

Editorial Manager(tm) for International

Organization

Manuscript Draft

Manuscript Number:

Title: The International Criminal Court: Benign, Malign or Futile? A Formal Analysis

Article Type: Research Article

Section/Category:

Keywords: International Criminal Court; human rights; international cooperation

Corresponding Author: Michael Gilligan,

Corresponding Author's Institution: New York University

First Author: Michael Gilligan

Order of Authors: Michael Gilligan

Manuscript Region of Origin:

Abstract: The International Criminal Court (ICC) is a very controversial institution. Some critics claim that the regime will be futile because it lacks power to apprehend the criminals it is meant to prosecute. Other critics say that the court will be absolutely harmful since it will induce atrocious leaders to hold on to power longer than they would if they could step down with immunity for past crimes. If these critics are right the ICC is a truly puzzling organization. Why would states create such a regime that is at best futile and at worst malign? The model in this paper provides a potential answer. I model the interaction between a leader in political trouble at home and a set of foreign states that have the option of offering that leader asylum. I then examine the effect of the creation of an ICC-like institution on that interaction. The model produces three main findings: 1) The leaders' reigns cannot be prolonged. 2) Although the institution has no enforcement power some leaders will be punished by it—those leaders who have such a high probability of being deposed that they would willingly surrender to the institution. In such a circumstance the foreign states have no incentive to offer the leader asylum. 3) The institution may deter some atrocities at the margin.

Acknowledgements

I thank Tanaz Moghadam, Sandy Gordon and Shanker Satyanath for helpful comments.

The International Criminal Court: Benign, Malign or Futile? A Formal Analysis

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 that was 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 not uncommon in international politics. Other examples include Jean-Claude "Baby Doc" Duvalier who lives comfortably in France. Duvalier is accused of torture and political assassinations and detentions in Haiti. 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

Cases like those described above are disturbing to human rights activists for at least two reasons. First there is the obvious normative/justice concern that perpetrators of human rights abuses are not being punished for their crimes. Second there is the more practical question of deterrence—if perpetrators of these crimes are able to get off scot-free what hope is there of deterring such crimes in the future? In part to help prevent the

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."

recurrence of cases like those described above a group of states prodded by a vast array of nongovernmental organizations created the International Criminal Court (ICC) in 1998 after a half century of on-again-off-again negotiations. Upon receiving its sixtieth ratification in the summer of 2002 the ICC has entered into force. In the summer of 2003 the Court's first Chief Prosecutor Luis Moreno-Ocampo was appointed and has begun proceedings in relation to alleged atrocities committed in the Ituri region of the Congo and Uganda.³

The Court has been the object of a great deal of criticism. I will review some of those criticisms in greater detail below. For purposes of introduction, however, I will mention two. The first criticism that this paper will address has to do with *enforcement*. Some legal scholars have argued that because the ICC is devoid of its own enforcement mechanisms it will fail to fulfill the goal of deterring atrocities. There is no international police force to capture war criminals.⁴ In many cases these criminals will be leaders who still hold state power. According to this argument, it will be impossible to bring these leaders to justice without armed intervention and the ICC has no such power. As long as such leaders hold on to power they will enjoy impunity for their crimes.

A second criticism of the court that this paper will address is that the ICC will give leaders who have committed atrocities incentive to hold on to power longer than they otherwise would if they could escape punishment. Persuading an atrocity-committing dictator to step down voluntarily is obviously less costly to the international community and the citizens who are ruled by the dictator than is forcing him out with

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 Indeed by one important measure the ICC has less authority than the *ad hoc* tribunals on Yugoslavia and Rwanda, which were created by the Security Council so their subpoenas are backed by the force of international law. The ICC does not have the same status (Henquet 1999).

civil unrest, strikes, coups, civil wars, bloodshed and so on. Offering outgoing dictators asylum is worthwhile according to this argument because it induces them to step down earlier than they otherwise would and without the chaos that would have ensued had a civil war or coup been necessary to depose them. Some commentators are concerned that if the ICC requires that such dictators pay for their crimes they will refuse to step down, their reigns will be prolonged and removing them will be more costly. According to its critics, then, the ICC will fail to deter leaders who have not yet committed atrocities but are considering doing so and it will prolong the reign of leaders who have already shown a propensity to abuse human rights.

