

Do Nationals in Leadership Positions Affect Cooperation with International Organizations? Evidence from the International Criminal Court using a Regression Discontinuity Design

Erik Voeten

Peter F. Krogh Associate Professor of Global Justice and Geopolitics
Edmund A. Walsh School of Foreign Service and Government Department
Georgetown University
Email: ev42@georgetown.edu

Abstract

Recent scholarship argues that even weak international organizations can influence state behavior by affecting information and accountability politics. A prominent national in a leadership position increases such leverage. This can be exploited by scholars interested in estimating whether international organizations have an independent effect on state behavior, especially when the selection of high-level officials occurs through an observable process such as rotation or competitive elections. I illustrate this point with an analysis of whether having a national elected as permanent judge increases cooperation with the International Criminal Court (ICC). Judges were chosen through competitive elections. I use a regression discontinuity design to obtain valid causal estimates; essentially comparing states that just did and just did not have a national elected as judge. The results show that states with nationals as permanent judges were less likely to undermine the ICC by signing non-surrender agreements with the United States and more likely to adopt a broad range of legislative measures that integrate the Rome Statute and cooperation with the Court into domestic law. This suggests that even indirect institutional participation that confers no meaningful influence can have meaningful behavioral effects.

(Interested but can't commit to submit to special issue)

I gratefully acknowledge data shared by Allison Danner and Beth Simmons. I appreciate suggestions and comments from Ellen Mastenbroek, Stefan Rummens, Bertjan Verbeek, Anna van der Vleuten, Angela Wigger and the participants in Georgetown University political economy luncheon. The usual disclaimer applies.

Do Nationals in Leadership Positions Affect Cooperation with International Organizations? Evidence from the International Criminal Court using a Regression Discontinuity Design

In most countries other than the United States it is a matter of some consequence to have a national in a leadership role in an international organization (IO) such as NATO, the World Bank, the European Union (EU), the International Criminal Court (ICC), or the United Nations (UN). Governments often exert considerable effort to acquire these positions for their nationals and put forth candidates with distinguished records, including former presidents, prime-ministers, and cabinet members. Governments presumably seek such appointments to gain policy influence, bargaining chips, prestige, and to reward fellow domestic elites. Yet, having a national in a high profile position could also backfire. IOs frequently want things from governments, such as policy change, financial contributions, participation in risky peacekeeping missions, or support against attacks by other state parties. A prominent national in a leadership position may alter the domestic politics of such requests, for example because he or she has more channels of influence and gets more media attention than another IO official would. It also potentially opens up additional avenues to hold governments accountable, for instance by highlighting inconsistencies between actions and statements of support that are inevitably part of campaigns for high-level appointments. This fits with a growing theoretical literature that argues that even weak international actors may alter state behavior by affecting the domestic salience of issues and through accountability politics (for overviews, see Cortell and Davis 2000; Simmons and Martin 2002).

Do high profile nationals in leadership positions encourage states to cooperate more with IOs? I examine whether having a national serve as a permanent judge of the International Criminal Court (ICC) makes states more likely to take costly cooperative measures, including refraining from signing a nonsurrender agreement with the United States and enacting legislation to embed the Rome Statute into domestic law. Answering this

question is not just interesting by itself but it is also a novel way to evaluate whether IOs exert influence over state behavior. First, virtually all existing quantitative studies analyze whether variation in IO membership correlates with state behavior, assuming equal entanglement among IO members. Instead, the question here is whether those member states with temporarily increased entanglement behave differently than those states that do not experience this treatment.

Second, studies of the effect of IO membership on state behavior are plagued by a problem that is well-understood but difficult to remedy: congruence between institutional purposes and state behavior may occur simply because those states already inclined to cooperate were also most likely to opt for IO participation (e.g. Downs et al 1996). This makes it difficult to evaluate whether commitment to the IO per se has any behavioral consequences. This study has a different selection problem: the same reasons that lead a government to seek and to be successful in finding a position of prominence for one of its nationals may also influence the behavior that these IOs wish to affect. For an important class of cases this selection process is observable. For example, some IOs distribute high-level posts based on rotation principles or, as in the case examined here, competitive elections. This allows for the use of a regression discontinuity (RD) design, which exploits situations where subjects are selected into a treatment if they meet some threshold, such as the majority of votes in an election. Because the selection process is observed, the RD design helps generate causal estimates (local average treatment effects) with greater internal validity than other quasi-experimental approaches. Indeed, with weak assumptions, RD designs are equivalent to randomized experiments for subjects reasonably close to the threshold values (Lee 2008), such as states whose nationals just won or just failed to win an election. Given that IR scholars lack the leverage to randomly assign states into control and treatment groups, this may be as close as scholars of international politics can get to a randomized experiment.

In what I believe is the first application of an RD design to the study of state behavior, I find that states with nationals as judges were less likely to undermine the ICC by signing a non-surrender agreement than states whose nationals just lost in the election for judgeships. Moreover, states with judges enacted more domestic legislative measures that embedded the Rome Statute and obligations to cooperate with the ICC into domestic law.

I proceed with a brief discussion of how nationals could affect state behavior and how these mechanisms relate to the broader theoretical literature. The remainder of the paper illustrates the ICC case, explains the RD design, shows the results, and discusses broader implications.

Mechanisms: Salience and Accountability Politics

A growing body of scholarship suggests that IOs, treaties, and transnational actors can have an impact on state behavior by affecting the domestic political process (e.g. Checkel 1999; Cortell and Davis 1996, 2000; Dai 2005; Haas, Levy and Keohane 1993; Milner 1997; Pevehouse 2005; Risse-Kappen 1994; Schimmelfennig 2001; Sikkink 1993; Simmons 2009).¹ The main sources of influence on the part of international actors are their ability to raise the domestic salience of their causes and to create new avenues for holding governments accountable. A national in a leadership position can temporarily increase both types of leverage. Such individuals may make requests for costly cooperation more difficult to ignore and more likely to become high stakes issue. Moreover, having a national in a leadership position plausibly increases the perceived commitment to an IO in various ways. I discuss these points in more detail below.

¹ This list is not meant to be comprehensive but serves the purpose of showing the diversity of theoretical approaches and issue contexts.

