

Does a Professional Judiciary Induce More Compliance?: Evidence from the European Court of Human Rights *

Erik Voeten
Dept. of Government
School of Foreign Service
Georgetown University
ev42@georgetown.edu

May 10, 2012

DRAFT - FIRST VERSION, prepared for presentation at the MultiRights Conference, Oslo June 1-2. Please contact author before citing or quoting (please check SSRN for updates). Suggestions for improvement are greatly appreciated.

Abstract

Are judgments written by professional judges more likely to elicit compliance? I exploit exogenous variation in the composition of European Court of Human Rights (ECtHR) chambers to evaluate this question. I find that judgments written by panels with a relatively high proportion of career judges are more likely to be implemented quickly than are judgments that come from panels with high proportions of former academics, legal diplomats, politicians, or private practitioners. This suggests that the performance of international courts could be improved by changing their professional composition. Theoretically, it implies that compliance is not merely a function of political incentives but also of the characteristics of those who interpret international law. Professional judges are both important compliance constituencies and consumers of legal reasoning. I build on sociological theories of the legal profession to explain why professional judges may be more amenable to judgments written by other professional judges.

*I thank Sharan Grewal for excellent research assistance.

1 Introduction

Are judgments written by professional judges more likely to induce compliance? Efforts to improve the performance of international and domestic courts frequently target the qualifications of judges as a particularly important area of reform. Yet, there is, to my knowledge, no systematic evidence that more qualified judges also produce judgments that are more likely to be implemented. Indeed, scholarship that examines the effectiveness of courts rarely concerns itself with the qualities of judges or judgments. Instead, the focus is almost exclusively on the capacities and incentives of those charged with implementing court decisions.

I investigate this issue on an international court with a great deal of variation in the background of its judges: the European Court of Human Rights (ECtHR). Some ECtHR judges are experienced former national judges whereas others have primarily had a career in academics, politics, legal diplomacy or private practice. I hypothesize that judgments written by panels with high proportions of career judges are more likely to be implemented quickly than those written by former politicians, diplomats, and academics.

I do not argue that professional judges necessarily write better judgments. Quality is in the eye of the beholder. Instead, I claim that they are more likely to write judgments that appeal to a constituency that both has significant authority to implement judgments and that pays attention to legal reasoning. Academics are consumers of legal reasoning but they are generally not important compliance constituencies for ECtHR judgments. Politicians and legal diplomats are compliance constituencies but they are more responsive to incentives and capacity constraints than legal argumentation. By contrast, professional judges are both important compliance constituencies and are likely to pay attention to how a judgment is

motivated. They have practice in how to write judgments that appeal to domestic compliance constituencies. To the extent that there is a transnational field of judicial professionals with shared ideas and practices about what constitutes an appropriate legal justification, those with practice in this field may be more competent in writing judgments that meet the expectations of national legal professionals.

I test this hypothesis using a dataset of 888 ECtHR lead cases and find strong evidence for it. The greater the proportion of career judges on a panel, the quicker a judgment is implemented. Since the proportion of former review judges is largely determined by exogenous processes (lot and rotation), I argue that this can plausibly be given a causal interpretation. There is also tentative support for the mechanism: career judges better embed their judgments in case law and this appears to encourage implementation. Naturally, I do not claim that this effect is all important. Increasing the percentage of career judges by one standard deviation (fifteen percentage points) increases the hazard of implementation by about 15%.

This paper focuses on an aspect of international governance that political scientists have largely avoided: the competence of international agents charged with interpreting and executing mandates. The main practical contribution of this paper is to provide rigorous social scientific evidence for how the quality of an international institution can be improved. The findings show that there are real negative consequences for the effectiveness of international courts if governments use international judicial appointments as a reward for politicians or legal diplomats who have proven loyal and/or who have outlived their domestic utility. At the same time, raising the qualifications of international judges is a promising target for

reforms that, while not easy, may often prove more feasible than other factors known to affect compliance, such as the level of democracy in target states. Indeed, there are successful examples of such reforms on both international and domestic courts. This paper offers not only empirical support for such reforms but also suggests what qualifications should be highlighted if implementation is the main concern of the reform effort.

The basic idea that more legitimate legal obligations are more likely to be implemented is widely shared across the range of compliance theories (Abbott et al. (2000); Franck (1990); Chayes and Chayes (1993); Simmons (1998)). Yet, it has been difficult to assess this claim in a non-tautological way (Hafner-Burton, Victor and Lupu (2012)). This paper is an attempt to test a more narrow but better identified variant of that claim: namely that international judges can affect state behavior through legal reasoning. I argue that this occurs not so much because states internalize a respect for international law or because politicians are swayed by the persuasive power of law. Rather, states are disaggregated actors in which professional (international) lawyers and judges have become important decision-makers (e.g. Finnemore (2004); Dezalay and Garth (2002)). I built on the sociological literature that analyzes the transnational judiciary as a “field” (e.g. Bourdieu (1986); Dezalay and Garth (1996)) to explain why these domestic legal interlocutors may be more inclined to accept international decisions if these come with acceptable legal justifications.

This claim differs from both standard rationalist and constructivist accounts. Rationalists tend to avoid explanations that rely on one group of actors being superior at some task. Constructivists have highlighted rational-legal authority as an important source of legitimacy for international institutions (e.g. Barnett and Finnemore (1999)). Yet, their focus has been

on organizational culture rather than agents or professions.

I proceed with an exposition of the theoretical claims followed by a description of the data gathered for the ECtHR case. The next section presents the main finding and plausibility probes that the effect is due to legal reasoning rather than leniency or experience. The conclusion speculates on how these findings can and cannot be generalized to other settings and discusses potential avenues for future research.

2 Theory

The theory is based on two claims. The first is that national judges have an important discretionary role in the implementation of ECtHR judgments. This point is widely accepted in the ECtHR literature (Keller and Stone Sweet (2009); Von Staden (2009); Hillebrecht (2009*a*)) and in the literature on compliance with international human rights law more generally (e.g. Simmons (2009); Koh (1998)).

All Council of Europe members have implemented the European Convention of Human Rights (hereafter the Convention) into national law. This makes the international obligations actionable in domestic courts. Yet, the ECtHR cannot on its own compel domestic courts to respect its jurisprudence. This gives domestic judges leeway in if, how, and when they accept the ECtHR judgments.

