The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects

Yonatan Lupu George Washington University

The effects of international institutions on state behavior make up a key research agenda in international-relations scholarship. Because states self-select into treaties, we cannot infer that these commitments have causal effects unless we address this selection effect. I explain the significant limitations of the methods used thus far to overcome this problem and argue that a more effective approach must take into account states' treaty preferences. I describe a novel combination of ideal-point estimation and propensity-score matching that can estimate the probabilities of treaty commitment and use them to test hypotheses. I use this procedure to test the effects of three key international human-rights treaties. My results provide significant new findings regarding the effects of these important agreements. I show that the Convention on the Elimination of All Forms of Discrimination against Women has significantly improved respect for women's rights, but that the Convention against Torture and the International Covenant on Civil and Political Rights have not had significant effects on human rights.

The effect of international institutions on state behavior is among the most important issues in international-relations scholarship. Theorists have argued that making international commitments allows states to eschew short-term gains for long-term interests, thus facilitating behavior they would not have undertaken absent these obligations (Keohane 1984; Martin 1992). More recently, scholars have extended these theories and empirically tested their implications with respect to the effects of commitment to institutions governing human rights (Hafner-Burton and Tsutsui 2005, 2007; Hathaway 2002; Hill 2010; Keith 1999; Landman 2005; Neumayer 2005; Poe, Tate, and Keith 1999; Simmons 2009), the environment (Mitchell 1994; Mitchell et al. 2006; Stokke and Hønneland 2007; Victor, Raustiala, and Skolnikoff 1998; Weiss and Jacobson 1998), and international economic relations (Simmons 2000; Simmons and Hopkins 2005; von Stein 2005).

As many scholars have recognized, the inference we can draw from the relationship between treaty commitment and state behavior is threatened by a selection effect. Because governments are free to choose whether or not to commit to treaties, we cannot infer that treaty commitment has a causal effect on their behavior unless we address this selection effect. Several recent studies have adopted sophisticated methods to address this problem (Hill 2010; Neumayer 2005; Simmons and Hopkins 2005; von Stein 2005), the most appropriate and promising of which appears to be the matching approach proposed by Simmons and Hopkins (2005). Yet causal inference is limited even using this approach to the extent variables that affect treaty commitment decisions are omitted from the matching model.

I argue that existing applications of matching in this context (Hill 2010; Simmons and Hopkins 2005) have omitted a key factor: states' treaty commitment preferences. I propose a methodology to estimate these preferences in order to include them in these models and make a significant improvement to the ignorability assumption. States' treaty commitment decisions, when analyzed systematically, provide a significant source of information that can reveal their preferences toward individual treaties and predict commitments to those treaties with a high degree of accuracy. I adopt a spatial-model-based method

Yonatan Lupu is Assistant Professor of Political Science, George Washington University, Monroe Hall, 2115 G St. NW, Washington, DC 20052 (ylupu@gwu.edu).

This research was supported in part by a Doctoral Dissertation Improvement grant from the National Science Foundation Division of Social and Economic Sciences (SES-1064108). For comments on previous drafts, I thank George Downs, Chris Fariss, James Fowler, Erik Gartzke, Emilie Hafner-Burton, Daniel Hill, Dan Hopkins, Kosuke Imai, Miles Kahler, David Lake, Paul Poast, Keith Poole, Christina Schneider, Erik Voeten, and Langche Zeng. Replication data and code are available on the AJPS Dataverse page.

American Journal of Political Science, Vol. 57, No. 4, October 2013, Pp. 912–925

 DOI: 10.1111/ajps.12033

designed for the analysis of legislative roll-call voting, W-NOMINATE (Poole and Rosenthal 1997), and demonstrate how it can be fruitfully applied in this context. After making a theoretical argument in favor of this methodology, I use Monte Carlo simulations to demonstrate that, in estimating treaty commitment probabilities, it outperforms a traditional model based on observable variables when significant unobservable factors affect treaty commitment decisions.

The article continues by demonstrating how this methodology can be applied to estimate the effects of commitments to three *key international human-rights agreements:* the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), the Convention against Torture (CAT), and the International Covenant on Civil and Political Rights (ICCPR). Using W-NOMINATE on an original data set of commitments to universal treaties, I estimate states' probabilities of commitment to each of these treaties. With these probabilities, I create matched samples that are well balanced on the probability of selection and use these samples to test the effects of treaty commitment. These are among the most important and frequently analyzed human-rights treaties, and empirical findings regarding their effects have found mixed—and often controversial—results. With respect to the CAT and ICCPR, several studies have found that treaty members are more likely to abuse the human rights of their citizens (Hafner-Burton and Tsutsui 2005; Hill 2010; Neumayer 2005), a counterintuitive result that has created an important puzzle regarding why and how commitment to these treaties might worsen respect for human rights.

My findings demonstrate that these prior results may have been due to an insufficient accounting for treaty commitment selection effects. With respect to both the ICCPR and CAT, I find that treaty commitment has not significantly affected human-rights practices. For several possible reasons, human-rights abusers appear to be more likely to join such treaties (Goodliffe and Hawkins 2006; Hathaway 2007; Vreeland 2008), yet joining these treaties does not appear to have a causal effect. In other words, there is a selection effect, but not a treatment effect. While this indicates that these treaties have not succeeded in improving human-rights practices, which is disappointing to proponents of international law, it also indicates that these treaties have not made matters worse, as several prior studies indicate. With respect to the CEDAW, I find robust results showing that treaty ratification has led to improvements in respect for women's political, economic, and social rights. These findings provide a strong indication, in contrast to prior results (Hill 2010), that commitments to this treaty have improved the lives of women across areas of human rights.

This article makes three significant contributions to the literature. Most importantly, it points to and provides a methodology for measuring an important factor that predicts treaty commitment decisions and that has been omitted from existing analyses. Including state preferences in the analysis of treaty effects has the potential to significantly improve our ability to empirically test theories of international institutions. Second, this article demonstrates how ideal-point estimation can productively be used outside of the legislative (or quasi-legislative) context and explains how it can be combined with matching to improve ignorability assumptions. To my knowledge, this is the first article to use these methods in this manner. Finally, by applying this approach to test the effects of the CEDAW, CAT, and ICCPR, this article provides robust empirical results that will hopefully reduce the existing uncertainty regarding the impact of these important treaties and improve our understanding of international law and institutions.

The Treaty Commitment Selection Effect

As Downs, Rocke, and Barsoom (1996) pointed out, and many social scientists have accepted, international institutions are often endogenous to state interests. The fact that states design and opt in to institutions creates a selection effect that threatens the validity of any inference we might draw from the relationship between institutions and states' behavior. There are two aspects to this problem: (1) both the nature of the underlying problem and state preferences affect the design of an institution; and (2) state behavior with respect to the activities required or proscribed by an international institution may result strictly from underlying state preferences and cannot be assumed to be the result of the institution itself. I am concerned here primarily with the second of these problems.

