

---

# Domestic Implementation of European Court of Human Rights Judgments: Legal Infrastructure and Government Effectiveness Matter: A Reply to Dia Anagnostou and Alina Mungiu-Pippidi

Erik Voeten\*

## Abstract

*This article responds to the valuable contribution by Dia Anagnostou and Alina Mungiu-Pippidi in which they analyse how nine countries implemented European Court of Human Rights judgments that found violations of Articles 8–11 of the European Convention on Human Rights. Their conclusion that capacity plays an important role in the implementation of ECtHR judgments is certainly correct. In this short response, I highlight various aspects of the authors' analysis where they make problematic choices with regard to data and statistical methods. First, I describe and use a more comprehensive dataset that allows us to reach more generalizable conclusions. Secondly, I show how survival analysis is a more appropriate framework than logit or linear regression for analysing these data. Thirdly, I argue that the difficulty of the implementation task needs to be accounted for in any analysis of cross-country variation in implementation. My re-analysis shows that low capacity countries attract judgments that are more difficult to implement. The analysis also uncovers a subtle relationship between time, institutional capacity, and checks and balances. High capacity helps willing politicians to implement judgments quickly. Yet, among judgments that have been pending longer, countries with higher capacity are no quicker to implement than lower capacity countries. By contrast, checks and balances initially slow down implementation but help to eventually ensure begrudging implementation.*

\* Erik Voeten is the Peter F. Krogh Associate Professor of Geopolitics and Justice in World Affairs at Georgetown University's Edmund A. Walsh School of Foreign Service and Department of Government. Email: [ev42@georgetown.edu](mailto:ev42@georgetown.edu).

## 1 Introduction

By 2012 international courts (ICs) had issued more than 30,000 legally binding judgments.<sup>1</sup> Over 90 per cent of these have come since 1990. What happens to these judgments? Are they implemented swiftly and completely or are they implemented belatedly, half-heartedly, or perhaps even ignored altogether? What explains why some judgments suffer the former fate while others meet with the latter? These questions are of obvious significance to lawyers and social scientists. Moreover the sheer number of judgments creates opportunities for quantitative analyses of implementation. Qualitative case studies enlighten us about the processes that determine the fate of individual or small groups of judgments but they are poor tools for making general inferences from large numbers of judgments.

Dia Anagnostou and Alina Mungiu-Pippidi offer a valuable contribution with their analysis of how nine countries implemented European Court of Human Rights (ECtHR) judgments that found violations of Articles 8–11 of the European Convention on Human Rights (ECHR).<sup>2</sup> They are correct to point out that the implementation of international court judgments depends not just on political willingness but also on the capacity of countries to do what international courts want them to do. Moreover, I am pleased that *EJIL* is embracing the idea that lawyers and legal scholars can learn from quantitative investigations of this issue.

In this short response, I highlight various aspects of the authors' analysis where they make problematic choices with regard to data and statistical methods. My goal here is not to undermine Anagnostou and Mungiu-Pippidi's main substantive conclusion. Indeed, I present analyses based on more comprehensive data and statistical methods, which corroborate their assertion that countries with greater legal and bureaucratic capacity on average implement judgments more quickly than countries with less capacity. Instead, I wish to help establish best practices for data collection and analysis and I will show how this helps us reach more nuanced substantive conclusions.

First, if we wish to ascertain what properties of countries (or governments) make them more or less likely to implement swiftly, then we need to take into account variation in the difficulty of the implementation tasks. I will show that low capacity countries attract judgments that require much more extensive reform than high capacity countries. I will show what aspects of judgments are correlated with quicker and slower implementation, which should be of interest to a legal audience.

Secondly, I will reveal a subtle relationship between time, institutional capacity, and constraints on the executive. High capacity helps willing politicians to implement judgments quickly. Yet, among judgments that have been pending for longer

<sup>1</sup> Alter, 'The Multiple Roles of International Courts and Tribunals: Enforcement, Dispute Resolution, Constitutional and Administrative Review', in J.L. Dunoff and M.A. Pollack (eds), *Interdisciplinary Perspectives on International Law and International Relations: The State of the Art* (2012).

<sup>2</sup> Anagnostou and Mungiu-Pippidi, 'Domestic Implementation of European Court of Human Rights Judgments: Legal Infrastructure and Government Effectiveness Matter'. 25 *EJIL* (2014) 205.