The relationship between domestic courts and the ECtHR can be competitive (Keller and Stone Sweet (2009)). This is not surprising. Individuals must exhaust domestic remedies before their ECtHR applications are admissible. Thus, when the ECtHR finds a violation, it often explicitly rejects national court rulings. Domestic courts sometimes accept the

ECtHR’s interpretations. Yet at other times, they explicitly reject or ignore relevant ECtHR case law. At other times, domestic judges give a narrow interpretation to ECtHR rulings, requiring the government to implement the individual measures required by judgments (such as to pay compensation) but not to take more general measures to prevent future violations. Or courts can defer implementation to the political sphere. This is the route the UK High Court took when it claimed that there are “no reasonable grounds in domestic law” for giving prisoners compensation for the government’s ban on prisoner voting, even though the ECtHR had found the UK’s ban a violation of the Convention, had granted prisoners compensation, and the Convention has been adopted into British law through the 1998 Human Rights Act.¹

Other domestic courts defer entirely to the ECtHR for interpretation of Convention rights. Keller and Stone Sweet (2009) conclude a large edited volume on the receptiveness of national courts to ECtHR rulings with the observation that most courts occupy “a mid-point, between conflict and deference, wherein courts seek to forge a cooperative relationship with the Strasbourg Court through dialogue and comity” (p. 705).

If national judges both have at least some authority *and* some discretion with regard to the implementation of ECtHR judgments, then it is at least plausible that more persuasive judgments lead to more judicial cooperation and thus faster implementation than do less persuasively reasoned judgments. The persuasiveness of judgments is, however, a highly complex and subjective concept. I focus on a narrower claim: that judgments written by professional judges are more likely to be accepted by other professional judges than are judgments predominantly written by judges with other professional backgrounds. The claim

¹See: “Prisoners’ vote compensation claims blocked by high court” *The Guardian*, February 18, 2011.

here is not that profession is the only factor that matters for the persuasiveness of legal reasoning or that judges always agree with each other. Rather, it is that professional judges share certain practices, norms, and modes of reasoning and that judgments that exhibit these modes of reasoning are more likely to be valued by other professional judges. This is particularly important for decisions by domestic judges to accept or reject an international judgment that potentially encroaches on domestic practices.

This claim requires that there is such a thing as a judicial profession whose norms and practices transcend borders. Professions are traditionally understood to be closely tied to the nation-state (e.g. Fourcade (2006)). Yet, there is increasing attention for their internationalization. The legal profession is a common example. Sociologists have shown how legal professionals have shaped and have been shaped by the creation of international jurisdiction in the fields of commercial arbitration (Dezalay and Garth (1996)), international criminal law (Hagan and Levi (2004)), European integration (Cohen and Vauchez (2007)), and, indeed, the ECtHR system (Madsen (2007)).

International jurisdiction creates competitive struggles over whose authority will prevail. Yet, while there may be competition between international and domestic judges, their common engagement in a judicial enterprise also creates potential alliances against non-judicial actors. Scholars of both the ECJ (Alter (1998); Weiler (1993)) and the ECtHR (Keller and Stone Sweet (2009)) have emphasized such strategic alliances between national and international judges. I add here that the degree to which such an alliance is attractive for national judges depends in part on the degree to which the international judgment persuasively distinguished itself from non-judicial realms of social life.

Anne-Marie Slaughter (1999) argues that the increased interaction between national and international judges as well as national judges across borders has led to “a deep sense of participation in a common global enterprise of judging, an awareness that provides a foundation for a global community of law” (p. 1104). Unlike Slaughter, my argument is less about networks and interactions among judges across borders than about shared practices that lead judges across borders to have similar ideas about what amounts to a properly reasoned legal argument. Sociologists frequently build on the idea that the judicial profession constitutes a field (Bourdieu (1986)). Bourdieu rejects the notion that the law is a self-contained system that is autonomous from its social and political environment, yet there are a set of unique qualities that separate legal practice from other activities and which limit the realm of specifically juridical solutions. As Bourdieu (1986) puts it:

“Even though jurists may argue with each other concerning texts whose meaning never imposes itself with absolute necessity, they nevertheless function within a body strongly organized in hierarchical levels capable of resolving conflicts between interpreters and interpretations. Furthermore, competition between interpreters is limited by the fact that judicial decisions can be distinguished from naked exercises of power only to the extent that they can be presented as the necessary result of a principled interpretation of unanimously accepted texts. Like the Church and the School, Justice organizes according to a strict hierarchy not only the levels of the judiciary and their powers, and thereby their decisions and the interpretations underlying them, but also the norms and the sources which grant these decisions their authority.”

In the ECtHR case, the hierarchy is more contested. This makes the norms and sources that grant decisions their authority even more important. In order to rationalize their judgments, judges resort to language of autonomy, neutrality, and universality. One need not accept Bourdieu’s broader theoretical framework or methodological dispositions to see the basic point that judges across societies face a similar problem in distinguishing legal justifications

from other justifications and use similar strategies to achieve this. Using more instrumental reasoning, Shapiro (1986) argues that in all modern societies judges face the same problems in legitimating their decisions to persuade audiences that the judges chose the outcome because the law required it not because the judges desired it. As Alec Stone Sweet points out: “judges in Strasbourg confront the same kinds of problems that their counterparts on national constitutional courts do; and they use similar techniques and methodologies to address these problems” (Stone Sweet (2009), p.1).

My claim then, is that such justifications are not simple exercises in cloaking political judgments in legal language but that they require both skill and practice.² Career judges have gotten, through education and practice, more proficient in constructing such arguments than other professionals. It may well be that they develop these skills even more while they serve on the international court, a possibility that I examine empirically later on. Yet, it is important to distinguish the professional skills of career judges from international judges with different backgrounds. Legal diplomats are well versed in representing the interests of their governments in foreign countries and international institutions. Politicians may develop persuasive skills targeted at voters or other politicians. Legal academics develop skills in legal reasoning but the contest for competence among academics nonetheless differs from that of judges. Professional judges may be more proficient in using autonomous, neutral, and universal language (Bourdieu (1986)). For example, they may be better at using relevant case law to illustrate that the law has been applied in similar way to others. They may also have more experience in what types of arguments are often rejected by bureaucrats and

²For a more general treatment of the practice approach in international affairs, see Adler, Pouliot et al. (2011).

politicians. Other judges or legal professionals charged with implementing judgments may be more likely to accept these arguments as falling within the realm of juridical solutions.

While arguments about the distinctiveness and the transnational nature of the judicial profession are widely accepted in sociology, there is little quantitative evidence that this affects outcomes. The literature on differences between professional and lay judges is almost exclusively concerned with criminal courts, where there has been a long-standing practice of mixed tribunals in many countries. This literature focuses not on legal reasoning but on the relative punitiveness of professional judges (e.g. Diamond (1990)) or the legitimacy effect of mixed tribunals (e.g. Ivkovic (2003)). Recent comprehensive surveys of the social science literature on international law confirm that there are no studies that examine the effect of court composition or legal reasoning on compliance (Hafner-Burton, Victor and Lupu (2012); Shaffer and Ginsburg (2012)). Indeed, Shaffer and Ginsburg (2012) contrast the ethnographic work on the legal profession inspired by Bourdieu with the more quantitative literature on compliance and judicial behavior. This is an attempt to reconcile these literatures.