Commitment decisions are the key explanatory variables when studying the effects of treaties, and unless we address this selection effect, we would be assuming treaty commitment decisions are random, which they surely are not.[1] With respect to some treaties, for example, a high rate of compliance may not be caused by the treaty but by the fact that compliance requires states to take actions they would take in any event (Downs, Rocke, and Barsoom 1996; Simmons 1998), meaning that the treaty has no independent effect on state behavior. In such a

[1] As Simmons and Hopkins argue, "Random assignment would imply a theory of frivolous commitment-making, hardly a model on which a useful theory of compliance with legal obligations can be developed" (2005, 624).

scenario, a model that does not account for this effect will tend to overestimate the effect of treaty commitment.

To address this problem, several important studies have adopted Heckman (1979) selection models (Neumayer 2005, 2007; von Stein 2005) or an instrumental-variables approach (Simmons 2009). These approaches are subject to several important problems in this context. First, Heckman models are generally highly sensitive to distributional assumptions and the specification of the selection equation (Liao 1995; Winship and Mare 1992). Second, the Heckman sample-selection model assumes incidental truncation (i.e., sample selection) and only addresses that truncation mechanism. The model is unnecessary (and often inappropriate) when the outcome variable can be observed for units that did not select into the sample (as is the case for many dimensions of government policy). Third, both the Heckman and instrumental-variables approaches require the analyst to specify a variable that is correlated with treaty commitment but is independent of the subsequent outcome. A Heckman model without such a variable is identified based only on assumptions about the distributions of residuals (which are generally not based on theory) and not about the variation in the explanatory variables (about which theories are based; (Sartori 2003)). In an instrumental-variables model, the need to specify such variables is the core of the approach. This problem becomes acute in the treaty commitment context, in which the decision to commit is made by governments, and the outcome variable is also a measure of government practice. Because the same actors often make both decisions, it is more likely that the same factors influence both decisions (Powell and Staton 2009), making finding a good instrument or exclusion restriction a significant problem. As Hill argues in the human-rights context, "States that commit are different from states that do not, and these differences contribute to the decision to select into a 'treatment.' These differences also contribute to their decision to repress their citizens, which makes it difficult to separate the effect of the treaty from the effect of the institutional features that led them to ratify" (2010, 1168).

An alternative approach proposed to address the treaty commitment selection effect is propensity-score matching (Simmons and Hopkins 2005). While it is infeasible to randomly assign treaty commitment, we can make causal inference with respect to the relationship between commitment and compliance if we can calculate and account for each state's probability of commitment. By first calculating the probability of selection, it is possible to match states into a sample that consists of sets of units that have and have not made the selection in question, with both sets having equivalent probabilities of having done so. Such a sample is designed to approximate random assignment to treatment (Ho et al. 2007). The propensity-score method calculates this probability based on a set of observed pretreatment characteristics that are theorized to affect the selection decision (Rosenbaum and Rubin 1983, 1984).

Matching has significant advantages over the methods discussed above, although it also has limitations. First, it can be used to control for bias on observable variables without the distributional assumptions required to sustain models such as that proposed by von Stein (2005). Second, matching is a useful method for creating a quasi-experiment by sampling similar treatment and control groups from a larger pool of such units (Morgan and Winship 2007).

Simmons and Hopkins (2005) advocate that propensity-score matching in this context be conducted using the following procedure. First, the analyst should theorize about which variables predict treaty commitment and use those to estimate the probability of treaty commitment. As a second stage, they suggest that treated (treaty members) and untreated (nonmembers) should be matched using methods informed by theory and the first-stage model. Using this procedure, the extent to which we can make causal inference depends in large part on the ability to correctly estimate the probability of treaty commitment. The estimation of the treatment effect in subsequent stages of the research design is highly sensitive to omitted variable bias in the estimated propensity scores (Arceneaux, Gerber, and Green 2006; Rubin 1997), and several studies have shown that the choice of underlying variables significantly affects the reliability of propensity-score analysis (Heckman et al. 1998; Heckman, Smith, and Clements 1997; Lechner 2000; Smith and Todd 2005). Simmons and Hopkins therefore recognize that the presence of unobservable determinants of treaty commitment creates a threat to inference, so "every effort should be made to theorize and to include in the commitment model all observables theory suggests are relevant, and an effort should be made to theorize and measure purported 'unobservables' as well" (2005, 627). This article attempts to do just that.

Estimating Treaty Commitment Preferences

I argue that the existing applications of matching to overcome the treaty commitment selection effect overlook a key factor that predicts these commitments: states' preferences with respect to treaties. The core of the question regarding the effects of international institutions is whether

they have the power to cause governments to undertake actions they would not otherwise undertake. Thus, when we test the effects of treaty commitment on some dimension of state behavior, we can only make causal inference from this finding if we account for the fact that some states have stronger underlying preferences for the treaty than others. This notion is implicit in the existing methods for overcoming the treaty commitment selection effect, yet these methods do not attempt to estimate preferences directly. Existing applications of propensity-score matching, for example, use more readily observable characteristics such as regime type and income to estimate the probability of treaty commitment (Hill 2010; Simmons and Hopkins 2005). Part of the intuition behind using these variables is the notion that democracies have different preferences from autocracies, rich countries different preferences from poor countries.

Existing studies have implicitly considered treaty commitment preferences unobservable, yet I argue that we can estimate them. Through their histories of treaty commitment decisions, states reveal significant information regarding their underlying preferences. This information can be used to estimate state preferences with respect to treaties and accurately predict state commitment to individual treaties. In turn, we can use these estimates to calculate predicted probabilities of treaty commitments and use these in a matching approach similar to that proposed by Simmons and Hopkins (2005).

In order to measure state preferences with respect to treaties, I rely on the spatial model of political choice (Davis, Hinich, and Ordeshook 1970; Downs 1957). The basic notion behind implementations of the spatial model is that, by observing the choices political actors make, we can measure their preferences relative to each other and relative to the options with which they are faced. In political science research, this model has largely been used to estimate the preferences of legislators with respect to roll-call voting decisions. By estimating the ideal points of legislators in this way, analysts can determine the probabilities of legislators voting for and against certain bills.

Both the ideal-point model (and its implementations using multidimensional scaling algorithms) and propensity-score matching are tools for reducing the dimensionality of data, so their use in combination is intuitive to some extent. The following is a more formal explanation of how these tools fit together. Suppose that a government makes treaty commitment decisions based on four sets of factors: (1) observable characteristics of the country (e.g., the country's regime type) (O_1); (2) observable characteristics of the treaty (e.g., the subject matter of the treaty) (O_2); (3) unobservable characteristics of the country (e.g., latent preferences for multilateral cooperation generally and latent preferences for certain types of treaties) (U_1); and (4) unobservable factors that are both country-specific and treaty-specific (U_2). Thus, in a given period of time, the decision Y of country i to ratify treaty j is based on the following model:

$$Y_{ij} = \beta_0 + \beta_1 O_{1i} + \beta_2 O_{2j} + \beta_3 U_{1i} + \beta_4 U_{2ij} + \varepsilon_{ij} \quad (1)$$

In a standard matching approach, both U_1 and U_2 will be omitted from the model, resulting in a strong ignorability assumption with respect to both such factors. My approach attempts to incorporate U_1 in order to improve this assumption and reduce omitted variable bias. An ideal-point estimation procedure attempts to reduce O_1, O_2, and U_1 to a specified number of dimensions (usually two) resulting in a set of ideal-point estimates θ. Ideal-point estimation is able to do this because all three of these factors are either country-specific or treaty-specific and thus their impact on treaty commitment decisions is systematically detectable in the treaty commitment data. By contrast, U_2 is both country- and treaty-specific, so its effects on treaty commitment decisions are too ad hoc to allow for dimensionality reduction. Formally, ideal-point estimation models the decision Y of country i ratifying treaty j as follows:

$$Y_{ij} = \alpha_j + \gamma_j \theta_i + \varepsilon_{ij}, \quad (2)$$

where α_j is an intercept for the treaty, γ_j is a discrimination parameter indicating how well the treaty discriminates between different types of countries, and the error includes the effects of U_{2ij} and any stochastic component. Having estimated the ideal points in this fashion, it is subsequently possible to estimate the probability of each country committing to each treaty, subject to the utility function we choose to impose upon the data (as discussed below).

