Is Enforcement Necessary for Effectiveness?
A Model of the International Criminal Regime
Michael J. Gilligan
New York University

Abstract
Recently scholars have questioned whether enforcement mechanisms are necessary to make regimes effective. This paper provides a model of the international criminal regime in which the regime changes state behavior even though it possesses no enforcement mechanisms. The paper also answers several prominent criticisms of the International Criminal Court (ICC). Critics claim that the ICC is at best futile because it lacks the power to apprehend the criminals it is meant to prosecute. Even worse, the ICC may be harmful because it will induce atrocious leaders to hold on to power longer than they would if they could step down with immunity for past crimes. The model in this paper suggests those criticism may be inaccurate. I model the interaction between a leader and a foreign state that has the option of offering that leader asylum. I examine the effect of the creation of an ICC-like institution on that interaction. The model produces three main findings: 1) Leaders’ reigns will not be prolonged as a result of the regime. 2) Although the institution has no enforcement power some leaders (those with such a high probability of being deposed that they would willingly surrender to the institution rather than try to stay in office) will be punished by it. In those circumstances the foreign state has no incentive to offer the leader asylum. 3) The institution may deter some atrocities at the margin.

* I would like to thank Tanaz Moghadam whose undergraduate honors thesis, which I advised, provides an earlier game-theoretic analysis of the Court and introduced me to the topic of the ICC. I am grateful to Lisa Martin and two anonymous reviewers who provided exceptionally helpful insights. I would also like to thank William Clark, Sandy Gordon, Dimitri Landa, James Morrow, Steve Ratner, Shanker Satyanath and Ken Scheve for their valuable comments. All errors remain my responsibility.
In August 2003 Charles Taylor escaped from Liberia, the chaotic country he had led for six years. Taylor was widely considered to be guilty of a variety of atrocities, and at the time of his escape had been indicted by a UN war crimes tribunal in Sierra Leone for arming bands of rebels, some members of which were children, who carried out a program of rape and mutilation during Sierra Leone’s civil war. Taylor is also considered to be guilty of atrocities in his own country. Taylor obtained *de facto* asylum in Nigeria as part of a deal brokered to help end the civil war in Liberia. Nigeria’s leadership has made it clear that it considers Taylor’s exile in Nigeria a political affair and will not release him for trial in Sierra Leone.¹ Cases like Taylor’s are unfortunately not uncommon. Other examples include Jean-Claude “Baby Doc” Duvalier who is accused of torture and political assassinations in Haiti and who lives comfortably in France. Alberto Fujimori, who is accused of various human rights abuses during the civil war in Peru, lives in Japan which has refused extradition. Ferdinand Marcos, the extraordinarily corrupt dictator of the Philippines, lived out his days without punishment in Hawaii, as did Idi Amin in Saudi Arabia. The list could go on.²

Such cases are disturbing for at least two reasons. First there is the obvious normative concern that perpetrators of atrocities are not being punished for their crimes. Second there is the more practical question of deterrence—if leaders who commit atrocities are able to get off scot-free what hope is there of deterring such crimes in the future? In part to help prevent the recurrence of cases like those described above a group of states created the International Criminal Court (ICC) in 1998 after a half century of on-

---

¹ “Charles Taylor: A wanted man.” See also “Charles Taylor - preacher, warlord and president” and “Taylor's new Nigerian home.”
² “No one writes to the tyrants.”
again-off-again negotiations. The ICC entered into force in the summer of 2002, and in the summer of 2003 the Court’s first Chief Prosecutor Luis Moreno-Ocampo was appointed.³

While the modal opinion in legal journals probably supports the creation of the ICC, the Court has been the object of virulent criticism from both legal scholars and policy makers. Unfortunately neither the Court’s proponents nor its detractors have been particularly rigorous in how they have made their arguments, so it is difficult to tell if the disagreement stems from faulty logic or simply from differences in unstated assumptions. This article attempts to contribute to the literature by offering a more formal argument about the possible implications of the creation of the ICC. I focus on two criticisms of the Court: 1) the ICC will fail to be effective because it lacks enforcement and 2) the ICC will prolong the reign of leaders who have abused human rights and make it more costly to get rid of them.

Goldsmith (2003, 92) offers a forceful example of the first critique:

“…the ICC is unlikely to punish the Husseins and future Milosevics of the world because it is unlikely to get a grip on them. The ICC has no inherent enforcement powers. It depends completely on member states to arrest and transfer defendants. So the efficacy of … prosecutions depends on the uncertain resolve of nations to use military or economic force to extricate the oppressive leader from his country”⁴

³ Information on current events at the ICC can be found on the Web at http://www.icc-cpi.int/otp/otp_events.html.
⁴ See Smidt (2001) for the same argument.
If this argument is correct then by implication the ICC cannot be expected to fulfill the goal of deterring atrocities.⁵ Krasner (2003) provides a particularly nice example of the second criticism.

“…Baby Doc Duvalier … ended up living on the French Riviera. … Idi Amin, … has lived in Saudi Arabia for the last twenty years now … [W]hen you think about justice and especially Idi Amin, who killed a lot of people, the question is this: was it better to offer him sanctuary in Saudi Arabia, or would it have been better to say to him when he was still in power, ‘We are going to prosecute you?’ The reaction of any autocratic ruler under those circumstances would be to hold on to power as long as he possibly can. … Threatening these actors with prosecution may make it more difficult to get rid of them.”⁶

I analyze these criticisms within the context of a game-theoretic model. I show that despite the lack of an international police force to enforce its indictments, the ICC may still deter leaders from committing atrocities at the margin because it offers states the option of refraining from offering asylum to leaders who are so likely to be deposed that they would willingly surrender to the ICC to avoid retribution from domestic political rivals (which I assume is worse than the punishment handed out by the Court). Since every leader knows that there is some chance (however remote) that he will be in that position the expected present value of committing atrocities is reduced and at the margin some leaders will refrain from committing them.

The model also suggests that the concern that the ICC will prolong tyrants’ reigns may be exaggerated. The ICC, like any international institution, must be self-enforcing. States will only comply with it if they have the incentive to do so. The ICC has no power to punish states that harbor atrocity committers so states will refrain from doing so only

---

⁵ Wippman (2000) and Smidt (2001) focus on ICC’s capacities (or lack thereof) to deter atrocities. Akhavam (2001) provides a more positive assessment.

⁶ The same argument can be found in Goldsmith (2003) and Goldsmith and Krasner (2003).
when it provides at least as high a payoff as not doing so. In this model they will only refrain from offering such leaders asylum when those leaders are so likely to be deposed that they are willing to surrender to the ICC, and therefore the reign of atrocity-committing dictators will not be prolonged as a result of the creation of the institution.

A second contribution of the paper concerns the role of enforcement in making international regimes effective. International relations scholars have known for some time that “trigger strategies” can induce states to comply with international regimes despite their lack of a higher authority to enforce those rules as long as the states’ interaction are repeated indefinitely and they do not discount future payoffs too heavily. According to other literatures these enforcement provisions are unnecessary for regimes to be effective. Compliance is either simply a management problem or is unnecessary because norms or new knowledge created by the regime induce states to comply with its rules.

This paper offers a rational choice model of an international institution that alters states’ behavior even though it is not enforced by trigger strategies or any other mechanism. Furthermore the model does not resort to “epistemic community” or norms arguments nor does it treat noncompliance as a management problem. There is no outside intervention either by states or international organizations to depose leaders who commit atrocities. More interestingly there is not even punishment for offering asylum to dictators who commit atrocities. Still states may refrain from offering asylum when the dictator’s hold on power is sufficiently tenuous because they know he would be willing to surrender to the ICC. States only comply with the institution’s indictments when it is costless to do so. They will regularly fail to comply with the regime modeled here, by
which I mean they will continue to grant asylum to leaders who have committed atrocities, however they will comply with it enough to deter atrocities at the margin.

In short, the paper contributes to both the policy debate surrounding the ICC and the theoretical literature on enforcement and international cooperation. While the model offers no hope that the creation of the ICC will bring about a world free of atrocities, it does offer a set of conditions under which there will be marginally fewer atrocities thanks to the creation of the institution. Furthermore, there is no downside to creating the institution in this model since the reign of atrocity-committing leaders is not prolonged as a result. Second, the model offers an interesting example of a rational choice model in which an international institution is effective even though no punishments (whether tit-for-tat, grim trigger etc.) are employed to enforce compliance.

In this next section I will review the debate on the necessity of enforcement for regime effectiveness and discuss the implications of the model for that debate. In the third section I will turn to a discussion of the literature on the ICC. Section four offers an informal discussion of the model and its results. I highlight a few features and extensions of the model in section five before concluding in section six.