If these arguments have any merit then the ICC is a truly puzzling institution. It is a criminal court with no police force to apprehend the criminals it is meant to prosecute. In order to accomplish its appointed task the ICC must rely on the cooperation of sovereign states—the very sovereign states who, as described in the opening paragraph, are prone to offer such dictators exile with impunity. Furthermore, participation in the ICC, while fairly broad, is far from universal. Even if the members’ compliance record with the ICC is perfect there are many states in the world that can legally continue to offer asylum to human rights abusers. The possibility that the ICC may actually prolong the reign of such leaders only deepens the puzzle. Given the spotty record of effectiveness of existing human rights treaties (Hathaway 2001) the critics can be excused for their skepticism, but if these arguments are correct then the creation of the ICC produces a puzzle: why would states create an organization that is at best futile—what one legal scholar (Goldsmith 2003) referred to as “the self-defeating international criminal court”—and at worst harmful?

The model in this paper provides a potential answer to this puzzle. It shows that despite the lack of an international police force to enforce its indictments, the ICC may still deter leaders from committing atrocities in some marginal cases. The international institution modeled in this paper serves as a coordination device in which states refrain from offering asylum to leaders who are so likely to be deposed that they would willingly surrender to the ICC rather suffer retribution from their domestic political rivals, which I assume is worse than the punishment handed out by the ICC. Since every leader knows that there is some chance (however remote) that he will find himself in that position the expected present value of committing atrocities is reduced and at the margin some leaders will refrain from committing them.

There is no outside intervention, either by states or international organizations, to depose leaders who commit atrocities in this model. More interestingly there is no punishment in this model for offering asylum to dictators that commit atrocities. Still states may refrain from doing so when the dictator's hold on power is sufficiently tenuous. States have no incentive to offer the leader asylum in such a case because they know he would be willing to step down even if it means surrendering to the ICC and getting punished for his past crimes. In such a case states are rid of the tyrant whether they offer him asylum or not. The behavior of states in this model is completely self-enforcing in the sense that states only comply with the institution's indictments when it is costless to do so.

The model in this paper also suggests that the concern that the ICC will prolong tyrants' reigns is exaggerated. The ICC is an international institution and like any other international institution it 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 offering such leaders asylum only when doing so provides them with 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, in the model in this paper, the reign of an atrocity-committing dictator will not be prolonged as a result of the creation of the institution.

In short, while the model presented in this paper 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 than there otherwise would have been. 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.

While the legal literature on the ICC is voluminous, the political science literature on the topic is sparser. Naturally the legal literature is less concerned with the puzzle that motivates this paper, namely why states would create such a seemingly futile regime, and instead focuses on normative questions of whether the court is a good idea or not and prescriptive questions concerning the Rome Statute. Within the political science literature the only article I am aware of is Rudolf (2001). Rudolph's (2001) article 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.”

Morrow (2001) provides a rational-choice-based case study of a related issue—treaties on treatment of prisoners of war. He explains the creation of those treaties as the rational response of states that need to recruit soldiers in the era of mass armies. His explanation is less applicable to the ICC which is meant to deter atrocities not only against enemy combatants but most especially against states’ own citizens. As Morrow indicates, states may refrain from violating POW treaties because if they violate them in time of war their enemies may retaliate by doing same—that is the game has the characteristics of the classic repeated prisoner’s dilemma. Morrow provides evidence of this process at work in Nazi treatment of British and American POWs during World War II. Such a game is not applicable to the case covered by this paper where a dictator is committing atrocities against a relatively powerless populace within his own country, and so nothing like the same enforcement mechanism exists.

This paper is my attempt to answer the puzzle of the ICC in a rational choice framework while at the same time commenting on the controversy in the legal literature. The paper is organized as follows. In the next section I will present a model of the bargain between a tyrant in political trouble at home and a foreign government that would like to facilitate a peaceful transition to a new government. Section III extends the model by adding a nonstrategic international actor, the Prosecutor, to whom the dictator can surrender. In section IV, I discuss the model and some extensions of it more informally and Section V provides some implications of the model for legal debates and criticisms concerning the ICC. Section VI concludes.

