to be able to perform credible causal analysis.\(^\text{143}\) To do so, scholars have focused on analyzing the impact of the ratification of an individual treaty\(^\text{144}\)—or participation in an international institution\(^\text{145}\)—on changes in policy.

1. Overlapping International Treaties

Isolating the causal effect of any individual treaty or institution, however, is complicated by the fact that countries are often bound by a number of international legal agreements that provide overlapping legal obligations and constraints. Given the rapid growth of international law in the last sixty years, states are likely to be subject to overlapping international treaty objections in many areas. For example, there were over 50,000 treaties on file with the UN Treaty System as of 2005.\(^\text{146}\) The result of this massive proliferation in the number of international legal agreements

\(^{142}\) See Posner, supra note 9.

\(^{143}\) Cf. Ho & Rubin, supra note 7, at 21-2 (explaining the importance of focusing on a single “treatment” when performing causal analysis).

\(^{144}\) See, e.g., Hathaway, supra note 4.

\(^{145}\) See, e.g., Simmons, supra note 11.

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

To briefly illustrate this point, we have collected data on how many of the treaties discussed in Part II.A.1 & Part II.B.1 states are party to. This data is presented in Figures 5 & Figure 6. As Figure 5 shows, 110 states are party to all six major human rights treaties. Moreover, Figure 6 shows that 120 states are party to the four treaties governing armed conflict that we have been analyzing. The reason that this is significant is that these treaties contain common elements, and efforts that states take in response to one agreement, may also fall under the scope of another agreement.

![Figure 5: Number of Major Human Rights Agreements that Countries Have Signed](chart)

---


148 For example, Article I of both the ICCPR and the ICESCR protect the “right of self-determination.” See ICCPR, supra note 91, at Art. I; ICESCR, supra note 92, at Art. I.
The reason that observational studies thus have difficulty isolating the effect of a single treaty when states are party to many similar agreements is that these constraints are often collinear. Collinearity occurs when a variable of interest is strongly correlated with another variable (or set of variables) that offer a potential explanation for a phenomenon.\textsuperscript{149} When this occurs, it is then difficult—if not impossible—to tease out the causal effects of these strongly related factors.\textsuperscript{150} For example, when there are multiple treaties that seek to protect minority rights, if there are similar ratification patterns for both treaties, it might be impossible to tell if either of the treaties is doing anything to protect minority rights. That is to say, if the set of countries have signed agreement X are also the same set countries that have signed agreement Y, the ratification of those two agreements are collinear, and thus it is impossible to know the causal effect of either treaty on an outcome of interest.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Number of Treaties on Armed Conflicts that Countries Have Signed}
\end{figure}

Fortunately, using experimental research designs can help researchers to directly address the problems posed by the presentences of overlapping legal constraints that confound observational studies. One way

\textsuperscript{149} As an informal definition, two variables can be said to be collinear when one variable can perfectly predict the other. For a discussion on this problem, see \textsc{King, Keohane, \& Verba, supra note 77}, at 122-24 \& 213-215.

\textsuperscript{150} For an excellent discussion of the problem that collinearity poses for testing whether there is a democratic peace, see \textsc{Tomz \& Weeks, supra note 20}, at 4-5.
that experimental designs can help to solve the problems caused by overlapping legal constraints is by varying information that would be collinear in observational data. That is to say, even if it is the case that nearly all countries that have ratified treaty X have ratified treaty Y—and thus 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). One problem that has plagued scholars trying to test this theory with observational data, however, is that the presence of democracy is often collinear with other variables that might explain peace (i.e. shared political interests). As a way 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 previously been hypothesized to both cause peace and be collinear by democracy. As a consequence of using this approach, Tomz and Weeks were able to provide new evidence that individuals are less supportive of war with countries that are democratic, even compared to autocracies that are similar in all relevant respects. As it should be clear, the relevance that this example has for sorting through the impact of overlapping legal constraints is that it 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.

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 the

---


153 See Tomz & Weeks, supra note 20.

154 If the research design used hypothetical countries, this will 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 250 – 252.
similarities and differences between international law.\textsuperscript{155} Scholars have long recognized that countries often have international laws and domestic laws that cover the same subjects.\textsuperscript{156} This can either be because countries select into international legal agreements that do not impose onerous new requirements,\textsuperscript{157} or because countries quickly change their domestic laws to mirror the international legal agreements they have signed.\textsuperscript{158} In either scenario, however, 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 any changes in state behavior.\textsuperscript{159} Moreover, this phenomenon may become increasingly common if international law continues to move towards regulating more areas that previously have been solely the sphere of national regulation.\textsuperscript{160}

\textsuperscript{155} This 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, e.g., Oona Hathaway & Scott J. Shapiro, \textit{Outcasting: Enforcement in Domestic and International Law}, 121 YALE L.J. 252 (2011) (arguing that international law is enforced 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 though a monopoly on the use of force within a territory); Jack Goldsmith & Daryl Levinson, \textit{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, \textit{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).