3 Data

3.1 The Implementation of ECtHR Judgments

The dependent variable is the time between an ECtHR judgment and the adoption of a final resolution by the Council of Europe’s Committee of Ministers, the supervisory organ for the execution of ECtHR judgments. The final resolution indicates that the judgment has been satisfactorily implemented according to the Committee. The sample is limited to “lead

cases,” which are cases that first reveal a new structural or general problem in a respondent state. Frequently, lead cases are followed by duplicate cases that involve different victims but identical or highly similar legal issues. Judgments on follow cases generally borrow their legal reasoning from the lead cases. Moreover, lead cases are often grouped with their follow cases in terms of the monitoring of their execution. As such, the follow cases cannot be evaluated as independent judgments.

There are 888 judgments on leading cases adopted between 1968 and 2006 for which we have full data on judicial composition of the panels and time until the final resolution. Of these, 20% remained pending as of June 8, 2011. Final resolutions were adopted anywhere between 48 days and 5139 days after the judgment with an average of 1384 days. For more details on data collection, I refer to Grewal and Voeten (2012).

This dependent variable measures only if and when the judgment is fully implemented, including potential legislative changes. It may well be that parts of the judgment, such as payment of just satisfaction, were completed earlier than the date of the final resolution. Such partial compliance is not examined here (see Hawkins and Jacoby (2009); Hillebrecht (2009b)). There are two justifications for this choice. First, it avoids subjective coding. Second, legal reasoning focuses mostly on justifications for why domestic actions, practices, or laws are a violation of the European Convention. If we want to assess whether legal reasoning affects behavior, we should evaluate whether these actions, practices, or laws have been modified to come into accordance with the Convention.

3.2 The Professional Make-Up of ECtHR Chambers

Critiques to the credentials of ECtHR judges have often focused on judicial experience. To take one extreme, the British paper *The Daily Mail* took the trouble of individually examining the credentials of each justice who decided against the UK in a prisoner voting rights case and concluded that:

The 17 who were appointed to the panel who rode roughshod over 100 years of British law to call for prisoners to get the vote included many with no judicial experience. Some were academics, some politicians and some human rights campaigners. Appointed by their home nation, often for political rather than for their legal expertise[.]. (Slack (2011))

Similar critiques have come from NGOs, academics and judges. The NGO Interrights found in a 2003 report that:

In practice, even in the most established democracies, nomination often involves a “tap on the shoulder” from the Minister of Justice or Foreign Affairs, and frequently rewards political loyalty more than merit. Nominees often lack the necessary experience and even fail to meet the very general criteria set out in the Convention (*Report on Judicial Independence* (2003), p.9).

Former Judge Loukis Loucaides concluded a reflection of his time at the Court with the sober conclusion that “not all members of the Court had the required competence” (Loucaides (2010)). A comprehensive study in which domestic interlocutors with the ECtHR were interviewed found that concerns about the quality of ECtHR judges and judgments was prevalent among domestic judges but rarely expressed by politicians or bureaucrats (Cali and Bruch 2011). To be sure, there are many ECtHR judges with extensive judicial experience but there is considerable variability.

The key independent variable is for each case the proportion of the ECtHR panel (chamber) whose former career was primarily that of a judge. Unlike on the U.S. Supreme Court, ECtHR judgments do not identify who actually wrote the opinion (although there is information about dissents). Thus, we have to make an assumption that the influence of profession on judicial reasoning is proportional to the number of judicial professionals on the Court.

Most cases are decided by a panel of seven judges. A Grand Chamber of seventeen judges decides cases deemed especially important by the Court or when a respondent government requests that a judgment by a regular chamber is reviewed. 36% of judgments in the database came from the Grand Chamber.

The number of ECtHR judges equals the number of member states. Since 1998, the Council of Europe's Parliamentary Assembly elects judges from a (sometimes rank ordered) list of three candidates proposed by each of the 47 member state governments. These governments submit standardized CVs for all candidates in French and/or English. In earlier years, governments proposed only one candidate. Biographies for these judges are available from the *Yearbook of the European Convention on Human Rights*. If the individual had not been a judge, I coded her as either an academic, private practitioner, politician, or a legal diplomat.

Many ECtHR judges fit multiple roles. I coded the role that was most prominent. Mostly this coincides with the last primary position an individual held before acceding to the ECtHR. Nevertheless, there were a few cases where a government's preferred candidate had conveniently been appointed to a high court just in time to be candidate for a European judgeship even though there was no evidence of a pre-existing judicial career. In these cases,

I coded the most prominent previous career (all of these cases concerned career politicians).

Each panel includes either the judge of a respondent government or an *ad hoc* judge appointed specifically for the case. Even though national judges are not formal representatives of their governments, they are still much more likely than other judges to vote against the finding of a violation (Voeten (2008)). This is even stronger among ad hoc judges (ibid). More troublesome, if those respondent governments that are more likely to appoint career judges to the ECtHR are also more likely to implement judgments quickly, then we could overestimate the effect of panel composition. I therefore exclude national judges from the calculation of panel composition (although the results hold when national judges are included). So, the government in charge of implementing a judgment has had no say over the appointment of any of the judges used to calculate the panel composition variable. Moreover, the results hold when the sample is limited to those cases where national judges dissent from findings of a violation (a frequent occurrence) and thus had no direct say in the writing of the majority judgment.

On average, 34% of the panels consists of career judges. The minimum is 0 and the maximum .88 (standard deviation .15). The other two most prominent categories are academics (25%) and diplomats (23%). Politicians constitute about 11% of all ECtHR judges and the remaining 7% were private practitioners. These descriptives are consistent with other efforts to code past professions of ECtHR judges (Bruinsma (2007); Madsen (2011)). There is no significant linear or non-linear trend in the proportion of judges on panels over time. The finding that the professional make-up of the Court has been fairly constant is also consistent with other research (Madsen (2011)).

In order to evaluate the possible exogeneity to the implementation of judgments, it is necessary to examine how judges are allocated to chambers. Each judge is a member of a Section which is “geographically and gender balanced and shall reflect the different legal systems among the Contracting Parties.”³ Judges rotate among sections at least once every three years. Within sections, judges are designated to Chambers on individual cases “by the President of the Section in rotation from among the members of the relevant Section.”⁴ In practice, the Section Registrar maintains the schedule for assigning judges.

The Grand Chamber includes the President of the Court and the President of all Sections. The remainder of the judges are assigned “by a drawing of lots by the President of the Court in the presence of the Registrar.”⁵

We cannot exclude with certainty that Section Presidents circumvent rotational principles and systematically stack former national judges on “easy cases,” although it is not obvious why they would do so. It may also be that some Sections have disproportionately high or low numbers of former review judges through random chance. This could affect balance because each Section has responsibility for cases coming from a fixed group of member states.