Equation Section 1

The Model

a. Set-up of the Model

This section will present a model of the interaction between a leader L_i , $i \in \{1, 2, 3, \dots\}$ and a set of foreign states, $\{S_1, \dots, S_k\}$, some or all of which may be willing to offer L_i asylum including, if necessary, impunity from punishment for past crimes in return for that leader stepping down. There are $k+2$ actors in the game at any given time, all of whom are infinitely lived: the S_i s, L_i , and a nonstrategic player N who selects the probability that a leader who commits atrocities is deposed in a given period and at some nodes of the game, also chooses whether that leader survives according to that probability. 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 L_i is deposed in the previous period N chooses L_{i+1} at the start of the subsequent period. I will index the leaders by the order in which they are chosen. The identity of the foreign states does not change at any point in the game. I assume both the leader and the foreign states have a one period discount rate of δ . 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_i 's per period payoff is 1, but when L_i commits atrocities it generates spillover costs that reduce S_i 's per period utility to $u_i < 1$.⁵ 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

⁵ Each S_i does not care whether L_i holds on to power or is deposed. It only cares whether atrocities are committed or not. If L_i does not commit atrocities S_i receives a payoff of one in that period whether L_i is deposed or not. If L_i commits atrocities S_i receives a payoff of u_i in that period whether L_i is deposed or not.

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 forever into the future.

As mentioned above, 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. For simplicity I assume there is no correlation between the current draw of a_t and previous draws. 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 Figures 1 and 2. Figure 1 illustrates what I call “the pre-atrocities subgame” which is the subgame up to and including the first period (if any) that L_i commits atrocities. Figure 2 illustrates the subgame for a given leader after the first period in which atrocities are committed, which I refer to as “the post-atrocities subgame.” To keep the narrative flowing I will offer only a brief description of the sequence of moves here, and refer readers to Figures 1 and 2 for a fuller illustration of the game tree.

Play proceeds over the course of discrete time periods $t \in \{1, 2, 3 \dots\}$. 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. S_1 (the indexing is arbitrary) then decides whether to accept the request. If S_1 accepts L_i steps down and the game ends for him, returning play to the beginning of the pre-atrocities subgame. If S_1 rejects the request, L_i can make the same request from S_2 and so on, if necessary all the way to S_k . If none of the foreign states' accept L_i 's request for asylum 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.⁶ If L_i does not survive 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 S_1 then decides whether to accept the request or not. If S_1 accepts L_i steps down the game ends for him, and play returns to the beginning of the pre-atrocities subgame. If S_1 chooses “reject,” L_i can keep requesting asylum from the remaining S_i 's as described above. If none of the foreign states accept L_i 's request, L_i faces a lottery over his political survival as described above.

In solving for the equilibria of this model, I will employ Markov perfect equilibrium (MPE) as a solution concept throughout (Fudenberg and Tirole 1991, chapter 13). I will highlight the results via a series of lemmas. Proofs of these statements in most

⁶ The game does not give L_i the opportunity to request exile after committing atrocities in the pre-atrocities subgame. Doing so would be superfluous, since 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.

cases follow from straightforward application of backward induction and are therefore omitted throughout (They are available from the author upon request).

b. Solving the post-atrocities subgame

In this subsection I will discuss the outcomes in the post-atrocities subgame. I will turn to the pre-atrocities subgame in the subsequent subsection. The only slight wrinkle in solving this subgame is S_i 's incentives at the last node of the game, which can be summarized as follows:

LEMMA 1: *If $a_i > 0$ and if there exists $a_{i+1} \in [0,1]$ such that L_{i+1} will refrain from committing atrocities S_i , $i \in \{1, 2, \dots, k-1\}$, has a weakly dominant strategy to accept a request for asylum and S_k has a strictly dominant strategy to accept a request for asylum.*

The intuition for the result is that if S_k accepts L_i 's request for asylum, then, with some probability $p > 0$, which I describe in greater detail below, S_k will obtain a payoff of one in the subsequent period and obtain a payoff of $u_k < 1$ with probability $1-p$. If, however, S_k rejects L_i 's request S_k will obtain a payoff of one only with probability $p(1-a_i)$. States S_1 through S_{k-1} have only weakly dominant strategies to accept such a request because they know that if they all reject L_i 's request S_k will accept it. The important point in this result is that some foreign state has a strictly dominant strategy to accept L_i 's request for asylum. The use of MPE is important to this result, because it assumes players respond only to payoff relevant information. As such by rejecting a request for asylum S_k

does not change future leaders' beliefs about what it (S_k) will do in the future, and therefore does not deter future atrocities.