I use the spatial model to estimate the preferences of states with respect to universal treaties. In my model, the options of committing and not committing to a treaty are represented by points in an n-dimensional policy space. Each state decides whether or not to commit to a treaty by weighing the distance between these points and its ideal point in this space. Simmons has recently suggested that this logic applies to treaty commitment decisions: "To use the language of spatial models, the nearer a treaty is to a government's ideal point, the more likely that government is to commit" (2009, 65, emphasis omitted). I adopt the language of Simmons (2009) in arguing that there are both false-negative and false-positive treaty commitment decisions. False negatives are occasions in which states do not commit to treaties that seem well aligned with their preferences. False positives are occasions on which states

do commit to treaties that seem to be contrary to their preferences. Some false positives may be empty promises, while others may result from factors such as uncertainty over long-term consequences. I follow Simmons (2009) in arguing that treaty commitments are sincere in the aggregate and therefore can reveal important information regarding preferences.

Thinking of treaty commitment decisions in this way allows for the use of spatial methods traditionally used to analyze other dichotomous choices, most importantly the methods used to study legislative roll-call voting. Specifically, I use the W-NOMINATE multidimensional scaling method to estimate states' treaty commitment preferences (Poole and Rosenthal 1997).[2] W-NOMINATE is a random-utility model of Euclidean spatial voting (Enelow and Hinich 1984; Hinich and Munger 1994, 1997) that assumes each actor assigns a utility to each of two options. This utility is determined both by the distance between the actor and the options as well as a stochastic error term.

Poole and Rosenthal (1997) created W-NOMINATE as a tool for analyzing legislatures and used it to study the history of roll-call voting in the U.S. Congress. Other scholars have used W-NOMINATE estimation to study such areas as the repeal of the Corn Laws (Schonhardt-Bailey 2003), the Confederate legislature (Jenkins 1999), the European Parliament (Hix 2001; Noury 2001), and various national legislatures (Londregan 2000; Morgenstern 2003). In addition, many scholars have used the distances between points in the W-NOMINATE space for various purposes, including analyzing party cohesion (Desposato 2008), testing ideological compatibility differentials on party membership (Desposato 2006), measuring party polarization (Howell and Lewis 2002), and measuring the benefits associated with the differences between voting options (Rothenberg and Sanders 2000). W-NOMINATE has also been applied in the international context, particularly to analyze voting by states in the United Nations General Assembly (Reed et al. 2008; Voeten 2000).

Most scholars who use ideal-point estimation focus on explaining the latent dimensions of the preference space and any cleavages that exist in that space (Poole and Rosenthal 1997; Voeten 2000). Another important aspect of ideal-point estimates is that they can be used to predict actors' choices based on the relative locations of the actors and choices in the preference space. In the legislative context, the closer a bill is to a legislator's ideal point, the more likely the legislator is to vote for the bill. In the treaty context, the closer a treaty is to a state's ideal point, the more likely the state is to ratify the treaty. These probabilities are analytically similar to propensity scores, although they are based on direct estimates of state treaty commitment preferences. W-NOMINATE places each treaty in two locations in the preference space: one indicating the location of treaty ratification and one indicating nonratification. Using these locations, W-NOMINATE calculates the probability that state i ratifies treaty j as follows:

$$P(Ratify)_{ij} = \frac{\exp[u_{ijr}]}{\exp[u_{ijr}] + \exp[u_{ijn}]}, \tag{3}$$

where u_{ijr} is the deterministic component of the state's utility from ratifying the treaty, and u_{ijn} is the deterministic component of the state's utility from not ratifying the treaty. These utilities are calculated as follows:

$$u_{ijr} = \beta \exp\left[\frac{-\omega^2 d_{ijr}^2}{2}\right] \tag{4}$$

$$u_{ijn} = \beta \exp\left[\frac{-\omega^2 d_{ijn}^2}{2}\right], \tag{5}$$

where d_{ijr} is the distance between the state's ideal point and the location of treaty ratification, d_{ijn} is the distance between the state's ideal point and the location of treaty nonratification, ω is a weight parameter estimated by W-NOMINATE, and β is a signal-to-noise ratio estimated by W-NOMINATE (Poole and Rosenthal 1997).[3]

These scenarios will often occur in the treaty commitment context. While scholars have learned a significant

[2] An important debate exists in the legislative-studies literature regarding the most appropriate method to use in that context. In part, the debate revolves around whether it is more appropriate to assume legislators have Gaussian utility functions (Poole and Rosenthal 1997) or quadratic utility functions (Clinton, Jackman, and Rivers 2004). This article is not intended to contribute to that debate. I am not aware of any theories that address the question of whether governments have normal or quadratic utilities or whether we can safely assume some distribution on their errors. As a result, it is not immediately apparent whether one ideal-point estimation method is more appropriate than another in the context of treaty commitment decisions. Ultimately, the choice of method may not significantly affect the results as IDEAL estimates are often very similar to W-NOMINATE estimates. Carroll et al. (2009) find, for example, that the ideal points estimated by W-NOMINATE and IDEAL for the 106th U.S. House of Representatives have a correlation coefficient of .996.