\textsuperscript{157} See supra Part II.E. See generally Downs, Rocke, & Barsoon, supra note 8, 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).

\textsuperscript{158} 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. \textit{Id.} at 237-45.

\textsuperscript{159} 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 15-16 (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 IMFs legal rules on state policy). The difficult, however, is that with large n-observational studies, it can be difficult to be confident of this causal link without more in depth research.

\textsuperscript{160} See generally Jacob Katz Cogan, \textit{The Regulatory Turn in International Law}, 52 HARV. INT’L L.J. 321 (2011) (arguing that international law has moved from just placing
In addition to the problem that exists when these agreements are collinear—as was discussed in the last section—another reason observational studies have difficulty accounting for overlapping domestic legal constraints is that simply being aware of, and accounting for, all of a state’s legal obligations 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, or even for a country’s constitutional laws, 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 degree is democratic—without including terms to account for all of a country’s relevant legal obligations.

In addition to providing a way to address problems of collinearity, experimental designs can also help to test whether international law has an effect beyond domestic law without the requirement of gathering huge amounts of country specific data. This is because experiments can be used to 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. To illustrate this point,

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).


163 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.

164 For a discussion of the difference between “additive” and “substitute” effects as it relates to international law, see Tomz, supra note 26, at 19 - 21. See also Chilton, supra
imagine a survey conducted that measured public support for the war on terror, and that one group of respondents were told that torture violated domestic laws; a second group of respondents that torture violated an international treaty that the United States had signed; and a third group of respondents that torture violated both domestic laws and an international treaty that the United States had signed. If public opinion on torture was the same after being told about domestic law, international law, or both (Group 1 = Group 2 = Group 3), then international law would have a “substitute” effect in this case.\textsuperscript{165} However, if information on international law changed opinion beyond information on domestic law alone (Group 3 > Group 1), then international law could be said to have an “additive” effect in this case.\textsuperscript{166} The import, of course, is that this kind of experimental design makes it possible to test for whether international law might influence policy outcomes, even when domestic law already exists on a given topic.\textsuperscript{167}

\textbf{D. Inadequate Dependent Variables}

As we have previously discussed, at the same time that scholarship on international law has become more empirical,\textsuperscript{168} scholars have shifted their attention to trying to understand whether international law causes states to change their behavior.\textsuperscript{169} 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.\textsuperscript{170} Simply put, a dependent variable is the outcome that researchers are trying to measure.\textsuperscript{171} For example, if a scholar was studying whether states comply with the laws of war, they might use civilian deaths during a given conflict as a dependent

\textsuperscript{26} (using an experimental research design that tests if the laws of war has an additive effect beyond moral arguments against targeting civilians).

\textsuperscript{165} That is to say that the information on international law merely substituted the effect that was created by simply being told about domestic law.

\textsuperscript{166} 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.

\textsuperscript{167} 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.

\textsuperscript{168} See generally Shaffer \& Ginsburg, supra note 1.

\textsuperscript{169} See Posner, supra note 9. See also Shaffer \& Ginsburg, supra note 1, at 1.

\textsuperscript{170} Cf. KING, KEOHANE, \& VERBA, supra note 77, at 109 (“[I]n social science, we must be careful to ensure that we are really interested in understanding out dependent variable, rather than the background factors that our research design holds constant.”)

\textsuperscript{171} See id. at 77. For a longer discussion of dependent variables, see id. at 107-9.
The foundational requirement of having an adequate dependent variable poses two significant problems for researchers hoping to study whether states comply with international law: 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 commitments they have made under the IMF Articles of Agreement by leaving their current accounts free from restriction. In this case, Simmons was able to use a dependent variable that was a direct, and clear, measure of non-compliance.

Of course, things are not always so straightforward. Instead, it is often the case that 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. Eric Posner has criticized this measure, however, for not necessarily measuring compliance with the ICCPR.

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. The difficulty in these

---

172 See, e.g., Valentino, supra note 5.
173 Simmons, supra note 11.
174 Id. at 833.
175 See SIMMONS, supra note 4, at 386.
176 Posner, supra note 9, at 6-7.
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 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.\(^\text{178}\)

A third problem is that the best available dependent variables for use might not prompt to measure compliance directly at all, but instead simply may be proxies for consequences that might result as a consequence of compliance. For example, when analyzing whether states change their policies as a consequence of ratifying the Convention on Political Rights of Women, Oona Hathaway uses the percentage of men in a country’s legislature as the dependent variable.\(^\text{179}\) As Hathaway admits, however, equal representation is not required by the convention. The dependent variable is thus a proxy for compliance, which then leaves the measure open for debate over whether it adequately captures compliance.\(^\text{180}\)

Experimental methods offer a way to directly solve all three of these problems. That is, experiments can be conducted that directly generate evidence of whether a state is in compliance with international law. For example, Baradaran et al. recently conducted what we believe to be the first field experiment on compliance with international law.\(^\text{181}\) In the experiment, the researchers sent requests for information on incorporating a shell company to 1,015 firms offering incorporation services in 182 countries.\(^\text{182}\) As the experimental treatment, the researchers randomized whether firms were given any facts on the status of international law.\(^\text{183}\) To be compliant with international law on international financial transparency, firms are required to demand proof of identity before providing incorporation services.\(^\text{184}\) Through the way the researchers worded the requests, they were able to directly measure whether firms were willing to

---


\(^{179}\) Hathaway, supra note 4, at 1976-76.