Turning now 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:

$$U_i = a_t(1+r+\delta U) - r \quad (1.1)$$

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

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

where $\bar{a} = \int_{\alpha}^1 f(a)a da$, and α 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. One might think of it as L_i 's EPV of being at the first node of the post-atrocities subgame.

For sufficiently high a_t and U the EPV in equation (1) will be greater than zero, in which case L_i will not request asylum, the EPV of which is normalized to zero. If however the EPV in equation (1) is less than or equal to zero L_i will request asylum, S_i will accept, and L_i will step down. Leaders will not step down voluntarily until a sufficiently low a_t is drawn. More formally if a leader is going to be willing to request asylum, which involves going into exile the following must be true:

$$a_t < \frac{r}{1+r+\delta U} \quad (1.2)$$

I call the cutoff where (1.2) holds with equality α .

This discussion demonstrates that the following are the equilibrium of the post-atrocities subgame.

LEMMA 2: *The equilibria to 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. Some S_i accepts and L_i steps down.*

Turning to some comparative statics on a_t and r , Figure 3 shows the effect of increasing the retribution cost to L_i . Larger retribution costs expand the range of a_t for which there is a possible asylum agreement. The increase in retribution costs from r' to r'' expands the range of a_t for which there are possible asylum agreements from zero to α' to zero to α'' . Thus if there was some way to increase these costs it could induce atrocity-committing dictators to step down for relatively higher probabilities of remaining in power. Higher retribution costs reduce L_i 's EPV of committing human rights abuses, both directly and through U , for any given value of a_t and therefore L_i 's EPV of committing such abuses will be lower for higher values of a_t .

c. Solving the pre-atrocities subgame

In the pre-atrocities subgame I begin, as I did in post-atrocities subgame, with S_i 's decision to accept or reject a request for asylum. In this subgame, S_i has a weakly dominant strategy to accept such a request.⁷ 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 now 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

⁷ S_i 's payoff from accepting a request for asylum is $1 + \delta\Omega$. The highest payoff that S_i can receive from rejecting a request for asylum in the pre-atrocities subgame is $1 + q(\delta\Omega) + (1-q)\delta\Omega = 1 + \delta\Omega$.

is higher than the EPV of not doing so. L_i 's EPV of not committing atrocities in the current period is:

$$Q = \text{Max} \left\{ \frac{q(1 + \delta(1 - p)U) + (1 - q)d}{1 - p\delta q}, 0 \right\} \quad (1.3)$$

where 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 in that period. L_i can guarantee himself 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 and/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 (1.1) is higher than the EPV in equation (1.3). This will be the case if:

$$a_t > \frac{r}{1 + r + \delta U} + \frac{1}{1 + r + \delta U} \cdot Q \quad (1.4)$$

I will call the cutoff indicated on the right side of (1.4) β . 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 $a_t > \beta$. Notice that $\beta = \alpha$ if the first term in the curly brackets in (1.3) is negative. In such a case L_i would commit atrocities only if $a_t > \beta = \alpha$, 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 to be precise) by not committing atrocities and stepping down.

In light of this discussion, the possible equilibria of the pre-atrocities subgame are as follows.

LEMMA 3: *Equilibria of the pre-atrocities subgame are:*

3. *If $a_t > \beta$, L_i commits atrocities and proceeds to the post-atrocities subgame.*

4. *If $a_t \leq \beta$, L_i does not commit atrocities and:*

$$4a. \text{ If } \frac{q(1 + \delta(1 - p)U) + (1 - q)d}{1 - p\delta q} < 0 \text{ } L_i \text{ requests asylum. } S \text{ accepts and } L_i$$

steps down, or

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

survives with probability q .

Equilibrium 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 if he does not commit atrocities is negative whether he commits atrocities or not. I will ignore this equilibrium for the remainder of the paper.

L_i knows that once he commits atrocities he "crosses the Rubicon" in the sense that he moves into the post-atrocities subgame and will suffer retribution costs if he is deposed. Therefore the leader is not going to be willing to commit atrocities for some one time increase in a_t unless that one time increase is quite large and/or the EPV of being in the post-atrocities subgame is relatively high compared to the EPV of being in the pre-atrocities subgame. L_i will be more likely to commit atrocities when U is large, which occurs when r is relatively small, or when q is small so that his chances of keeping office

if he does not commit atrocities are low, or if d is close to $-r$ so that L_i has little to lose by committing atrocities.