Saliency and Information Politics

Not all causes deemed worthy by international actors resonate in domestic political debates. The same causes that are ignored in some countries lead to vigorous debates in others and quickly become accepted norms in yet other countries. Variation in the domestic saliency of international norms is frequently invoked as a key explanation for why some states do and other states do not cooperate with IOs, although saliency is rarely measured independently from the behavior that the norm supposedly explains (Cortell and Davis 2000). As such, it is difficult to infer whether international actors independently influenced state behavior or whether there are other reasons (e.g. culture, material interests) that make an issue more likely to become salient domestically and that also affect state behavior.

The most important aspect of saliency is the intensity by which an issue appears in domestic political discourse (Cortell and Davis 2000, p.71). Such discourse encourages more informed publics, which increases the chances that governments are constrained by domestic political processes (Dai 2005). Cortell and Davis (2000, p.71) argue that this is also the measure of saliency that is most likely to be conflated with other factors that may lead states to act in accordance with the wishes of international actors.

Nationals in leadership positions can have an exogenous impact on the intensity of public discourse surrounding a request for cooperation if their selection into leadership is random or can be considered “as if” it was random.² High-level IO appointees were almost always prominent players in domestic politics or public life. They have built up stature, understanding of how a system works, and connections that provide access to media and elites. They could use this political capital to advance the interests of their new employer. For example, before Jaap de Hoop Scheffer became NATO’s Secretary-General in 2004, he had been Dutch minister of foreign affairs and a CDA party leader (the party of prime-minister

² Why this condition holds in the regression discontinuity design is discussed later in this paper.

Jan-Peter Balkenende). In 2005 NATO asked the Dutch government to take a lead role in a dangerous mission in Afghanistan (Uruzgan). De Hoop Scheffer was highly visible throughout the process. For instance, when one of the three coalition government parties (D'66) came out against Dutch participation, De Hoop Scheffer was cited on the front pages of all major Dutch newspapers as saying that NATO allies “would not understand” a negative decision and that it could lead to a “domino effect.”³ He also intervened in the debate over the renewal of the mission saying that “he could not imagine” The Netherlands leaving Afghanistan.⁴ The chair of the parliamentary commission for defense, Hans van Baalen, noted that “as a Dutchman, he better understood how that strange discussion in our country went and thus understood the “do’s and don’ts.”⁵

The preceding posits that nationals become attached to an IO’s causes, occasionally leading them to go against the wishes of their national governments (or a part of that government). This is more likely to happen when the appointment and re-appointment is done through IO state parties as opposed to cases where a national government is entitled to appoint and re-appoint a representative. For example, UN General Assembly representatives are diplomats appointed by and presumably loyal to their states. An example of an in-between case would be the World Bank, whose president is traditionally appointed by the United States but still requires approval by the Bank’s board. Yet, even if a national is not actively involved in promoting an IO’s cause, they may raise the intensity of a debate on an issue. To continue the example, it is plausible that Dutch media are especially interested if a Dutch NATO Secretary-General meets the US President or is praised or critiqued by world leaders. To the extent that a public wants the national to succeed, this may constrain government behavior. At the very least, it makes it more difficult for governments to ignore

³ December 17 2005.

⁴ “Kritiek op Uruzgan-uitspraken NAVO-chef” *NRC Handelsblad* September 20, 2007.

⁵ Quoted in: “Nederland voortrekken deed hij beslist niet: NAVO vertrekkend secretaris-generaal De Hoop Scheffer zorgde voor ‘reputatiewinst’ van Den Haag. *NRC Handelsblad*, July 7 2009.

costly IO requests for cooperation. This effect on domestic salience in some countries but not in others can be exploited by researchers.

Commitment and Accountability Politics

A second mechanism prominent in the literature is that commitments to international organizations are not just cheap talk. For example, Simmons (2009) argues that when governments ratify human rights treaties, they risk being held accountable for the promises entailed in these commitments. Treaties give domestic coalitions new tools and arguments to point out inconsistencies or divisions in government behavior. This may affect behavior if governments are sensitive to audience costs, for example because they face competitive elections or are otherwise susceptible to demands by civil society or opposition forces (e.g. Fearon 1997).

The logic of this argument applies not only to the signing of treaties but also to other ways in which governments increase their entanglement with a cause or an institution. During campaigns for their nationals, governments frequently make promises or otherwise proclaim their devotion to an IO's cause. Such statements provide additional opportunities for pointing out inconsistencies between past promises and current behavior. Moreover even in the absence of explicit campaign promises, governments may be accused of disloyalty or of internal divisions if they reject a public appeal for cooperation by a prominent national. To continue the example, opposition parties would have feasted on the public schism if the CDA had gone against its former leader. This provided strong incentives for the largest government party to go along with the proposed Dutch role in Afghanistan.

There could also be damage to the international reputation of states if they are exposed for acting in ways that are inconsistent with past promises. Such international and domestic "naming and shaming" mechanisms are central in the theoretical literature on the effects of formally weak human rights institutions on state behavior (see Hafner-Burton 2008

for an overview) and in international political economy (see Tomz 2007). Such concerns about international reputation or standing may also enhance incentives not to undermine a national. For example, while making the argument that De Hoop Scheffer's tenure had increased Dutch standing among NATO allies, defense minister Van Middelkoop remarked that: "Suppose De Hoop Scheffer had performed badly. We would have noticed immediately."⁶

Empirical tests of the accountability argument also face a selection problem. Those states that are willing to commit to (human rights) institutions may well be different in important and difficult to observe ways from those that are unwilling to make such a commitment. I suggest that nationals in leadership positions can under certain conditions be a source of exogenous variation in the strength of the commitment a government makes to an IO.