To check this, I ran linear regressions of *ex ante* characteristics of cases and respondent governments on the proportion of former national judges that serve on a panel. Figure 1 reports regression coefficients from fifteen individual regressions. There is no evidence that either the number of Convention articles at stake (indicating possibly more complicated cases) or cases that deal with specific articles of the Convention have relatively high or low proportions of former national judges. The figure highlights the most common violations

³*Rules of the Court*, 25(2)

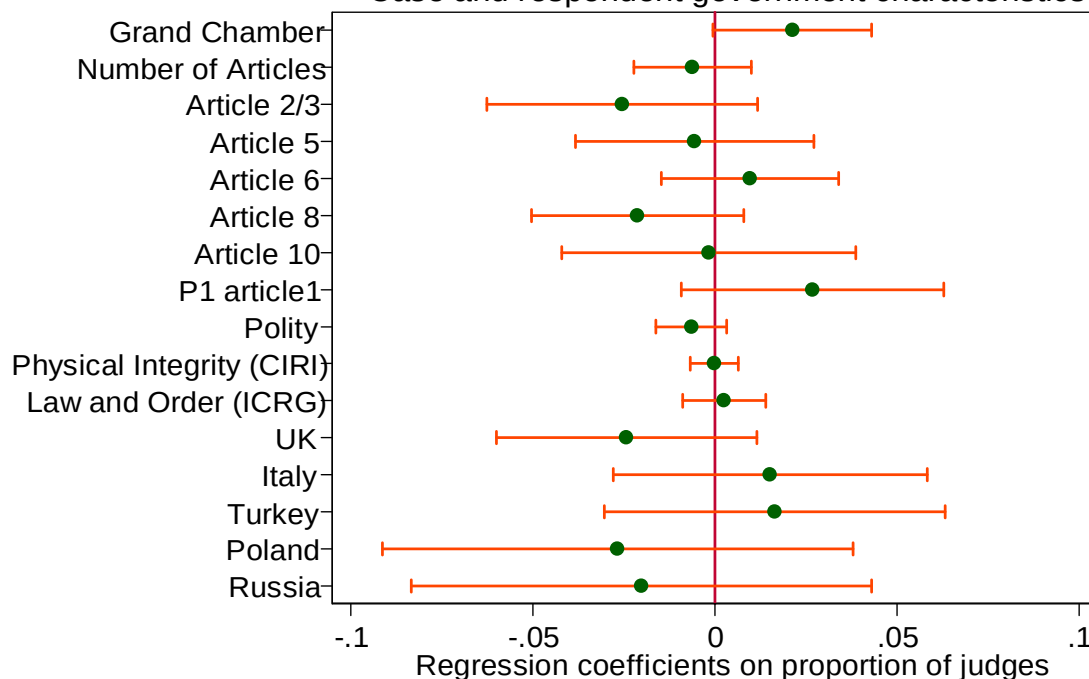
⁴*Rules of the Court*, 26 (1)

⁵*Rules of the Court*, 24 (2e)

found by the ECtHR: death or inhumane treatment/torture (article 2 and 3), prisoner rights (article 5), the right to an independent and impartial trial (article 6), privacy (article 8), freedom of speech (article 10), and property rights (Protocol 1, article 1). There is some evidence that judgments decided by the Grand Chamber have relatively higher proportions of career judges. This is due to the fact that Section Presidents are somewhat more likely to be career judges. Nevertheless the effect is small (on average the Grand Chamber includes 2,5% more professional judges) and the 95% confidence level includes zero (although barely).

There is also no evidence that cases dealing with more difficult respondent governments attract panels with relatively few former national judges. There is no correlation between the proportion of former judges and the democracy levels, human rights violations or the levels of law and order in respondent governments. Moreover, cases from frequent violators and known poor implementers (the UK, Poland, Italy, Russia, and Turkey) are not decided by panels with relatively few or many career judges. Moreover, career judges are not different ideologically from their counterparts (Voeten (2007)).

Figure One: Correlates of Panel Composition
Case and respondent government characteristics



The process of panel assignment is partially organized through processes that are entirely exogenous to implementation (lot and rotation). Still, we cannot exclude with certainty that strategic actors had some control over panel assignment. It is not obvious why these actors should assign former national judges to cases that are easy to implement and I know of no tales that claim this. Moreover, there are few if any observable differences between cases decided by panels with many or few former national judges. While we cannot have the same assurances over random assignment as we would in a well-executed field experiment, this is about as close as it gets to exogenous variation in observational studies of international affairs.

4 Findings

4.1 Methods

The dependent variable is the duration in days between the ECtHR judgment and a resolution by the Committee of Ministers that the judgment was implemented satisfactorily. The dependent variable is censored: 20% of all judgments remain pending at the last time of measurement. These features make the use of event-history (or “duration” or “survival”) models most appropriate (e.g. Box-Steffensmeier and Jones (2004)).

The primary method of estimation is the semi-parametric Cox proportional hazards model. The Cox model is attractive because it leaves the distributional form of baseline hazard rates unspecified. The hazard rate for the i th judgment at time t (since judgment) is:

$$h_i(t) = h_{0j}(t) \exp(\beta * Judges) \quad (1)$$

$h_o(t)$ is the baseline hazard function, which is assumed to be unknown and is left unparametrized. Given that there are multiple judgments for each country and that there are vast differences in the implementation rates of countries, I estimate a stratified Cox model, which allows the form of the hazard function to vary for each country j .⁶ Since *Judges* is measured as a proportion, β informs us how much more likely it is that at any given time a judgment is implemented when it is written by a panel of all professional judges ($Judges = 1$) versus a panel that contains no professional judges ($Judges = 0$). Since panels

⁶Aside from addressing heteroscedascity, stratifying also helps with reducing violations of the proportional hazards assumption discussed in the following paragraph.

of all professional judges are nonexistent and panels of no judges are quite rare in the data, I will instead focus inference on interpreting changes that amount to a standard deviation shift in the proportion of judges (.15). In the tables, coefficients are transformed into hazard ratios (the ratio of hazard rates for $Judges = 1$ and $Judges = 0$.)

If we are willing to believe that the proportion of judges is exogenous to implementation, then equation 1 generates an unbiased estimate of β without introducing other covariates. Introducing pre-treatment covariates can improve the efficiency of estimates and can correct for small imbalances, although conditioning could also introduce bias. I present both unconditional estimates and estimates that are conditional on the case and country characteristics presented in figure 1.

For the Cox models to generate unbiased estimates, we must assume that the effect of covariates is consistent at all times t (Box-Steffensmeier and Zorn (2001)). Throughout, I employ tests based on Schoenfeld residuals to assess whether this assumption is met and include interactions with time when it is not.

4.2 The Effect of Professional Judges on Implementation

Table one present hazard ratios from unconditional Cox regressions as well as Cox regressions that include the pre-treatment case and respondent government characteristics introduced in figure 1. The effects of having a panel composed of former review court judges are consistently significant. On average, the estimated hazard ratio is about two. This means that the risk that a judgment is implemented at any time is about twice as high for a judgment written by only professional judges compared to a judgment written a panel that consists of no

professional judges. However, this is a counterfactual that goes outside the bounds of the observed variation in panel composition. A one standard deviation increase in the proportion of professional judges (.15) increases the hazard of implementation by about 15%. This is a modest but consequential difference.