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 penalty of being deposed not having committed atrocities*, which in this model is 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 in fact 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 6 illustrates this difference between two hypothetical states. Country 1 is a country where the norm for political succession is assassination. Country 2 is a country where the norm for political succession is well-ordered transfers of power (e.g. following an election). Let us assume that the penalty for committing atrocities in both states is for some fixed leader i of this outcome is $-r < 0$. As illustrated in Figure 4 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 leader i from committing atrocities is identical, $-r$, but d is lower in country 1 because of the two

different norms of succession in the two different states. As such the leader would have more incentive to commit atrocities in country 1 than in country 2.

A few simulations may clarify some of these comparative statics. Figure 5 illustrates the effect of changes in r and d on the number of periods in which atrocities are committed over a twenty period time span. In all of the simulations I assumed that $F(a)$ is uniform on the zero-one interval and $\delta = 0.9$. I simulated 1000 twenty period time spans to determine in how many of those twenty periods atrocities were committed on average.⁸ I simulated each of these twenty year times span 1000 times. The downward slope of all of the curves indicates that as r increases the periods in which atrocities are committed is reduced. Furthermore Figure 5 shows that atrocities are decreasing in d .⁹ The lowest value of d in the figure, -0.5 , generates the highest number of periods of atrocities on average for each level of r . As d increases the number of periods in which atrocities are committed is reduced until eventually at levels of d equal to 0.2 and 0.3 there are no atrocities committed over any of the twenty year periods in any of the simulations for high enough values of r . In simulations using $d = 0.4$ (not shown) there were no periods of atrocities regardless of the value of $r \geq 2$.

Figure 6 illustrates the results from simulations in which I varied q , the probability that L_i remains in power if he does not commit atrocities. As before I assumed that $F(a)$ is uniform on the zero-one interval and $\delta = 0.9$. I simulated one thousand twenty period time spans. I assumed a value of $d = 0$ for all the simulations in this figure and simulated three values of q , 0.45 , 0.5 and 0.55 . The figure shows that the number of

⁸ The *Mathematica* notebooks used for the simulations are available upon request.

⁹ Strictly speaking the number of periods of atrocities is weakly decreasing in d . Once $\beta = \alpha$ L_i will commit atrocities when $a_i > \alpha$ which is independent of d .

periods of atrocities is decreasing in q . For each value of r the number of periods of atrocities is higher when $q = 0.45$ than when either $q = 0.5$ or $q = 0.55$. In fact for $q = 0.55$ no atrocities are committed for values of $r \geq 5$.

Equation Section (Next)

The Effect of an International Criminal Regime

Now suppose that some proper subset of the foreign states create an international institution that is staffed by an international nonstrategic actor P (the prosecutor) who automatically indicts any leader who commits atrocities.¹⁰ The pre-atrocities subgame is unaltered by this change, but the post-atrocities subgame is amended as indicated in Figure 7. I will call this subgame the “post-indictment subgame” to distinguish it from the post-atrocities subgame discussed in the previous section.

The post-indictment subgame begins just as the post-atrocities subgame did with a draw of a_t 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 then presents S_1 with a choice—to accept or reject that request. If S_1 chooses to reject, L_i can continue requesting asylum from S_2 and so on until either some S_i accepts his request or all k of the reject it. If some S_i accepts L_i 's request L_i steps down and the game ends for him as described previously. If no S_i accepts L_i 's request then L_i must choose whether to surrender to P and receive, with certainty, a one time penalty of $-m < 0$.¹¹ If L_i surrenders he steps down, returning play to the beginning

¹⁰ 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.

¹¹ Think of m as the maximum punishment that P can impose. Admittedly there may be some uncertainty as to whether L_i will actually be convicted after he turns himself in to P and there may also be some ambiguity *a priori* as to what L_i 's punishment will actually be in the end, but I abstract from those complications here. I also do not consider the possibility that P may negotiate a “plea bargain” with L_i in which L_i surrenders

of the pre-atrocities subgame. Alternatively at this point, L_i can choose not to surrender in which case he takes his chances at holding on to office according to the usual lottery.

