Explaining Commitment: States and the Convention against Torture

Jay Goodliffe  Brigham Young University
Darren G. Hawkins  Brigham Young University

Why do states commit to international human rights treaties that may limit state sovereignty? Existing arguments focus on either the fear of domestic democratic instability or on international norms. We focus instead on the variation in three kinds of costs that states must pay to commit: policy change, unintended consequences, and limited flexibility. We use a discrete time-duration model to test all of these explanations on state commitment to the international Convention Against Torture, one of the most important international human rights treaties. We find strong evidence for the importance of norms and all three types of costs, but no evidence supporting state desires to lock in the benefits of democracy in the face of domestic democratic instability.

In the months following 11 September 2001, key Bush administration officials spent substantial time and effort assessing their legal commitments regarding torture and war crimes.1 The Department of Justice, the Attorney General, the White House Counsel’s office, the Department of Defense, and President Bush himself all wrote and signed memos analyzing and interpreting U.S. commitments. International treaty commitments constituted one of their chief concerns. Many of the memos went to considerable lengths to minimize U.S. commitments to international treaties. The well-known memo produced by Justice Department’s legal counsel office in August 2002 took particular aim at the international Convention Against Torture and Other Cruel, Inhuman, and Degrading Treatment or Punishment (better known as the Convention Against Torture, or CAT) and the U.S. legislation that implemented that treaty in domestic law. The memo famously argued that the definition of torture set by CAT and the implementing legislation was so high that only “serious physical injury, such as organ failure, impairment of bodily function, or even death” would qualify as torture. Further, the memo argued that even if U.S. officials undertook such measures, they would still be protected from prosecution by the president’s commander-in-chief powers and by legal doctrines of self-defense and necessity.

Although this memo and others argued that the costs of CAT were low, their production appears to be driven by a fear that those costs could be high. Indeed, at the same time that some officials were minimizing the importance of CAT and other international legal commitments, other U.S. officials were arguing that those treaties and accompanying domestic legislation placed substantial constraints on U.S. interrogation procedures. Administration officials as high as Defense Secretary Donald Rumsfeld involved themselves in the details of what kinds of interrogations were allowed. One of Rumsfeld’s key memos required that he give direct approval for interrogation techniques not on the list he provided. The documentary history suggests that those interrogation techniques were shaped by careful study of CAT and related rules. Government officials have been so worried about the costs of CAT that Vice President Dick Cheney spent substantial political capital and risked his personal reputation fighting against references to CAT in con-


© 2006 Southern Political Science Association  ISSN 0022-3816
gressional legislation, which prompted the Washington Post to label him as the “Vice President for Torture.”

It is important not to overstate the constraints imposed by international treaties like CAT. States remain sovereign entities and most international treaties lack enforcement mechanisms. Yet, a variety of scholars have shown multiple ways in which even relatively “toothless” treaties can influence state behavior (Checkel 2001; Cortell and Davis 1996; Hawkins 1997). Some have argued that commitments create political dynamics of their own that increase pressure for compliance (Keck and Sikkink 1998; Risse and Sikkink 1999). Other times, the same factors that lead states to commit to treaties can also lead to compliance (von Stein 2005). Abbott and Snidal (2000), Downs, Rocke, and Barsoom (1996), and others have argued that commitment costs can be high enough to deter many states from making strong international legal commitments. In the case of CAT, 55 states have been deterred from ratification since the treaty was adopted in 1984, and many of the 139 ratifying states did so only after delays of many years. This paper takes up these issues and asks the following questions. What kinds of states are more likely to commit and why? Why do some states commit sooner than others?

Previous scholars attempting to explain commitments to human rights treaties have focused either on the benefits of those treaties or on the concept of norms. In an innovative article, Moravcsik (2000) argued that states are likely to commit to international human rights treaties when they are unstable democracies and need to “lock in” the benefits of human rights norms. Finnemore and Sikkink (1998) argued that states are likely to commit to human rights standards when other states have already committed and such commitment seems to be generally expected. While theoretically appealing, these explanations have not been subjected to much empirical testing. One important contribution of this paper is to operationalize measures for these arguments and test them on state commitments to CAT.

The other difficulty is that scholars have not systematically identified or tested the different costs that accrue to states from their international treaty commitments. We attempt to remedy that problem in this paper. We develop three types of costs—policy change, unintended consequences, and limited flexibility—and argue that states weigh the benefits of lock-in and normative conformity against the costs of those commitments. Generally speaking, the lower the costs, the more likely states are to commit. We also operationalize these measures and test them on state commitments to CAT.

We find that normative conformity and costs both matter in whether and when states sign and ratify CAT. However, our results show no support for the idea that states sign and ratify to lock in human rights benefits. This paper proceeds by first discussing CAT in context with other international treaties. We then lay out the theoretical arguments and measures, explain our methodology and findings, and conclude.

The Torture Treaty and Commitment

The Convention Against Torture was adopted in December 1984 and went into effect in June 1987. It can be classified in a growing group of treaties with medium to high levels of obligation and delegation, two of the components of “legalization” (Abbott et al. 2000; Goldstein et al. 2000). Obligation refers to the extent to which an international rule is legally binding while delegation refers to the extent to which states have granted authority to an international institution to take action. Many international rules have low levels of obligation and delegation, simply laying out general principles to which states proclaim their adherence, with little mention of clear legal obligation or third-party delegation.

A growing number of international rules, however, have become increasingly legalized through increasing levels of obligation, delegation, and precision. Higher levels of legalization imply somewhat higher costs for states who commit to them, though these are presumably offset by increasing benefits of legalization (Abbott and Snidal 2000). The seven core human rights treaties, a group that includes CAT, all have moderate levels of obligation, delegation, and precision. As treaties, they are more legally binding than many international rules. In most of them, states

---


5The treaty texts, as well as their classification as core treaties, may be found at the Office of the United Nations High Commissioner for Human Rights, http://www.ohchr.org/english/law/, accessed 15 April 2005.
have delegated some authority to international committees who monitor state treaty compliance through state self-reports and NGO commentary.