The Link between Accountability and Salience

It is both unwise and impossible to discriminate between the accountability and the salience mechanisms. Without salience, there is little opportunity for actors to hold a government accountable. Without accountability mechanisms, the salience of an issue may not matter much. Indeed both scholars with constructivist and rationalist inclinations combine both types of arguments, as they should. For example, Keck and Sikkink (1998) stress both information and accountability politics, as does Dai (2005, 2007) in game-theoretic analyses of compliance with international institutions. Neither accountability nor salience is necessarily decisive. At times large amounts of public debate on an issue may backfire. Governments are known to occasionally disregard past promises and public pleas. The argument advanced here is simply that nationals in leadership positions make it more likely

⁶ "Parting NATO Secretary-General De Hoop Scheffer caused reputational gains for The Hague but did not pull any strings for the Netherlands" *NRC Handelsblad* p.3 July 7, 2009 (translation by author).

that a request for costly cooperation enters into the public debate and that it increases opportunities to hold leaders accountable for their commitment to an IO. This changes the political calculus surrounding causes advanced by IOs. As Bram Boxhoorn, chairman of the Atlantic Commission (a Dutch thinktank) put it: “With a Dutchman as Secretary-General it is just a little harder to ‘say no’ when you are asked to stay an additional two years in Uruzgan.”⁷

National Judges and Cooperation with the ICC

The ICC presents one of the clearest cases where states were asked to revisit a previous commitment to an IO (Kelley 2007). The United States requested each state to sign an agreement not to surrender Americans to the Court.⁸ An indication of how serious the U.S. took this issue is the adoption of the American Service-Members' Protection Act, sometimes labeled “The Hague Invasion Act”, which authorizes the President to use “all means necessary and appropriate to bring about the release of any US or allied personnel being detained or imprisoned by, on behalf of, or at the request of the International Criminal Court” and obliges the US to withdraw military aid from ICC state parties that did not sign nonsurrender agreements (with the exception of NATO and some other allies).⁹

These agreements were widely viewed as attempts to undermine the ICC’s authority. In an analysis of why states accepted or refused to ratify such agreements, Kelley (1997) shows convincingly that states were not driven purely by materialist concerns. Most notably, her analyses show that states with strong domestic rule of law were less likely to give in to U.S. demands. Kelley suggests two plausible interpretations of this result. First, states with developed legal systems may have rejected the nonsurrender agreements because of their

⁷ Ibid.

⁸ These agreements are called Bilateral Immunity Agreements (BIA) or Article 98 agreement (after an article in the Rome Statute).

⁹ <http://www.state.gov/t/pm/rls/othr/misc/23425.htm>

appreciation of the moral value of the court. Second, states with strong domestic rule of law may attach greater value to the commitment they made to the ICC. Using a traditional research design it is impossible to credibly distinguish the effects of entanglement with the ICC per se from the general inclinations that may lead states to support the ICC.¹⁰

I suggest that having a national serve as a permanent judge works as a shock that boosted the salience of cooperation with the Court and intensified the commitments of some states but not others. There is some anecdotal evidence for this. Indeed, Kelley raises it in her case study of Costa Rica. Costa Rica's president blocked the nomination of Elizabeth Odio-Benito for political reasons (she had been a vice-president for the opposition party). Consequently, Odio-Benito was nominated by Panama and was elected. The blockage of Odio-Benito's nomination spurred domestic protest and increased public support for the ICC, thus tying the government's hands when faced with U.S. demands to undermine the court (Kelley 2007, p583-4).

Another interesting example is Bolivia, which signed an agreement in July 2003, thereby receiving a six-month waiver from U.S. sanctions. After failing to ratify the treaty,¹¹ the U.S. cut military assistance in 2004 (Ribando 2006).¹² The fact that Bolivia had a national on the court (former minister of Justice René Blattmann) featured prominently in the ratification debates and in appeals to the Bolivian Chamber of Deputies by NGOs, such as

¹⁰ Kelley acknowledges this in her discussion of alternative approaches to estimation, such as Heckman selection models, arguing that there are no obvious variables that can be excluded from the ratification equation. In the article, Kelley stresses the commitment interpretation, partially because domestic rule of law does not significantly predict ratification of the ICC Treaty. This is not a full assurance, however, because ratification was not simply about the moral value of the court but had much to do with the consequences of the commitment to the ICC (see Danner and Simmons 2009). For example, the US may well appreciate the moral value of the Court (as exemplified by its leading role in establishing the ICTY) but was less enthusiastic about potential consequences.

¹¹ The Senate ratified but the Chamber of Deputies did not.

¹² Note that all of this happened before Evo Morales was elected in December 2005 on an anti-American platform.

Amnesty International¹³ and the Coalition for the International Criminal Court.¹⁴ Sacha Llorenti, a parliamentarian and president of Bolivia's National Human Rights Assembly claimed that: "Bolivia would be the only country in the world to agree to such a pact that also has a judge on the court."¹⁵ Even in the only country with a judge on the court that did actually ratify a nonsurrender agreement, Ghana, the fact that it had a judge was emphasized by the minority in parliament: "It will be the hallmark of double standards for Ghana to ratify the Rome Statutes that established the International Criminal Court, nominate its Vice-President¹⁶ and turn around to ratify an agreement that obviously undermines the integrity of the Court."¹⁷

By contrast, ten of the twenty-five states whose candidates failed in their bids signed non-surrender agreements with the United States. This difference in proportions is suggestive of an effect. Yet, it is plausible that the Assembly of State Parties was simply successful in electing judges from countries that were already committed to the Court and therefore not inclined to undermine the Court by signing a nonsurrender agreement. The regression discontinuity design investigates this possibility.

Broader Cooperation with the Court

In addition to refrain from signing nonsurrender agreements, governments were also asked to take positive steps to enhance their cooperation with the ICC. First, the ICC's functioning is based on the complementarity principle, which leaves first responsibility for war crimes trials to states. For this to work properly, states must adopt the crimes articulated

¹³ See: <http://www.amnesty.org/en/library/asset/AMR18/002/2004/en/45f6cfe7-d630-11dd-ab95-a13b602c0642/amr180022004en.html> (accessed June 19, 2009).

¹⁴ <http://www.iccnw.org/documents/BoliviareisistBIA01Jun04.pdf>

¹⁵ Quoted in: "U.S. Threatens Bolivia in Effort to Secure Criminal Court Immunity." Pacific News Service, March 3 2005. http://news.pacificnews.org/news/view_article.html?article_id=40d8f93957008266edbc544c21df75be (accessed June 19, 2009).

¹⁶ Ghanaian judge Kuenyahia is Vice-President of the ICC. She is a well-known academic and human rights activist.