\(^{180}\) For other criticisms about Hathaway’s selection of dependent variables, see Goodman & Jinks, supra note 12 (arguing that Hathaway’s selection of dependent variables do not “account for strategies often adopt in response to improved enforcement of a norm”).

\(^{181}\) Baradaran et al., supra note 24, at 7.

\(^{182}\) Id. at 37.

\(^{183}\) Id. at 38-39.

\(^{184}\) Id. at 43.
violate international law in their responses.\textsuperscript{185} As a result, this experimental approach was able to bypass any complaints that could be levied against existing data sources for not adequately measuring compliance with international law, because the responses to their experiment itself was evidence of non-compliance. In other words, they did not have to argue that their DV was a good measure of compliance, because it was an original, direct measure of compliance.

2. Data on Compliance Is Not Available

In addition to cases where the existing data sources are not ideal measures of non-compliance, a related problem exists when data on whether compliance is occurring is not available. In cases where dependent variables that measure compliance simply do not exist, scholars’ only option if they still wish to research the topic is to expend the effort to collect original data on whether states have complied with their treaty obligations. It is worth noting that despite the time and hassle that may be involved with this process, we do believe that scholars provide an incredibly valuable service when they take the time to engage in original data collection, and that this step should be taken and the data made public more often.\textsuperscript{186}

That said, obviously original data collection on whether countries are compliant with international law is not possible in all cases. This may be the case for several reasons. The first is that it may require extensive research into the practices of a large number of individual countries. For example, if field research were required to determine if a state is compliant with an international agreement, it is likely not possible for a researcher to visit enough countries to conduct a large-n study. A second reason is that that data on whether a state is compliant with a particular international agreement may simply not be publically available. For example, it might simply not be possible to collect data on whether states comply with the Convention Against Torture.\textsuperscript{187} Moreover, even if it is possible to find some national level data, there might be considerable subnational variation in actual practice that was unaccounted for in the high level data.

\textsuperscript{185} Id. at 43–44.

\textsuperscript{186} For a discussion of the value of posting and sharing data, see Gary King, Ensuring the Data Rich Future of the Social Sciences, 233 SCIENCE 719, 720 (2011). 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 123 (collecting data on the date the United States took steps to comply with adverse rulings in the WTO).

\textsuperscript{187} See CAT, supra note 95.
In addition to these two obstacles to original data collection, one fact that is often overlooked by scholars using empirical methods to research compliance with international law is that international treaties often guarantee the protection of a large number of rights.\(^ {188}\) As a result, even if there is a dependent variable 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 this point, we have collected information on the number of individual rights protected in the six major human rights treaties.\(^ {189}\) This information is presented in Figure 7. As Figure 7 shows, the average number of different rights protected in the human rights agreements is 35. In fact, the ICCPR protects 60 different rights, ranging from the right to not be imprisoned due to debt\(^ {190}\) to the right to marry and found a family.\(^ {191}\) As a result, it may be thus difficult to reliably test whether states comply with a specific treaty when data may only be available to use as dependent variables for a few of the rights contained within a treaty.

![Figure 7: Number of Rights Protected by Major Human Rights Agreements](image)

Conducting an original experiment can provide an appealing alternative to observational studies in cases where it is not possible to collect data on compliance. In addition to using the approach used by

\(^{188}\) See Posner, *supra* note 9, at 7 (criticizing Simmons (2009)’s dependent variables because human rights treaties “collectively contain dozens or maybe even hundreds of provisions”).

\(^{189}\) See *supra* notes 91 – 96.

\(^{190}\) ICCPR, *supra* note 91, at Art. 11.

\(^{191}\) ICCPR, *supra* note 91, at Art. 23.
Baradaran et al. discussed in the previous section to collect direct evidence of non-compliance, another way that experimental methods can be used in cases where it is impossible to collect data on non-compliance is to conduct an experiment that tests 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 Torture might have an influence on America’s use of torture in the war on terror. Wallace’s approach was to study whether international law exerts an independent influence on state behavior by directly examining one theoretical mechanism by which this might occur—by altering domestic mass political views. Wallace conducted two survey experiments on national samples of American adults, and the experimental treatment was that he varied information provided on international law, before gauging views on the acceptability of the use of torture. Wallace’s experiment produced evidence that international law changes support for the use of torture by roughly 6 percentage points, which provides evidence that international law might change policies on torture without having to have a dependent variable that directly measures torture itself.