[3] In the online supporting information, I use Monte Carlo simulations to demonstrate that ideal-point estimation is an effective method of estimating the probability of states committing to treaties. The simulations demonstrate that, only under very limited conditions, estimation of treaty commitment probabilities based only on observable variables known to affect those decisions is the more efficient method. One such situation is when we know and can measure all of the predictors of treaty commitment without error or bias. A second is when the unobservable, unmeasurable, or unknown predictors explain relatively little of the variance in those decisions. As the explanatory power of unobservable variables increases, the estimates of treaty-ratification effects based on the observable variables become significantly more biased, demonstrating the risk of using such a model.

amount about the factors that drive treaty commitment, we should not assume that we can fully—or nearly fully—model these decisions. One problem is that factors that have not yet been theorized may be driving certain treaty commitment decisions; it is a strong assumption to argue that there are no "unknown unknowns." The accuracy of treaty commitment probabilities estimated using observable data drops quite rapidly when unobservables become increasingly significant, meaning that, by assuming unobservables do not exist, we take a significant chance of producing unreliable estimates. A second problem is that, even when we can theorize about a particular concept that affects treaty decisions, we may not have access to direct measures of it. Simmons and Hopkins (2005), for example, note that political will is an important factor that explains treaty commitment and which is difficult to observe or measure. They use three reasonable proxy measures for political will, yet it is unlikely they have measured this factor completely. Third, even when we have direct measures of a certain concept known to affect treaty commitment, it is often the case in the international-relations data that we have reason to question whether the measures are free of error or bias. Finally, correct specification of the model predicting treaty commitments necessitates not only including the correct variables but also specification of the correct functional form, including the inclusion of any polynomial and interaction terms. Yet theory often does not provide sufficient guidance with respect to how to configure these variables (Smith and Todd 2005). In sum, it is unrealistic and risky to assume that we can know, observe, and properly measure enough of the factors that explain treaty commitment decisions to meet the conditions under which using these variables to estimate treaty commitment probabilities is more efficient than using the spatial-modeling approach.

I therefore generally recommend using the ideal-point estimation method in this context. It is also possible to combine matching based on other observables with the procedure outlined here. For example, if we believe that regime type predicts treaty commitments in a way not captured by the ideal-point estimates, we could use both factors to match treaty members to nonmembers. In certain situations, doing so may further reduce omitted variable bias and improve the ignorability assumption. There are certain risks to this approach. First, an incorrectly specified variable would likely increase the bias in the matching model. Second, any bias in the measurement of this variable would also translate into bias in the matching model. Finally, adding such a variable to a parametric matching model may introduce bias if the functional form is incorrectly specified.

Consequently, I recommend that analysts generally follow a three-step procedure in this context (for a similar procedure in a broader context, see Guo and Fraser 2010). First, the probability of each country's commitment to a treaty should be estimated for the full sample of applicable country-years using ideal-point estimation. Only treaties open to signature by the full set of countries should be included in the analysis. In the second stage, matching should be performed based on these probabilities of treaty commitment. In some situations, it may be preferable to include additional observable variables known to affect commitment to the applicable treaty in the model, keeping in mind the risks of doing so. Third, postmatching analysis should be conducted on the matched sample, generally using multivariate regression analysis. Such a model should include controls for other variables believed to affect the outcome variable.[4]

This methodology can be used with respect to any set of multilateral treaties. The most straightforward application is to universal treaties, which are open to all states, but the method can be used to measure the preferences of states over other multilateral treaties so long as all states in the sample are de jure eligible to join all of the treaties. Multilateral treaties offer a particularly interesting area in which to develop and test theories of international institutions. As many scholars have argued, compliance is more difficult to achieve in the multilateral setting. For example, it is much more difficult to rely on tit-for-tat strategies in the multilateral setting because it is unclear which of the many members of the treaty is charged with retaliation (Oye 1986). In such a setting, if one state violates the agreement, others will fear that if they retaliate (and bear the cost of doing so), other states will free ride instead of also retaliating (Goldsmith and Posner 2005). Second, multilateral treaties often include states with many different characteristics, allowing for a comparison of the effects of these treaties on different types of states. Finally, multilateral treaties include many of the most controversial and salient areas of international cooperation, and it is, therefore, critical for political scientists to better understand their effects.

Applications: Three Human-Rights Treaties

To demonstrate how this methodology can be used, I test the effects of commitment to three human-rights treaties:

[4] Some matching algorithms, such as optimal matching, will require special regression adjustments in this stage.

the CEDAW, CAT, and ICCPR. The CEDAW, which was adopted in 1979 by the United Nations General Assembly and came into force in 1981, prohibits discrimination against women with respect to a broad set of political, economic, and social rights. The CAT specifically addresses government-sponsored torture. It was adopted in 1984 and came into force in 1987. The ICCPR, by contrast, protects individuals from a broad range of government abuses, including various personal-integrity rights. It was adopted in 1966 and came into force in 1976.

Prior Results. Empirical findings regarding the relationship between commitment to the CEDAW and women's rights have been mixed. Hafner-Burton and Tsutsui (2005) found that commitment to the CEDAW is significantly correlated with an increase in the level of state repression (broadly defined to include other human-rights violations). Gray, Kittilson, and Sandholtz (2006), however, found that CEDAW commitment is associated with higher levels of female life expectancy and lower levels of female illiteracy. Using the instrumental-variables approach, Simmons (2009) found that CEDAW commitment results in a more even ratio of girls to boys in primary and secondary education and in greater access to family-planning resources. Most recently, using a matching approach based on observable predictors of treaty commitment, Hill (2010) found that CEDAW commitment has a positive and significant effect on state respect of women's political rights, but no effect on social and economic rights.

Tests of the effects of the CAT have produced arguably the most surprising and controversial findings in this literature. Beginning with Hathaway (2002), several studies have consistently found that CAT ratification is associated with higher rates of torture (Neumayer 2005), although Landman (2005) found the opposite result. Most recently, Hill (2010) also found a negative effect of CAT ratification in an analysis that used a measure of torture provided by the Cingranelli-Richards Human Rights Data Project (CIRI 2009). These counterintuitive results have led to two puzzles researchers have attempted to address. The first is why governments of states that conduct torture would choose to ratify the CAT (Goodliffe and Hawkins 2006; Hathaway 2007; Vreeland 2008). The second, far more challenging, puzzle is how, if this relationship is truly causal, ratification of the CAT leads to higher rates of torture (Hollyer and Rosendorff 2011).

Findings with respect to the effects of ICCPR ratification have also been mixed. While Keith (1999) and Hathaway (2002) found no significant effects of ICCPR membership, and Landman (2005) found it led to improvements, Hafner-Burton and Tsutsui (2005) found that ICCPR members were more likely to conduct personal-integrity-rights violations. This finding, indicating that membership in this treaty may actually make human-rights practice worse, was supported by Neumayer (2005), which uses a Heckman selection model. Using a matching technique, Hill (2010) likewise finds that ICCPR members become more likely to violate their citizens' personal-integrity rights, as measured by the index of physical-integrity rights provided by CIRI.

Selection Effects. While part of the reason for these mixed findings may be differing choices regarding estimation model and dependent variable, the results are also due in part to selection effects. Using the approach outlined above, I test the effects of ratification of these treaties. I first collected a data set of states' decisions with respect to commitment to universal treaties. I collected the data from the United Nations Treaty Collection (UNTC), an online database that provides information regarding all treaties deposited with the UN Secretary-General. I analyzed the set of treaties hosted by the UNTC to determine which are de jure open to all states and which are limited to a specific set of states, such as on a regional basis. The latter are excluded from the analysis. The UNTC includes conventions, treaties, protocols to treaties, and treaty amendments. If a list of signatories is included for any of these types of instruments, I include it in my data set as a separate choice. I do this because each item reflects a separate decision made by states, regardless of whether the item amends a previous choice. For simplicity, I will refer to each such item as a "treaty" in this article.[5] For each treaty, I have thus created a matrix consisting of all of the states in the international system and an indication of whether or not they ratified the treaty. If a state has ratified a treaty as of a given year, I have coded that state as a "1" with respect to that treaty; otherwise the state is coded as a "0." Using these data, I created a matrix for each year between 1950 and 2007 that indicates, for each treaty then in effect, which states then in existence had ratified the treaty as of the end of the year. The supporting information provides certain summary statistics of the treaty data set.