**Is Enforcement Necessary for Regimes to be Effective?**

The main implication of this research for broader international relations theory concerns enforcement and the efficacy of international regimes.\(^7\) The cooperation problem modeled here shares a variety of features that are common in international institutions more generally—the lack of a police force, reliance on sovereign states and

---

\(^7\) I would like to thank the editor and the two anonymous reviewers for pointing out this contribution of the model.
non-universal membership. One of the main endeavors of international relations scholarship over the last two decades has been an answer to the question of how international agreements can have any effect on state behavior under such circumstances.\footnote{This discussion is necessarily brief. For a fuller discussion see Martin and Simmons (1998), Simmons (1998) Raustiala and Slaughter (2002), and Simmons and Martin (2002).}

The game-theoretic approach provides one answer via the “folk theorem.” If the interaction between states is repeated indefinitely states can enforce their treaty commitments by employing trigger strategies. Cooperation is enforced by the credible threats (either implicit or explicit) of the other members of the agreement to revert to noncooperation for some subsequent amount of time in punishment for any deviation from cooperation. Even in cases where noncompliance is a dominant strategy in the one-shot interaction states will comply with their commitments. Without such enforcement mechanisms states would fail to comply with agreements and the agreements will have no effect on state behavior.

More recently scholars have questioned whether these types of enforcement mechanisms are the best way to understand states’ compliance with their treaty commitments and whether they are necessary or even desirable for international agreements. One such group of scholars has been dubbed the “managerialists” by Downs et. al. (1996)\footnote{Examples of managerialist works include Chayes and Chayes (1993, 1995), Haas, Keohane and Levy, eds. (1993), Mitchell (1993, 1994) and Young (1994)} who provide a nice summary of managerialist thinking:

Its message is that (1) compliance is generally quite good (2) this high level of compliance has been achieved with little attention to enforcement (3) those compliance problems that do exist are best addressed as management rather than enforcement problems; and (4) the management rather than the enforcement approach holds the key to future regulatory cooperation in the international system (p. 379).
Downs et. al. are quite critical of this argument. They question whether the relatively good record of compliance indicates that enforcement is unimportant since in equilibrium regimes will be designed to bring about compliance. One cannot infer from the relatively sterling record of compliance whether enforcement measures are unnecessary or whether they are very necessary but also very effective. Furthermore if enforcement provisions are weak or nonexistent as managerialists contend high rates of compliance are probably due to shallow treaties. They argue that we should not be sanguine about the combination, touted by managerialists, of weak or non-existent enforcement and high rates of compliance because that combination implies shallow treaties.\(^\text{10}\)

A second school of thought that regards enforcement as unimportant for regime effectiveness is the normative approach (often associated with constructivism). According to this approach credible punishments for noncompliance are not what change state behavior. Instead regimes change states’ preferences away from noncooperation toward cooperation by establishing new norms that alter states perceptions of what behaviors are acceptable.\(^\text{11}\) A third literature that considers enforcement unnecessary for regime efficacy points to “epistemic communities” and learning.\(^\text{12}\) According to this literature communities of experts associated with the regime generate new knowledge about the benefits of international cooperation. Enforcement becomes unnecessary as

\(^{10}\) Also see Simmons (1998, 80) for a discussion of this debate. Tallberg (2004) takes the middle ground, arguing that within the EU both approaches are used and both are valuable in promoting compliance.


states learn that cooperation is in their best interests. Finally Dai (2005) offers a rational choice model in which states comply with their international commitments despite the absence of international enforcement mechanisms. In her model states comply with their commitments not to avoid punishment from other states but because they are pressured to do so by powerful domestic constituencies that prefer compliance.

Like the accounts described in the preceding paragraph, the regime modeled in this paper has no enforcement provisions either explicit or tacit, and states are not punished for noncompliance. Unlike those accounts neither norms nor learning play any role in the model, nor does the model treat compliance as a simple management problem or rely on pressure from domestic constituencies. States in the model often will strictly prefer noncompliance (i.e. offering asylum to indicted leaders) and will fail to comply in such situations. Indeed contrary to the accounts described above noncompliance will be quite frequent in this model and yet it is still effective in that states comply with it enough that it does deter atrocities at the margin. The regime modeled here is crucial to this change in state behavior because without it the foreign state would continue to offer asylum to dictators in those cases where it refrains from doing so after the creation of the regime.

Indeed this discussion raises the issue of whether a good record of compliance is even necessary for regimes to be effective. The two concepts, compliance and effectiveness, are sometimes conflated in the literature. However, as Raustiala and Slaughter (op. cit.) point out “[R]egimes can be effective … even if compliance is low. And while high levels of compliance can indicate high levels of effectiveness they can also indicate low, readily met and ineffective standards.” Downs et. al. are clearly
concerned about the latter phenomena. Depending on one’s definition of “effectiveness” the model in this paper provides a set of conditions that produce the former.\textsuperscript{13}

In summary, the model provides a rational choice account of a regime that changes state behavior but where enforcement is unnecessary and perhaps even undesirable because it may introduce commitment problems and prolong atrocious leader’s reigns. Members of the regime frequently will fail to comply with the regime but they will occasionally do so, which is enough to deter atrocities in some cases. Harsher punishments for members who offer asylum to indicted leaders would induce more compliance but may come with the added cost of prolonging the rule of atrocious dictators.

Previous Research on the ICC

The legal literature on the ICC is voluminous, so a thorough review of it is beyond the scope of this paper. The political science literature on the topic is sparser, and can be divided into theoretical and empirical contributions. Turning first to the theoretical contributions, the only other game game-theoretic treatment of the ICC of which I am aware is Moghadam (2004), who argues that the ICC may deter of atrocities. The source of this deterrent effect is an intervention to punish human rights abusers by “the international community,” and as such her model fails to address one of the main criticisms of the Court, namely that such interventions are unlikely. Furthermore her paper does not address whether the ICC will prolong the reign of leaders who have committed atrocities.

\textsuperscript{13} See also Keohane et. al. (1993) Young (1994) and Victor et. al. (1998)
In a rational-choice-based case study of a related topic—treaties on treatment of prisoners of war—Morrow (2001) argues that states refrain from violating POW treaties in time of war because if they did their enemies may retaliate by doing same. Thus the game has the characteristics of the classic repeated prisoners’ dilemma. Morrow provides evidence of this process at work in Nazi treatment of British and American POWs during World War II. Morrow’s analysis highlights the particular difficulties inherent in an international criminal regime which is meant to deter atrocities committed by leaders against their own citizen, because nothing like the same enforcement mechanism exists.

Meernik (2004a, 356) also raises some of the questions raised here:

“To what extent should we expect the international community to hold individuals accountable? Will individual leaders comply with international law? Most importantly, what factors will influence the level of enforcement of and compliance with international law?”

Rather than developing a theoretical model, however, Meernik’s approach is to suggest covariates that will correlate with these behaviors. Meernik’s answers are of course speculative since the ICC has yet to indict a single a person. He suggests *inter alia* that states that have shown a commitment to international law as indicated by their support for human rights treaties in the past and democratic states should be more supportive of the ICC ceteris paribus.

Turning to empirical studies, Meernik and Shairick (2005) examine the kind of states that have acceded to the ICC and find that states with high degrees of civil liberties are more likely to do so. They also find that states with a greater dependence on foreign aid and more foreign debt are more likely to sign the Rome Statute which they suggest is evidence that economically strong states with a commitment to the ICC have been able to
use economic leverage to induce smaller states to sign the treaty. Kelley (2005) examines the types of countries that have entered into “Article 98” agreements with the United States. She finds that states that are relatively large, powerful, democratic, and that have strong norms of rule of law are less likely to sign such agreements.

Rudolph (2001) was published before the Rome Statute entered into force so his discussion is understandably highly qualified. Applying a constructivist norms-based approach to case studies of three ad hoc international criminal tribunals and the negotiations over the Rome Statute he infers: “The evidence suggests that expanding liberal norms of state conduct and protecting human rights certainly explain the existence of tribunals in locales with little strategic or material importance.”

In short research on the ICC has been dominated by empirical studies, both qualitative and quantitative, on the negotiations of the Rome Statute and of the types of countries that have signed it and most closely adhered to the spirit of it by refraining from Article 98 agreements. Given the youth of the regime it is understandable that none of this work offers much insight into the prospects for an effective ICC. What theoretical work there has been either assumes away the problem of enforcement by including an “international community” that intervenes to impose the Court’s indictments or addresses war crimes treaties, which (as Morrow has convincingly argued) plausibly possess the self-enforcing features common in rational choice models of international cooperation. Thus this paper contributes to the literature on the ICC by offering a model of the regime

---

14 In these agreements states pledge to refrain from surrendering each other’s citizens to the ICC. The agreements are allowed under Article 98 of the Rome Statute (thus the name). The American Service Members’ Protection Act of 2001 requires such agreements before member states of the ICC are eligible for US military aid, including US participation in peacekeeping missions in such countries. As of this writing the United States has concluded 92 of these agreements.
that explicitly addresses its inherent lack of enforcement in a rational choice framework. I now turn to that model.