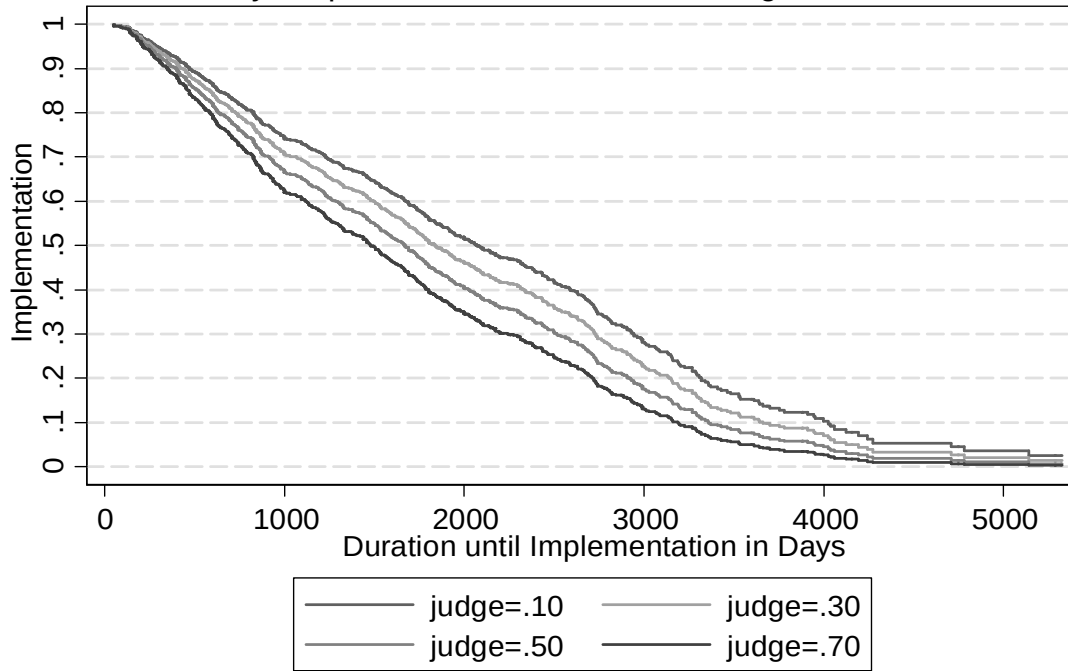
Table 1: Stratified Cox Models (coefficients are hazard ratios, standard errors in brackets)

Model	1	2	3	4
Proportion professional judges	2.36 (.63)***	2.03 (.55)***	1.85 (.53)**	3.08(1.21)***
Proportion academics	-	-	-	1.34(.57)
Proportion legal diplomats	-	-	-	1.61(.74)
Case characteristics	No	Yes	Yes	No
Government characteristics	No	No	Yes	No
N	888	888	825	888
Log-Likelihood	-1976.93	-1908.28	-1788.09	-1976.39

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure two presents the estimates graphically at varying proportions of former national judges. A judgment written by a panel that consists of only 10% of professional judges has a 50% chance of being implemented after about 2000 days. This is 500 days shorter for a judgment written by a panel that consists of 70% professional judges. Even more modest variation in the proportion of judges (.3 versus .5) is associated with more than 100 days more to implement 50% of judgments. These are consequential differences in the delay of justice.

Figure 2 Cox Proportional Hazard Estimates
by Proportion of former National Judges on Panel



Models 2 and 3 include the case and country characteristics from figure 1. I include interactions with time in order to correct for violations of the proportional hazards assumption for the number of articles, and dummies for article 5 and 6 violations. Cases that involve alleged violations of many Convention articles and cases that deal with politically sensitive issues (such as alleged torture or inhumane treatment) take longer to implement. Respondent government characteristics are all measured in the year of the judgment. After stratification, no respondent government characteristic significantly correlates with more or less speedy implementation of judgments. For a more detailed discussion of the correlation of case and respondent government characteristics, I refer to Grewal and Voeten (2012). The important thing here is that the effect of panel composition is robust to the inclusion of

these pre-treatment covariates even though their inclusion is not necessary if we believe that the proportion of professional judges is exogenous to the processes that lead to a speedier implementation of judgments.

Model 4 shows that the finding is not an artifact of the phenomenon that variation in the proportion of professional judges affects the professional composition of panels in other ways. The finding is robust to the inclusion of the proportion of former legal diplomats and academics.⁷ Neither of these variables significantly affect judgment implementation, either when introduced together or separately. The findings are robust to estimating a shared frailty Cox regression and to excluding the pre 1990 cases from the analysis. If we limit our attention to judgments issued since 1990, the estimated hazard ratio is 2.17 (s.e .63), nearly identical to the finding for the entire period.

4.3 Why?

The previous section established that there is a robust statistically significant and substantively important relationship between the proportion of professional judges who write ECtHR judgments and the speed with which these judgments are implemented. Given that the proportion of professional judges is plausibly exogenous to implementation, this can be interpreted as a causal effect. Understanding precisely why this causal effect may occur is both important and extraordinarily difficult as it requires consideration of factors that are difficult to measure and that are likely endogenous to implementation (see Imai et al. (2011)).

This subsection is devoted to a series of plausibility probes regarding the role of legal

⁷Note that the size of the hazard ratio in model 4 is different because the reference category is no longer the same.

reasoning, experience, and leniency as the driving forces behind this effect.

4.3.1 Legal Reasoning

The theory section privileges legal reasoning as the reason why panels with more professional judges are better able at inducing implementation. This argument is based on the notion that national judges are an important compliant constituency and are consumers of legal justifications. To the extent that there is a transnational field of judicial professionals with shared ideas and practices about what constitutes an appropriate legal justification, those with practice in this field may be more competent in writing judgments that meet the expectations of national judges.

Evaluating this mechanism proceeds in at least two steps, each with its own hurdles. The first step is to examine whether panels with more judges are indeed more likely to motivate their judgments in ways that meet the expectations of other professional judges. The main obstacle is that it is extremely difficult to observe judicial reasoning. The dominant way to evaluate the quality of judges, courts, and judgments is to examine citations (Caldeira (1983); Landes, Lessig and Solimine (1998)). A similar approach is possible here based on a network analysis of citations within the ECtHR (Lupu and Voeten (2012)). Panels with more professional judges may well embed their judgments more carefully in the the case law of the Court. Opinions that more carefully argue that a decision follows from a consistent application of the law t may more successfully stake the claim that the government lost because the law required it to lose rather than because a particular set of international judges thought the government deserved to lose.