(presumably due to unwilling politicians) those against countries with higher capacity are implemented no more quickly. By contrast, checks and balances on executive authority initially slow down implementation. Yet, among the set of judgments that have been pending for three years or more, those that involve countries with adequate checks and balances are most likely to be implemented. Lack of capacity may be why Romania and Turkey implement few judgments rapidly, but it is not the reason why some of their judgments have been pending for over a decade. By contrast, countries with high levels of constraints on the executive do implement judgments eventually.

## 2 Data and Methodological Issues

### A Data Issues

There are two limitations in the way the authors select cases. First, they restrict their analysis to judgments on Articles 8–11 in nine countries. The geographical constraint is especially problematic. The authors are primarily interested not in explaining variation across judgments but in explaining variation across countries. Capacity is relatively stable within countries. Thus, we have only nine sources of variation. Moreover, the authors also wish to control for potentially confounding variables that are also measured only at the country level. This leaves us with few degrees of freedom. The promise of quantitative analysis is that it allows the analyst to look at large samples. Given that data collection is relatively straightforward, I do not see a pertinent reason for this limitation. It is always possible to highlight a few countries for deeper qualitative study even if the quantitative study is based on the full spectrum of cases from all CoE member states.

Secondly, the authors take all cases as independent observations. In reality, however, a large number of cases are so-called ‘follow-on cases’, which raise the same legal issue as the ‘lead case’ and are usually (although not always) implemented at the same time. For example, a general measure such as legislation may implement dozens of judgments in one go. Including these judgments as individual observations amounts to double-counting.

Sharanbir Grewal and I gathered data on the implementation of all ECtHR judgments for all Articles and all countries.<sup>3</sup> Our primary restrictions were that we collected data only on lead cases, we eliminated friendly settlements (which raised few implementation challenges), and we restricted our attention to judgments adopted by 31 December 2006. This left us with 1,056 cases, of which 846 (80.1 per cent) had been finally resolved by 22 September 2012. We also collected information on characteristics of the judgments, such as which articles they concerned, how many follow-up cases there were, whether legislative or other general measures were required for implementation, and whether a judgment was delivered by the Grand Chamber.

<sup>3</sup> S. Grewal and E. Voeten, ‘Are New Democracies Better Human Rights Compliers?’, *Int’l Org* (forthcoming) available at SSRN: <http://ssrn.com/abstract=2187428> or <http://dx.doi.org/10.2139/ssrn>.

This dataset provides a more solid basis for generalizable results than the data collected by Anagnostou and Mungiu-Pippidi.

## B *Methodological Issues*

The dependent variable in Anagnostou and Mungiu-Pippidi's analysis is time to implementation (in months), with the complication that some judgments have not (yet) been implemented. The authors run two separate analyses. First, they run a (logit) regression on a binary indicator of whether the judgment was implemented. Secondly, they look only at the set of implemented judgments and run a regression on the number of months it took for implementation to take place.

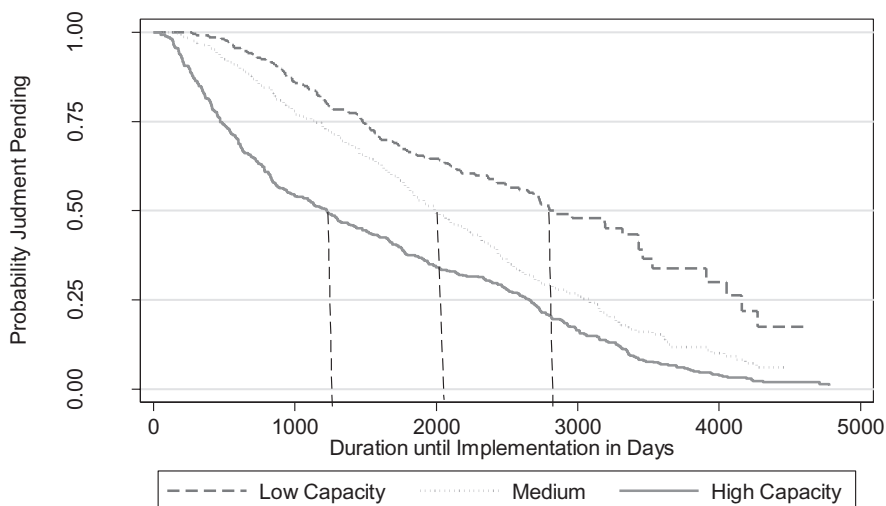
This departs from conventional approaches in the social, medical, and biological sciences, which analyse this type of data using statistical models called 'event-history', 'duration', 'survival', or 'hazard' models. The term survival model comes from the example where the event of interest is death. Alas, death becomes more likely as time passes. Yet, individuals engage in behaviours that may increase (e.g., smoking) or decrease (e.g., moderate consumption of dark chocolate) the likelihood of this event. Survival models estimate whether covariates indeed reduce/increase the risk of an event happening, given that the likelihood of the event also changes with age/time.