An important feature of the game is that it does not give P any means of enforcing its indictment on S_i . S_i does not pay any costs in the model from offering asylum to L_i . Without any higher authority to enforce S_i 's compliance with the new institution all commitments that S_i has made by joining the new institution must be self-enforcing. In practice that means that S_i cannot obtain higher utility from accepting L_i 's request for asylum than it does from rejecting it. As I will show the creation of the regime may change S_i 's behavior compared to the pre-regime scenario, however S_i 's membership in this new institution is actually irrelevant to its behavior in the sense that S_i has the same incentives whether it is a member of the regime or not. 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 countries have yet to sign the Rome Statute so they are in no way obligated to comply with its indictments.

a. 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 all S_i s have turned down his request for asylum. L_i 's EPV of not surrendering at that point is:

and receives a payoff (punishment) less than zero but still greater than $-m$. As I will address below, there are circumstances in this model where allowing such a possibility will enhance the deterrent effect of the institution.

$$\tilde{U}_t = a_t(1+r+\delta\tilde{U})-r \quad (2.1)$$

Unsubscripted \tilde{U} in equation (2.1) is L_i 's EPV in the next period (i.e. before a_{t+1} is known) of being in the post-atrocities subgame in this new situation following the creation of this new institution. It might be thought of as the EPV of being at the first node of the post-indictment subgame. \tilde{U} is a constant, the value of which depends on L_i 's beliefs about what the S_i 's will do in the future as I discuss below.

I now define a new cutoff point, which I will call, $\tilde{\alpha}$, which is the value of a_t below which L_i would willingly surrender to P. Since L_i 's EPV from surrendering to P is $-m$, $\tilde{\alpha}$ corresponds to the level of a_t for which L_i 's EPV was equal to $-m$, so that:

$$\tilde{\alpha} = \frac{r-m}{1+r+\delta\tilde{U}} \quad (2.2)$$

I also define:

$$\alpha' = \frac{r}{1+r+\delta\tilde{U}} \quad (2.3)$$

as the level of a_t for which L_i would be willing to step down if he could obtain asylum from some S_i in this new situation following the creation of the regime. It is the value of a_t for which L_i 's EPV is zero in this new situation. The only difference between α and α' is that α' has \tilde{U} in the denominator rather than U . If $a_t < \tilde{\alpha}$, L_i would be willing to surrender to P if all S_i s rejected his request for asylum. If $a_t > \tilde{\alpha}$ L_i would not be willing to surrender to P but would be willing to step down if he received asylum if $a_t \leq \alpha'$. Obviously if $a_t > \alpha'$ L_i would be unwilling to step down even if granted asylum by S_i .

Turning to the S_i 's' decisions of whether to accept L_i 's request, the forgoing discussion implies that while S_k still has a strictly dominant strategy to accept L_i 's request

when $a_i \in [\tilde{\alpha}, \alpha']$, it no longer has a strictly dominant strategy to accept L_i 's request for asylum when $a_i < \tilde{\alpha}$, because L_i would be willing to surrender to P. S_k is indifferent between rejecting and accepting L_i 's request under those circumstances because, regardless of whether it does or not, L_i will step down. Therefore L_i can no longer be certain that S_k will accept his request for asylum. Notice that this is true *regardless* of whether S_k is actually a member of the agreement 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 equilibrium in the post-indictment subgame as described in the following lemma.

LEMMA 4: *The possible equilibria 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 $a_i < \alpha'$, L_i requests asylum. Some state S_i accepts, and L_i steps down.*
- 3'. *If $a_i < \tilde{\alpha}$, L_i requests asylum. All S_i s refuse. L_i surrenders to P.*

The significance of outcome 3i is that, despite P's lack of enforcement power, some atrocity committers surrender to him. For low enough values of a_i , all S_i s will be indifferent between accepting L_i 's request for asylum and rejecting it because they know that even if all of them reject the request L_i will surrender to P and step down. This is in contrast to the outcome in the previous section where S_k clearly had a dominant strategy to accept L_i 's request.