CAT actually goes a little farther than the other six treaties in one important way. It endorses an enforcement mechanism known as universal jurisdiction, defined as the principle that a state’s jurisdiction is based on the nature of the crime rather than factors such as where the crime occurred or the nationality of the alleged perpetrator or victim (Boulesbaa 1999, 204–205; Ratner and Abrams 2001, 160–62; Rodley 1999, 48–50, 129–30). This principle allows Nigeria, for example, to prosecute a crime committed in Germany by an American against an Indonesian.

Historically, states applied the principle to universally repudiated crimes that occurred outside any state’s formal territorial jurisdiction, such as piracy (Boulesbaa 1999, 204–205). More recently, states have endorsed universal jurisdiction for the crimes of hijacking, hostage taking, and crimes against diplomats (Lippman 1994, 316). It is important to note that international treaties grant extraterritorial jurisdiction only to states ratifying the treaties. As a result, jurisdiction over such crimes is not truly “universal” in the sense of applying to all states. While the term “universal” may not be precisely accurate, it is conventionally and widely used to refer to broad extraterritorial jurisdictional grants based on the nature of the crime (Ratner and Abrams 2001), and we apply it in this way.

Although CAT did not invent universal jurisdiction, it applied the principle to human rights abuses for the first time (Boulesbaa 1999, 204–205; Randall 1988, 819; Rodley 1999, 48–50, 120–32). Article 5 of the convention allows states to establish jurisdiction over crimes involving torture not only when the crime occurs within its territory, but also when the alleged offender is a national of that state, when the victim is a national of that state, or when the alleged offender is present in its territory and the state decides not to extradite the accused. Article 7 requires states to adopt such jurisdiction rather than simply permitting them to do so. Previously, universal jurisdiction covered crimes that had some transnational character or crimes that were not sponsored by the state, or both (Randall 1988). The Geneva Conventions, for example, adopted universal jurisdiction for war crimes—but these crimes were generally committed by one state’s armed forces against foreign nationals and thus had a clear transnational character. In approving the torture convention, states explicitly opened their own officials to prosecution by other states. In essence, prosecuting authority is handed to third-party states for crimes that are both state-sanctioned and targeted against a state’s own citizens.

Skeptics might suggest that none of this really matters since states are unlikely to utilize universal jurisdiction. Yet one recent study found that 109 states had incorporated universal jurisdiction into their domestic legislation (Hawkins 2003). Of those, 14 have actually tried court cases based on the principle, and courts have upheld the law in 12 of them. Virtually all of this activity has occurred since the early 1990s. The principle is not widely practiced, but it is widely recognized in domestic law and occasionally practiced. For example, former Chilean president and dictator Augusto Pinochet’s decision to sign and ratify the torture convention eventually led to his political demise. Ten years after he signed the treaty, he was himself held to its provisions when British and Spanish courts exercised universal jurisdiction for the crime of torture (Davis 2003). British and Spanish efforts against Pinochet helped trigger a flood of court cases in Argentina and Chile prosecuting individuals accused of human rights abuses and have prompted other Latin American governments to move toward such prosecutions and to otherwise penalize responsible individuals (Lutz and Sikkink 2001).

As various scholars have pointed out, legalization creates costs for states. CAT creates slightly higher potential costs than most other human rights treaties due to universal jurisdiction, yet it belongs generally among treaties with moderate levels of legalization. This makes CAT a nice test case for theories of state commitment because it allows for variation in costs and benefits. In fact, 139 states have ratified CAT and those ratifications have been spread widely across the past 20 years, offering a wide range of variation in the outcome to be explained.

States can commit to CAT at two different levels. First, they can sign the convention, indicating an official state intention to ratify the treaty and informally binding the state to abide by the treaty’s provisions. Second, they can ratify or accede to the treaty, meaning that they are fully and legally bound to abide by the treaty. Ratification and accession are functionally similar; ratification simply means that states have previously signed the treaty whereas accession means that states had not previously signed it.

States and individuals can make formal promises (commitments) to do something without following through (compliance) and can also comply with rules without first committing. These forms of behavior are analytically distinct and empirically important. A variety of scholars routinely make this distinction between commitment and compliance, often arguing
that commitments help alter future behavior in com-
pliance-oriented ways (Abbott and Snidal 2000; Moravcsik 2000; Pierson 1996; Pollack 2003; Simmons 2000). We wish to explain the timing and level at
which states commit to the treaty.

Why Commit?

In this section, we review three sets of causal factors
that can lead states to commit to international human
rights treaties: international norms, the domestic ben-
efits of locking in to favorable policies, and the sove-
erignty costs of international commitment. It is quite
possible and even likely that different states commit
during different reasons. As noted, CAT does not have
equal benefit or cost to all countries. We hope to iden-
tify this variation with the variables introduced below.
If there is merely random variation across countries,
we would be surprised to pick up any systematic
effects. We also discuss how we operationalize these
factors.6

Norms

In recent years a variety of scholars have argued
that international norms can drive state behavior.
Finnemore and Sikkink (1998) argue that as new
norms emerge and are adopted by a few leader states,
a “cascade” effect results. As more states commit to a
norm, other states feel pressured to commit as well.

There is some debate around the question of
which causal logic connects norms to behavior. Many
norms scholars have focused on the “logic of appro-
priateness,” in which states adopt international norms
because they become convinced that such behavior is
appropriate for any actor claiming statehood
(Finnemore 1996). In this view, norms become con-
stitutive of statehood such that states routinely adopt
them as part and parcel of their identities as states.
Sociologists adopting a world society approach argue
in a similar vein that institutional forms spread
throughout states because they are legitimized inter-
nationally as part of what it means to be a state (Meyer
et al. 1997). Other scholars have adopted a “logic of con-
sequences” to argue that states adopt norms
because of the rewards associated with that behavior
and the penalties imposed for norm violation (Schim-

melfennig 2001). According to this logic, states
commit to norms in order to establish their credibil-
ity on a given issue (Simmons 2000). With credibility
established, other states and third-party actors (cor-
porations, NGOs) reward that state through invest-
ment, trade, aid, and positive political relationships.