With these data, I use a two-dimensional W-NOMINATE model to estimate the probabilities of ratification of the CEDAW, CAT, and ICCPR on a country-year basis. For each treaty, I perform nearest-neighbor matching using the MatchIt package in the R programming language (Ho et al. 2011) to match the country-years that ratified the treaty with those that did not do so, following the procedure described by Ho et al. (2007). For the

[5] A full list of these treaties is available from the author upon request.

TABLE 1 Balance Statistics

	CEDAW		CAT		ICCPR	
	Full	**Matched**	**Full**	**Matched**	**Full**	**Matched**
Sample Size	4368	1652	3519	1580	4368	1966
Treatment Units (Treaty Members)	3139	826	2252	790	2947	983
Control Units (Nontreaty Members)	1229	826	1267	790	1421	983
Mean Pr(Ratification) - Treatment Group	0.911	0.662	0.760	0.394	0.882	0.647
Mean Pr(Ratification) - Control Group	0.484	0.683	0.230	0.358	0.481	0.654
Percentage Improvement in Balance	94.90%		94.49%		98.23%	

CEDAW and ICCPR, I include in the full sample country-years from 1981 to 2007. The first year for which data on the dependent variables are available (as discussed below) is 1981. For the CAT, I begin the full sample in 1987, which is the year the treaty took effect. Table 1 sets forth the results of the matching stage. The online supporting information provides additional information regarding the matched samples. In the full samples, the mean probability of commitment for the treatment group is much higher than for the control group. In the matched samples, however, these probabilities are significantly more well balanced.

Dependent Variables. As dependent variables, I follow Hill (2010) in using the measures of human-rights practices provided by CIRI. For women's rights, I use the CIRI measures of women's political, economic, and social rights. The political-rights measure is based on factors such as women's rights to vote, run for office, and petition government officials. The economics-rights measure is coded based on rights such as nondiscrimination in the workplace and equality in hiring, promotion, and pay. The social-rights measure includes rights such as equal inheritance, marriage and divorce rights, and education. These measures are coded on ordinal scales that range from 0 to 3, with higher scores indicating greater levels of respect for rights. A score of 3 indicates the country provides the full set of rights, a 2 indicates the country provides some rights that are effectively enforced, a 1 indicates the country provides some rights that are unenforced, and a 0 indicates the country does not provide these rights. With respect to the CAT, I use the CIRI measure of torture, which is coded on an ordinal scale ranging from 0 to 2. A score of 2 indicates there were no incidents of state-sponsored torture in the country that year, a 1 indicates there were few such incidents, and a 0 indicates torture was practiced frequently. To test the effects of ICCPR on personal-integrity rights, I use the Physical Integrity Index provided by CIRI. This index is an additive scale of measures of four personal-integrity-rights violations: torture, extrajudicial killings, political imprisonment, and disappearances. Each component is coded in a manner analogous to the coding for torture (i.e., an ordinal scale from 0 to 2), and thus the values of the resulting index range from 0 to 8.

Controls. A perfectly balanced sample approximates random assignment to treatment and therefore simple t-tests can often be used on such samples. My samples are not perfectly balanced with respect to the probability of assignment to treatment and are not completely balanced with respect to several additional factors that may influence human-rights practices, as shown in Table 2. To address remaining imbalance between the treatment and control groups, I use ordered-probit models to test the effects of treaty commitment, while controlling for these variables. With respect to all three treaties, I control for several factors believed to affect human-rights practices. Independent domestic courts can perform important enforcement functions with respect to human rights, particularly when protections have been incorporated into domestic law (Keith 2002; Keith, Tate, and Poe 2009; Powell and Staton 2009; Simmons 2009). As a measure of judicial independence, I adopt the data provided by CIRI (JUDICIAL INDEPENDENCE), which are coded as 0 for "not independent," 1 for "partially independent," and 2 for "generally independent." A series of studies has found that democracies are more likely to respect a range of human rights (Davenport 1995, 1999, 2007; Poe and Tate 1994; Poe, Tate, and Keith 1999). Using data from the Polity IV project (Marshall and Jaggers 2002), I therefore control for regime type (POLITY). Human-rights practices may also vary between old and new regimes, and I follow Hafner-Burton and Tsutsui (2007) by controlling for this factor using the Polity IV data (REGIME DURABILITY). States experiencing either internal or external wars may

Table 2 Balance Statistics—Controls, with Mean Values Reported for Treatment and Control Groups

	CEDAW		CAT		ICCPR	
	Treatment	Control	Treatment	Control	Treatment	Control
Judicial Independence	1.10	1.18	0.95	1.07	1.03	1.09
Polity	−0.93	−1.34	0.78	0.08	0.17	−0.87
Regime Durability	20.75	29.34	21.07	15.41	16.53	24.46
Civil War	0.23	0.22	0.25	0.18	0.25	0.19
External War	0.04	0.04	0.01	0.03	0.03	0.04
GDP Per Capita (logged)	7.04	7.69	6.99	7.21	6.84	7.48
Population (logged)	15.71	15.76	15.98	15.69	15.54	15.88
INGOs	464.16	539.68	484.58	477.50	372.88	543.47

be more likely to repress human rights (Poe and Tate 1994; Poe, Tate, and Keith 1999), so I control for this using data from the Correlates of War Project. The number of international nongovernmental organizations (INGOs) active in a country may affect the government's human-rights practices, so I control for this factor using the data provided by Hafner-Burton and Tsutsui (2005). More economically developed states may be less likely to repress their citizens (Poe and Tate 1994; Poe, Tate, and Keith 1999), and I control for this using a measure of per capita GDP provided by the World Bank. I use the natural log of this measure because this effect is likely nonlinear (Davenport 2007). To address potential differences among states of different sizes and potential monitoring biases based on this factor, I follow much of the literature in including a control for the natural log of a state's population, using data provided by the World Bank. To address serial correlation, I include lags of the applicable dependent variable for years $t-1$ and $t-2$ for women's rights. With respect to torture and personal-integrity rights, I include a lag for year $t-1$. Lagrange multiplier tests indicate that additional lags are not necessary to address serial correlation.[6]

Not surprisingly, there were many observations with missing data among these variables. Because the underlying reasons for the missingness of the data are likely nonrandom, listwise deletion of these observations may result in biased inference (Little and Rubin 1987). I therefore follow Hill (2010) and others in imputing the missing values using the Amelia II Program (Honaker, King, and Blackwell 2008).[7] In all models, I report standard errors that are robust toward arbitrary heteroskedasticity and that take into account clustering by country. All of the models also include fixed effects for the year of the observation. Table 3 reports the results of these models.