**Informal presentation of the model and results**

In this section I will first provide a brief overview of the logic of the argument. I will then turn to a fuller discussion of the players and their preferences. Following that I will discuss the timing of the model and then present the results in somewhat more detail and offer some intuition and comparative statics for the results. The model is presented more formally in the appendix.

**Logic of the Argument**

The logic of the argument can be summarized as follows:\(^\text{15}\):

1. Dictators’ payoffs (in expected present value) of committing atrocities are increased if they can expect asylum when they have low odds of retaining power and they face domestic retribution that they could avoid by asylum.
2. Foreign countries prefer not to have atrocity-committing dictators in power in other countries and so cannot credibly commit to deny asylum to an asylum-seeking human rights abuser.
3. Voluntary surrender to the ICC is worse for a human-rights-abusing dictator than asylum, but may in some cases be better than attempting to stay in power because of the risk of retribution from his domestic political rivals (which I assume is worse than the punishment imposed by the ICC) if he is deposed.

\(^{15}\) I thank an anonymous reviewer for this particularly cogent summary of my argument.
4. The existence of the ICC allows foreign countries to credibly reject requests for asylum in these cases, because the dictator will voluntarily step down peacefully and surrender to the ICC.

5. Thus the existence of the ICC lowers the expected present value of committing atrocities in the first place, since there is a chance a dictator will find himself in the position of preferring to surrender to the ICC, when without ICC he would have obtained asylum in a foreign land in such a case. This is the *ex ante* deterrence effect.

6. Since states that are potential safe havens for dictators can ignore the ICC without penalty, the regime does not reduce their incentive to grant asylum except in cases where the dictator prefers surrendering it than staying in power, and since surrendering to the ICC is completely voluntary, the option does not increase the dictator’s incentive to hold on to power to avoid ICC prosecution. Thus the regime does not increase dictators’ incentive to cling to power to avoid prosecution.

In a nutshell, the existence of the regime modeled here allows states to credibly refuse asylum in some cases when, without it, they would prefer to offer it and it does so without prolonging the reign of atrocity-committing dictators.

*Players and their preferences*

There are 3 players in the model at any given time: the leader \(L_i, i \in \{1, 2, 3, \ldots\}\), the foreign state \(S\)\(^{16}\) and nature \(N\). Nature chooses new leaders as the old ones are deposed and also randomly chooses the probability that a leader who has committed

---

\(^{16}\) The model generalizes straightforwardly to multiple foreign states.
atrocities is deposed. I call this randomly drawn probability, \( a_t \), (the \( t \) subscript denotes the relevant period). For simplicity the probability that a leader survives the given period if he does not commit atrocities is fixed throughout his reign. There is only one leader at any given iteration of the stage game, but new leaders will be selected as old ones are deposed or step down. If leader \( L_i \) is deposed in the previous period Nature chooses the next leader, \( L_{i+1} \), at the start of the subsequent period. Leaders are indexed in the order they are chosen. The identity of the foreign state (S) does not change at any point in the game. The leader receives a payoff of one each period he is in power and a payoff of zero each period he is not in power. If a leader is deposed having committed atrocities he suffers a one-time cost of \(-r\). If he is deposed not having committed atrocities he receives a one-time payoff \( d \in (-r, 1) \). In other words, I assume that the costs after being deposed are strictly worse for a leader who has committed atrocities than for a leader who has never committed atrocities. Leaders care only about staying in power, which they strictly prefer to not being in power, and the punishment they suffer if deposed. The foreign state is assumed to strictly prefer a leader who has never committed atrocities. An important feature of the model is that there is no foreign intervention to depose leaders. Leaders are deposed only by domestic political rivals.

**Timing of the model**

The extensive form of the game is presented in Figure 1. The timing in the model is as follows. A new leader (\( L_i \)) is chosen by Nature (N). Nature then chooses the probability that that leader \( L_i \) will be deposed if he commits atrocities and knowing that probability the leader chooses whether or not to commit atrocities. If he does not commit atrocities
he can request asylum abroad or he can take his chances on surviving in office. If he survives the game repeats and a new survival probability is drawn. If however the leader commits atrocities he faces a lottery over his political survival. If he survives until the following period he enters the post-atrocities subgame. In each iteration of that subgame a new $a_t$ is drawn. Knowing that probability, the leader decides whether to request asylum or to take his chances on the lottery of his political survival. If the leader requests asylum, the foreign state, S, has the choice of granting it or not granting it. If the foreign state refuses to grant asylum the leader faces a lottery over his survival as described above. Committing atrocities has an irreversible effect in that once the leader commits atrocities his probability of survival is given by the random variable $a_t$ in each period thereafter and if he is deposed he suffers a retribution punishment regardless of whether he commits atrocities in that period or not.

FIGURE 1 HERE

In short the leader’s decision of whether or not to commit atrocities is a trade-off between the increase in his probability of surviving politically that committing atrocities may afford versus the extra costs of the more severe punishment that he will suffer if he is deposed. The possibility of asylum abroad mitigates this latter concern because the leader knows that if his probability of surviving in office becomes too low he has the option of requesting asylum and (assuming that request is honored) going into exile with impunity.

Results
In this section I will describe the results in somewhat more detail and provide the intuition for them. Readers should turn to the appendix for the complete explication of the model. First I will describe the outcome in the model before the international prosecutor is created. I will then turn to the outcomes in the model after the creation of the international prosecutor to show the effects of the institution on the leaders’ incentive to commit atrocities and to show that leaders have no increased incentives to hold on to office longer than they would without the institution.

**Pre-institution**

In the pre-institution setting the foreign state, S, has a dominant strategy to accept the leader’s request for asylum. The foreign state is assumed to suffer disutility from there being a leader in power who has committed atrocities. Once a leader commits atrocities, the foreign state would prefer that that leader be removed from power because there is some chance (depending on Nature’s draw of $a_{t+1}$) that the subsequent leader will not commit atrocities. Thus by accepting the current leader’s request for asylum the foreign state will get rid of an atrocious dictator for sure. However if the foreign state rejects the leader’s request, it is rid of the current leader only with probability $1-a_t$.

The leader’s equilibrium behavior is equally straightforward. The leader will commit atrocities if doing so increases his probability of remaining in office by a sufficient degree. Leaders will not commit atrocities for a small increase in their probability of remaining in office because once a leader commits atrocities he “crosses the Rubicon” in the sense that he will have to pay the retribution costs for his past atrocities whether he continues to commit atrocities or not. Once a leader commits atrocities he will play the lottery over his political survival in subsequent periods as long
as Nature draws a high enough probability that he will survive. If, however, Nature draws too low a probability the leader will choose to request asylum. In such a case his expected present value from exile with asylum is higher than from attempting to hold on to political office where there is a chance he may be deposed and have to pay the retribution cost. In fact there is a cutoff level of \( a_t \), which I call \( \alpha \), for which the leader is indifferent between requesting asylum and playing the lottery over his survival. For values of \( a_t \) lower than \( \alpha \) the leader strictly prefers asylum in a foreign land to attempting to stay in office, and as I already discussed the foreign state has a dominant strategy to accept such a request. Therefore in such a situation the leader will request asylum and it will be granted.

Leaders commit atrocities in this model when doing so increases their chances of surviving politically and when the penalty they suffer if they are deposed having committed atrocities \((-r)\) is small relative to the payoff of being deposed not having committed atrocities, \((d)\). The italicized qualification is important because casual empiricism would suggest that the penalty from being deposed in states most prone to having atrocity-committing leaders are quite high, and may include assassination or execution possibly preceded by torture. The point is that the penalty of being deposed not having committed atrocities is also quite high in these states—it may also involve torture and death. By comparison, being deposed in states not prone to atrocities is not that unpleasant. Financially lucrative book deals and speaking engagements are the norm; assassination and detention are quite rare.

Figure 2 illustrates this difference between two hypothetical states. In country 1 the norm for political succession is assassination. In country 2 the norm for political
succession is well-ordered transfers of power with the loser going into august retirement. In other words \( d \) is considerably higher in country 2. Assume that the retribution costs of being deposed having committed atrocities is the same in both states. As illustrated in Figure 2 the difference in the penalty of being deposed after committing atrocities compared to the penalty of being deposed after not committing atrocities is quite small in country 1. In country 2, by contrast, the difference between these two penalties is quite large. In both cases the penalty to the leader from committing atrocities is identical but \( d \) is lower in country 1 because of the two different norms of succession. As such the leader would have more incentive to commit atrocities in country 1 than in country 2.