E. Selection Bias

The barrier to reliable causal inference that has received the most attention in international law is selection bias. Although at least one other scholar has explained how experiments can help to overcome the problems caused by selection bias, we will still mention them briefly. Selection bias occurs when the observations that have received a particular treatment is systematically related to a potential outcome. For example,

---

192 See supra text accompanying notes 181 – 185.
193 CAT, supra note 95.
194 Wallace, supra note 26.
195 Id. at 1-2.
196 Id. at 16-19.
197 Id. at 20.
198 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 scholarship on international law drawing attention to selection effects, see Downs, Rocke, & Barsoon, supra note 10 (arguing that previous claims that international law is generally complied with ignore a variety barriers to inference, including selection effects).
199 See, e.g., Tomz, supra note 26, at 7.
200 For a clear statement of the definition of selection bias, see ALAN S. GERBER & DONALD P. GREEN, FIELD EXPERIMENTS: DESIGN, ANALYSIS, AND INTERPRETATION 37 (2012). See also KING, KEOHANE, & VERBA, supra note 77, at 135-37.
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 a test preparation course (receive the treatment) may be more conscientious students that were systematically likely to score higher (the potential outcome) than other students regardless of whether they were enrolled in a test preparation course. In the study of international law, producing reliable causal estimates of international agreements and institutions has been consistently frustrated by the fact that countries that choose to participate in international legal regimes may be systematically different than the countries that do not.\textsuperscript{201} This reality is a major problem for observational studies on international law\textsuperscript{202} that experimental research can help to 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.\textsuperscript{203} 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.\textsuperscript{204} The difficulty, however, is that whether a given state chooses to commit to an international treaty is not random.\textsuperscript{205} Instead, there are a number of strategic calculations that determine whether states choose to agree to international treaties.\textsuperscript{206} As a result, it is reasonable to think that

\textsuperscript{201} 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).

\textsuperscript{202} But see Hafner-Burton et al., supra note 1, at 89 n.231 (documenting studies that have used “sophisticated methods” like instrumental regression and matching to obtain valid inferences on the effects of international law).

\textsuperscript{203} Cf. Simmons, supra note 1.

\textsuperscript{204} See Posner, supra note 9, at 5. For example, 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).

\textsuperscript{205} For a discussion of this issue inferential problem caused by the fact that states do not randomly choose to sign treaties, see Tomz, supra note 26, at 7.

\textsuperscript{206} 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); Oona Hathaway, Why Do Nations Join Human Rights Treaties?, 51 J. CONFLICT RES. 588 (2007) (examining why states consent to
the states that select into international treaties are systematically different from those states that do not. This selection bias thus presents an inferential problem that has made it difficult for scholars of international law to study the effects of international agreements.207

Using 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.208 For example, in a recent experimental study, political scientists Tingley & Tomz tested Americans’ willingness to take retaliatory actions against a country that has increased its consumption of fossil fuels.209 Half of the respondents were told that the country had said that it would not increase its use of fossil fuels, and the other half were told that the country had signed a treaty promising that it would not do so.210 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%).211 Since the assignment of the experimental treatment (being told a treaty had been signed) was randomized, the survey experiment thus suggests that leaders may have strong incentives to treat countries that have signed international agreements differently than those that have not. If the authors had instead simply looked at US responses to polluting countries that had signed treaties using observational data, it might not be possible 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 have ratified into the treaty may be the same states that US citizens are more willing to take actions against in general—and thus induce bias. The use of experimental methods thus helps to overcome problems created by selection bias and directly measures the causal inference of treatments of inference.

---

207 See, e.g., von Stein, supra note 12 (discussing the problems posed by selection 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 notes 13 – 14. See also Hafner-Burton et al., supra note 1, at 89 n.231.

208 See, e.g., Tomz, supra note 26 (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).

209 Tingley & Tomz, supra note 26.

210 Id. at 26-27.

211 Id. at 28.
2. Selection Bias & International Litigation

A related problem that poses difficulties for international law scholars is that litigation in international courts suffers from selection bias.\(^{212}\) Simply put, disputes that parties choose to litigate until a 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.\(^{213}\) For example, disputes that reach a judgment may be systematically different than 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 may be difficult to draw inferences about the effects of litigation on the settlement of disputes more generally.

Once again, experimental research can be used to overcome these selection biases. For example, if researchers were interested in understanding whether countries are likely to change their policies as a result of WTO disputes,\(^{214}\) even if observational research showed that countries changed their policies as a result of adverse WTO decisions, it might not be reasonable to conclude that they would do so as a result of any adverse WTO decision. Instead, complainant states may only bring complaints in situations where they believed a country would be responsive to an adverse ruling—for example, when the issue was not too politically sensitive. Experimental methods could be used to test this, however, by telling participants about 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

\(^{212}\) 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 the 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, & Heiner Schultz, _The European Court of Justice, national Governments, and Legal International in the European Union_, 52 INT’L ORG. 149, 151 (1998) (discussing their efforts to develop a case identification strategy that will help to minimize selection bias).