The results of these models differ in important ways from those of prior studies. The clearest comparison may be to the results reported by Hill (2010) based on models that use a similar matching procedure but do not balance based on treaty commitment preferences. With respect to the CEDAW, while Hill (2010) found that it improves respect for women's political rights, I find that it also improves respect for women's economic and social rights. These findings are based on the fullest accounting to date for treaty commitment selection effects and thus provide a good indication that the treaty has made an important positive impact on women's lives. Figure 1 reports the marginal effects of CEDAW ratification on women's rights, based on the models reported in Table 3. The figure indicates that not only is the impact of CEDAW membership statistically significant, but it also has a fairly substantial impact on respect for women's rights. Across all three categories of rights, CEDAW members become much more likely to provide either limited or full and enforceable rights. Similarly, across all categories, CEDAW members become less likely to only provide unenforced rights or no rights at all. The results are therefore encouraging both for those who seek to improve global respect for human rights and those who believe international institutions can be effective tools for accomplishing these goals.

With respect to the CAT, my results contradict those of most previous studies, including Hill (2010). My finding is that ratification of the CAT does not significantly affect torture rates, which stands in sharp contrast to prior findings indicating that such ratification may increase torture. These prior results may have been due to a less

[6]The results reported below with respect to torture and personal-integrity rights are robust to the additional inclusion of a lagged dependent variable for year $t-2$.

[7]I conducted the imputation procedure using the full set of country-years (rather than the matched sample) because including the full data allows for more accurate imputation.

TABLE 3 Effects of Treaty Ratification: Ordered-Probit Models

	CEDAW			CAT	ICCPR
	Political Rights	Economic Rights	Social Rights	Torture	Physical Integrity
Treaty Ratification	0.234**	0.171**	0.265***	−0.091	−0.032
	(0.078)	(0.066)	(0.064)	(0.077)	(0.062)
Judicial Independence	0.0386	0.114*	−0.0313	0.136	0.195***
	(0.059)	(0.057)	(0.056)	(0.071)	(0.0501)
Polity	0.013*	0.013*	0.030***	0.009	0.024***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.005)
Regime Durability	0.000	0.004**	0.004**	0.000	0.003
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Civil War	0.024	−0.227**	−0.157*	−0.426***	−0.903***
	(0.072)	(0.085)	(0.075)	(0.099)	(0.099)
External War	−0.085	−0.047	−0.264	0.033	−0.071
	(0.137)	(0.216)	(0.160)	(0.376)	(0.132)
GDP Per Capita (logged)	−0.087*	0.039	−0.001	0.091**	0.062*
	(0.035)	(0.037)	(0.034)	(0.039)	(0.031)
Population (logged)	−0.022	−0.047	−0.030	−0.186***	−0.127***
	(0.029)	(0.028)	(0.026)	(0.030)	(0.024)
INGOs	0.000*	0.000	0.000	0.000	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Rights_{t-1}	1.145***	0.765***	0.953***	1.008***	0.481***
	(0.102)	(0.090)	(0.079)	(0.073)	(0.021)
Rights_{t-2}	0.595***	0.480***	0.559***	–	–
	(0.097)	(0.061)	(0.074)		
Fixed Effects for Year	Yes	Yes	Yes	Yes	Yes
n	1652	1652	1652	1580	1966

Note: Robust standard errors in parentheses.
*p < 0.05, **p < 0.01, ***p < 0.001.

complete accounting for selection effects than the procedure I have outlined provides. In other words, prior results appear to have found that governments of states that ratify the CAT are more likely to torture, yet my results indicate that this relationship is not likely to be causal. While this finding certainly leaves us with the puzzle of why states that conduct torture would more often choose to ratify the CAT, it renders moot the more difficult puzzle of how treaty ratification might cause more torture.

On a similar note, my results contradict several recent findings that ICCPR ratifications are associated with increases in personal-integrity-rights violations (Hafner-Burton and Tsutsui 2005; Hill 2010; Neumayer 2005). Those findings had left us with a similar puzzle as the findings with respect to CAT: why and how could treaty ratification actually cause governments to increase these abuses? My results indicate that these prior findings may have been picking up on a selection effect, rather than a treatment effect. Of course, this still leaves us with a puzzle as to why abusing states tend to select into the

FIGURE 1 Marginal Effects of CEDAW Ratification on Women's Rights

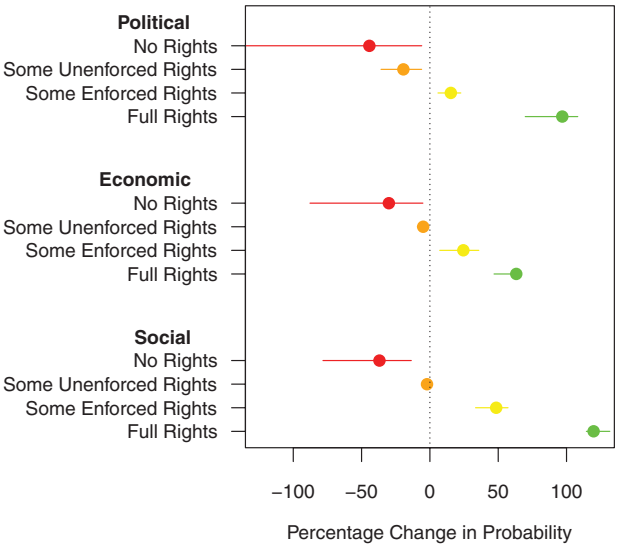

Table 4 Summary of Treaty Commitment Coefficients

Treaty	DV	Matching On	Coef.	s.e.	p	n
CEDAW	Pol. Rights	Observable Covariates	0.306	0.159	0.056	1642
CEDAW	Pol. Rights	Treaty Commitment Preferences	0.233	0.078	0.003	1652
CEDAW	Pol. Rights	Both	0.274	0.074	0.000	1592
CEDAW	Econ. Rights	Observable Covariates	0.039	0.119	0.747	1642
CEDAW	Econ. Rights	Treaty Commitment Preferences	0.171	0.066	0.009	1652
CEDAW	Econ. Rights	Both	0.203	0.065	0.002	1592
CEDAW	Soc. Rights	Observable Covariates	0.15	0.12	0.211	1642
CEDAW	Soc. Rights	Treaty Commitment Preferences	0.265	0.064	0.000	1652
CEDAW	Soc. Rights	Both	0.272	0.070	0.000	1592
CAT	Torture	Observable Covariates	−0.65	0.123	0.000	1642
CAT	Torture	Treaty Commitment Preferences	−0.091	0.077	0.237	1580
CAT	Torture	Both	−0.098	0.079	0.217	1552
ICCPR	Phys. Int.	Observable Covariates	−0.405	0.101	0.001	1643
ICCPR	Phys. Int.	Treaty Commitment Preferences	−0.030	0.062	0.608	1966
ICCPR	Phys. Int.	Both	0.010	0.062	0.869	1870

Note: Results based on matching on observable covariates only are from Hill (2010).
Results based on matching on treaty commitment preferences are as reported in Table 3.
Results based on matching on both sets of variables are as reported in the online supporting information.

treaty more often, but this is arguably a less problematic puzzle than the question of why such commitments would cause increases in abuses.