The key dependent variable is the hub score at the time of judgment. The hub score is determined by the number of other cases a judgment cites as well whether the cited cases have greater “authority,” meaning that the cited cases have themselves been cited a lot by other cases. A larger hub score suggests that the judgment may be more authoritative as a matter of law. Large hub scores may simply be a function of case characteristics or the presence of sufficient precedent. Yet, they could also be shaped by agency: professional judges may better utilize the precedent that exists than especially non-legal professionals such as politicians.

The mean hub score in this sample is 104 with a standard deviation of 145. The hub score is left-censored at zero so I follow Lupu and Voeten (2012) in estimating a tobit model. The model includes the same case characteristics as in figure 1 with others found influential in Lupu and Voeten (2012): whether a government filed a preliminary objection that a case should not be evaluated on its merits and how many cases (on a given article and in general) there were in the network at the time of the decision.

Table two reports the results both with and without respondent government fixed effects. There is support for the notion that panels with high proportions of judges embed their judgments better in case law than politicians (the reference category) but the differences between judges and academics are not significant and the difference between judges and legal diplomats are not robust. The strongest difference is between panels with high proportions of judges and high proportions of politicians (see model 3 and 4). So, while the coefficients indicate that panels with more professional judges on average embed their judgments better in case law, the major difference is between legally trained and non-legally trained individu-

als. Nevertheless, when non-judge legal professionals are grouped together in models 3 and 4, the difference between these legal professionals and professional judges is significant at the 5% level.

Table 2: Tobit regressions on hub scores (robust standard errors clustered on respondent governments in parentheses)

	1	2	3	4
Judges	102.33(39.01)***	80.70(40.16)**	45.90(19.69)**	34.48(17.59)**
Academics	88.49(51.15)*	68.18(44.92)	-	-
Legal Diplomats	35.93(36.00)	52.77 (39.15)	-	-
Politicians	-	-	-97.69(44.55)**	-76.39 (44.49)*
Country Fixed Effects	No	Yes	No	Yes
Issue Characteristics	Yes	Yes	Yes	Yes
N	885	885	885	885

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The second, even bigger, challenge is to investigate if variation in legal reasoning leads to variation in implementation. Even if we had a perfect measure of the theoretically relevant concept of legal reasoning, it may be that judges exert more effort on cases where implementation is expected to be problematic. Indeed, Lupu and Voeten (2012) find that ECtHR judges are more careful to embed judgments in its existing case law when the expected value of persuading domestic judges is highest. For example, they find that cases concerning torture and inhumane treatment are more carefully embedded in case law than are other cases. These cases are also slow to be implemented. This might bias us against finding a relationship between hub scores and implementation.

Given that the professional composition of panels is plausibly exogenous, it may be tempting to think we can use this as an instrument. The trouble is that the proportion of judges likely influences implementation through other channels than citations. If so, then this would violate the exclusion restriction necessary to identify causal effects using instrumental variables.

Model 1 in table 3 assumes that the endogeneity issue can be addressed by including the case characteristics from table 2 as covariates. The hub score is positively correlated with implementation: the better a judgment is embedded in case law, the quicker a judgment is implemented. A one standard deviation increase in the hub score increases the hazard ratio by about 10 percent. However, introduction of the hub score does not reduce the coefficient on the proportion of judges, suggesting that the proportion of judges also affects implementation in other ways.

Table 3: Stratified Cox Models (coefficients are hazard ratios, standard errors in brackets)

Model	1	2
Proportion professional judges	2.25 (.63)***	2.12(.63)**
Hub score	1.0007(.0003)**	-
Hub score (instrumented)	-	1.45(.20)***
Case characteristics	Yes	Yes
N	885	885
Log-Likelihood	-1904.643	-1901.020

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Model 2 uses the proportion of politicians as the instrument for the hub score. The

proportion of politicians is also plausibly exogenous, not significantly correlated with the proportion of judges (Pearson $R=.08$), and strongly negatively correlated to the hub scores of judgments. Moreover, we have no strong theoretical reasons to believe that the proportion of politicians on a panel directly influences implementation. The first stage equation for Model 2 includes the independent variables from Model 3 in table 2 in a linear regression on the natural log of the hub score. Taking the natural log ensures that the endogenous covariate is relatively normally distributed. I then insert the predicted hub scores in the equation estimated by Model 1. Standard errors are obtained through bootstrapping.⁸

The coefficient on the hub score is again positive and significant (at the .01 level) but does not affect the coefficient on judges, suggesting that this is indeed not the only route through which judges affect implementation. The size of the coefficient is larger than before: a one standard deviation increase in the hub score (.5 for the instrument) is now associated with a more than 20% increase in the hazard rate. This suggests that perhaps the coefficient in table 1 is downwardly biased because judges exert more effort embedding difficult issues in case law. Nevertheless, the validity of these results depend not only on the validity of the proportion of politicians as the appropriate instrument but also on some some distributional assumptions, given the non-linearity of the Cox regression.

Panels with a high proportion of professional judges do appear to cloak their judgments better in case law than panels with many former politicians. Depending on our willingness to accept certain assumptions, there is also some evidence that this matters for implementation. Yet, it is unlikely that citations fully capture the legal reasoning mechanism or even that our

⁸I know of no published literature on instrumental variables regression with duration models and know of no statistical package that implements this, thus necessitating the two-stage approach.

hub score measure perfectly captures how well judgments are embedded in case law. Moreover, the citations link does not sharply distinguish judges from especially legal academics. Furthermore, there is no evidence here that legal reasoning matters because national judges are more likely to accept better reasoned judgments. The appropriate conclusion then is that this subsection has provided evidence that the legal reasoning mechanism is plausible. Yet, the evidence is not conclusive.

4.3.2 Experience

A somewhat different interpretation is that it is not so much professional norms but levels of experience that matter. Judges that have served on the court for a long time may be better able to understand how judgments should be written and what judgments are likely to be complied with. Panels with many professional judges are on average slightly more experienced (a one standard deviation increase in the proportion of judges is associated with an average .36 additional years on the court). Thus, it is useful to examine whether the effect found in section 4.1 is an artifact of that correlation.

Throughout the period of analysis, judges were appointed for six year renewable terms (this was recently reformed to nine year nonrenewable terms). This raises an obvious concern: judges with more experience were necessarily reappointed (or introduced for reelection) by their own governments. Thus, a chamber with more experienced judges may be one that is on average more predisposed towards taking government interests into consideration. If we find an effect of panel experience on implementation, we may wonder whether this is because the judgment is better reasoned or because the judges are more lenient in their judgments.

Unlike IACtHR judges, ECtHR judges do not write implementation requirements in their judgments. They simply decide whether or not a specific article of the Convention has been violated. The evaluation of whether implementation was satisfactory is up to the Committee of Ministers rather than the Court. Yet, conceivably judges can signal lenience through their reasoning.