\(^{213}\) Cf. Anna Harvey & Barry Friedman, _Ducking Trouble: Congressionally Induced Selection Bias in the Supreme Court’s Agenda_, 71 J. POLITICS 574 (2009) (empirically testing how the selection bias introduced by the fact that the Supreme Court has discretionary jurisdiction influences the Court’s Agenda).

\(^{214}\) See, e.g., Brewster & Chilton, _supra_ note 123 (analyzing United States’ compliance with adverse WTO decisions).
result, in addition to their other benefits that we have discussed, experimental methods also present one way to overcome the inferential problem posed by selection bias in international litigation.

III. CONDUCTING EXPERIMENTAL RESEARCH

As we have argued, scholars of international law should begin to embrace experimental methods. We believe that although experimental research methods are not appropriate ways to study every question, they are currently 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.\textsuperscript{215}\textsuperscript{216}

A common schema of experiment types breaks experiments into three categories: laboratory experiments, survey experiments, and field experiments.\textsuperscript{217} Laboratory experiments range considerably in what they entail, but share a focus on experimentation in highly controlled laboratory settings.\textsuperscript{218} Survey experiments embed manipulations and randomized frames into standard public opinion surveys.\textsuperscript{219} Field experiments turn to naturally occurring behavior but manipulate features of the environment in order to make inferences.\textsuperscript{220} A forth category of experiments are natural experiments. Natural experiments take advantage of naturally occurring randomizations, such as arbitrarily drawn legal borders\textsuperscript{221} or the distribution

\textsuperscript{215} For good textbook introductions to experimental methods, see MORTON & WILLIAMS, supra note 35 (providing an overview on how to conducting experimental research); GERBER & GREEN, supra note 200 (providing an overview specifically on field experiments).

\textsuperscript{216} 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.

\textsuperscript{217} 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 (Druckman, Green, Kuklinski, & Lupia, eds., 2011) (“[M]ost experiments have been implemented in one of three contexts: laboratories, surveys, and the field.”).

\textsuperscript{218} Id. at 6-7.

\textsuperscript{219} Id. at 7.

\textsuperscript{220} Id.

\textsuperscript{221} 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 dependent on the
of international election monitors during elections, in order to make inferences. Although natural experiments can be used in legal research, we will not consider them in depth in this article so that we can focus our attention on experimental methods where the researcher controls the randomization and other experimental processes. We suspect that survey and field experiments will be most likely in studies of international law and laboratory experiments less common, but we discuss all for completeness. In discussing each of these types of experimentation we also talk about how issues of randomization, challenges to internal, external, and construct validity play out.

In this part, we hope to provide a minimal foundation in experimental methods for that to occur as well as point scholars to further resources. Specifically, we discuss different experimental techniques and their virtues/vice with respect to applications in international law. Depending on the research questions or context, some types of experiments will be

---

222 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 to reduce fraud at the polling stations they visit).

223 For guidance on finding natural experiments, see Gregory Robinson, John E. McNulty, & Jonathan S. Krasno, Observing the Counterfactual? The Search for Political Experiments in Nature, 17 POL. ANAL. 341 (2009). For additional advice on establishing whether an event can be considered a natural experiments, 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).


225 See, e.g., Hafner-Burton et al., supra note 25 (conducting an experiment in a laboratory setting to test the behavioral influences of international cooperation).

226 Internal validity is defined as: “[t]he approximate truth of the inference or knowledge claim within a target population studied.” MORTON & WILLIAMS, supra note 35, at 188.

227 External validity is defined as: “[t]he approximate truth of the inference or knowledge claim for observations beyond the target population studied.” Id. at 188.

228 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.
more applicable or easier to apply. For example, it may be easier to use survey experiments to study how international law affects public opinion, 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. 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. But 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. In the social sciences, psychology far and away uses laboratory experiments the most. 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 of subjects or by targeted recruiting of subjects with specific characteristics (like gender or work experience). Many psychology departments have laboratories with subject pools, and some universities have broader social science laboratories. Experimental sessions begin with a consent and

---

229 See, e.g., Wallace, supra note 26 (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).

230 See, e.g., Findley et al., supra note 24 (conducting a field experiment to evaluate whether private firms violate international law when responding to requests to incorporate shell companies).

231 See infra text accompanying notes 262 - 263.

232 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 73 (Druckman, Green, Kuklinski, & Lupia, 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).

233 Cf. Falk & Heckman, supra note 34, 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.”).

234 For an extended discussion on subject recruitment, focusing on the use of students as subjects, see MORTON & WILLIAMS, supra note 35, at 237-258.