Although these logics are quite different, they can
coexist and reinforce each other in practice. States
may adopt norms both because it is appropriate and
because they reap rewards from the practice. Because
these two logics can reinforce each other, it is often dif-
ficult to tell which causal mechanism is at work or
which is more important. Some scholars, such as Risse
and Sikkink (1999), combine the two logics, arguing
that they coexist but that they are generally sequenced
in terms of importance. At early stages, when states are
strongly norm resistant, the logic of consequences is
more important. As states enter into dialogue with
each other and become willing to question their own
assumptions, then a logic of appropriateness may
kick in.

Operationalizing norms is a tricky business. As
abstract entities that operate at the level of social psy-
chology or intersubjective understandings, norms are
difficult to identify and their influence is hard to track.
Simmons (2000) pioneered an innovative approach
that we adopt in this paper. She identifies two groups
of states among which norms are likely to exist: all
states in a global community and states sharing the
same geographical region. Following Simmons, we
measure the strength of the norm by examining the
extent to which other states in these two groups have
committed to CAT. Following Simmons (2002), we
utilize a 0–2 scale, where 0 is no action, 1 is signature,
and 2 is ratification or accession. We create one annual
measure for the average commitment score for all
states in the world, Global Score, and another annual
measure for the average commitment score of states
within a particular geographical region, Regional
Score, always leaving out the commitment level of the
state whose commitment we wish to explain. To the
extent that states desire good reputations among peers
or to the extent that states identify themselves as part
of a region rather than part of the global system, they
should be particularly concerned about states in their
own region. We expect the higher the global and
regional commitment level to CAT, the more likely a
state is to commit to it as well.7

6 More details on our operationalizations and sources of our meas-
ures are available in our appendix found at the journal website
(http://www.journalofpolitics.org) or at http://goodlife.byu.edu/
papers.

7 We also estimated models that follow with a (time-constant)
dummy variable for each region instead of a (time-varying)
regional score. The results are qualitatively similar.
Lock-In

In an innovative and influential article, Moravcsik (2000) proposed that new and unstable democracies create human rights regimes to lock in democratic principles in the face of domestic uncertainty. He argued that the primary proponents of reciprocally binding international human rights institutions are the governments of newly established or unstable democracies. This is because these governments fear the possible return to authoritarian government and are seeking to “lock in” democratic principles. In this logic, established democracies will only offer lukewarm support for binding international human rights treaties because the costs of reduced sovereignty outweigh the benefits of the commitment. In essence, established democracies receive no benefits because human rights are already respected, and they do not need to restrain any threatening domestic actors. Additionally, authoritarian governments will not support human rights regimes for the obvious reason that they are the states most likely to violate these norms.

For Moravcsik, new and unstable democracies contain one or more undemocratic opposition groups (such as the military, communists, or religious fundamentalists) that threaten the democratic government. His logic adopts the view that politicians delegate political power to other entities, such as domestic courts and administrative agencies, to constrain the behavior of future national governments (Moravcsik 2000, 228). By extension, unstable democracies might also delegate constraint functions to international treaties, “locking in” democratic rule through the enforcement of human rights. By placing power in the hands of independent authorities, governments seek to restrain future nondemocratic governments or elected governments that may subvert democracy from within.

One difficult issue is how to measure new and unstable democracies. Scholars have not tested the lock-in logic in large-n studies, though Simmons (2002) has helped blaze the trail. Following her lead, we employ three measures, each capturing a different dimension of Moravcsik’s argument. The Polity data set is the most widely used and most comprehensive measure of a country’s level of democracy. We identify a country as a New Democracy when it first achieves a score of 7 or above on the Polity2 scale, which ranges from −10 to 10. It remains a New Democracy until it either drops below 7 on the scale or else stays at 7 or above for at least 10 years. Although Moravcsik (2000, 231–33) suggests that states remain new democracies for 30 years after their initial transition, we adopt a standard that emphasizes the quality of democracy within a country and so utilize a shorter time period. This emphasis on quality is consistent with the literature on democratic consolidation, defined as the point at which democracy becomes “the only game in town” (Linz and Stepan 1996, 5). The best measures of consolidation are based on behavioral patterns (Schedler 2001), a dimension that is captured by the Polity data. High levels of democracy (7 and above) suggest that no other actors are seriously challenging the existing institutions and 10 years is long enough to give them a chance to try.9

Moravcsik’s argument is not only about new democracies but also unstable democracies. Democracies can be unstable, even if they fall into the “full democracy” (7–10) range on the Polity2 scale, when powerful domestic actors prefer more authoritarian rule and these threats are not picked up by coding methods that examine existing behavior. Hence, we identify Unstable Democracies in dichotomous fashion as those states which have achieved a positive Polity2 score since 1975 (coinciding with the beginning of the “third wave” of democracy) but who also experienced at least a three-point drop in their score at any point since achieving it. Such drops measure the existence of real threats to democracy that have created instability in the country and could do so again. Consistent with Moravcsik’s logic, we expect unstable democracies to have particularly strong interests in international commitment when they become democratic; hence, we interact this measure with the Polity2 score. The more democratic unstable states become, the more likely they should be to lock in human rights norms.

Finally, it is possible that democracies are unstable not because they have previously achieved democracy and regressed, but rather because politics in that country are particularly volatile and government shifts a frequent occurrence. Hence, we measure Regime Volatility as the standard deviation of a country’s Polity2 score from 1975 to 2002. As with democratic instability, we do not expect volatility itself to predict commitment; rather, it should matter when interacted with Polity2 scores, so that volatile coun-

---

8The “Polity2” score is more appropriate than “Polity” for time-series analysis because it imputes values for states experiencing severe polity change and would otherwise be coded as missing.

9The Polity data is highly correlated with alternative measures. See Russett and Oneal (2001, 45–46).
tries who become more democratic are the most likely to commit to human rights treaties.10 These last two measures, regime volatility and democratic instability, focus on democratic characteristics during the entire period from 1975 to 2002 as explanations for state behavior in any given year during that period. Although some may object to using future information to predict past events, Moravscik explicitly argues that fears of future instability are one of the key factors that drive democratic actors to lock in human rights commitments. Hence, we measure democratic instability and volatility over the entire period.