The average experience of panels follows a normal distribution with mean 6.6 years (SD=2.76). Unlike the professional composition of panels, the experience of judges varies temporally. The creation of the new permanent court in 1998 created a massive inflow of new judges: in the new Court the average experience of the judges is 4.3 years less than it was before. Since the new court also lowered the barrier for applications (and thus potentially alters implementation), it is necessary to control for this in regression analyses.⁹

Table 3 presents the findings from a stratified Cox model. The regressions are identical to those in table one except that they now include indicators for experience in years and controls for the new court. There is no evidence that more experienced panels induce more implementation. This finding is robust to using an alternative measure of experience based on the number of lead cases a judge has decided.

Model 3 includes an interaction between the average experience of a panel and the proportion of judges. The findings suggest that only experienced panels with a large number of career judges induce more implementation. So, it may be that judges both need the international and domestic experience in order to write compelling judgments. This is speculative, of course, given that it is based on an assessment of panel level average experience and aver-

⁹I do this by including a dummy for the new court and the number of years since the creation of the new court at the time of the judgments. Coefficients are omitted from the table.

age professional composition. Yet, it is consistent with the theme that professional practice matters and can influence the effectiveness of an international court.

Table 4: Stratified Cox Model Including Average Experience of Panels (hazard ratios and standard errors)

	1	2	3
Proportion of judges	2.00(.55)**	1.89(.53)**	.26(.21)
Experience (years)	1.006(.024)	1.029(.025)	.94 (.04)
Experience*Judges	-	-	1.34 (.15)***
Case characteristics	No	Yes	Yes
N	888	888	888

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

4.3.3 Leniency

An alternative mechanism is that professional judges are more lenient towards respondent governments. Professional judges are not less likely to vote against findings of violations or to exercise greater self-restraint in general (although legal diplomats do, see: Voeten (2007, 2008)). Yet, it may be that they write judgments that are somehow less demanding to implement. ECtHR judges do not write detailed instructions with standards that make explicit when a judgment is implemented satisfactorily. This is left to the Committee of Ministers. Yet, the text of judgments could guide how that body sets its standards.

I examine the plausibility of this mechanism in three ways. First, some judgments only require individual measures such as restitution of property, release from jail, or payment of just satisfaction. Other judgments require general measure by administrative actors, judicial,

or even legislatures. Grewal and Voeten (2012) find that especially judgments that require legislative reform take longer to implement. I ran regressions on these indicators with and without controls for case characteristics. The proportion of judges did not significantly reduce the likelihood that general measures were required or that legislative, executive, or judicial action in particular were needed to implement the judgment.

Second, some judgments have many follow on “clone” cases whereas others have none or only a few. Grewal and Voeten (2012) show that judgments with many clone cases are implemented much more slowly. However, there is no evidence that judgments with many follow-up cases were written by panels with relatively few professional judges. The proportion of former judges is insignificant in tobit regressions and linear regressions, regardless of whether case characteristics were included in the model.

Third and finally, leniency could be revealed by a reduced willingness to demand monetary compensation from states. In the data, there were 574 cases in which the ECtHR demanded that governments pay monetary compensation to victims of human rights abuses. The mean sum was \$95K but the median demand was only \$11K, indicating a skewed distribution. Linear regressions and regressions on a logarithmically transformed variable reveal no evidence that panels with higher proportions of review judges demand less or more just satisfaction. This holds regardless of whether controls for case characteristics and respondent governments are included in the analysis.

In short, there are no observable links between the proportion of career judges on a panel and the tendency of that panel to hand down judgments that would be less costly to implement.

5 Conclusion

ECtHR judgments written by panels with a high proportion of career judges are implemented more quickly than judgments written by panels dominated by other professionals. Since the distribution of career judges on panels is plausibly exogenous to implementation, this finding can be given a causal interpretation.

This provides empirical support for reform efforts that target increased qualifications for international judges. For example, the design of the International Criminal Court was heavily shaped by dissatisfaction with the qualifications of the judges at the temporal tribunals for Yugoslavia and Rwanda. In response, ICC judges are elected from two lists. One list has candidates with “established competence in criminal law and procedure, and the necessary relevant experience, whether as judge, prosecutor, advocate or in other similar capacity, in criminal proceedings.” The second list requires candidates to have “established competence in relevant areas of international law such as international humanitarian law and the law of human rights, and extensive experience in a professional legal capacity which is of relevance to the judicial work of the Court.”

These qualifications emphasize professional legal experience and led to a reduction in the proportion of legal academics in comparison to the temporary tribunals (Danner and Voeten (2010)). What makes this system work is that elections for judgeships are competitive and prestigious (Voeten (2009)). The ECtHR’s system does not provide governments with sufficient incentives to propose highly qualified candidates given that one of their three proposed candidates is assured election. Although the Council of Europe’s Parliamentary Assembly occasionally sends back candidate lists for lack of qualifications, such acts are rare

and politically costly. Thus, any reform must not only consider the qualifications ECtHR judges ought to have but also the incentives for governments to seek the most qualified judges they can muster. Perhaps adding a small number of judges chosen through competitive elections may help both with the enormous backlog in cases as well as with improving the quality of judgments.

The extent to which this finding is generalizable to other settings, including domestic courts, depends on the precise mechanism through which the effect operates. Unfortunately, finding conclusive evidence for causal mechanisms is notoriously difficult and this paper is no exception. I argue that the most plausible interpretation is that the more career judges there are on a panel, the more likely it is that the panel writes a judgment that meets standards of judicial reasoning that matter to domestic legal professionals charged with implementing judgments. I find some evidence for this, in that panels with more judges embed their judgments more thoroughly in case law and case law citations are positively correlated with implementation. I also find some evidence that the effect of having more career judges increases the longer on average the panel has served on the Court. This suggests that the combination of domestic and international experience is what matters. There is no evidence that panels with large proportions of career judges are more lenient towards governments.

This mechanism ought to be transferable to other international courts for whom domestic judges are important compliance constituents. It may also be relevant for domestic high courts who struggle to have their decisions accepted by lower courts. Nevertheless, this interpretation provides less ground for optimism for domestic or international courts where the main issue is persuading political actors to accept their judgment. This distinction may

provide some leverage on the debate as to whether courts with independent or dependent judges are most likely to elicit compliance (see Posner and Yoo (2005); Helfer and Slaughter (2005)). The findings here strongly favor the independent judges side of the argument but the conclusion could play out differently for courts that more purely depend on political actors for compliance.

The question how, if at all, international judges and bureaucrats can use their agency to affect outcomes deserves more attention from scholars not just because it can provide empirical foundations for institutional reform but also because it steers us into new theoretical directions. The literature on linkages between international and domestic institutions is voluminous but it has so far largely neglected the role played by professions. Domestic judges can use linkages with international judges to expand the influence of the juridical sphere domestically. Yet, similar arguments can also be applied to other internationalized professions, such as economists (Fourcade (2006)). This is a potentially useful area for future research.