235 For example, Harvard University runs the Harvard Decision Science Laboratory (HDSL) (http://decisionlab.harvard.edu/), Princeton University runs the Princeton Laboratory for Experimental Social Science (PLESS) (http://pless.princeton.edu/), 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.\textsuperscript{236}

Laboratory experiments are useful for a number of reasons.\textsuperscript{237} The ability to precisely control the experimental environment means that effects are most likely due to the experimental intervention.\textsuperscript{238} 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 the domestic policies. Were this intervention not randomized, and we just correlated awareness of the international law with preferences, then any number of variables might commonly cause both awareness and preferences.

The controls afforded in laboratory experiments let researchers create more variation that exists 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 dealing with the fact that certain treaties have been essentially universally adopted).\textsuperscript{239} In the real world of international law we believe that \textit{X} has an impact on \textit{Y}. But the variation in \textit{X} is quite small, making it difficult, perhaps impossible, to detect such a relationship. By manipulating \textit{X} in a hypothetical scenario we can begin to explore this relationship.

Laboratory experiments also enable very clear and direct measurement of subject decision-making or responses to stimuli. Ambiguity 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. Or, some experimenters in the American politics tradition have even sought to replicate the consumption

\textsuperscript{236} For a discussion on compensating subjects, and implications that it may have on experimental validity, see MORTON & WILLIAMS, supra note 35, at 259-291.

\textsuperscript{237} For a general discussion on the advantages of laboratory experiments, see MORTON & WILLIAMS, supra note 35, at 225-226, 305-307.

\textsuperscript{238} For a discussion on the advantage that laboratory experiments afford for controlling variation and the experimental environment, see Falk & Heckman, supra note 34, at 535.

\textsuperscript{239} See supra PART II.A.1.
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. In some social science experiments, researchers have begun to draw on measurement techniques used in the medical sciences such as galvanic skin response and functional magnetic resonance imaging, or measuring subject specific physical attributes like their genetic profiles. Such physiological recordings would be very difficult outside of the laboratory.

The decision to use a laboratory experiment depends on the type of stimulus/situation under investigation. A general rule is that 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 but also timing. For example, the implicit association test (IAT) measures prejudices and biases by comparing reaction times to different stimuli at the millisecond level. Greater control is also required when researchers 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 biased estimates if

---


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


the level of focus differs across experimental conditions.

Another common use of laboratory experiments is when researchers want to allow for interactions between subjects. For example, nearly all experiments testing game theoretic models are implemented in the laboratory, though this is changing with the advent of web based platforms. 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, predictions from formal game theoretic models can be compared to actual human behavior.

Experiments of this nature could be important in testing views on international law that emphasize strategic interaction between decision-makers.

While experiments with deception remain common in psychology, most economic labs eschew or formally prohibit deception on the grounds that deception could contaminate future experiments that rely on subjects not suspecting any deception. There exists considerable debate on this subject. In his classes on experimental political science, the author Tingley suggests avoiding deception if possible but not letting it get in the way of interesting research questions if one can identify a laboratory that will allow its use.

Laboratory experiments face a series of challenges. Perhaps the

---

248 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).
250 See, e.g., Engel, supra note 25 (using a laboratory experiment to show how expectations help shape how customary international law develop).
251 See DAVIS & HOLT, supra note 246, at 24 n.28.
252 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).
253 For an extended discussion on the ethical implications of conducting experiments that involve deception, see MORTON & WILLIAMS, supra note 35, at 362-378. See also McDermott, supra note 39, at 41.
254 See McDermott, supra note 39, at 40-41 (summarizing the challenges and
biggest are construct and external validity, due to that fact that most laboratory experiments involve convenience samples of college students (though this is changing, see below) in highly artificial settings.\textsuperscript{254} Simply put, the laboratory setting may not aptly capture the real world situation that the researcher is interested in. Another challenge is that experimenters may inadvertently induce bias by creating demand effects.\textsuperscript{255} 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 wanted to make such a choice, but because they wanted to conform to what the experimenter wants. Ultimately these limitations are the flip side of advantages.\textsuperscript{256} Careful control of the experimental setting helps to establish causal relationships and explore relationships impossible to examine in the real world but which we think are important.

B. Survey Experiments

A survey experiment takes surveys administered to individuals and inserts various types of randomizations.\textsuperscript{257} These can include randomized textual prompts/frames, question wordings, response options, or other stimuli like video feeds or audio clips.\textsuperscript{258} These manipulations let researchers explore how survey responses differ across different settings. Additional, non-experimental, questions of course can also be asked so that researchers have additional subject specific covariates.\textsuperscript{259} Survey disadvantages of experimental research).

\textsuperscript{254} Id.
\textsuperscript{255} Id. at 33-34.
\textsuperscript{256} Id. at 38-39.
\textsuperscript{259} 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
experiments can be delivered via professional polling firms260 or researcher created surveys using online platforms.261 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.262 Survey experiments are an efficient means for researchers to understand how individuals respond to information and stimuli, often with relatively large sample sizes compared to those of laboratory experiments.263 Many experiments that are conducted

---

260 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 field them to respondents through a variety of means for a fee.