Varying Costs

Norms and lock-in approaches consider the benefits of international treaties. But states are likely to calculate not only the benefits of locking in particular policies but also the costs of doing so. Moravscik (2000, 228) considers some costs, but assumes they are either constant or else randomly distributed for all states. In contrast, we argue that sovereignty costs vary substantially from one state to the next in systematic ways.11 We identify three main kinds of costs that are likely to influence a state's commitment decision: policy change, unintended consequences, and limited flexibility.12

First, when states commit to new international treaties, they often have to change domestic policies, practices, laws, and even institutions in order to credibly comply with those commitments. As Downs, Rocke, and Baroom (1996) argue, most governments prefer to retain complete sovereignty and so only commit to international agreements when they have already adopted the relevant policies. As a result, the higher the congruence between a state's policy and the international treaty, the lower its policy costs in committing to the treaty and the more likely it is to commit sooner. Hathaway (2002, 2003) and Vreeland (2003) find evidence to support this argument, reporting that more democratic states are more likely to commit to CAT, but democratic states with poor torture records are less likely to participate in the treaty.

For CAT, we expect that countries with high levels of democracy, low levels of human rights abuse and high values on the rule of law are more likely to commit sooner. Note that this logic runs contrary in some respects to that of Moravscik, who expects stable democracies to have the same costs as new democracies, but no incentive to act. We measure level of democracy with the Polity2 score. We measure human rights with two indices provided by Cingranelli and Richards (2004). The first is the Physical Integrity Rights Index, which measures the extent to which states abuse physical integrity rights (disappearance, killings, political prisoners, torture), on a 0–8 scale, with 8 as the best record and 0 as the worst abuses. The second is the Empowerment Rights Index, which measures the extent to which states respect empowerment rights (freedom of association, movement, speech, political participation, religion), on a 0–10 scale, with 10 as best and 0 as worst.13 For Rule of Law, we utilized the PRS Country Risk Guide data, which uses experts to rank states on a 1–6 point scale, with strong rule of law at the high end.14 These operationalizations may capture concepts besides costs. As a result there may be multiple ways to interpret the results. Empowerment rights and rule of law, for example, may have more to do with a state's identity than with policy change costs.

Second, unintended (and sometimes unanticipated) consequences can create costs for states—as Pinochet discovered. States can calculate rough probabilities that a new treaty or any new commitment will be used against them in unintended and unwanted ways. We measure the possibility of unintended consequences in two very different ways. Internationally, we focus on state power, defined in terms of resources. Generally, powerful states have resources that can be used against weak states and that can be used to escape punishment should others attempt to inflict it. With respect to the torture treaty, powerful states have well-developed and well-financed judicial and police

10Another operationalization is to interact unstable and volatile democracies with a dummy variable for high-scoring democracies (7 or more on the Polity2 scale) rather than the Polity2 score itself. The results are qualitatively similar.


12In all cases, we offer objective measures of costs rather than subjective costs as perceived by policymakers. Such subjective costs are important and difficult to measure, and should be informed by the objective categories we establish. Nevertheless, our choice imposes limitations and future research should address policymakers' perceptions of costs.

13As an alternative to the physical integrity abuse index, we also examined the widely used "political terror scale" data set, which codes state abuses on a 1–7 scale (where 7 is the most abusive) by examining both U.S. State Department and Amnesty International reports (Gibney 2004). The results are qualitatively similar.

14The World Bank's "Governance Matters" database provides a more comprehensive rule of law index by compiling several different indices, but it only covers 1996–2002. Only the PRS index provides relatively comprehensive coverage for most countries back to 1980 (PRS Group).
systems that can be used to prosecute citizens of other states under the universal jurisdiction provisions. On the flip side, they can bring power resources to bear to help their citizens escape universal jurisdiction should others decide to exercise it. Hence, for powerful states, the likelihood of unintended consequences is lower than for weak states. We measure this with the natural logarithm of real GDP. On the domestic side, different kinds of institutions create different risks for states. In particular, Common Law Judicial Systems open up more possibilities for judges to apply international treaties in unintended ways. In common law systems, judges can create law through rulings and can more easily draw on legal sources beyond a statutory code written by the legislature. Hence, we expect states with common law legal systems to be more reluctant committers.

The third type of cost is the cost of foreclosing policy options in the face of uncertainty and threat; it is costly because it limits flexibility. Substantial evidence exists to suggest that states abuse human rights when it helps them achieve goals like power and wealth (Davenport 2000; Poe, Tate, and Keith 1999). States are most likely to benefit from human rights abuse when they face significant security threats. In such situations, they do not want to tie their hands so that they cannot use torture or other kinds of abuses to defeat their enemies. We measure threats both internationally and domestically. Internationally, we use the hostility level score for each country from the “Militarized Interstate Dispute” data, ranging from zero, no militarized dispute, to five, war. Other data sets on international conflict are less useful because they tend to focus on war, which is relatively rare and is the last step in a continuum of violence captured by the notion of militarized interstate disputes. In the absence of a similar measure for domestic conflicts, and because poverty is often associated with internal violence and civil war, we examine per capita real GDP as a measure of threat to the government.

**Findings**

We are interested in which countries commit to CAT, as well as their level of commitment. We use a duration model to see which of the above factors affect different states’ commitments. A duration model not only takes into account which counties sign or ratify, it examines why different countries commit more quickly than others. Furthermore, it takes into account the fact that countries almost never unsign or deratify the treaty. Finally, the duration model takes into account the fact that some countries have not yet signed or ratified (but can in the future), and that some countries ceased to exist before they made some commitments.

We examine the decision to sign or ratify separately. The opportunity to sign CAT starts after it is proposed at the end of 1984 (in the language of duration analysis, this is when the time “at risk” begins). For countries that came into existence after 1984 (e.g., Uzbekistan), their opportunity to sign starts when they begin to exist. Because some data were only available up to 2000, our analysis examines the likelihood of commitment from 1985 to 2000.