FIGURE 2 HERE

Figure 3 illustrates the effect of increasing the retribution costs. Higher retribution costs reduce leaders’ expected present values of committing atrocities. As such leaders’ expected present values of committing such abuses will be lower for any given value of \( a_t \) and leaders prefer asylum to attempting to stay in power for relatively higher values of \( a_t \). Leaders will step down with higher probabilities of surviving in office and they are less likely to commit atrocities in the first place (that is Nature must draw higher values of \( a_t \) before they are willing to commit atrocities).

FIGURE 3 HERE

Post-institution

The institution modeled in this paper creates a nonstrategic actor, the prosecutor, who automatically indicts any leader who commits atrocities.\(^{17}\) If the foreign state

\(^{17}\) Obviously this assumption is not general. Allowing the prosecutor to choose which atrocity committers to indict and which to ignore is a possible extension of the current model.
becomes a member of the regime it pledges not to offer asylum to any leader under indictment. This latter assumption is not a necessary feature of the model because, as we shall see, the foreign state’s behavior is the same regardless of whether it is a member of the regime or not, however the assumption does allow a discussion the foreign state’s compliance with the regime. The introduction of the prosecutor changes the extensive form of the game (which is illustrated in Figure 4) in one respect—the leader can now surrender to the prosecutor. If the leader surrenders to the prosecutor he will suffer a punishment with an expected present value of \(-m \in (-r, 0)\).

The expected present value of asylum is normalized to zero, however the leader may still be willing to surrender to the prosecutor if the foreign state turns down his request from asylum because by doing so he avoids retribution from his domestic political rivals. Indeed there is a cutoff value of \(a_i\), which I call \(\alpha\), for which the leader is indifferent between surrendering to the prosecutor and attempting to remain in power. I also define another cutoff value of \(a_t\), \(\alpha_t\), which is the value of \(a_t\) for which the leader is indifferent between stepping down and not doing so if he receives asylum from the foreign state. Obviously \(\alpha_t < \alpha\), that is the probability of survival for which the leader would be willing to step down and be punished must be lower than the probability for which he is willing to step down with impunity.

An important feature of the game is that it does not give the prosecutor any means of enforcing its indictment on the foreign state, \(S\). The foreign state does not pay any

---

18 Admittedly there may be some uncertainty as to whether the leader will actually be convicted after he turns himself in and there may also be some ambiguity a priori about what the leader’s punishment will actually be in the end. I abstract from those complications here. I also do not consider the possibility that the prosecutor may negotiate a “plea bargain” with the leader in which the leader surrenders and receives a payoff with and expected present value less than zero but still greater than \(-m\). As discussed below there are circumstances where allowing such a possibility will enhance the deterrent effect of the institution.
costs for granting the leader asylum. Obviously this is an extreme assumption however it does have some basis in fact. The Rome Statute contains several loopholes that allow members fairly wide latitude as to whether they must comply with the Court’s requests, and the ICC has no explicit enforcement provisions. The Court must rely on the cooperation of its members (Henquet 1999). Most importantly many states have yet to sign the Rome Statute so they are in no way obligated to comply with its indictments.

Without any higher authority to enforce the foreign state’s compliance with the new institution all commitments that the foreign state has made by joining the new institution must be self-enforcing—the foreign state cannot obtain higher utility from accepting the leader’s request for asylum than it does from rejecting it. Still the addition of the prosecutor alters the foreign state’s incentives in important ways. First, while the foreign state still has a strictly dominant strategy to accept the leader’s request for asylum when \( a_i \geq \alpha \), the foreign state no longer does when \( a_i < \alpha \) because when the leader’s probability of survival is that low the leader will actually surrender to the prosecutor if the foreign state rejects his request. Therefore when the leader’s probability of survival is that low the foreign state is rid of the leader whether it accepts the request or not. Note that if the foreign state complies with the prosecutor’s indictment it is not doing so to avoid punishment from the regime’s other members because there are no costs to noncompliance in this model. Instead the foreign state complies simply because it is costless to do so. In fact, since it is costless to comply with the prosecutor’s indictment, when the leader’s probability of survival is that low the foreign state may as well do so regardless of whether it is a member of the regime or not.
The effect of this change on the foreign state’s behavior may in turn affect the leader’s behavior. I say “may” because the effect depends on the leader’s beliefs about what the foreign state will do if a value of \( a_t < \alpha \) is actually drawn. If the leader does not believe that the foreign state will reject his request in those circumstances then the regime will have no effect on his behavior. However if the leader believes that the foreign state will comply with the regime when it is costless to do so, his expected present value of committing atrocities will be reduced. Before the creation of the regime the leader’s expected present value if Nature drew \( a_t < \alpha \) was zero. Under such beliefs, the leader’s expected present value if Nature draws \( a_t \) in that range is lower as a result of the regime, (namely \( -m < 0 \)). The possibility that the leader may be forced to accept punishment for atrocities if Nature draws a sufficiently low value of \( a_t \) reduces his expected present value of committing atrocities, and therefore makes it less likely that he will do so in the first place. Specifically, the value of \( a_t \) that Nature must draw to make it worthwhile for the leader to commit atrocities is higher and therefore less likely to be drawn.

Figure 5 illustrates what might be called the “total” effect of the new regime on the leader’s expected present value of committing atrocities when the leader believes the foreign state will reject his request for asylum when it is costless to do so. There is a direct effect and an indirect effect. The direct effect is that for values of \( a_t < \alpha \) the leader will surrender to the prosecutor and receive a punishment of with an expected present value of \( -m \) rather than zero as it did in prior to creation of the institution. The indirect effect is the reduction in the leader’s expected present value of being in the post-indictment subgame, as compared to the post-atrocities subgame before the creation of the regime. This indirect effect in turn causes an increase in the value of \( a_t \) for which the
leader will request asylum from $\alpha$ to $\alpha$ and a downward shift in the leader’s expected present values for draws of $a_i > \alpha$. The total change in the leader’s expected present value as a result of the direct and indirect effects can be found by integrating the leader’s expected present value in each of the two scenarios (pre-and post-regime) over all possible values of $a_i$. As shown in Figure 5 the total expected present value of committing atrocities is smaller as a result of the institution by the amount of the gray shaded area, and therefore the leader is less likely to do so.

FIGURE 5 HERE

Discussion

A few points are worth highlighting to clarify the model and its results. First the model points to a problem with the criticism, mentioned in the introduction, that the ICC will fail to deter because it has no enforcement power. That criticism assumes that foreign military or economic intervention is the only way human rights abusers may be brought to justice, but of course that is not the case. As Akhavam (2001, 7) put it:

…political … fortunes change, and the seemingly invincible leaders of today become the fugitives of tomorrow. Whether their downfall comes through political overthrow or military defeat the vigilance of international criminal justice will ensure that their crimes do not fall into oblivion undermining the prospect of an easy escape.

The mechanism modeled this paper is akin to the one that Akhavam describes. The ICC has no means of sanctioning either atrocity-committing leaders or the states that offer them sanctuary. Despite the extreme weakness of enforcement powers for both the ICC and foreign states the institution may, in the right circumstances, still have some deterrent effect because leaders may recognize that a time may come when their best
option is surrendering to the ICC. Under those circumstances they may have little hope of obtaining asylum because states have no incentive to offer it. Indeed the record of the International Criminal Tribunal on the Former Yugoslavia provides some examples of this process at work. As of 2003, of the 69 indicted war criminals in the custody of the ICTY or on provisional release, a plurality (27) surrendered voluntarily (International Criminal Tribunal on the Former Yugoslavia 2003).

A second implication of the model is that the number of periods in which atrocities are committed will not increase as a result of the creation of this institution. This result is in stark contrast to the concerns mentioned in the introduction that fear of prosecution will induce some atrocity-committing leaders to hold on to office longer than they otherwise would have. The institution potentially reduces leaders’ expected present values of committing atrocities which in turn means that leaders are less likely to do so in the first place (Nature will have to draw a higher $a_i$, higher for them to choose to do so). In doing so, however it does not prolong the reign of leaders who have committed atrocities. As I show in the appendix the cutoff point of $a_i$ (leader’s probability of survival) below which the leader will step-down cannot be lower with the institution than without it and in general it will be larger. In other words the institution modeled here will not increase the number of periods of atrocities and may reduce them.

The motivation for this criticism is the realization that leaders will be less willing to step down if they are going to be punished for their crimes than if they will not be punished. Those incentives are illustrated by a comparison of $\alpha$, the maximum value of $a_i$ for which the leader would still be willing to surrender to the prosecutor, and $\alpha$, the

---

19 Of the remainder 24 were detained by international forces, 11 were arrested by Serbia and seven were turned over by other states. Eighteen remain at large.
maximum value of \( a_t \) for which the leader would be willing to step down if granted immunity. The former is unambiguously lower than the latter; meaning that range of \( a_t \) for which the leader would be willing to step down and be punished is smaller and therefore less likely to be drawn from. The point is that surrendering to the prosecutor is not the only way that a leader who has committed atrocities might leave office even after the creation of the new regime. Foreign states may still agree to grant asylum if it is the only way to get rid of an atrocious dictator.