¹⁷ Parliamentary debate, October 29, 2003. Quoted in: http://www.iccnw.org/documents/HighOfficialQuotes_Current.pdf.

in the Rome Statute into their domestic penal codes and must identify criminal responsibility and fair trial guarantees in accordance with the Rome Statute. Second, states must enact measures that ensure their cooperation with the ICC. This includes the ratification of the Agreement on the Privileges and Immunities of the Court (APIC), which ensures access and immunity for ICC employees, and cooperation legislation, which guarantees cooperation in investigations and access for the ICC. The willingness to incorporate the ICC domestically is an important indication of commitment, as domestic laws and institutions are generally more difficult to circumvent than international ones (Danner and Simmons 2007).

I construct a scale of cooperation by including two indicators that measure complementarity and cooperation each. First, Amnesty International has closely monitored the adoption and drafting of both cooperation and implementation legislation. I code whether states had adopted or drafted each type of legislation in January 2006, just before the second ICC election (Van der Pas 2006). Second, Danner and Simmons (2007) coded whether states had adopted crimes against humanity in domestic penal codes by June 2005. Third, I coded whether states had signed and ratified the APIC Treaty by January 2006.¹⁸ For each indicator, a draft/signature confers one point and an enactment/ratification two points. Together with the nonsurrender agreements, we obtain a 10-point scale that is roughly normally distributed (mean 5.8, SD 2.8) with a high reliability (Cronbach's alpha is .76). I use this composite scale of cooperation as the second dependent variable in the analysis.

The Regression Discontinuity Design

Researchers have used various techniques to deal with selection issues in estimating the effect of institutional participation on behavior. These include Heckman selection models (e.g. Mitchell et al 2007, Von Stein 2005), matching on observables (Simmons and Hopkins

18

http://www.minbuza.nl/verdragen/en/searchtreaties/detailspage?verdragId=19147&datasource=/content/verdrag/en/en/documents/zoek-verdragen/verdragen_resultaten.xml (June 29, 2009).

2005), and instrumental variables approaches (e.g. Simmons 2009). These techniques are certainly appropriate in some contexts but they often rely on strong assumptions. Structural models like Heckman selection rely on difficult to motivate exclusion restrictions and/or distributional assumptions while matching demands the presence of observables that fully capture the selection process. Good instrumental variables are difficult to find for the type of problem that interests scholars of international institutions. Moreover, imperfectly exogenous instruments can wreak serious havoc with statistical inferences (e.g. Bartels 1991).

The RD design is a special case of matching where the selection variable is precisely observed and thus does not need to be estimated based on observable covariates (Heckman et al. 1999). This gives RD an internal validity advantage over other quasi-experimental approaches (Imbens and Lemieux 2008; Lee 2008). The classic application is in educational research (Thistlewaite and Campbell 1960), where students sometimes enter a program if they achieve a target test score. The effect of the program can then be evaluated by examining if a discontinuity in student performance occurs at the threshold level for admittance. The design also has a shorter but very active pedigree in economics (for overviews, see Imbens and Lemieux 2008; Lee and Lemieux 2009).

The basic intuition is straightforward and can be illustrated with minimal mathematical notation. The key requirement is the presence of a forcing variable X that selects subjects (states) with a value of X greater than or equal to a cutoff value c into a treatment T . In our example, the treatment is having a national serve as a permanent judge on a war crimes tribunal. The forcing variable is the vote share a judicial candidate receives in elections, where candidates that receive more than 50% of the vote are selected into the treatment.¹⁹ Figure 1 gives an ideal-typical example with simulated data. If a propensity to

¹⁹ In their overview of RD applications, Lee and Lemieux (2009) cite five studies that similarly use vote shares as a forcing variable.

cooperate matters in the selection of judges, then we would expect that the more votes the states' candidate in this example receives, the more likely the state is to cooperate with the institution. In the bottom panel of figure 1, this is precisely the relationship we witness. The top panel of figure 1 is identical but there is also a discontinuity precisely at the cutoff point where judges are elected. This indicates an effect of winning election on cooperation. In the top half the treatment effect (actually having a judge elected) is positive whereas in the bottom panel it is zero. Note, however, that if we would not have had the information on the number of votes a state's candidate received, we may well have estimated a positive treatment effect with the data from the bottom panel. This effect would have been mistakenly attributed to the treatment (a national as judge) whereas it would have been the result of the manner by which judges were selected.

If the underlying relationship between the forcing and outcome variables is linear (which it is in the simulated example), then the treatment effect γ can be estimated with a simple linear regression:

$$Y = \alpha + \beta X + \gamma T + \epsilon \quad (1)$$

If the outcome variable is dichotomous (as it is in our first example), equation (1) can be estimated as a logit model (Berk and DeLeeuw 1999). The RD design does not require that the relationship between the forcing variable and the outcome variable is linear; it just has to be continuous. Yet, non-linear relationships do have to be modeled to avoid biased estimates of the treatment effect. It is therefore good practice to estimate non-parametric models, kernel regressions, or other forms of non-linear regressions in addition to the simple linear model (e.g. Hahn et al. 2001). Yet, extensive modeling of functional form is difficult in our context given the small sample.

The validity of inferences in the RD design do not depend on an assumption that states do not manipulate the number of votes they get for a prestigious office. It does,

however, require that states cannot *precisely* control the number of votes they get around the majority threshold (Lee 2008). Given that most IO membership elections have secret ballots this is an unlikely objection. If these assumptions are satisfied, RD designs are equivalent to and can be analyzed as a local randomized experiment (Lee 2008).²⁰ The term local refers to the fact that the inferences are limited to subjects that are reasonably close to the cut-off value. One would have to rely on stronger assumptions to draw inferences about the population of states as a whole. So, the results in this paper should not be interpreted to mean that the Sudan would have behaved differently if it had only had an ICC judge. Instead, they estimate the effect for the subpopulation of states that are ICC members, have an interest in having a judge, and had at least a modicum of support for that.