It is possible for a country to ratify the treaty without signing it through accession; thus, we do not treat the ratification decision as contingent on the signing decision. Most of the data are aggregated at an annual level, so we use a discrete-time duration model, also known as a grouped duration model (Beck, Katz, and Tucker 1998). We incorporate time-varying covariates, when possible, to allow changes in the country to affect its likelihood of commitment. Because the likelihood of commitment by one country may affect regional (or global) commitment and the country’s human rights practices, we lag the regional and global commitment scores as well as the human rights scores by one year to avoid reverse causation or

---

15 As an alternative measure of power, we also examined the Composite Index of National Capabilities, from the Correlates of War project, version 3.02 (Singer 1987). It was not statistically significant in the models that follow.

16 We used data provided by LaPorta et al. (1999). It is possible that states with common law legal systems avoid international human rights treaties for other reasons, such as British cultural heritage.

17 We also used average dispute levels during the entire period for which data is available as explanations for state behavior in any given year during that period under the assumption that governmental actors in each country are aware of real future threats that manifest themselves during this period. In addition, we examined other operationalizations of disputes using dummy variables, moving averages, etc. The results are qualitatively similar.

18 We tried using state failure data (Goldstone et al. 2000), which measures genocide, ethnic conflict, revolution, and adverse regime change, but the results were never statistically significant in any model for any operationalization (raw score, dummy variables, moving averages, overall average). The biggest difficulty of the state failure data is that it does not measure threats to the peace that fall short of these violent events while the militarized interstate dispute data does. Per capita GDP is obviously far from an ideal measure, but it does pick up latent domestic threats; instability and poverty go hand in hand.

19 For an excellent introduction to duration models in political science, see Box-Steffensmeier and Jones (2004).
simultaneity (called “rate dependence” in duration models; see Goodliffe 2003; Yamaguchi 1991).\textsuperscript{20} We do not think the other independent variables are affected by the likelihood of commitment.

In duration models, an important concept is the hazard rate, which gives the probability of commitment given that a commitment has not yet been made. We use the discrete-time equivalent of the Cox proportional hazard model (the complementary log-log model), which places no constraints on whether the hazard rate goes up or down (or both) over time.\textsuperscript{21} One way to do this is to include an indicator variable for each year since the duration began.\textsuperscript{22} An alternative method that we use is to include cubic smoothing splines as recommended by Beck, Katz, and Tucker (1998),\textsuperscript{23} which capture the hazard shape with fewer variables.\textsuperscript{24}

In Table 1, we assess the hypotheses discussed above through the discrete-time duration model.\textsuperscript{25} A positive coefficient indicates that as that independent variable increases the country is more likely to commit in general and also more likely to commit earlier. Any statement that a country is more likely to commit is also a statement that the country will commit earlier (and vice versa).\textsuperscript{26}

First, we find fairly strong support for the norms argument. The level of commitment to CAT by a country’s neighbors affects the likelihood that the country will sign or ratify: If most neighbors have signed or ratified, then that country is more likely to sign or ratify, and do so sooner. Although global norms do not affect the likelihood that a country will sign, they do affect whether the country will ratify.\textsuperscript{27}

There is little evidence that supports the lock-in logic. The three variables that we have used to measure lock-in benefits—New Democracy, Unstable Democracy × Polity2 Score, and Regime Volatility × Polity2 Score—are individually and jointly insignificant for both signing and ratifying. Furthermore the result does not depend on the operationalization used here—when using a threshold approach to level of democracy, the effect is neither statistically nor substantively significant.\textsuperscript{28}

Arguments about the importance of sovereignty costs receive stronger support. The first category of costs is policy change costs, which refer to the extent to which a country has to change its policies if it commits to a treaty, where more extensive changes should make states less likely to commit. Two of the four measures provide support for this argument. Countries that score higher on the Empowerment Rights Index, a measure of civil liberties like free speech and free association, are more likely to sign and ratify. Further, countries that follow the rule of law are more likely to sign the treaty. Surprisingly, a country’s score on the Physical Integrity Rights Index does not affect its likelihood of signing or ratifying.\textsuperscript{29} This echoes Hathaway’s (2003, 1849) finding that non-democratic regimes with higher rates of torture are more likely to commit to CAT. Vreeland (2003) addresses this puzzle, arguing that regimes with higher rates of torture are more likely to need the normative cover provided by committing to CAT.\textsuperscript{30} And controlling for other variables, more democratic countries (Polity2 Score) are less likely to ratify, although the substantive effect is negligible, as we show below.

The second category of costs is unintended consequences, which consists of two variables, both of which are significant. Countries that have a common law judicial system are less likely to sign and ratify, in line with our hypothesis that government officials in

\textsuperscript{20}The results that follow are qualitatively similar when we do not lag the scores.

\textsuperscript{21}We discuss the hazard shape, which is not monotonic, in the appendix found at the journal website (http://www.journalofpolitics.org) or at http://goodliffe.byu.edu/papers.

\textsuperscript{22}We also estimated a model that included an indicator variable for each calendar year, rather than year since the duration began, to pick up any year-specific effects (such as when Pinochet was arrested). The calendar year dummies were not statistically significant.

\textsuperscript{23}We used Tucker’s (1999) program to generate the cubic splines.

\textsuperscript{24}The results that follow are qualitatively similar when using temporal dummy variables to control for duration dependence.

\textsuperscript{25}We use robust standard errors, clustered by country. The results are qualitatively similar for other assumptions on standard errors. In addition, controlling for country heterogeneity with a random-effects model or a population-averaged model yields qualitatively similar results.

\textsuperscript{26}For both commitments, the duration dependence variables—Years at Risk and the cubic splines—are jointly statistically significant.

\textsuperscript{27}We also tested for a threshold effect at one-third of the states signing and ratifying, as hypothesized by Finnemore and Sikkink (1998), but it was not significant.

\textsuperscript{28}Regime volatility (without regard to level of democracy) appears to increase likelihood of ratifying, but not signing.

\textsuperscript{29}We carried out further analysis to ensure this finding was robust. The Physical Integrity Index is not highly correlated with any of the other explanatory variables and is not significant even in leaner specifications of the model that leave out the key predictors.