This criticism appears to be based on one of two assumptions neither of which is particularly plausible. The first is that states will comply with the Court’s indictments even when it is not in their interests to do so. Under this assertion, mutually beneficial asylum agreements between beleaguered dictators and foreign states will be foregone because they would be contrary to the requirements of the ICC. In this model states only comply with the institution’s indictments when it is costless to do so, that is when doing so offers them at least as high an expected present value as not doing so. On occasion (when \( a_t < \alpha \)) states may refuse to offer asylum but only because they know that the leader will surrender to the prosecutor anyway. The scenario in which the reign of atrocity-committing leaders is prolonged cannot occur in this model because in such a case states would simply not comply with the indictment and would offer the dictator asylum.

Alternatively this criticism could be based on the assumption that the ICC has some enforcement power—that it can impose costs on states that do not comply with its indictments (or somehow makes its easier for other states to do so). If the ICC had such power then the kind of asylum agreements between a leader and a foreign state modeled
here would become impossible because they would no longer be credible.\textsuperscript{20} The leader would know that the foreign state would not comply with its asylum agreement and therefore would not request one in the first place. The reign of leaders would indeed be prolonged as such leaders would cling to power even when their probabilities of surviving in office fell below the cutoff point.

The assumption that the ICC has such enforcement power is problematic. Indeed the ICC’s lack of enforcement power is one of the central criticisms of the regime as discussed above. More importantly, membership in the ICC while fairly broad is far from universal. States that are not members of the ICC have not committed themselves to comply with the prosecutor’s indictments and have no legal responsibility do so. Of course individual states and the UN Security Council can always impose sanctions or other punishments on a country for harboring atrocity committers, but those entities had that option before the creation of the ICC and the creation of the ICC does make it any easier or less costly to avail themselves of that option. The point is not that commitment problems do not exist but that there is nothing to indicate that the ICC will make them more severe.\textsuperscript{21}

Third the model has implications for the criticism that ICC will create a system of “Cadillac justice” in which leaders are tried fairly by the ICC and low-level perpetrators

\textsuperscript{20} The timing of the costs is important. If the foreign state were to pay one-time costs for making such an agreement then agreements would still be possible if those costs were not too high. Such costs would be paid once at the time the leader went into asylum and therefore would be irrelevant to any future decision making. The scenario discussed above is a case where the ICC is somehow able to impose costs after that dictator has gone into asylum. the foreign state cannot credibly accept a request for asylum in those circumstances because it has a dominant strategy to turn the leader over to avoid those costs.

\textsuperscript{21} Suppose however, that the creation of the ICC does, despite all of the indications to the contrary, make punishing states that harbor atrocity committers easier—including states that did not even sign the Rome Statute. The effect of the ICC is ambiguous in such a case. The deterrent effect of the ICC would make atrocities less likely. That effect would be offset by the effect of prolonging the reign of atrocity committers. Which of these effects will dominate is unclear.
are subjected to harsh and unfair trials at home. The point of this criticism, which was raised by Roth (2002) among others, is mainly normative—such a result is unfair on its face. However, such a system would also be detrimental to deterring atrocities. If the ICC were able to compel domestic political agents to turn over deposed atrocities perpetrators to the ICC the deterrent effects described in the model would disappear. The crux of the issue is the severity of punishments. The ICC is barred from handing down a death sentence—something many atrocity-committing leaders would face from domestic political rivals. If a leader knew that his domestic opposition would be forced to turn him over to the ICC if he were deposed, the effect on deterring atrocities and on the leader’s incentives to step down would be the same as reducing the value of \( r \), the comparative statics of which I explored in Figure 3, and the ICC could actually lead to more atrocities according to this analysis.

Fourth, the institution in this model does not really do anything. The prosecutor merely waits for human-rights-abusing dictators to surrender to him. As such this analysis may beg the question: *why did the world need an ICC at all?* It is true that what the institution modeled here accomplishes could be accomplished simply by states pledging not to harbor human rights abusers unless they pay some agreed-upon punishment (recognizing that will states will fail to comply with those pledges when \( a_r \geq \alpha \)). However such a treaty would still leave the practical problem of whom the leader would surrender to and who would extract the “payment” of \(-m\). These tasks could have been delegated to a state member of the agreement, but the creation of a

Prosecutor’s office seems as good a choice as any.

---

22 This criticism was raised about the Rwandan tribunal in particular. Because of this and other problems with the tribunal the Rwandan government actually voted against the establishment of the tribunal (Akhavam 1996, Henquet 1999)
Fifth, the model has implications for a potential problem that participation in the regime is far from universal. *Will non-members of the ICC make the regime irrelevant by continuing to offer atrocity-committing dictators asylum?* This model suggests that there is no reason for them to do so as long as those leaders’ probability of being deposed is sufficiently high. The effect of the regime in this model is not the creation of some enforcement mechanism that induces the regime’s members to comply with their treaty obligations. Rather the effect of the regime is the creation of an international agent to whom atrocity-committing dictators can surrender and who can hand out punishments to those dictators. The regime fulfills a jurisdictional gap that previously existed in international criminal law. A country that is not a member of the regime has no less incentive to refrain from offering a leader asylum than does a member of the regime.

A sixth point is that it may appear that *the problem of atrocities is too easily solved in this model.* If the international community wants to eliminate atrocities all it has to do is set the punishment handed out be the prosecutor \((m)\) at a very high level, which will produce a large reduction in the expected present value of being in the post-atrocities subgame and atrocities will be deterred. However that interpretation of the role of \(m\) in the model would be incorrect; \(m\) is the cost of punishment to the outgoing leader. The prosecutor does not really have that much control over that parameter. The prosecutor has limitations on how harsh a punishment it can set, and, more importantly, the prosecutor cannot alter the leader’s valuations of certain outcomes. Regarding the ICC in specific, life in prison is the worst punishment that it can hand down, and it is likely that that punishment is simply not that bad compared to what many leaders could expect to receive
from his domestic opposition. I might also mention that the model implies that making \( m \) very large would lead to more offers of asylum.\(^{23}\)

Finally, the model suggests a possible role for *plea bargaining*. Prior to the creation of the institution in this model, leaders who draw \( a_i \) less than \( \alpha \) enter into an asylum agreement and receive a payoff with a present value of zero. Such leaders are receiving a surplus from asylum. They would have actually been willing to “pay” in some sense for asylum but are not required to do so. The lower their value of \( a_i \) is the greater their surplus from an asylum agreement. Following the creation of the institution leaders with sufficiently low values of \( a_i \) are required to “pay” through the imposition of punishment cost \( m \). It is the prospect of this “payment” that reduces the expected present value of committing atrocities and thereby deters some marginal human rights abusers. Granting the prosecutor some ability to make “plea bargains” could actually enhance the deterrent effect of the institution in this model. If the prosecutor were allowed to make such agreements he or she could tailor the punishment to each leader’s value of \( a_i \) thereby obtaining some punishment from every leader who draws a sufficiently low \( a_i \) that his expected present value of holding office is less than zero. Via these plea-bargained punishments leaders would be required to “pay” so that they obtained no surplus from the plea agreement. The prosecutor would be playing a role akin to that of a price-discriminating monopolist. All leaders who drew an \( a_i \) that offered an expected present value of staying in power less than zero would be punished to the maximum extent possible while still inducing them to step down, so that the expected present value of

\(^{23}\) I thank Lisa Martin for pointing this implication of the model out to me.
committing atrocities in the first place would be even lower, and the deterrent effect would be greater.\textsuperscript{24}

**Conclusion**

The model in this paper considers the effects of an ICC-like institution on the strategic interaction between an atrocity-committing leader and a foreign state that would like to facilitate the transition to a more palatable leader. Prior to the creation of the institution foreign states strictly prefer accepting any request for asylum from an atrocity-committing leader, and following the creation of the regime they no longer do in certain circumstances. If the leader is so likely to lose power that he would willingly accept punishment by the regime in return for protection from domestic political rivals the foreign state has no reason to offer the leader asylum because whether it does or not the atrocity-committing leader will step down.

The analysis answers certain criticisms of the ICC in the literature. First, it provides some conditions under which the ICC may deter atrocities at the margin despite its lack of enforcement power. Second the institution does not prolong the reign of atrocious leaders as some commentators have worried. Leaders who would step down if granted asylum but who would not step down if forced to be punished for their crimes are granted asylum in this model. The only leaders who are actually punished in this model are those who willingly accept punishment in order to avoid retribution from domestic political rivals. As such their reigns are not prolonged.