References

- Abbott, K.W., R.O. Keohane, A. Moravcsik, A.M. Slaughter and D. Snidal. 2000. "The concept of legalization." *International Organization* 54(3):401–420.
- Adler, E., V. Pouliot et al. 2011. *International practices*. Cambridge Univ Press.
- Barnett, M.N. and M. Finnemore. 1999. "The politics, power, and pathologies of international organizations." *International Organization* 53(4):699–732.

- Bourdieu, P. 1986. "Force of Law: Toward a Sociology of the Juridical Field, The." *Hastings LJ* 38:805.
- Box-Steffensmeier, J.M. and B.S. Jones. 2004. *Event history modeling: A guide for social scientists*. Cambridge Univ Pr.
- Box-Steffensmeier, J.M. and C.J.W. Zorn. 2001. "Duration models and proportional hazards in political science." *American Journal of Political Science* pp. 972–988.
- Bruinsma, F.J. 2007. "The Room at the Top: Separate Opinions in the Grand Chambers of the ECHR (1998-2006)." *Recht der Werkelijkheid* 28:7–24.
- Caldeira, G.A. 1983. "On the reputation of state supreme courts." *Political Behavior* 5(1):83–108.
- Cali, B. Koch, A. and N. Bruch. 2011. *The Legitimacy of the European Court of Human Rights: The View from the Ground*.
- Chayes, A. and A.H. Chayes. 1993. "On compliance." *International Organization* 47(2):175–205.
- Cohen, A. and A. Vauchez. 2007. "Introduction: Law, lawyers, and transnational politics in the production of Europe." *Law & Social Inquiry* 32(1):75–82.
- Danner, A. and E. Voeten. 2010. "Who Is Running the International Criminal Justice System." *Who Governs the Globe* pp. 35–71.
- Dezalay, Y. and B.G. Garth. 1996. *Dealing in virtue: International commercial arbitration and the construction of a transnational legal order*. University of Chicago Press.

- Dezalay, Y. and B.G. Garth. 2002. *Global prescriptions: the production, exportation, and importation of a new legal orthodoxy*. Univ of Michigan Pr.
- Diamond, S.S. 1990. "Revising images of public punitiveness: Sentencing by lay and professional English magistrates." *Law and Social Inquiry* 15(2):191–221.
- Finnemore, M. 2004. *The purpose of intervention: changing beliefs about the use of force*. Cornell Univ Pr.
- Fourcade, M. 2006. "The Construction of a Global Profession: The Transnationalization of Economics1." *American Journal of Sociology* 112(1):145–194.
- Franck, T.M. 1990. *The power of legitimacy among nations*. Oxford University Press, USA.
- Grewal, S. and E. Voeten. 2012. "The Politics of Implementing European Court of Human Rights Judgements.".
- Hafner-Burton, E.M., D.G. Victor and Y. Lupu. 2012. "POLITICAL SCIENCE RESEARCH ON INTERNATIONAL LAW." *AMERICAN JOURNAL OF INTERNATIONAL LAW* 106(1):47–97.
- Hagan, J. and R. Levi. 2004. "Social skill, the Milosevic indictment, and the rebirth of international criminal justice." *European Journal of Criminology* 1(4):445.
- Hawkins, D. and W. Jacoby. 2009. Partial Compliance: A Comparison of the European and Inter-American American Courts for Human Rights. In *Workshop on Law, Politics and Human Rights, Emory University*. URL: http://mailer.fsu.edu/whmoore/garnet-whmoore/hrconf/2009_program.html.

- Helfer, L.R. and A.M. Slaughter. 2005. "Why States Create International Tribunals: A Response to Professors Posner and Yoo." *California Law Review* pp. 899–956.
- Hillebrecht, C. 2009*a*. "The Cost of Compliance: Signaling, Credible Commitments and Compliance with International Human Rights Tribunals."
- Hillebrecht, C. 2009*b*. "Rethinking Compliance: The Challenges and Prospects of Measuring Compliance with International Human Rights Tribunals." *Journal of Human Rights Practice* 1(3):362–379.
- Imai, K., L. Keele, D. Tingley and T. Yamamoto. 2011. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review* 105(4).
- Ivkovic, S.K. 2003. "An Inside View: Professional Judges and Lay Judges Support for Mixed Tribunals." *Law and Policy* 25(2):93–122.
- Keller, H and A. Stone Sweet. 2009. "A Europe of rights."
- Koh, H.H. 1998. "How is international human rights law enforced." *Ind. LJ* 74:1397.
- Landes, W.M., L. Lessig and M.E. Solimine. 1998. "Judicial influence: A citation analysis of federal courts of appeals judges." *J. Legal Stud.* 27:271.
- Loucaides, L.G. 2010. "Reflections of a Former European Court of Human Rights Judge on his Experiences as a Judge." *Roma Rights Quarterly* (1):61.
- Lupu, Y. and E. Voeten. 2012. "Precedent on International Courts: A Network Analysis

- of Case Citations by the European Court of Human Rights.” *British Journal of Political Science* .
- Madsen, M. R. 2011. “The Geopolitics of International Law: An Evolutionary Analysis of European International Courts.” *Paper presented at the annual meeting of the The Law and Society Association San Francisco CA* .
- Madsen, M.R. 2007. “From Cold War instrument to supreme European court: The European Court of Human Rights at the crossroads of international and national law and politics.” *Law & Social Inquiry* 32(1):137–159.
- Posner, E.A. and J.C. Yoo. 2005. “Judicial independence in international tribunals.” *Cal. L. Rev.* 93:1.
- Report on Judicial Independence.* 2003.
- Shaffer, G. and T. Ginsburg. 2012. “The Empirical Turn in International Legal Scholarship.” *AMERICAN JOURNAL OF INTERNATIONAL LAW* 106(1):1–46.
- Shapiro, M. 1986. *Courts: a comparative and political analysis*. University of Chicago Press.
- Simmons, B.A. 1998. “Compliance with international agreements.” *Annual Review of Political Science* 1(1):75–93.
- Simmons, B.A. 2009. “Mobilizing for human rights: international law in domestic politics.”.
- Slack, J. 2011. “Named and shamed: The European human rights judges wrecking British law.” *Daily Mail* .
- Slaughter, A.M. 1999. “Judicial globalization.” *Va. J. Int’l L.* 40:1103.

- Stone Sweet, A. 2009. "On the Constitutionalisation of the Convention: The European Court of Human Rights as a Constitutional Court." *Revue trimestrielle des droits de l'homme* 80.
- Voeten, E. 2007. "The Politics of International Judicial Appointments: Evidence from the European Court of Human Rights." *International Organization* 61(4):669.
- Voeten, E. 2008. "The Impartiality of International Judges: Evidence from the European Court of Human Rights." *American Political Science Review* 102(4):417–432.
- Voeten, E. 2009. "Politics of International Judicial Appointments, The." *Chi. J. Int'l L.* 9:387.
- Von Staden, A. 2009. Shaping Human Rights Policy in Liberal Democracies: Assessing and Explaining Compliance with the Judgements of the European Court of Human Rights PhD thesis Princeton University.