\textsuperscript{30}This answer of course raises further questions about why states that violate empowerment rights would not need the same type of normative cover and why some kinds of costs matter more than others. We leave these puzzles for future research.
those countries are concerned about how judges will interpret the treaty. And more powerful countries (as measured by GDP) are more likely to ratify, again supportive of our argument that powerful states believe they have sufficient resources to avoid negative unintended consequences.

The third category of costs consists of concerns about limiting a government’s flexibility. We find evidence that countries involved in interstate disputes are less likely to commit in any way. But poor states (as measured by GDP/capita) are not more or less likely to commit.

To compare the substantive significance of different variables, we compare the median survival time, changing one independent variable at a time. The survivor function gives the probability of not committing up to a certain time. In this study, the median survival time gives the time at which the cumulative probabil-

Table 1  Discrete-time Duration Model of Commitment to Convention against Torture

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Signature</th>
<th></th>
<th>Ratification/Accession</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (s.e.)</td>
<td>p-value</td>
<td>Coefficient (s.e.)</td>
<td>p-value</td>
</tr>
<tr>
<td>Regional Score (lagged)</td>
<td>1.574 (.468)</td>
<td>.001</td>
<td>1.216 (.399)</td>
<td>.002</td>
</tr>
<tr>
<td>Global Score (lagged)</td>
<td>-.749 (.841)</td>
<td>.373</td>
<td>3.392 (1.137)</td>
<td>.003</td>
</tr>
<tr>
<td>New Democracy</td>
<td>-.099 (.516)</td>
<td>.848</td>
<td>-.603 (.369)</td>
<td>.102</td>
</tr>
<tr>
<td>Unstable Democracy × Polity2 Score</td>
<td>.006 (.046)</td>
<td>.904</td>
<td>.056 (.046)</td>
<td>.219</td>
</tr>
<tr>
<td>Regime Volatility</td>
<td>.012 (.008)</td>
<td>.159</td>
<td>.012 (.008)</td>
<td>.127</td>
</tr>
<tr>
<td>Unstable Democracy × Polity2 Score</td>
<td>.181 (.333)</td>
<td>.586</td>
<td>-.068 (.381)</td>
<td>.857</td>
</tr>
<tr>
<td>Regime Volatility</td>
<td>.060 (.053)</td>
<td>.254</td>
<td>.126 (.060)</td>
<td>.034</td>
</tr>
<tr>
<td>Polity2 Score</td>
<td>-.025 (.043)</td>
<td>.562</td>
<td>-.082 (.044)</td>
<td>.059</td>
</tr>
<tr>
<td>Empowerment Rights Index (lagged)</td>
<td>.155 (.072)</td>
<td>.030</td>
<td>.174 (.072)</td>
<td>.013</td>
</tr>
<tr>
<td>Physical Integrity</td>
<td>-.071 (.074)</td>
<td>.339</td>
<td>-.080 (.074)</td>
<td>.266</td>
</tr>
<tr>
<td>Rights Index (lagged)</td>
<td>.249 (.131)</td>
<td>.056</td>
<td>.132 (.124)</td>
<td>.287</td>
</tr>
<tr>
<td>Rule of Law</td>
<td>.127 (.141)</td>
<td>.266</td>
<td>.214 (.099)</td>
<td>.031</td>
</tr>
<tr>
<td>ln(GDP)</td>
<td>-.813 (.300)</td>
<td>.007</td>
<td>-.627 (.288)</td>
<td>.030</td>
</tr>
<tr>
<td>Common Law</td>
<td>-.142 (.093)</td>
<td>.125</td>
<td>-.161 (.077)</td>
<td>.037</td>
</tr>
<tr>
<td>Judicial System</td>
<td>.178 (.177)</td>
<td>.315</td>
<td>.244 (.198)</td>
<td>.219</td>
</tr>
<tr>
<td>Militarized Interstate Dispute Score</td>
<td>-.674 (.243)</td>
<td>.006</td>
<td>-12.460 (2.172)</td>
<td>.000</td>
</tr>
<tr>
<td>ln(GDP/capita)</td>
<td>-.749 (.468)</td>
<td>.373</td>
<td>3.392 (1.137)</td>
<td>.003</td>
</tr>
<tr>
<td>Duration Dependence</td>
<td>-.099 (.516)</td>
<td>.848</td>
<td>-.603 (.369)</td>
<td>.102</td>
</tr>
<tr>
<td>constant</td>
<td>.006 (.046)</td>
<td>.904</td>
<td>.056 (.046)</td>
<td>.219</td>
</tr>
<tr>
<td>Number of Countries</td>
<td>115</td>
<td>116</td>
<td>115</td>
<td>116</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>798</td>
<td>1,075</td>
<td>798</td>
<td>1,075</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>-216.2</td>
<td>-245.0</td>
<td>-216.2</td>
<td>-245.0</td>
</tr>
<tr>
<td>$\chi^2 (p &gt; \chi^2)$</td>
<td>134.7 (p = .000)</td>
<td>136.1 (p = .000)</td>
<td>134.7 (p = .000)</td>
<td>136.1 (p = .000)</td>
</tr>
</tbody>
</table>

Notes: Dependent variables are signing the Convention Against Torture and ratifying/accessing the Convention. Coefficients are complementary log-log regression estimates with robust standard errors clustered by country in parentheses; p-values are for two tails.

*Duration Dependence represents the Years at Risk variable and three cubic spline variables; the p-value is for a joint significance test.
### Table 2  Change in Median Survival Time of Signing the Convention Against Torture