261 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).

262 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 JUDGMENT & DECISION MAKING 5 411 (2010).

263 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. Charbris, Garga Chatterjee, & Jermeby B. Wilmer, Is the Web as Good as the Lab? Comparable Performance from Web and Lab in Cognitive/Perceptual Experiments, 10 PSYCHONOMIC BULL. REV. 847 (2012). For research documenting the performance of Mechanical Turk for economic research, see Horton et al., supra note 247 (providing a review of MTurk 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 MTurk 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
in a laboratory can be done in a survey experiment fashion, though 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 example by Tomz discussed above, 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. All respondents were randomly assigned being receiving various pro and con arguments about the policy. One of these pieces of information stated that the US had signed an international treaty prohibiting it from breaking off trade with Burma under international law. This piece of information was randomly assigned to subjects. Tomz found that individuals receiving information about international law were 17% points more likely to oppose restricting trade with Burma. Tomz also reports a follow-up study using British members of Parliament. 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 they were told the country had not signed the NPT. Finally, 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 and higher expectations in the control group. Hence the effect of international law on its own—not factors that might change the propensity of signing on to international law—were shown to have 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 forms. 

---

264 See Tomz, supra note 26. See supra text accompanying notes 103 – 104.
265 Tomz, supra note 26, at 13-21.
266 Id. at 18.
267 Id. at 24-28.
268 Id. at 27.
ways and a respondent answers questions about the issue. Another form of survey experiment is a “conjoint” survey popular in marketing research. In this type of survey, subjects are asked to evaluate a product, or some other quantity like an international agreement, but where attributes of the product are randomized with the subject evaluating the same quantity but under multiple different profiles of attributes. Reactions to multiple different attribute profiles enables a vast amount of information to be collected compared to the case where individuals only evaluate one profile of attributes. Finally, survey experiments can also be used to test the mechanisms that are hypothesized to link the experimental treatment/frame with a respondent’s stated preferences. This can be done by asking additional questions that measure changes in these intermediate variables. For example, Tomz and Weeks use a survey experiment to explore the connection between political regime type and willingness to go to war. 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.

As discussed before, survey experiments provide a cost effective way to collect a substantial amount of data. Survey experiments though face many of the same problems that standard surveys face. These include the fact that creating representative samples can be costly, it can be difficult to create panels in order to track an individual over time, and subjects might give socially desirable answers rather than their true opinions.

---


272 Id. at 19-28.

273 Tomz & Weeks, supra note 20.

274 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
In addition to these general problems, survey experiments potentially face some specific problems. If there are multiple experimental manipulations in a survey, then it is possible that there can be spill-over effects where treatment status in one part of the survey impacts responses to a different experiment. Construct validity concerns can of course also be an issue for survey experiments. Online surveys that ask questions about politics can only approximately simulate considerations about politics that the researcher is interested in. 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. 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 might be competing considerations that the survey experiment has abstracted away from that in the real world are present.


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 drawn from them. Field experiments minimize some of these concerns by examining behavior in the setting which it would naturally take place in.\footnote{For a comprehensive guide on how to conduct field experiments, see Gerber \& Green, supra note 200.} We suspect that most scholars of international law will gravitate towards using field experiments because of the realism they afford and connectedness to behavior directly relevant to international law.\footnote{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 24. 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 52.}

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.\footnote{See generally Gerber \& Green, supra note 200, at 131-210. See also Morton \& Williams, supra note 35, at 116.} While one can randomize the treatment conditions, it is sometimes difficult to ensure that the condition an individual is placed in is the one that they actually participate in. This might arise from conscious subterfusion or something more innocuous. In these settings researchers can leverage econometric techniques like instrumental variables estimation\footnote{See Morton \& Williams, supra note 35, at 106-108.} or report alternative quantities of interest, such as the intention to treat (ITT).\footnote{For more information on using ITT as an alternative quantity of interest, see Gerber \& Green, supra note 200, at 141-43.}

Field experiments have been remarkable successful in the study of American politics, especially in the study of voter turnout.\footnote{See generally Gerber \& Green, supra note 18 (summarizing 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 42 (conducting a large-scale field experiment on voters in Michigan to show that social pressure can increase voter turnout).} However, in recent years there have been field experiments that directly speak to topics