\textsuperscript{24} Unfortunately the plea bargain would have to be tailored to the leader’s probability of political survival and not to the severity of his crimes.
In addition to these policy considerations, the paper makes a theoretical contribution to our understanding of the necessity of enforcement for the effectiveness of international regimes. The model provides an example of a regime that is efficacious despite the fact that none of the usual trigger strategies are employed to insure compliance. The institution deters atrocities under specified conditions even though foreign states never intervene to depose atrocity-committing dictators. Most interestingly, foreign states occasionally comply with the indictments of the prosecutor even though there is no punishment for noncompliance. Furthermore the model does not rely on norms arguments or learning, or pressure from domestic constituencies, nor does it assume away the enforcement problem by asserting that it is a management problem.

A natural extension of the argument would be to better model the domestic political milieu—in particular the process by which atrocities increase leaders’ ability to hold on to power. Recent work by Bueno de Mesquita et. al. (2004, chapter 8) has addressed this topic although, obviously, without reference to the ICC. Incorporating the features of their model into the one in this paper may provide more refined predictions about the likely effects of the ICC, especially with regard to the effect of domestic political institutions.

The main remaining question, of course, is how substantively important the potential deterrent effect discussed in this paper will be. As Cameron (2004, 92-93) has pointed out there is little evidence that making punishments harsher lowers crime rates (i.e. that deterrence is effective) even within relatively well-functioning domestic criminal justice systems. What hope is there then of a substantively important impact from the ICC? Ultimately that question is an empirical one and can only be answered
prospectively by parameterizing the model from real-world cases, which is the subject of ongoing empirical work. In the meantime I hope that this paper will inform the continuing legal debate such that the ICC is no longer characterized as a either a cure-all or a catastrophe-in-waiting. If this model is any guide, the ICC will be neither of those things. It will not be an institution that does absolute harm and it may do some good. The extent of that good remains to be determined.

Appendix

Basic Set-up of the Model

Players are described in main text. I assume the leader (L_i) and the foreign state (S) have identical one-period discount rates of \( \delta \). L_i derives per period utility of one from holding office and per period utility of zero from not being in office. In the absence of atrocities by L_i, S’s per period payoff is 1, unless L_i has committed atrocities in which case S’s per period utility is \( u < 1 \). S does not care whether L_i holds on to power or is deposed, only whether atrocities have been committed or not.

If L_i commits atrocities and is subsequently removed from office he suffers a one-time retribution cost of \( -r \leq 0 \), leaves office and the game ends for him. If L_i is deposed not having committed atrocities he receives a one-time payoff of \( d \in (-r, 1) \) and the game ends for him. If L_i goes into exile the game ends for him, but he avoids the retribution cost, \( -r \), and does not receive the one-time payoff from being deposed, \( d \). Once L_i leaves office, either by being deposed or stepping down, the game ends for him, he has no prospect of returning to power and he receives a payoff of zero each period from that point forward.
Nature (N) chooses the probability that $L_i$ will keep power in a given period if he commits atrocities. I call this probability $a_t$ where $t$ indexes the current period. This probability is chosen from a known distribution $F(a)$, with support $[0,1]$ and associated density $f(a)$. The value of $a_t$ is common knowledge among all players. The probability that the leader holds on to power if he does not commit atrocities is fixed at $q$.

The stage game, which corresponds to the complete reign of a given leader, $L_i$, and as such may cover several periods, is illustrated in Figure 1. In Figure 1, the subgame for a given leader after the first period in which he commits atrocities, which I refer to as “the post-atrocities subgame” is shaded gray. The remainder of the game in Figure 1 illustrates “the pre-atrocities subgame” which is the subgame up to and including the first period (if any) that $L_i$ commits atrocities.

Play proceeds over discrete time periods $t \in \{1, 2, 3 \ldots\}$. In the pre-atrocities subgame the following moves occur sequentially in each time period. N chooses the probability that $L_i$ will be deposed if he commits atrocities. $L_i$ then decides whether or not to commit atrocities. If $L_i$ decides not to commit atrocities he can request asylum. the foreign state then decides whether to accept the request or not. If the foreign state accepts $L_i$ steps down and the game ends for him, returning play to the beginning of the pre-atrocities subgame. If the foreign state rejects the request, $L_i$ faces a lottery in which he keeps power with probability $q$ and is deposed with probability $1-q$. If instead $L_i$ decides to commit atrocities he faces a lottery over his political survival.\(^{25}\) If $L_i$ does not survive

\(^{25}\) A straightforward alteration of the game would be to allow $L_i$ to request exile after committing atrocities in the pre-atrocities subgame. Doing so would be superfluous, however because if $L_i$ is planning to step down and go into exile in the first period of his reign there is no reason to commit atrocities before doing so.
the lottery play returns to the beginning of the pre-atrocities subgame. If \( L_i \) survives that lottery play proceeds to the post-atrocities subgame.

Sequence of play is similar in the post-atrocities subgame. \( N \) chooses the probability that \( L_i \) will be deposed. \( L_i \) then decides whether to request asylum or not. If he does not request asylum he faces a lottery over his political survival—he survives with probability \( a_t \) and is deposed with probability \( 1-a_t \). If \( L_i \) does request asylum the foreign state then decides whether to accept the request or not. If the foreign state accepts \( L_i \) steps down the game ends for him, and play returns to the beginning of the pre-atrocities subgame. If the foreign state chooses “reject,” \( L_i \) faces a lottery over his political survival as described above.

I employ Markov perfect equilibrium (MPE) as a solution concept (Fudenberg and Tirole 1991, chapter 13). MPE is appropriate because there is a stochastic component to the game—the realization of future draws on the dictator’s probability of political survival—but no asymmetric information about the realizations of that variable once drawn. I use MPE purposefully to illustrate the efficacy of the regime without the use of trigger strategies. Trigger strategies are not possible in a MPE because players use only payoff relevant information, not information about previous moves.

I present the results via a series of lemmas. Proofs of these statements follow from straightforward application of sequential rationality and are therefore omitted.

**Solving the post-atrocities subgame**

The only slight wrinkle in solving the post-atrocities subgame is \( S \)’s incentives at the last node of the game, which can be summarized as follows:
LEMMA 1: If \( a_t > 0 \) and if there exists \( a_{t+1} \in (0,1] \) such that \( L_{t+1} \) will refrain from committing atrocities \( S \) has a strictly dominant strategy to accept \( L_i \)'s request for asylum.

The intuition for the result is that if \( S \) accepts \( L_i \)'s request for asylum, then, with some probability \( p > 0 \), which I describe in greater detail below, \( S \) will obtain a payoff of one in the subsequent period and obtain a payoff of \( u < 1 \) with probability \( 1-p \). If, however, \( S \) rejects \( L_i \)'s request \( S \) will obtain a payoff of one only with probability \( p(1-a_t) \).

The important point is that \( S \) has a strictly dominant strategy to accept \( L_i \)'s request for asylum prior to the creation of the regime. MPE is crucial to this result, because it assumes players respond only to payoff relevant information. As such by rejecting a request for asylum \( S \) does not change future leaders’ beliefs about what it will do in the future, and therefore does not deter future atrocities.

Turning to \( L_i \)'s decision at the second node of the post-atrocities subgame, \( L_i \)'s expected present value (EPV) of not requesting asylum at that node is:

\[
U_i = a_t \left(1 + r + \delta U\right) - r
\]

(1)

Unsubscripted \( U \) in equation (1) is a constant, the formula for which is:

\[
U = \int_{a}^{1} f(a) a(1 + r + \delta U) - r \ da = \frac{\hat{\alpha}(1 + r) - (1 - F(\alpha))r}{1 - \delta \hat{\alpha}}
\]

where \( \hat{\alpha} = \int_{a}^{1} f(a) a \ da \), and \( \alpha \) is defined below. \( U \) is \( L_i \)'s EPV from the next period (before \( a_{t+1} \) is known) forward of being in the post-atrocities subgame—that is \( U \) is \( L_i \)'s EPV at the first node of the post-atrocities subgame.
If the EPV in equation (1) is less than or equal to zero $L_i$ will request asylum, the EPV of which is normalized to zero, $S$ will accept, and $L_i$ will step down. Leaders will not step down voluntarily until a sufficiently low $a_i$ is drawn. If a leader is going to be willing to request asylum the following must be true:

$$a_i < \frac{r}{1 + r + \delta U} \equiv \alpha$$

(2)

**Lemma 2:** In any MPE of the game the outcomes of the post-atrocities subgame are:

1. If $a_t \geq \alpha$, $L_i$ does not request asylum, and survives with probability $a_t$.
2. If $a_t < \alpha$, $L_i$ requests asylum. $S$ accepts and $L_i$ steps down.