Implementation of survival analysis is straightforward in the court context. The event of interest is implementation rather than death. Implementation becomes more likely as time passes. A judgment adopted in 1981 surely has a greater likelihood of being resolved in 2008 than one adopted in 2006. Yet, the logit model presented by Anagnostou and Mungiu-Pippidi (Table 3, at 219) does not include time. Moreover, the regression on duration is performed on a sample that excludes the judgments that have been pending the longest (namely those that have not yet been implemented) as we do not yet know the duration of the implementation process, thus leading to biased coefficients. By contrast, survival models estimate a single model on the full sample, which estimates the probability of implementation given how long a judgment has already been pending.

Figure 1, below, offers the simplest possible estimates of the probability of implementation as a function of time and whether a country has high, medium or low levels of bureaucratic and legal capacity. The measure is a combination of the Inter-Country Risk Guide (ICRG)'s bureaucratic and law and order variables. The two variables are so strongly correlated that it would make little sense to separate them (Anagnostou and Mungiu-Pippidi use both government effectiveness and law and order).<sup>4</sup> Countries are grouped in the high category if they score in the top third of the combination of these variables and in the low category if they rank in the bottom third (among CoE member states).<sup>5</sup>

<sup>4</sup> Based on Grewal-Voeten data described in previous subsection.

<sup>5</sup> For more on the data see Howell, 'ICRG Methodology' (2011), available at: [www.prsgroup.com/ICRG\\_Methodology.aspx](http://www.prsgroup.com/ICRG_Methodology.aspx) (accessed 30 Nov. 2012). Both of these variables are used to construct the World Bank scale preferred by Anagnostou and Alina Mungiu-Pippidi, but the ICRG data range further back in time. I could not infer from the article how the authors assigned capacity scores for judgments before 1996, when the World Bank data began.



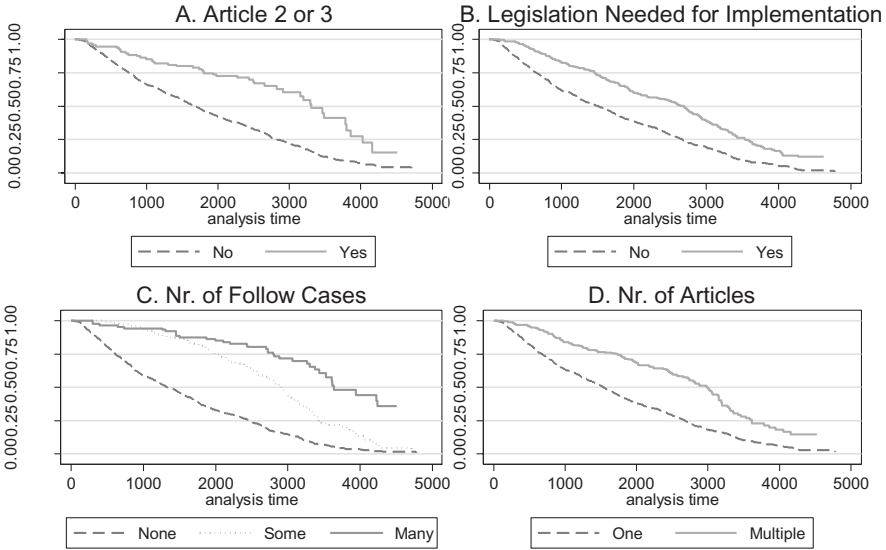
**Figure 1.** Kaplan-Meier Survival Estimates for ECtHR Judgments by Capacity

The probability that a judgment remains pending drops sharply the more time passes following the issue of a judgment. Yet, the figure also offers evidence for Anagnostou and Mungiu-Pippidi's hypothesis. After about 1,200 days, a judgment issued against a high capacity country has a 50 per cent chance of being implemented. Countries of medium capacity need about 2,000 days to implement 50 per cent of judgments, and countries with low capacity about 2,800 days to reach this landmark (see the reference lines in [Figure 1](#)). These are meaningful delays in the exercise of justice.