In the above models, I estimate the probabilities of treaty commitment using only states' treaty commitment preferences (as estimated using W-NOMINATE). It is possible, however, that these preferences do not capture the full model of commitment. Indeed, other observable factors may affect the probability of commitment, including the factors that ultimately affect states' respect for human rights (Powell and Staton 2009). As a robustness test, I match states using a model that includes the probability of commitment calculated by W-NOMINATE as well as all of the controls listed in Table 2. I then estimate a set of ordered-probit models on the matched samples that are identical to the models reported in Table 3. The balance statistics of this sample and the regressions results are reported in the online supporting information. These robustness tests produce findings substantively similar to those reported above.

Table 4 summarizes my results and compares them with those in Hill (2010). For each treaty and dependent variable, Table 4 shows the results from Hill (2010) based on matching on observable treaty commitment covariates, followed by the results reported in Table 3 based on matching on treaty commitment preferences, and then followed by the robustness test results reported in the supporting information based on matching on treaty commitment preferences and observable treaty commitment covariates. In summary, the results of my models differ from previous findings with respect to four of the five dependent variables. This indicates that a fuller accounting for treaty commitment selection effects can significantly impact the results of subsequent hypothesis tests of the effects of treaty commitment.

Conclusions

Whether or not international institutions have the power to constrain states has been the subject of a central research agenda in the international-relations literature for over two decades. Scholars have especially paid attention to determining the effects of multilateral treaties on state behavior. Yet the fact that states self-select into these treaties threatens our ability to determine whether they act as constraints or simply as screening devices. By using methods created in other substantive areas, scholars have attempted to overcome the selection-effect problem. Yet these methods have several important limitations that limit their utility in this context.

In this article, I have argued that these methods have a significant substantive shortcoming when used to study treaty effects: an insufficient accounting for underlying state preferences. I have outlined an approach to the treaty commitment selection-effect problem that estimates states' preferences with respect to treaties by using

the spatial model and, subsequently, uses these estimates to match treaty members to comparable nonmembers. This approach has the potential to substantially improve our ability to test theories and advance our knowledge of the causal effects of international institutions.

To my knowledge, this is the first article to use W-NOMINATE outside of the legislative or quasi-legislative context and the first to combine it with propensity-score matching in this manner. I hope that scholars will find the combination of these frequently employed tools useful in other areas. An example might be in the legislative context in which W-NOMINATE is often used. Suppose we are interested in determining whether voting for a particular bill affects legislators' probability of being assigned to a particular committee, seeking reelection, winning reelection, or some other outcome. To perform such an analysis, we could use a procedure similar to that outlined above: begin by estimating the probability of legislators voting for the bill, create a matched sample based on these probabilities (and possibly additional factors), and run further statistical tests using the outcome variable.

I have also provided novel empirical findings regarding the effects of commitment to the CEDAW, CAT, and ICCPR. The results indicate that ratification of the CEDAW has led to improvements in the lives of women across a broad range of political, economic, and social rights. By contrast, commitments to the ICCPR and CAT have not significantly affected human-rights practices. These results are an important contribution to the debate on the effects of international institutions and will hopefully serve to clear up empirical puzzles, especially with respect to existing controversies regarding the possible negative effects of ICCPR and CAT ratifications. Yet the results also point us to new puzzles by contrasting one case, the CEDAW, in which international law has been a useful tool for improving human-rights practices, with other cases, ICCPR and CAT, where this does not appear to be the case. Theoretical work on international institutions suggests several possible explanations for these contrasting results. It may be the case that the CEDAW is more effective at improving human-rights practices by virtue of the design of its legal requirements. Alternatively, the answer may lie not in the treaty design but in the types of violations these treaties address. I have analyzed the effects of the CAT and ICCPR on physical-integrity rights, whereas the CEDAW addresses other civil rights. As Lupu (2013) and others have argued, the mechanisms for the enforcement of human-rights treaties may be less effective with respect to physical-integrity rights. Regardless of the explanation, the results point to the need for nuanced theories about when international law is effective and when it is not.

References

Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment." *Political Analysis* 14(1): 37–62.

Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. 2009. "Comparing NOMINATE and IDEAL: Points of Difference and Monte Carlo Test." *Legislative Studies Quarterly* 34(4): 555–91.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data: A Unified Approach." *American Political Science Review* 98(2): 355–70.

Davenport, Christian. 1995. "Multi-Dimensional Threat Perception and State Repression: An Inquiry into Why States Apply Negative Sanctions." *American Journal of Political Science* 39(3): 683–713.

Davenport, Christian. 1999. "Human Rights and the Democratic Proposition." *Journal of Conflict Resolution* 43(1): 92–116.

Davenport, Christian. 2007. "State Repression and Political Order." *Annual Review of Political Science* 10(1): 1–23.

Davis, Otto A., Melvin J. Hinich, and Peter C. Ordeshook. 1970. "An Expository Development of the Mathematical Model of the Electoral Process." *American Political Science Review* 64(2): 426–48.

Desposato, Scott. 2006. "Parties for Rent? Ambition, Ideology, and Party Switching in Brazil's Chamber of Deputies." *American Journal of Political Science* 50(1): 62–80.

Desposato, Scott. 2008. "The Impact of Electoral Rules on Legislative Parties: Lessons from the Brazilian Senate and Chamber of Deputies." *Journal of Politics* 68(4): 1018–30.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. Boston: Addison Wesley.

Downs, George W., David M. Rocke, and Peter N. Barsoom. 1996. "Is the Good News about Compliance Good News about Cooperation?" *International Organization* 50(3): 397–406.

Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting*. New York: Cambridge University Press.

Goldsmith, Jack, and Eric Posner. 2005. *The Limits of International Law*. New York: Oxford University Press.

Goodliffe, Jay, and Darren G. Hawkins. 2006. "Explaining Commitment: States and the Convention against Torture." *Journal of Politics* 68(2): 358–71.

Gray, Mark M., Miki Caul Kittilson, and Wayne Sandholtz. 2006. "Women and Globalization: A Study of 180 Countries, 1975–2000." *International Organization* 60(2): 293–333.

Guo, Shenyang, and Mark W. Fraser. 2010. *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: Sage.

Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2005. "Human Rights in a Globalizing World: The Paradox of Empty Promises." *American Journal of Sociology* 110(5): 1373–1411.

Hafner-Burton, Emilie M., and Kiyoteru Tsutsui. 2007. "Justice Lost! The Failure of International Human Rights Law to Matter Where Needed Most." *Journal of Peace Research* 44(4): 407–25.

Hathaway, Oona. 2002. "Do Human Rights Treaties Make a Difference?" *Yale Law Journal* 111(8): 1935–2042.

Hathaway, Oona. 2007. "Why Do Countries Commit to Human Rights Treaties?" *Journal of Conflict Resolution* 51(4): 588–621.

Heckman, J., J. Ichimura, J. Smith, and P. E. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66(5): 1017–98.

Heckman, J., J. Smith, and N. Clements. 1997. "Making the Most out of Social Experiments: Accounting for Heterogeneity in Program Impacts." *Review of Economic Studies* 64(4): 487–536.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153–61.

Hill, Daniel W. 2010. "Estimating the Effects of Human Rights Treaties on State Behavior." *Journal of Politics* 72(4): 1161–74.