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Change in Independent Variable</th>
<th>Median Survival Time</th>
<th>Change in Median Survival Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td>1.7 years</td>
<td></td>
</tr>
<tr>
<td>Common Law Judicial System</td>
<td>0 to 1</td>
<td>14.6</td>
<td>12.9 years</td>
</tr>
<tr>
<td>Regional Score (lagged)</td>
<td>.5 to 0</td>
<td>14.1</td>
<td>12.4</td>
</tr>
<tr>
<td>Empowerment Rights Index (lagged)</td>
<td>5 to 0</td>
<td>13.9</td>
<td>12.2</td>
</tr>
<tr>
<td>Militarized Interstate Dispute Score</td>
<td>0 to 4</td>
<td>9.2</td>
<td>7.5</td>
</tr>
<tr>
<td>Rule of Law</td>
<td>3 to 1</td>
<td>7.0</td>
<td>5.3</td>
</tr>
<tr>
<td>Global Score (lagged)*</td>
<td>.7 to 1.2</td>
<td>3.8</td>
<td>2.1</td>
</tr>
<tr>
<td>GDP/capita (in thousands of real dollars)</td>
<td>2.6 to .8</td>
<td>2.6</td>
<td>.9</td>
</tr>
<tr>
<td>Unstable Democracy*</td>
<td>0 to −1</td>
<td>2.6</td>
<td>.9</td>
</tr>
<tr>
<td>Regime Volatility</td>
<td>3.2 to 0</td>
<td>2.5</td>
<td>.8</td>
</tr>
<tr>
<td>GDP (in billions of real dollars)</td>
<td>23.5 to 5.7</td>
<td>2.4</td>
<td>.7</td>
</tr>
<tr>
<td>Physical Integrity Rights Index (lagged)*</td>
<td>5 to 7</td>
<td>2.2</td>
<td>.5</td>
</tr>
<tr>
<td>Regime Volatility × Polity2 Score (including effects of Polity2 Score)</td>
<td>0 to −32.0</td>
<td>2.1</td>
<td>.4</td>
</tr>
<tr>
<td>New Democracy*</td>
<td>0 to 1</td>
<td>2.0</td>
<td>.3</td>
</tr>
<tr>
<td>Unstable Democracy × Polity2 Score (including effects of Unstable Democracy and Polity2 Score)</td>
<td>0 to 10</td>
<td>1.7</td>
<td>.0</td>
</tr>
</tbody>
</table>

**Notes:** Median survival time of signing the convention is calculated from the coefficients in Table 1. The baseline case sets each independent variable to its median value, except for Polity2, which is set to 0, and interactive variables which are set to the product of the medians of the two constituent variables. For ease of interpretation, we convert logged GDP values back to their original values.

**Example:** The baseline country signed CAT after 1.7 years. The baseline country with a common law judicial system signed CAT after 14.6 years, an increase of 12.9 years over the baseline case.

*The effect is opposite of the hypothesized direction.*

*Unstable Democracy can only change to +1, but we change it to −1 to compare the effects in the same direction. A change to +1 decreases the median survival time to 1.2 years (a change of −.5 years).*

The change in median survival time for signing CAT are in Table 2. The independent variables that cause the greatest changes in the median survival time relate to costs and norms. The results suggest that lock-in variables are not only statistically insignificant; they are also substantively of much less importance. Moving to a common law judicial system (an unintended cost variable) increases the median time of signing by almost 13 years. Reducing the number of nearby countries that have signed or ratified (a norm variable) increases the median time of signing by 12.4 years. Similarly, moving from a median empowerment regime to a low empowerment regime (a policy cost variable) increases the median time of signing by 12.2 years. In addition, moving from a country with no interstate disputes to one which is currently using

---

31 We hold the regional and global scores constant across time, so that we can compare the effect of changing those variables against the effect of changing other variables. Of course, we allow the duration dependence variables—Years at Risk and the cubic spline variables—to vary across time.

32 For positive coefficients, we move the independent variable from the 50th to the 10th percentile. For negative coefficients, we move the independent from the 50th to the 90th percentile.

33 Although the survivor function is calculated annually, we estimate where the function crosses .5 assuming the function changes linearly between years. If the survivor function is assumed to change discretely, then one rounds up each median survival time.

---
TABLE 3 Change in Median Survival Time of Ratifying the Convention Against Torture

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Change in Independent Variable</th>
<th>Median Survival Time</th>
<th>Change in Median Survival Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td></td>
<td></td>
<td>1.7 years</td>
</tr>
<tr>
<td>Global Score (lagged)</td>
<td>.7 to 0</td>
<td>16+</td>
<td>16+</td>
</tr>
<tr>
<td>Empowerment Rights Index (lagged)</td>
<td>6 to 0</td>
<td>15.2</td>
<td>13.5</td>
</tr>
<tr>
<td>Regional Score (lagged)</td>
<td>.5 to 0</td>
<td>3.4</td>
<td>1.7</td>
</tr>
<tr>
<td>Militarized Interstate Dispute Score</td>
<td>0 to 4</td>
<td>3.3</td>
<td>1.6</td>
</tr>
<tr>
<td>Common Law Judicial System</td>
<td>0 to 1</td>
<td>3.3</td>
<td>1.6</td>
</tr>
<tr>
<td>New Democracy</td>
<td>0 to 1</td>
<td>3.1</td>
<td>1.4</td>
</tr>
<tr>
<td>Regime Volatility × Polity2 Score^*</td>
<td>0 to 31.3</td>
<td>2.6</td>
<td>.9</td>
</tr>
</tbody>
</table>

(including effects of Polity2 Score)

Regime Volatility                                     3.1 to 0                        2.5                   .8
GDP (in billions of real dollars)                       28.0 to 5.8                     2.3                   .6
Unstable Democracy × Polity2 Score (including effects of Unstable Democracy and Polity2 Score) 0 to 10 2.3  .6
GDP/capita (in thousands of real dollars) 3.2 to .8 2.3  .6
Rule of Law                                            3 to 1                          2.1                   .4
Physical Integrity Rights Index (lagged)^*              5 to 8                          2.1                   .4
Unstable Democracy^*                                    0 to 1                          1.8                   .1

Notes: Median survival time ratifying the convention is calculated from the coefficients in Table 1. The baseline case sets each independent variable to its median value, except for Polity2, which is set to 0, and interactive variables which are set to the product of the medians of the two constituent variables. For ease of interpretation, we convert logged GDP values back to their original values.

Example: The baseline country ratified CAT after 1.7 years. The baseline country with a common law judicial system ratified CAT after 3.3 years, an increase of 1.6 years over the baseline case.

^*The effect is opposite of the hypothesized direction.

force (a flexibility cost variable) increases median time of signing by 7.5 years. Finally, moving from a state with median rule of law to one with low rule of law (a policy cost variable) increases the median time of signing by about five years. The other independent variables do not approach statistical significance, and, in comparison, are substantively insignificant. Substantively then, regional norms and the three types of costs are most important for signing, and lock-in is least important.