### 3 Analysis

#### A Case Characteristics

An important limitation of [Figure 1](#) is that countries with inadequate bureaucratic and legal capacity may attract judgments that are more difficult to implement. For example in this database, 15 per cent of violations against low capacity countries concern politically sensitive violations of Article 2 and/or 3, which prohibits the state from engaging in 'deprivation of life' and 'inhuman or degrading treatment or punishment'. This compares with 5 per cent of judgments against high capacity countries. Judgments may also demand more general measures, such as legislation, rather than just individual measures, such as payments of just satisfaction. Thirty-six per cent of cases against low capacity countries require legislative reform against 27 per cent for high capacity countries. Judgments may also be more complex. For instance, judgments against low capacity countries on average attract 13 follow-on cases against only six for high capacity countries. Finally, the average judgment against a low capacity country concerns more articles of the Convention than those against a high capacity country.



**Figure 2.** Kaplan-Meier Survival Estimates for ECtHR Judgments by Judgment Characteristics

Figure 2 is identical to Figure 1 except that it plots how the probability that a judgment remains pending correlates with the judgment characteristics highlighted above. Clearly, all of the above-mentioned factors are associated with slower implementation. For example, the line for a judgment that involves Article 2 or 3 always remains above judgments that do not involve either article, indicating that a smaller percentage of Article 2 or 3 judgments are implemented at any time. Only after 3,500 days does an Article 2 or 3 case have 50 per cent probability of being implemented, almost twice as long as judgments that concern other violations. If legislation is required, then it takes about 1,000 days longer to reach the 50 per cent threshold than if no legislation is needed for implementation. Judgments that do not attract follow-on cases and that deal with only one article of the Convention are on average implemented much more quickly than judgments with many follow-up cases or that involve several Convention articles.

Anagnostou and Mungiu-Pippidi collected very little information on judgment characteristics. This is problematic because, as we have seen, low capacity countries attract more judgments of the type that are implemented more slowly by all countries. Thus, it is possible that the findings from Figure 1 are driven not so much (or at least not just) by differences in the capacities of countries but by differences in the difficulties of the implementation task.

The most straightforward way to address this issue is to run a regression analysis that controls for judgment characteristics.<sup>6</sup> I use a Cox proportional hazard model for

<sup>6</sup> Grewal and Voeten, *supra* note 3, use a more sophisticated methodology based on matching, which goes beyond the scope of this note.

this purpose (a survival model). This model adopts a very flexible approach to how the likelihood of implementation varies with time, but also allows us to test the influence of many other variables. I estimated a shared frailty Cox model to deal with the fact that most countries have multiple judgments in the data set and that these are not independent observations.

Table 1 shows the results of this analysis. If a coefficient is significantly larger than zero, then a higher value on the variable on average is associated with faster implementation. If a coefficient is smaller than zero, then an increase in the variable is associated with slower implementation. Model 1 estimates the model with only legal capacity. Model 2 includes a battery of judgment characteristics: varying from the Articles of the Convention to which the judgment applies to whether the judgment invited a separate opinion, whether it came from the Grand Chamber, how many Articles of the Convention were violated, how many follow-up cases there were, and whether the judgment required general measures (legislative, executive, or judicial) or just individual measures (such as paying just satisfaction).<sup>7</sup>

Even after controlling for judgment characteristics, the coefficient for the capacity variable is significantly larger than 0.<sup>8</sup> Yet, the coefficient is considerably lower in model 2 than in model 1, which does not control for judgment characteristics. This suggests that judgment characteristics are confounding: meaning that lower capacity countries attract more difficult to implement judgments and that some of the simple association between capacity and implementation is a consequence of this. There are ample judgment characteristics that are associated with faster or slower implementation times. Aside from the ones discussed in Figure 2, judgments that just require individual measures are implemented much faster than judgments that require legislative or judicial action. Protocol 1-1 cases (property rights) are implemented slowly on average.