The equivalence with randomized experiments also means that baseline covariates do not need to be included in regression models. However, as with experiments, covariates improve the efficiency of estimates; an important issue in applications to IO elections where the sample size tends to be small. There are several extensions to the RD design for cases where the forcing variable is either not observed perfectly or doesn't perfectly force subjects into treatments. For these and other statistical issues, I refer the interested reader to the literature (e.g. Imbens and Lemieux 2008; Lee and Lemieux 2009). Here, I focus on the practical issues involved in applying the RD design to the study of IOs and the potentially interesting substantive and theoretical lessons that can be drawn from such applications.

Empirical Analysis

ICC Judgeships and Nonsurrender Agreements

The first ICC elections were held on February 7, 2003. On that day, a few non-surrender agreements were already in force but none that involved the 43 states that had put forth

²⁰ There is a lively discussion in the economics literature about the utility of such local average treatment effects (e.g. Deacon 2009; Imbens 2009).

candidates for the 18 judgeships.²¹ The electoral rules were complex. To be elected, a judge had to obtain a two-thirds majority of the votes (56) of the Assembly of States Parties. Each state could issue as many votes as there were judgeships to be awarded in a given round of voting.²² Moreover, states were required to elect a minimum number of candidates from each regional group, gender, and legal expertise (criminal law and international law).²³ In the first round, the vote totals for seven candidates passed the majority threshold. It took thirty-three rounds of voting to elect all 18 judges.²⁴ No judge received more than 78% or less than 10% of the vote, another indication of the competitiveness of the election. Elected judges were assigned by lot to terms of three, six, and nine years. All analyses focus on behavior that occurred between the first and second elections (January 26 2006).²⁵

The forcing variable is the maximum number of votes a judicial candidate achieved in any of the electoral rounds. If that number surpassed the threshold, then the candidate became a judge. Nevertheless, it may well be that the first ballot results more accurately reflect the degree to which state parties perceive another state's commitment to the ICC. I therefore also estimate a model that includes the results of the first ballot outcome.²⁶ With a small sample such as this, base-line covariates can help improve the efficiency of the estimates. I therefore include two covariates that are known to affect the signing of nonsurrender agreements (Kelley 2007)²⁷: logged GDP per capita and civil and political liberties as measured by

²¹ Originally there were 45 candidates but two countries withdrew their candidates before the election. Results are available here: http://untreaty.un.org/cod/icc/elections/results/1st_round.htm (accessed July 2, 2009).

²² So, in the first round each state could issue 18 votes. In the second round 18 minus the number of candidates that achieved the majority threshold in the first round.

²³ For more detail see: ICC/ASP/1/4 "Elections of the Judges of the International Criminal Court": <http://www.iccnw.org/documents/NoteElectionJudges200212Eng.pdf> (accessed July 2, 2009).

²⁴ The last elected judge was Claude Jorda (France) who defeated Nigerian candidate Adolphus Karibi-White (the other candidates had been withdrawn); a judge who obtained a measure of notoriety after consistently falling asleep on the bench while serving at the ICTY (Danner and Voeten, n.d.).

²⁵ Five of the six judges that sat for reelection were reelected in 2006, the sixth (the Samoan judge) lost. There were only 10 candidates for the judgeships, perhaps suggesting that most states did not believe there was much of a chance to replace judges after a three-year term.

²⁶ This should be a "fuzzy" RD model but I do not implement it as such here.

²⁷ I do not include "domestic rule of law" as it would lead to fewer observations, which would defeat the point of improving efficiency.

Freedom House, with higher values indicating worse respect for rights (other measures of democracy and/or human rights yield similar results).

Table one reports the results from a rare event logistic regression, which is more efficient for small samples than regular logit (King and Zeng 2001, Tomz, King, and Zeng 2003).²⁸ The results show that states whose candidates were elected were indeed significantly less likely to sign bilateral agreements than states whose candidates were not elected. The signs and significance of the estimates for the covariates are as expected. Yet the forcing variable has a small positive effect on signing non-surrender agreements, suggesting that state parties were not motivated by or not successful in voting for judges from countries who were particularly committed to the ICC (when removing the election variable from the equation, estimate on the votes variable becomes 0). Indeed, this strongly suggests that having a judge on the court is what led states to subsequently behave differently, not some feature of the selection process.

The estimated effect is very large although imprecisely estimated due to the small sample size. Holding the other variables at their means,²⁹ the predicted probability that a state who has no judge signs a nonsurrender agreement is .58 (CI [.33;79]) whereas this is .009 (CI: [.00-.57]) for states with a judge. Another useful comparison may be to look at states whose candidates received large proportions of the vote but were not elected. Among the top half (12) of states whose candidates failed to be elected (between 25 and 55% of the vote), five ratified a non-surrender treaty, as opposed to one out of the 18 states whose judges acquired between 67 and 78% of the vote.

²⁸ The result holds in a regular logit or probit.

²⁹ In many RD designs this would be an odd thing to do but it is possible here given that the forcing variable has no impact on the outcome.

Broader Cooperation with the Court

Table two reports the linear regression results on the composite cooperation scale. Aside from GDP and democracy, the model also adds two covariates that have proven important in the literature on the adoption of domestic war crimes legislation.³⁰ First, an indicator for whether a country experienced an international or an internationalized armed conflict on its territory between 1988 and 2003, acknowledging that among ICC state parties the experience of war is a powerful incentive to enact legislation to prevent recurrence of atrocities (Danner and Simmons 2007).³¹ Second, countries with common law systems are generally much less reluctant (or slower) in signing human rights treaties and adopting legislation, presumably because once these are signed they have greater consequences as judges tend to be more powerful, independent, and have a broader interpretive role (e.g. Simmons 2009). As before, the covariates are not needed to avoid bias but they do advance the precision of the estimates.

The results reveal a significant and sizeable effect of having a national elected as judge. Countries with nationals as judges had more than two points extra on the ten-point scale, amounting to an additional piece of legislation enacted in support of the ICC. The other covariates have the expected signs. As before, the forcing variable is not significant. This implies either that state parties were not very successful in identifying those most likely to cooperate with the ICC or were not interested in this objective when electing judges. This could be so because they are electing individuals rather than states, although the general perception is that, with a few exceptions, these elections are more about states rather than the qualifications of candidates (Danner and Voeten n.d.). Yet, it is not unusual in these types of elections to choose states based on things other than their devotion to an IO's cause, the

³⁰ I also estimated a model that included membership in the group of "like-minded states" (Kelley 2007) but this variable was not significant and did not alter inferences.