{\footnotesize \begin{itemize}
  \item[280] For a comprehensive guide on how to conduct field experiments, see Gerber \& Green, supra note 200.
  \item[281] 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 24. 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 52.
  \item[282] See generally Gerber \& Green, supra note 200, at 131-210. See also Morton \& Williams, supra note 35, at 116.
  \item[283] See Morton \& Williams, supra note 35, at 106-108.
  \item[284] For more information on using ITT as an alternative quantity of interest, see Gerber \& Green, supra note 200, at 141-43.
  \item[285] See generally Gerber \& Green, supra note 18 (summarizing 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 42 (conducting a large-scale field experiment on voters in Michigan to show that social pressure can increase voter turnout).
\end{itemize}
that interest scholars of international law.\textsuperscript{286} One prominent example is the influence of international observers on election fraud.\textsuperscript{287} In the first field experiment on democracy promotion, political scientist Susan Hyde conducted a large-scale experiment during the 2004 Indonesian presidential election to determine how international observers influenced voting.\textsuperscript{288} Hyde worked with the Carter Center to gain permission to randomly assign which villages their teams of election monitors would be sent to.\textsuperscript{289} These teams then monitored the voting conduct at 147 individual polling stations.\textsuperscript{290} After the election, Hyde was able to compare the results of randomly monitored polling places with the overall election results, and counter intuitively show that the incumbent candidate actually did better at polling stations that received election monitoring.\textsuperscript{291} Hyde’s experiment is interesting because it 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 hypothesis (like the belief that monitors would help the challenger) may not bear out after being tested in the field.

Another more recent example that most clearly engages with issues related to international law is the forthcoming work by Findley et al.\textsuperscript{292} As we previously mentioned, they conducted a field experiment by sending emails to 1,015 firms providing incorporation services that expressed an interest in anonymously incorporating a shell corporation.\textsuperscript{293} The experiment randomly varied whether information provided on the status of international law. Firms that responded by fail to request information on the senders’ identity were in violation of international law. The experiment is thus of interest because it was not only directly able to measure compliance, but also because it is a rare example of studying compliance at international law by private actors.

\textsuperscript{286} For a general overview of how field experiments can be used to study international relations, see Susan D. Hyde, \textit{The Future of Field Experiments in International Relations}, 628 \textit{ANNALS AM. ACADEMY POL. & SOC. SCI.} 72 (2010).

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

\textsuperscript{288} See Hyde, \textit{supra} note 52.

\textsuperscript{289} Id. at 516-17.

\textsuperscript{290} Id. at 517.

\textsuperscript{291} Id. at 517-20.

\textsuperscript{292} See \textit{supra} text accompanying notes 181 — 185.

\textsuperscript{293} Baradaran et al., \textit{supra} note 24, at 7.
Additionally, the work by Findley et al. is interesting from a practical standpoint because it 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, offer a practical means of engagement without the high expenses of direct face-to-face contact necessary in other field experiment designs. Of course, one

294 See, e.g., David E. Broockman, Black Politicians Are More Intrinsically Motivated To Advance Blacks' Interests: A Field Experiment Manipulating Political Incentives, AM. J. POL., SCI. (forthcoming 2013) (finding that black state legislators were more likely to respond to a request for assistant from a putatively black citizen that did not reside in their district than white state legislators); Daniel M. Butler, Christopher Karpowitz, & Jeremy C. Pope, A Field Experiment on Legislators' Home Style: Service versus Policy, 74 J. POLITICS 474 (2012) (finding that legislatures are more likely to respond to letters with service concerns than policy concerns); Gwynth McClendon, Co-ethnicity and Democratic Governance: A Field Experiment with South African Politicians, Working Paper (2012) (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) (providing 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) (finding that sending letters providing state legislatures with information from a public opinion survey on their constituents’ opinions affected their voting); Daniel M. Butler, Monitoring Bureaucratic Compliance: Using Field Experiments to Improve Governance, PUBLIC SECTOR DIGEST 41 (2010) (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) (submitting letters to the editor at 100 newspapers and determining 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) (working with a newspaper to randomly create websites devoted to the activities of a randomly selected group of delegates and finding no evidence that increased transparency impacts delegate performance in authoritarian Vietnam). 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 Gwynth H. McClendon, Ethics of Using Public Officials as Field Experimental Subjects, 5 EXPERIMENTAL POL. SCIENTIST 13 (2012).
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 non-response might be thought of as a quantity of interest. 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.

Along these lines, when deciding between experimental methods, researchers will have to 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. There are two reasons why researchers of international law might be interested in measuring mass opinion. The first is that 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. The second is that there is a growing body of research suggesting that elite opinions mirror mass opinions on questions of international affairs, and that thus mass surveys can be an excellent way to predict elite views. If, however, scholars wish to directly study elite behavior, they will likely either have to conduct a survey to a sample of elites or conduct field experiments. As we have previously mentioned,
field experiments have recently been used in exciting ways to directly test the views and responses of elites, including public officials, and may be a promising way forward for scholars of international law.

CONCLUSION

Social sciences and legal researchers are increasingly using 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, as we discussed in PART II, international law faces many of these same issues to the extreme, as well as problems that are unique to the field. As a result, although the empirical revolution in international law has been a welcome development, it is time that international law undergoes an experimental revolution as well.

* * *

British Parliament); Hafner-Burton et al., supra note 26 (conducting a survey on elites over trade preferences).

300 See supra note 294.