In short, it would be a mistake to ignore judgment characteristics. Even if capacity still mattered, we would overestimate its importance without controlling for judgment characteristics. Moreover, understanding what aspects of a judgment are correlated with slower or faster implementation should be of interest to lawyers.

## B Capacity, Constraints, and Time

The analysis in the previous subsection confirms for a larger sample of judgments and a larger group of countries that greater bureaucratic and legal capacity is indeed on average positively correlated with quicker implementation of judgments. This finding is not surprising. Most political scientists view managerial explanations, which highlight capacity, and explanations that emphasize the political incentives for implementation as complementary rather than competitive.<sup>9</sup> Future research should go beyond

<sup>7</sup> More details are in *ibid.*

<sup>8</sup> This is not the case in a stratified Cox model, which estimates a separate base-line hazard for each country. This may be too heavy medicine for this type of data but all other results from Table 1 do hold in a stratified model.

<sup>9</sup> See, e.g., Tallberg, 'Paths to Compliance: Enforcement, Management, and the European Union', 56 *Int'l Org* (2002) 609; Beach, 'Why Governments Comply: an Integrative Compliance Model that Bridges the Gap between Instrumental and Normative Models of Compliance', 12 *J European Public Policy* (2005) 113.

**Table 1.** Shared Frailty Cox Proportional Hazard Models on Implementation of EctHR Judgments

<b>Models</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>Main</b>	<b>Interaction with Time</b>
<i>Country Characteristics</i>					
Political constraints				-8.914*** (2.753)	1.297*** (0.385)
Capacity	1.687*** (0.356)	1.080*** (0.306)		13.32*** (2.204)	-1.755*** (0.306)
<i>Judgment Characteristics</i>					
Separate opinion		0.217** (0.0846)		0.170** (0.0851)	
Importance		0.129* (0.0741)		0.152** (0.0751)	
Grand Chamber		0.0882 (0.125)		0.0799 (0.126)	
Individual measure		0.247** (0.113)		6.211*** (0.729)	-0.872*** (0.107)
Judicial measure		-0.227** (0.114)		-0.287** (0.116)	
Legislation		-0.262*** (0.0927)		-0.344*** (0.0946)	
Practical measure		-0.282* (0.162)		-0.430*** (0.165)	
Executive/administrative measure		-0.0102 (0.0940)		-0.0126 (0.0949)	
Number of articles		-0.220 (0.135)		-2.581*** (0.842)	0.321*** (0.112)
Number of other cases (ln)		-0.511*** (0.0535)		-3.320*** (0.660)	0.373*** (0.0853)
Article 2 or 3		-0.474** (0.198)		-0.444** (0.199)	
Article 5		0.102 (0.149)		3.078*** (0.921)	-0.423*** (0.130)
Article 6		0.193 (0.122)		2.530*** (0.732)	-0.336*** (0.102)
Article 8		-0.261* (0.143)		-0.282* (0.144)	
Article 10		0.231 (0.164)		3.233*** (1.122)	-0.422*** (0.160)
Article 13		-0.256 (0.210)		-0.346 (0.214)	
PI_1		-0.450*** (0.168)		-0.489*** (0.170)	
Observations	1,001	1,001		1,001	1,001
Number of countries	38	38		38	38

Entries are coefficients with S.E. in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$



the question whether capacity matters and investigate how capacity relates to more political explanations for compliance.

I advance a simple attempt here. A commonly highlighted political explanation centers on checks and balances, which provide constraints that make it more difficult for executives to ignore international legal obligations. Constraints are a way eventually to force what Courtney Hillebrecht labelled 'begrudging compliance'.<sup>10</sup> This is quite the opposite of capacity, which helps willing executives implement their obligations quickly.

I therefore pose two simple hypotheses: high capacity should help executives implement judgments quickly, but among the set of judgments that remain pending longer, those in high capacity countries are no more likely to be resolved quickly than those in low capacity countries. By contrast, checks and balances make quick policy change more difficult, but they force executives to comply eventually. Thus, among the set of judgments that remain pending longer, those in countries with effective checks and balances should be most likely to be resolved.

I use Witold Henisz' measure of political constraints POLCONIII.<sup>11</sup> This measure combines information on the number of independent branches of government with veto power and the distribution of preferences within those branches. Thus, countries with many institutions that can exercise checks and balances and where those institutions are controlled by actors from different political parties receive higher scores. The correlation between capacity and constraints is 0.39 in this dataset.