³¹ PRIO/Uppsala Armed Conflict Dataset. Danner and Simmons interact this variable with democracy. Doing so does not generate different results here.

election of Libya to chair the UN Human Rights Commission serving as a particularly inauspicious example (see Lebovic and Voeten 2006).

Figure two plots the residuals from a regression on the cooperation scale with the covariates other than vote-share and election status. This figure is the equivalent of figure one, with actual data. The results are obviously much noisier than the simulated data. Cyprus and Italy, a notoriously poor complier even with European institutions such as the EU and the European Court of Human Rights, are the countries with judges who show much less cooperation than expected by the base-line model. The bottom panel plots the same figure with votes on the first ballot on the axis. It is instructive to compare some similar countries who received roughly equal numbers of votes on the first ballot but for whom election outcomes differed, such as Switzerland and Germany or Bolivia and Venezuela. In both cases, the state with the judge exhibited higher levels of cooperation.

Conclusion

Implications from this Study

The analysis demonstrates with a high degree of internal validity that states who had one of their nationals elected as an ICC judge were more likely to take potentially costly measures to cooperate with the Court than states who nominated a national but were ultimately unsuccessful. A cautionary note about this result is the small sample size and the resulting inability to allow for more flexible modeling of the relationship between the forcing and outcome variables. Yet, given that there appears to be no relationship between these variables, this objection seems less serious.

The estimated effect could be the consequence of a “middle finger effect,” in that states whose candidates were rejected took affront and cooperated less as a result. There is no anecdotal evidence for such an effect and the theoretical mechanisms (other than sheer frustration) are also unclear. Moreover, there is no statistically significant difference in

signings of nonsurrender agreements between states whose candidates were rejected and states that did not put forth a candidate. Thus, I interpret the evidence as implying that having a national judge increases cooperation.

A bigger issue is how these results generalize to other states. It would be unwarranted to assume that if we randomly assigned judges to states in the international system, we would find the same impact. It is more plausible that the effects would be similar for other ICC State Parties. The decision whether or not to nominate a judge was largely done in cooperation with other states within regions. There are no statistically significant differences in terms of levels of democracy and development among those states that decided to nominate a judge and those that did not. It is not implausible, although it cannot be shown, that a national judge would have similarly affected the states that did not decide to nominate a candidate.

The case study could be extended to other situations involving the Court. For example, in response to the indictment of the Sudanese president Bashir by the ICC, Libyan leader Gaddafi has called upon all African leaders to withdraw from the ICC for what he said was "warped justice in favor of Europe."³² It would be interesting to see whether African countries with judges are more likely to resist such calls.

Theoretically, the findings demonstrate that even indirect institutional participation that confers no meaningful influence on states can have meaningful behavioral effects. Nationals who serve as ICC judges are not representatives of their states. Even if they behaved as such, they could do little to advance the interests of their states. Yet, states appeared more likely to do consequential things as a result of having a judge, including defying the United States by not signing a non-surrender agreement and changing their penal codes. If even a relatively modest increase in the salience and accountability opportunities

³² Africa News, July 2, 2009 "The ICC Represents Hope, Says Kofi Annan." (Lexis Nexis).

can have an impact on state behavior, then it would seem at least plausible that the types of campaigns that are central to the human rights literature may have similar effects.

Other Applications of Regression Discontinuity Designs

A second purpose of this paper has been to introduce the regression discontinuity design to the study of international institutions. It is a central feature of international law that states have the right to voluntarily opt in and out of treaties and institutions. The motivations that lead states to participate in institutions typically also influence the behavior that these institutions seek to affect. As such, studies of the effect of international institutions and law on state behavior have been particularly plagued by selection problems. The RD design can improve the internal validity of causal estimates in the subset of cases where entry into a treaty or an institution occurs after some threshold on a forcing variable is achieved. Other institutions where members are chosen through competitive elections, such as the UN Human Rights Council and the Security Council, may be an obvious extension of the current research. Or, one could ask whether EU member states are more likely to comply with directives issued by a national who serves as a commissioner? But there are other potential applications. For example, some international courts make decisions by majority vote. One could study the impact of courts by evaluating whether policy change was more likely to occur on those issues that narrowly passed versus those that were narrowly rejected. Or, some states use scoring systems to determine whether a country is entitled to development aid (such as the Millennium Challenge Account). One could compare the effect of aid on countries that just passed or failed to pass the threshold score.

Finally, this study hopefully serves as an encouragement for scholars interested in the effects of institutions to look beyond membership and ratification. States' entanglements with IOs vary, sometimes for reasons that they have no precise control over. Such situations could be exploited to investigate the effects of institutional participation on state behavior.

References

- Axelrod, Robert, and Robert O. Keohane. 1985. Achieving cooperation under anarchy: Strategies and institutions. *World Politics*
- Barnett, Michael N. and Martha Finnemore. 1999. The Politics, Power, and Pathologies of International Organizations. *International Organization* 53:699-732
- Bartels, Larry. 1991. Instrumental and “Quasi-Instrumental” Variables. *American Journal of Political Science*. 35(3): 777-800.
- Berk, R. and Jan De Leeuw. 1999. An Evaluation of California’s Inmate Classification System Using a Generalized Regression Discontinuity Design. *Journal of the American Statistical Association* 94(448):1045-52.
- Cook, T., 2008., “Waiting for Life to Arrive”: A History of the Regression-Discontinuity Design in Psychology, Statistics, and Economics, *Journal of Econometrics*. 142(2):636-54
- Dai, Xinyuan. 2005. Why Comply? The Domestic Constituency Mechanism. *International Organization* (2005), 59 : 363-398
2006. The Conditional Nature of Democratic Compliance. *Journal of Conflict Resolution*, 50:690-713.
- Cortell, Andrew P. and James W. Davis, Jr.. 1996. How Do International Institutions Matter? The Domestic Impact of International Rules and Norms *International Studies Quarterly*, 40(4):451-478
- ___ 2000. Understanding the Domestic Impact of International Norms: A Research Agenda. *International Studies Review*, 2(1): 65-87