**Solving the pre-atrocities subgame**

I begin the discussion of this subgame with $S$’s decision to accept or reject a request for asylum. For ease of exposition define $S$’s EPV from node 1 forward as $\Omega$ and define $S$’s EPV from node 9 forward as $W$. $\Omega$ is $S$’s EPV when there is a new leader and $W$ is $S$’s EPV when it is already in the post-atrocities subgame. More formally:

$$W = \int_0^a f(a)(u + \delta \Omega)da + \int_1^a f(a)(a(u + \delta W) + (1 - a)(u + \delta \Omega))da = \frac{u + \delta(1 - \hat{a})\Omega}{1 - \delta \hat{a}}$$

(3)

and

$$\Omega = p(1 + \delta \Omega) + (1 - p)(u + \delta W) = \frac{p + (1 - p)(u + \delta W)}{1 - p\delta}$$

(4)

where $p$ is the (constant) probability that $N$ will draw a sufficiently low $a_i$ in the next period that $L_i$ will choose to refrain from committing atrocities in that period. Equations
(3) and (4) can be combined to solve for values of $W$ and $\Omega$ solely as functions of the constants $p$, $u$, $\bar{a}$ and $\delta$, and therefore $W$ and $\Omega$ are themselves constants.

If $L_i$ has yet to commit atrocities, $S$ has a weakly dominant strategy to accept a request for asylum. $S$’s payoff from accepting such a request is $1+\delta\Omega$. The highest payoff that $S$ can receive from rejecting a request for asylum in the pre-atrocities subgame is $1 + q(\delta\Omega) + (1-q)\delta\Omega = 1 + \delta\Omega$. This is the substantively uninteresting case where $L_i$ does not commit atrocities and steps down in the first period of his reign, so I simply stipulate that all such requests are accepted.

Turning to $L_i$’s decision at the third node of the pre-atrocities subgame, $L_i$ will commit atrocities if the EPV of doing so is higher than the EPV of not doing so. $L_i$’s EPV of not committing atrocities in the current period is:

$$Q = \text{Max} \left\{ \frac{q(1 + \delta(1 - p)U) + (1 - q)d}{1 - p\delta q}, 0 \right\}$$

$L_i$ can guarantee an EPV of zero by stepping down in the first period. If the first term in the curly brackets is negative (which could happen if $d$ or $U$ were negative and large enough in magnitude) then the highest payoff that $L_i$ can receive when he does not commit atrocities is zero, which he obtains by stepping down. In such a case $L_i$ will either not commit atrocities and step down, or if $a_t$ is high enough, $L_i$ will commit atrocities.

A necessary condition for $L_i$ to choose to commit atrocities is that the EPV in equation (1) is higher than the EPV in equation (5). This will be the case if:

$$a_t > \frac{r}{1 + r + \delta U} + \frac{1}{1 + r + \delta U} \cdot Q = \bar{\alpha}$$

As indicated this is only a necessary condition—obviously it must also be true that $a_t > q$ since $L_i$ would not choose to commit atrocities if it lowered his probability of keeping
office. $L_i$ will commit atrocities if $a_t > \bar{a}$. Notice that $\bar{a} = \underline{a}$ if the first term in the curly brackets in (5) is negative. In such a case $L_i$ would commit atrocities only if $a_t > \bar{a} = \underline{a}$, which by definition is the value of $a_t$ for which the EPV of committing atrocities is equal to zero. $L_i$ will never commit atrocities in the pre-atrocities subgame if the EPV of doing so is negative because he can guarantee a higher EPV (zero) by not committing atrocities and stepping down. Incidentally $p$ is the probability that $a_t \leq \bar{a}$.

**Lemma 3:** In any MPE of the game the outcomes of the pre-atrocities subgame are:

3. If $a_t > \bar{a}$, $L_i$ commits atrocities and proceeds to the post-atrocities subgame.

4. If $a_t \leq \bar{a}$, $L_i$ does not commit atrocities and:

   4a. If \[ \frac{q(1 + \delta(1 - p)U) + (1 - q)d}{1 - p\delta q} < 0 \] $L_i$ requests asylum. $S$ accepts and $L_i$ steps down, or

   4b. If \[ \frac{q(1 + \delta(1 - p)U) + (1 - q)d}{1 - p\delta q} \geq 0 \] $L_i$ does not request asylum and survives with probability $q$.

Outcome 4a is the substantively uninteresting case where $L_i$ steps down and goes into exile in the first period of his reign. This occurs when $L_i$’s EPV of remaining in office is negative whether he commits atrocities or not. I will ignore this outcome for the remainder of the paper.
**Adding an International Criminal Regime**

Now suppose an international institution that is staffed by a nonstrategic actor \( P \) (the prosecutor) is created. \( P \) automatically indicts any leader who commits atrocities. The pre-atrocities subgame is unaltered by this change, but the post-atrocities subgame is amended as indicated in Figure 5. I will call this subgame the “post-indictment subgame” to distinguish it from the post-atrocities subgame discussed above.

The post-indictment subgame begins just as the post-atrocities subgame did with a draw of \( a_i \) by N. \( L_i \) then chooses whether or not to request asylum. If \( L_i \) chooses not to request asylum he is then faced with a lottery over his political survival in the usual way. If, however, \( L_i \) requests asylum he presents \( S \) with a choice to accept or reject that request. If \( S \) accepts \( L_i \)’s request \( L_i \) steps down and the game ends for him as described previously. If \( S \) does not accept \( L_i \)’s request then \( L_i \) must choose whether to surrender to \( P \) and receive, with certainty, a one time penalty of \( -m \in (-r, 0) \). If \( L_i \) surrenders he steps down, returning play to the beginning of the pre-atrocities subgame. Alternatively \( L_i \) can choose not to surrender in which case he takes his chances at holding on to office according to the usual lottery.

**Solving the post-indictment subgame**

I will begin the discussion of equilibria in this new situation with \( L_i \)’s decision at the last node of this subgame where \( S \) has turned down his request for asylum. \( L_i \)’s EPV of not surrendering at that point is:

\[
\tilde{U}_i = a_i (1 + r + \delta \tilde{U}) - r
\]

(7)
Unsubscripted $\hat{U}$ in equation (7) is $L_i$’s EPV in the next period (before $a_{i+1}$ is known) of being in the post-atrocities subgame in this new situation following the creation of the regime. It is $L_i$’s EPV at the first node of the post-indictment subgame. $\hat{U}$ is a constant, the value of which depends on $L_i$’s beliefs about what $S$ will do in the future as discussed below.

Define:

$$\alpha = \frac{r}{1 + r + \delta U}$$

(8)

as the level of $a_i$ for which $L_i$ would be indifferent between stepping down and not stepping down if he could obtain asylum from $S$ in this new situation following the creation of the regime. It is the value of $a_i$ for which $L_i$’s EPV is zero in this new situation. The only difference between $\alpha$ and $\alpha_c$ is that $\alpha_c$ has $\hat{U}$ in the denominator rather than $U$. I also define a new cutoff point:

$$\alpha_c = \frac{r - m}{1 + r + \delta U}$$

This is the value of $a_i$ below which $L_i$ would be indifferent between surrendering to $P$ and attempting to stay in power. $L_i$’s EPV from surrendering to $P$ is $-m$, so $\alpha_c$ corresponds to the level of $a_i$ for which $L_i$’s EPV of attempting to stay in power was $-m$. If $a_i < \alpha_c$, $L_i$ would prefer surrendering to $P$ rather than attempt to stay in office if $S$ rejects his request for asylum. If $a_i \in [\alpha_c, \alpha]$ $L_i$ would not be willing to surrender to $P$ but would be willing to step down if he received asylum. If $a_i > \alpha_c$ $L_i$ would be unwilling to step down even if granted asylum by $S$. 
Turning to S’s decision of whether to accept L_i’s request, the forgoing discussion implies that while S still has a strictly dominant strategy to accept L_i’s request when \( a_i \geq \alpha \), it no longer has a strictly dominant strategy to accept L_i’s request for asylum when \( a_i < \alpha \), because L_i would be willing to surrender to P. S is indifferent between rejecting and accepting L_i’s request under those circumstances because, regardless of whether it accepts or not, L_i will step down. Therefore L_i can no longer be certain that S will accept his request for asylum. This is true regardless of whether S is actually a member of the agreement or not—that is, regardless of whether S has formally committed itself to comply with P’s indictments or not. To summarize, the same two outcomes that occurred in the post-atrocities subgame without the regime also occur, mutatis mutandis, in the post-indictment subgame, however, in addition there is a new outcome in the post-indictment subgame as in which L_i surrenders to P. These outcomes are formalized in the following lemma.

**Lemma 4:** In any MPE of the game the possible outcomes of the post-indictment subgame are:

1’. If \( a_i \geq \alpha \), L_i does not request asylum, and survives with probability \( a_i \).

2’. If \( \alpha \leq a_i < \alpha \), L_i requests asylum. S accepts, and L_i steps down.

3’. If \( a_i < \bar{\alpha} \), L_i requests asylum, and either.