To test these hypotheses, I estimate an interaction effect between the constraints and the capacity variables and time. Interactions with time are commonly included in Cox regressions to deal with violations of the proportional hazard assumption, but they also have substantively interesting interpretations.<sup>12</sup> I also include interactions with time for judgment characteristics based on tests with Schoenfeld residuals.

Model 3 in Table 1 shows the results, which are consistent with the hypotheses: capacity has a positive effect on implementation, but this effect decreases with time (coefficient of the interaction with time is negative). The opposite is true for constraints. Both effects are significant but the interactive coefficients are difficult to interpret. Figure 3 therefore plots how the coefficient changes with time. Coefficients smaller than zero indicate a negative effect on implementation.

Countries with high levels of political constraints are initially no quicker to implement judgments. Indeed, the more veto points there are the more difficult it is to achieve political change, thus slowing down implementation. Yet, as judgments remain pending longer, the coefficient becomes positive; meaning that among the set of judgments that remain pending for about 1,200 days or more, those that concern countries with high levels of political constraints are more likely to be implemented quickly.

<sup>10</sup> Hillebrecht, 'Implementing International Human Rights Law at Home: Domestic Politics and the European Court of Human Rights', 13 *Hmn Rts Rev* 279, at 284–285.

<sup>11</sup> Henisz, 'The Institutional Environment for Economic Growth', 12 *Economics & Politics* (2000) 1.

<sup>12</sup> Box-Steffensmeier and Zorn, 'Duration Models and Proportional Hazards in Political Science', 45(4) *Am J Political Science* (2001) 72.

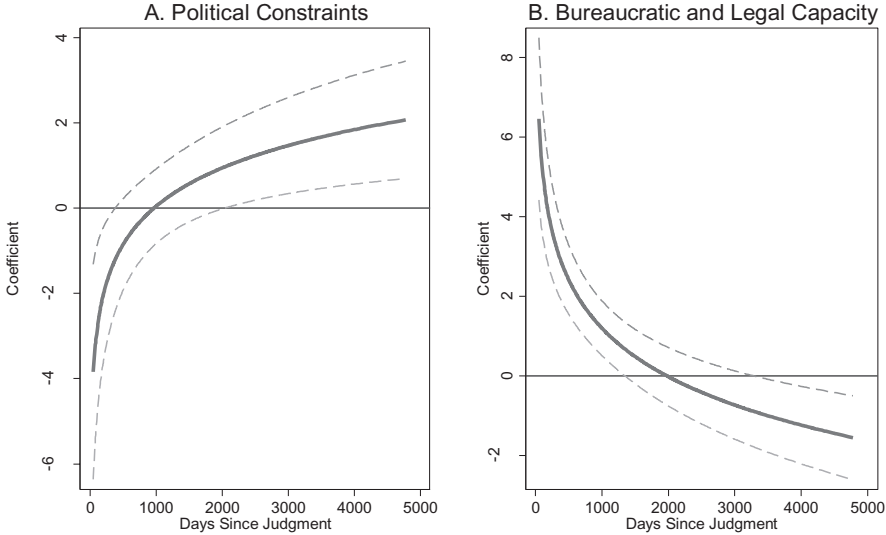


Figure 3. Effects of Constraints and Capacity by Time since Judgment

The effect on capacity is the opposite: high levels of bureaucratic and legal capacity help willing executive to implement judgments quickly. However, among the subset of judgments that remain pending after 1,000 days, those that concern countries with high legal and bureaucratic capacity are no more likely to be implemented swiftly than those from low capacity countries.

### 4 Conclusion

The large number of judgments issued by international courts opens up the possibility of learning about their effects through quantitative analysis. Anagnostou and Mungiu-Pippidi have provided a very useful first step in this regard. But further progress hinges on more interdisciplinary collaboration between legal scholars and social scientists. Such endeavours have so far been more common in the theoretical than the empirical realm. I have only scratched the surface in this note on what such collaborations might offer. They could yield valuable theoretical insights that increase our understanding of the actual effects of international court judgments. Yet, understanding why some judgments are implemented more quickly than others may also yield practical insights that aid the ECtHR in improving its effectiveness.