- Danner, Allison Marston and Beth Simmons. 1997. Sovereignty Costs, Credible Commitments, and the International Criminal Court. Manuscript.
http://ducis.jhfc.duke.edu/wp-content/uploads/archive/documents/Simmons_paper.pdf
- Danner, Allison Marston and Erik Voeten. n.d. Who is Running the International Criminal Justice System? Forthcoming in Avant, Finnemore and Sell *Who Are the Global Governors?* Cambridge University Press.
- Deaton, Angus. 2009. Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development. *NBER Working Paper #14690*
- Downs, George David M. Rocke and Peter N. Barsoom. 1996. Is the Good News About Compliance Good News About Cooperation? *International Organization*, 50(3): 379-406
- Fearon, James D. 1997. Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs. *Journal of Conflict Resolution* 68-90.
- Gourevitch, Peter. 1978. The Second Image Reversed: The International Sources of Domestic Politics. *International Organization*: 32(4): 881-912 Haas, Peter M., Levy, Marc A. and Keohane, Robert O. (eds.) (1993) *Institutions for the Earth: Sources of Effective International Environmental Protection*. Cambridge, Mass.: MITPress.
- Hafner-Burton, Emilie M. 2008. Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem. *International Organization* 62:04, 689
- Hahn, Jinyon, Petra Todd, and Wilbert Van der Klaauw. 2001. Identification and Estimation of Treatment Effects with a Regression Discontinuity Design. *Econometrica* 69(1):201-9.

- Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith. 1999. The Economics and Econometrics of Active Labor Market Programs. In: Orley Ashenfelter and David Card (eds) *Handbook of Labor Economics* Vol 3A Elsevier Science.
- Imbens, Guido W. 2009. "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)"
http://www.economics.harvard.edu/faculty/imbens/files/bltn_09apr10.pdf
- Imbens, Guido W., and Jeremy. Wooldridge. 2009. Recent Developments in the Econometrics of Program Evaluation, forthcoming, *Journal of Economic Literature*.
- Imbens, Guido and Thomas Lemieux. 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics* 142(2): 615-635.
- Johnston, Alastair Iain. 2001. Treating International Institutions as Social Environments. *International Studies Quarterly*, 45(4):487-515
- Keck, Margaret, and Kathryn Sikkink. 1998. *Activists Beyond Borders: Advocacy Networks in International Politics*. Cornell University Press.
- Kelley, Judith. 2007. Who Keeps International Commitments and Why? The International Criminal Court and Bilateral Non-Surrender Agreements. *The American Political Science Review* 101(3):573-89.
- Gary King and Langche Zeng. 2001., "Logistic Regression in Rare Events Data," *Political Analysis* 2001.
- Lebovic, James and Erik Voeten. 2006. The Politics of Shame: The Condemnation of Country Human Rights Practices in the UNHRC. *International Studies Quarterly* 50 (4): 861-88
- Lee, David S. 2008. Randomized Experiments from Non-Random Selection in U.S. House Elections. *Journal of Econometrics* 142(2):675-97.

- Lee, David S. and Thomas Lemieux. 2009. Regression Discontinuity Designs in Economics. NBER Working Paper 14723.
- Lesik, S. A. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *Review of Higher Education*, 30, 1-19.
- Mitchell, Sara and Paul Hensel. 2007. International Institutions and Compliance with Agreements. *American Journal of Political Science*,
- Nichols, Austin. 2007. rd: Stata module for regression discontinuity estimation. <http://ideas.repec.org/c/boc/bocode/s456888.html>
- Pevehouse, Jon. 2005. *Democracy from above: regional organizations and democratization*
- Pierson, Paul. 1996. The path to European integration: a historical institutionalist analysis. *Comparative Political Studies*: 29 (2):123-163
- Ribando, Clare M. 2006. Article 98 Agreements and Sanctions on U.S. Foreign Aid to Latin America. CRS Report for Congress (Order Code RL33337) available at: https://www.policyarchive.org/bitstream/handle/10207/2788/RL33337_20060410.pdf?sequence=1
- Risse-Kappen, Thomas. 1994. Ideas do not float freely: transnational coalitions, domestic structures, and the end of the cold war, *International Organization*, 48(2): 85-214
- Schimmelfennig, Frank. 2001. The Community Trap: Liberal Norms, Rhetorical Action, and the Eastern Enlargement of the European Union. *International Organization* 55:47-80.
- Sikkink, Kathryn. 1993. Human Rights, Principled Issue-Networks, and Sovereignty in Latin America. *International Organization*, 47(3):411-441
- Simmons, Beth. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Forthcoming Cambridge University Press.

- Simmons, Beth, and Lisa Martin. 2002. International Organisations and Institutions. *Handbook of International Relations* (London et al.: Sage, 2002) Walter Carlsnaes, Thomas Risse, and Beth Simmons (eds.),
- Thistlewaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51, 309-317.
- Tomz, Michael, Gary King and Langche Zeng. 2003. *Journal of Statistical Software* 8, no. 2 (2003).
- Tomz, Michael. 2007. *Reputation and International Cooperation: Sovereign Debt across Three Centuries*. Princeton, NJ: Princeton University Press.
- Van der Pas, Sara. 2006. Progress Report on Implementing the Rome Statute into National Law *The International Criminal Court Monitor* Issue 32, p.5/ 32 • May 2006 http://www.iccnw.org/documents/Monitor32eng_WebFinal.pdf (accessed June 29, 2009).
- Voeten, Erik. 2008. "The Impartiality of International Judges: Evidence from the European Court of Human Rights," *American Political Science Review* 102(4): 417-433
- _____. 2007. "The Politics of International Judicial Appointments: Evidence from the European Court of Human Rights". *International Organization* 61(4):669-701
- Von Stein, Jana. 2005. Do treaties constrain or screen? Selection bias and treaty compliance. *American Political Science Review*,

Table 1: Rare Event Logistic Estimation on Whether State Signed BIA Agreement

	<i>Model 1</i>	<i>Model 2</i>
Elected	-4.94 (2.65)**	-4.35 (2.05)**
Votes (highest total)	.11 (.06)	-
Votes on first ballot	-	.11 (.06)
Log GDP	-1.01 (.53)**	-1.17 (.62)**
Civil/Political Liberties	.58 (.30)**	.57 (.33)**
Constant	4.58 (5.18)	6.21 (6.13)
N	43	43

Robust standard errors in parentheses, * significant at 10%, ** significant at 5%; *** significant at 1% (one-tailed tests).