3a’. S accepts, and L_i steps down, or

3b’. S refuses, and L_i surrenders to P.
The significance of $3b'$ is that, despite P’s lack of enforcement power, some atrocity committers may surrender to him. For low enough values of $a$, S will be indifferent between accepting L’s request for asylum and rejecting it because S knows that even if it rejects the request L will surrender to P and step down. This is in contrast to the outcome in the previous section where S clearly had a dominant strategy to accept L’s request.

Note that there is some ambiguity about the equilibrium when $a < \alpha$. S is indifferent between accepting and rejecting L’s request for asylum in those circumstances since either way L will step down. In a MPE, which I use, S has no reason to reject L’s request for asylum for the purpose of sending a signal to leaders L_{i+1} and beyond because in MPE players only make decisions based on payoff-relevant information. If I had assumed that S gets some very small utility increment, $\epsilon$, from seeing atrocity committers punished it would eliminate this ambiguity and make rejecting such a request a dominant strategy for values of $a$ in the neighborhood of $\alpha$ and below. However, I do not make that assumption because it seemed *ad hoc*.

*Solving the pre-atrocities subgame*

The structure of the pre-atrocities subgame is identical to that of the subgame prior to the creation of the regime. As in the previous case I will set aside the substantively uninteresting question of S’s decision to offer asylum in this subgame, if asked.\(^ {26} \) I turn instead to L’s incentives to commit atrocities. Differences in L’s decision

\(^ {26} \) This is the case in which L requests asylum in the first period of his reign. As before the foreign state has a weakly dominant strategy to accept such a request and so to dispense with this case I stipulate that one of them always does.
(if any) arise only because of changes in his EPV of committing atrocities now that there is the possibility that he could be punished for doing so if N draws a sufficiently low \( a_t \). That EPV will depend in turn on \( L_i \)’s beliefs about what S will do if \( a_t < z \). I will dispense with the obvious case first: if \( L_i \) believes that S will grant him asylum when \( a_t < z \) (in other words he believes that equilibrium 3b’ will not occur) then the effect of the new regime on the equilibria in the pre-atrocities subgame will be nil. \( L_i \)’s behavior will be identical to those described in the previous section.

If however \( L_i \) believes that outcome 3b’ will occur if N draws a low enough \( a_t \) then his behavior in the pre-atrocities subgame will be quite different because his EPV of committing atrocities will be lower and this in turn will affect the cutoff values of \( a_t \) for which he would be willing to commit atrocities. At the third node of the pre-atrocities subgame \( L_i \)’s EPV of choosing to commit atrocities is given by equation (7) where, given these beliefs:

\[
\hat{U} = -m \int_0^q f(a) \, da + \int_{a_t}^1 f(a) \left[ a(1 + r + \tilde{U}) - r \right] \, da = \frac{\hat{a}(1 + r) - (1 - F(\hat{a}))r - F(z)\tilde{m}}{1 - \delta \hat{a}}, \quad \text{and}
\]

\[
\hat{a} = \int_{a_t}^1 f(a) a \, da \text{ analogous to } \hat{a} \text{ in the previous section.}
\]

\( L_i \)’s EPV of not committing atrocities at that node is:

\[
\hat{Q} = \max \left\{ \frac{q(1 + \delta(1 - \tilde{p})\tilde{U}) + (1 - q)d}{1 - \tilde{p}\delta q}, 0 \right\} \tag{9}
\]

where \( \tilde{p} \) is the (constant) probability that N will draw a sufficiently low \( a_t \) in the next period that \( L_i \) will choose to refrain from committing atrocities. The new regime has no effect on \( L_i \)’s payoffs if he never commits atrocities because \( d \) and \( q \) remain fixed. The
only difference in \( L_i \)'s EPV of not committing atrocities compared to the earlier case is in the change from \( U \) to \( \tilde{U} \) and from \( p \) to \( \tilde{p} \).

As before, \( L_i \) will commit atrocities if the EPV of doing so exceeds the EPV of not doing so—that is if equation (7) is greater than equation (9). I define \( \tilde{\alpha} \) as the value of \( a_t \) that makes these two equations equal, analogous to \( \bar{\alpha} \) in the previous section. According to the same line of argument in the previous section, \( L_i \) will commit atrocities if \( \tilde{N} \) draws and \( a_t > \tilde{\alpha} \). I summarize these possible outcomes in Lemma 5.

**Lemma 5:** If \( L_i \) believes that equilibrium 3b' will occur when \( a_t < \underline{\alpha} \), the possible outcomes of the pre-atrocities subgame following the creation of the regime are:

4'. If \( a_t > \tilde{\alpha} \), \( L_i \) commits atrocities and proceeds to the post-indictment subgame.

5'. If \( a_t \leq \tilde{\alpha} \), \( L_i \) does not commit atrocities and:

5a'. If \[ \frac{q(1+\delta(1-\tilde{p})\tilde{U})+(1-q)d}{1-\tilde{p}\delta q} < 0 \] \( L_i \) requests asylum. \( S \) accepts and \( L_i \) steps down, or

5b'. If \[ \frac{q(1+\delta(1-\tilde{p})\tilde{U})+(1-q)d}{1-\tilde{p}\delta q} \geq 0 \] \( L_i \) does not request asylum and survives with probability \( q \).

Finally, the analysis shows that the reign of a leader who has committed atrocities will not be prolonged as a result of the regime modeled in this paper. Prior to the creation of the institution, \( L_i \) would leave power whenever \( a_t < \underline{\alpha} \) and after the creation of the institution \( L_i \) will leave office whenever \( a_t < \bar{\alpha} \). Since \( \tilde{U} \leq U \), a comparison of equations
(2) and (8) shows that $\alpha \geq \alpha$. 
References


“Charles Taylor: A wanted man.”

“Charles Taylor - preacher, warlord and president.”


Krasner Stephen D. 2003. Interview: Conversations with History; Institute of
International Studies, March 31, 2003, UC Berkeley. Transcript available at:
http://globetrotter.berkeley.edu/people3/Krasner/krasner-con0.html.

the Art on an Art of the State. *International Organization*. 40(4): 753-75.


Martin, Lisa L. and Beth A. Simmons. 1998. Theories and Empirical Studies of

Meernik, James. 2004. Reaching Inside the State: International Law and Superior

Meernik, James and Jamie Shairick. 2005. Promoting International Humanitarian Law:

and International Environmental Law (RECIEL)*. 2: 327-34.

________. 1994. Regime Design Matters: Intentional Oil Pollution and Treaty

of the International Criminal Court. Undergraduate Senior Honors Thesis in International
Relations. New York University. Copy on file with author.


“No one writes to the tyrants.” *Guardian Unlimited*. January 20, 2003. Available at:
http://www.guardian.co.uk/Iraq/Story/0,2763,878739,00.html

Price, Richard. 1998. Reversing the Gun Sights: Transnational Civil Society target Land

Raustiala Kal and Anne-Marie Slaughter. 2002. International Law International Relations
and Compliance. In Walter Carlsnaes, Thomas Risse and Beth A. Simmons, eds.


Figure 1: Extensive form of the stage game, pre-institution

1. N
Choose leader

2. N
draw $a_i$

3. $L_i$

4. $L_i$
Request asylum
Do not request asylum

8. N
$L_i$ survives 1, $u$
$L_i$ deposed $r$, $u$
return to node 1

9. N
draw $a_i$

10. $L_i$
Request asylum
Do not request asylum

11. S
reject
accept

12. N
$L_i$ survives 1, $u$
return to node 9
$L_i$ deposed $r$, $u$
return to node 1

13. N
$L_i$ survives 1, $u$
return to node 9
$L_i$ deposed $r$, $u$
return to node 1

Post-atrocities subgame

5. S

7. N
reject
accept

6. N

3. $L_i$
atrocities
no atrocities

14. $N$

15. $N$

16. N
Figure 2: Differences in scale arising from different norms of succession in two different states

Country 1

Country 2
Figure 3: Comparative statics on $a_t$ and $r$

$\alpha' = \frac{r'}{1+r' + \delta U'}$

$\alpha'' = \frac{r''}{1+r'' + \delta U''}$

Caption: Larger retribution costs expand the range of leaders’ probabilities of survival ($a_t$) for which there is a possible asylum agreement. Increasing retribution costs from $r'$ to $r''$ expands the range of $a_t$ for which there are possible asylum agreements from zero to $\alpha'$ to zero to $\alpha''$. 
Figure 4: Extensive Form of the stage game with institution.
Figure 5: Total effect of the regime when $L_i$ believes $S$ will comply when $a_t < \alpha$

$L_i$’s expected present value prior to the creation of the regime

$L_i$’s expected present value following the creation of the regime

S will not accept $L_i$’s request for asylum for values of $a_t$ in this range.

$L_i$ will not request exile for $a_t$ in this range.