Hinich, Melvin J., and Michael C. Munger. 1994. *Ideology and the Theory of Political Choice*. Ann Arbor: University of Michigan Press.

Hinich, Melvin J., and Michael C. Munger. 1997. *Analytical Politics*. Ann Arbor: University of Michigan Press.

Hix, Simon. 2001. "Legislative Behaviour and Party Competition in the European Parliament: An Application of Nominate to the EU." *Journal of Common Market Studies* 39(4): 663–88.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3): 199–236.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2011. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42(8): 1–28.

Hollyer, James R., and B. Peter Rosendorff. 2011. "Why Do Authoritarian Regimes Sign the Convention against Torture? Signaling, Domestic Politics and Non-Compliance." *Quarterly Journal of Political Science* 6(3–4): 275–327.

Honaker, James, Gary King, and Matthew Blackwell. 2008. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45(7): 1–47.

Howell, William G., and David E. Lewis. 2002. "Agencies by Presidential Design." *Journal of Politics* 64(4): 1095–1114.

Jenkins, Jeffrey. 1999. "Examining the Bonding Effects of Party: A Comparative Analysis of Roll-Call Voting in the U.S. and Confederate Houses." *American Journal of Political Science* 43(4): 1144–65.

Keith, Linda Camp. 1999. "The United Nations International Covenant on Civil and Political Rights: Does It Make a Difference in Human Rights Behavior?" *Journal of Peace Research* 36(1): 95–118.

Keith, Linda Camp. 2002. "Judicial Independence and Human Rights Protection around the World." *Judicature* 85(4): 195–200.

Keith, Linda Camp, C. Neal Tate, and Steven C. Poe. 2009. "Is the Law a Mere Parchment Barrier to Human Rights Abuse?" *Journal of Politics* 71(2): 644–660.

Keohane, Robert O. 1984. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.

Landman, Todd. 2005. *Protecting Human Rights: A Comparative Study*. Washington, DC: Georgetown University Press.

Lechner, M. 2000. "An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany." *Journal of Human Resources* 35(2): 347–75.

Liao, Tim Futing. 1995. "The Nonrandom Selection of Don't Knows in Binary and Ordinal Responses: Corrections with the Bivariate Probit Model with Sample Selection." *Quality and Quantity* 29(1): 87–110.

Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: Wiley and Sons.

Londregan, John. 2000. *Legislative Institutions and Ideology in Chile*. Cambridge: Cambridge University Press.

Lupu, Yonatan. 2013. "Best Evidence: The Role of Information in Domestic Judicial Enforcement of International Human Rights Agreements." *International Organization*. Forthcoming.

Marshall, Monty, and Keith Jaggers. 2002. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2002." University of Maryland.

Martin, Lisa L. 1992. *Coercive Cooperation*. Princeton, NJ: Princeton University Press.

Mitchell, Ronald B. 1994. *Intentional Oil Pollution at Sea: Environmental Policy and Treaty Compliance*. Cambridge, MA: MIT Press.

Mitchell, Ronald B., William C. Clark, David W. Cash, and Nancy M. Dickson. 2006. *Global Environmental Assessments: Information and Influence*. Cambridge, MA: MIT Press.

Morgan, Stephen L., and Christoher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Morgenstern, Scott. 2003. *Patterns of Legislative Politics: Roll-Call Voting in Latin America and the United States*. Cambridge: Cambridge University Press.

Neumayer, Eric. 2005. "Do International Human Rights Treaties Improve Respect for Human Rights?" *Journal of Conflict Resolution* 49(6): 925–53.

Neumayer, Eric. 2007. "Qualified Ratification: Explaining Reservations to International Human Rights Treaties." *Journal of Legal Studies* 36(2): 397–430.

Noury, Abdul G. 2001. "Ideology, Nationality and Euro-Parliamentarians." *European Union Politics* 3(1): 33–58.

Oye, Kenneth A. 1986. "Explaining Cooperation under Anarchy: Hypotheses and Strategies." In *Cooperation under Anarchy*, ed. Kenneth A. Oye. New York: Macmillan, 1–24.

Poe, Steven, and C. Neal Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88(4): 853–72.

Poe, Steven, C. Neal Tate, and Linda Camp Keith. 1999. "Repression of the Human Right to Personal Integrity Revisited: A Global Cross-National Study Covering the Years 1976–1993." *International Studies Quarterly* 43(2): 291–313.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Powell, Emilia Justyna, and Jeffrey K. Staton. 2009. "Domestic Judicial Institutions and Human Rights Treaty Violation." *International Studies Quarterly* 53(1): 149–74.

Reed, William, David H. Clark, Timothy Nordstrom, and Wonjae Hwang. 2008. "War, Power, and Bargaining." *Journal of Politics* 70(4): 1203–16.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 41–55.

Rosenbaum, Paul R., and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79(387): 516–24.

Rothenberg, Lawrence S., and Mitchell S. Sanders. 2000. "Legislator Turnout and the Calculus of Voting: The Determinants of Abstention in the U.S. Congress." *Public Choice* 103(3–4): 259–70.

Rubin, Donald B. 1997. "Estimating Causal Effects from Large Data Sets Using Propensity Scores." *Annals of Internal Medicine* 127(8): 757–63.

Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions." *Political Analysis* 11(2): 111–38.

Schonhardt-Bailey, Cheryl. 2003. "Ideology, Party and Interests in the British Parliament of 1841–1847." *British Journal of Political Science* 33(4): 581–605.

Simmons, Beth A. 1998. "Compliance with International Agreements." *Annual Review of Political Science* 1(1): 75–93.

Simmons, Beth A. 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94(4): 819–34.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics.* Princeton, NJ: Princeton University Press.

Simmons, Beth A., and Daniel J. Hopkins. 2005. "The Constraining Power of International Treaties: Theory and Methods." *American Political Science Review* 99(4): 623–31.

Smith, J. A., and P. E. Todd. 2005. "Does Matching Overcome LaLonde's Critique of Non-experimental Estimators?" *Journal of Econometrics* 125(1): 305–53.

Stokke, Olav Schram, and Geir Hønneland. 2007. *International Cooperation and Arctic Governance: Regime Effectiveness and Northern Region Building.* London: Routledge.

Victor, David G., Kal Raustiala, and Eugene B. Skolnikoff. 1998. *The Implementation and Effectiveness of International Environmental Commitments: Theory and Practice.* Cambridge, MA: MIT Press.

Voeten, Erik. 2000. "Clashes in the Assembly." *International Organization* 54(2): 185–215.

von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99(4): 611–22.

Vreeland, James R. 2008. "Political Institutions and Human Rights: Why Dictatorships Enter into the United Nations Convention against Torture." *International Organization* 62(1): 65–101.

Weiss, Edith Brown, and Harold K. Jacobson. 1998. *Engaging Countries: Strengthening Compliance with International Accords.* Cambridge, MA: MIT Press.

Winship, Christopher, and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18(1): 327–50.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

- **Treaty Data**
- **Matching Results**
- **Robustness Tests**
- **Monte Carlo Simulation**