Table 2: Linear Regression on Cooperation with ICC

	<i>Model 1</i>	<i>Model 2</i>
Elected	2.86 (1.69)**	2.19 (1.08)**
Votes (highest total)	-.04 (.05)	-
Votes on first ballot	-	-.03 (.03)
Log GDP	.94 (.38)***	.98 (.38)***
Civil/Political Liberties	-.54 (.18)***	-.51 (.17)***
Conflict	1.05 (.70)*	1.03 (.70)*
Common Law	-1.08 (.81)*	-1.04 (.81)
Constant	-1.71 (4.63)	-2.55 (4.47)
R ² _{adj}	.55	.55
N	43	43

Robust standard errors in parentheses, * significant at 10%, ** significant at 5%; *** significant at 1% (one-tailed tests).

Figure 1: Example of Regression Discontinuity Design

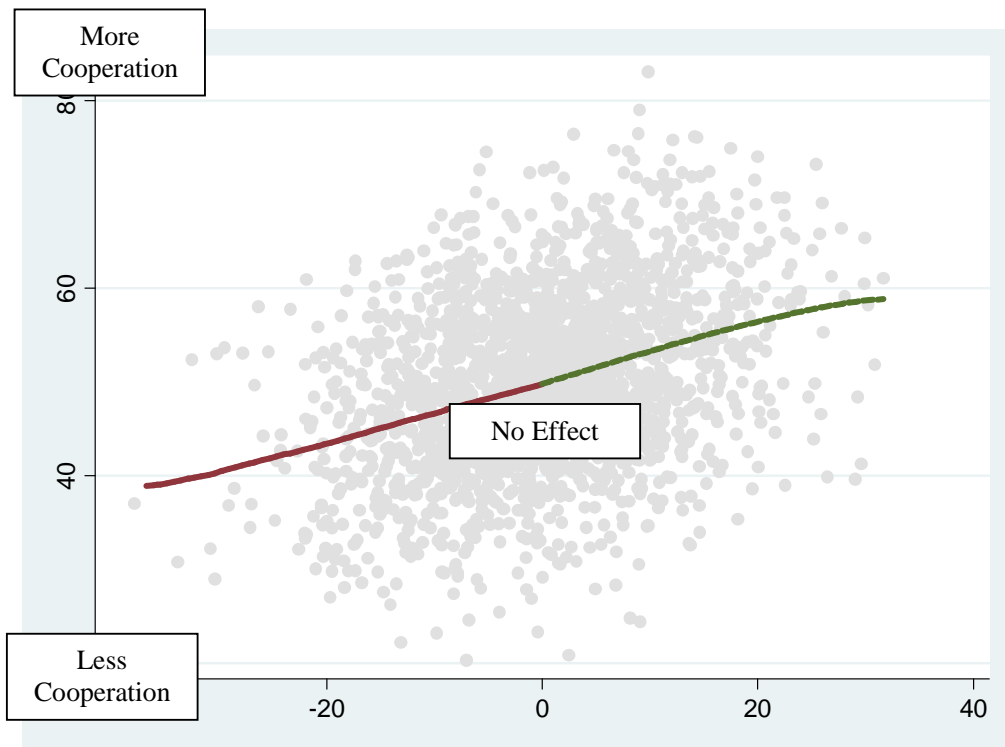
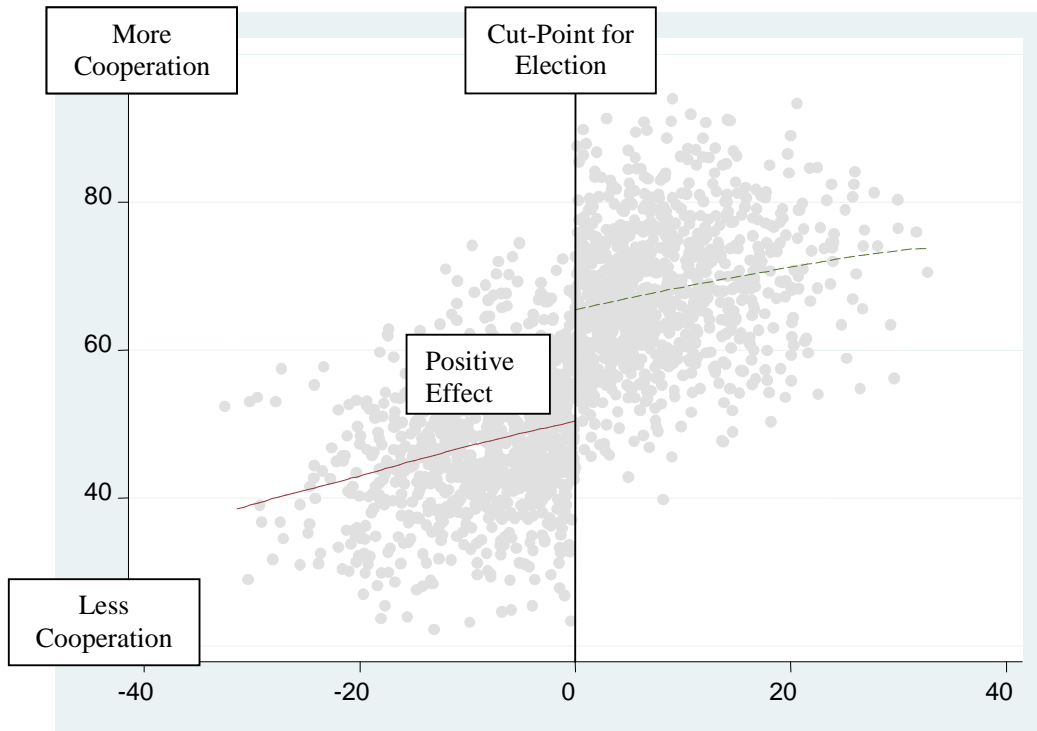


Figure 2: Residuals of Regression on Cooperation against Vote Share and Election Status