The changes in median survival time for ratifying CAT are in Table 3. Similar to the changes in signing CAT, the strongest effects are for norm and cost variables, although there are some differences. By far the strongest variable is the number of countries in the world that have signed or ratified (a norm variable). When one reduces the number of countries that have signed or ratified, the median time until ratification extends far beyond the 16-year window of the data. Moving from a median empowerment regime to a low empowerment regime (a policy cost variable) increases the median time of ratifying by 13.5 years. These effects are substantively much stronger than other effects. Like the effects on signing, reducing the number of nearby countries that have signed or ratified increases the median ratifying time, but only by 1.7 years. Moving from a country with no interstate disputes to one which is currently using force increases the median time of ratifying by 1.6 years. And moving to a common law judicial system also increases the median ratifying time by about 1.6 years. There are other variables that are statistically significant (e.g., GDP), but they are substantively unimportant. Similar to signing, the lock-in variables are substantively unimportant, and are in the wrong direction. Global norms are substantively most important for ratifying, policy costs are next in importance, then regional norms, unintended costs and flexibility costs, and finally, lock-in is least important.

Overall, in terms of substantive and statistical significance, norms and sovereignty costs are more important than lock-in benefits.

Conclusions

Our findings suggest three broad conclusions. First, not all states experience the same costs when considering commitment to an international treaty, and cost variation can substantially influence commitment
patterns. Hence, Moravcsik’s (2000, 228) assumption that costs are either constant or randomly distributed is untenable (though it may have been useful for the development of lock-in theory and an initial analysis of its plausibility). Theoretical arguments that focus on the benefits of state lock-in without considering the variation in costs are likely to be misleading and misspecified. Moreover, it is worth noting that cost variables were statistically and substantively significant even when tested against norms. As we noted, norms may operate either by creating logics of appropriateness or by creating costs and benefits for states. If one adopts the latter position, then our methodology actually underestimates the costs by including one type of cost (reputational costs of refusing to commit to a norm against torture) in the category of norms. This paper therefore creates a difficult test for our costs categories.

Scholars have been slow to think about the importance of the costs of international commitments or to identify the different kinds of costs. In her extensive review of the literature, Hathaway (2003) identified one set of scholars arguing that the costs of international commitments are uniformly high and another set arguing that states are attracted by norms regardless of costs. A third set of scholars, typified by Downs, Rocke, and Barsoon (1996), does indeed argue that costs influence states, yet these scholars have only just begun identifying the range of costs. Hathaway (2003) focuses almost exclusively on policy change costs related to existing human rights practices and the level of democracy.

We have identified and found evidence for three different kinds of costs encompassing four variables. Among policy change costs, both the Empowerment Rights Index and the Rule of Law were statistically and substantively significant. States that already respect individual liberties like free speech are more likely to both sign and ratify earlier, at the median, by margins of 12.2 and 13.5 years, respectively. Rule of Law does not have such dramatic effects and is only significant for signing, but it does move up the median time to signing by five years. More puzzling is the fact that the Physical Integrity Index does not affect state commitment, a finding that deserves further attention in future research. States also seem concerned about unintended consequences, our second cost category. A Common Law Judicial System was significant for both signing and ratifying, with important substantive effects that altered the median signing time by nearly 13 years and median ratifying time by 1.6 years. GDP, in contrast, was statistically significant for ratifying, but not substantively important. States also appear to factor in the extent to which a commitment will tie their hands in unwanted ways, the third category of costs. In particular, a state’s level of hostility internationally appears to affect both signing and ratifying, delaying them by a median score of 7.5 years and 1.6 years, respectively.

Second, we do not find any evidence for the lock-in hypothesis that is so prominent in the literature but has not been widely tested. We measured lock-in variables in three main ways, with variations on those three measures (e.g., using thresholds for the Polity2 variable), but failed to find any of them statistically significant in any model. Substantively, they also seem unimportant, affecting state commitments by less than a year. The only lock-in variable that seems to matter substantively is New Democracy, but it is signed in the wrong direction (new democracies will take longer to ratify CAT). It is possible that CAT simply is not strong enough to provide reassurances to new democracies and that this explanation would work better for even more highly legalized international institutions.

Third, the strong evidence for regional and global norms is fairly striking and is generally supportive of arguments that states observe other states and do what they do (Finnemore and Sikkink 1998). These norms arguments have not been tested much in large-n data; thus, these findings are notable. The finding also reinforces recent work on state commitment in the context of international finance. Simmons (2000) found that state commitment to the International Monetary Fund is influenced by both Global and Regional Commitment Scores of other states. We found largely the same results, suggesting that the effects of norms are robust across issue areas. Further, the substantive importance of global and regional norms is quite large. At the median, the Global Commitment Score influenced state ratification by far more than 16 years. The causal mechanism, however, remains unclear. Rationalist scholars have argued that states commit to norms to avoid paying reputational costs. In contrast, constructivist scholars have suggested that states commit to norms because it fits with their identities or their social psychological learning processes. We cannot disentangle these competing causal mechanisms, but our results are somewhat suggestive. The strong effects of norms in the face of a variety of different kinds of costs suggests, in a more speculative vein, that perhaps norms play a role above and beyond the costs and benefits they provide. Given the robust findings and uncertain causal mechanisms, further systematic research on the effects of norms certainly seems to be in order.
Acknowledgments

We would like to thank Celest Allred for providing excellent research assistance and contributing substantially to earlier versions of this paper. We would also like to thank Lawrence Broz, Gary Bryner, Scott Cooper, Dan Nielson, Ken Stiles, Erik Voeten, the anonymous reviewers, and participants at various presentations for comments on previous drafts. Grants and professional leaves from Brigham Young University aided in the research and writing.

Manuscript submitted 6 June 2005
Manuscript accepted for publication 21 October 2005

References


Jay Goodliffe is associate professor of political science, Brigham Young University, Provo, UT 84602-5545. Darren G. Hawkins is associate professor of political science, Brigham Young University, Provo, UT 84602-5545.