Note that there is some ambiguity about the equilibrium when $a_t < \tilde{\alpha}$ because the S_i s are indifferent between accepting and rejecting L_i 's request for asylum in those circumstances since either way L_i will step down. Under the assumption of MPE which I make in this paper the S_i 's have no reason to reject L_i 's request for asylum in for the express 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 the S_i 's get some very small utility increment, ε , from seeing atrocity committers punished it would eliminate this ambiguity and make rejecting such a request a dominant strategy for values of a_t in the neighborhood of $\tilde{\alpha}$ and below. However, I do not make that assumption in this paper because it seemed *ad hoc*.

b. 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_i 's decision to offer asylum in this subgame, if asked.¹² I turn instead to L_i 's incentives to commit atrocities. Differences in L_i 's decision (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 , and that EPV will depend in turn on L_i 's beliefs about what the S_i 's will do if $a_t < \tilde{\alpha}$. I will dispense with the obvious case first: if L_i believes that some S_i will grant him asylum when $a_t < \tilde{\alpha}$ (in other words he believes that equilibrium 3i will not occur) then the

¹² This is the case in which L_i requests asylum in the first period of his reign. As before all S_i s 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.

effect of the new regime on the equilibria in the pre-atrocities subgame will be nil—the equilibria of the pre-atrocities subgame after the creation of the new regime will be identical to those described in the previous section.

If however L_i believes that outcome 3i 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 some time t at the third node of the pre-atrocities subgame L_i 's EPV of choosing to commit atrocities given these beliefs is given by equation (2.1) where, given this set of beliefs:

$$\tilde{U} = -m \int_0^{\tilde{\alpha}} f(a) da + \int_{\alpha'}^1 f(a) [a(1+r+\tilde{U})-r] da = \frac{\tilde{\alpha}(1+r) - (1-F(\tilde{\alpha}))r - F(\tilde{\alpha})m}{1-\delta\tilde{\alpha}}, \text{ and}$$

$$\tilde{\alpha} = \int_{\alpha'}^1 f(a)a da .$$

L_i 's EPV of not committing atrocities at that node is:

$$Q' = \text{Max} \left\{ \frac{q(1+\delta(1-\tilde{p})\tilde{U}) + (1-q)d}{1-\tilde{p}\delta q}, 0 \right\} \quad (2.4)$$

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 in that period in this new situation following the creation of the regime.

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} . In

other words the only effect of the new regime is to change L_i 's EPV if he commits atrocities in future periods.

As before L_i will commit atrocities if the EPV if doing so exceeds the EPV of not doing so—that is if equation (2.1) is greater than equation (2.4). I define β' as the difference of these two equations, analogous to β in the previous section. According to the same line of argument presented in the previous section, L_i will commit atrocities if N draws and $a_t > \beta'$. I summarize these possible equilibria in Lemma 5.

Lemma 5: *If L_i believes that equilibrium 3i will occur when $a_t < \tilde{\alpha}$, the possible equilibria of the pre-atrocities subgame following the creation of the regime are:*

4'. *If $a_t > \beta'$, L_i commits atrocities and proceeds to the post-atrocities subgame.*

5'. *If $a_t \leq \beta'$, 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. Some S_i 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 .

Figure 8 illustrates what might be called the “total” effect of institution on L_i 's EPV of committing atrocities in this scenario. There is a direct effect and an indirect effect. The direct effect is that for values of $a_t < \tilde{\alpha}$ L_i will surrender to P and receive an EPV of $-m$ rather than zero as it did in prior to creation of the institution. This direct

effect produces the indirect effect—the reduction from U to \tilde{U} —which in turn causes an increase in the value of a_t for which L_i will request asylum from $\alpha = \frac{r}{1+r+\delta U}$ to

$\alpha' = \frac{r}{1+r+\delta\tilde{U}}$ and a downward shift in L_i 's EPV for values $a_t \in [\alpha', 1]$. The total

expected change in L_i 's EPV as a result of the direct and indirect effects can be derived by integrating L_i 's EPV in each of the two scenarios (pre-and post-regime) over all possible values of a_t . As shown in Figure 8 the total EPV of committing atrocities is smaller as a result of the institution by the amount of the gray shaded area.

Figure 9 illustrates the effect of the regime on L_i 's incentives to commit atrocities in simulations. The graphs indicate the average number of periods of atrocities before and after the creation of the institution in two different scenarios. In the first scenario, illustrated in part A of Figure 9, d is -0.2 . In the second scenario, illustrated in part B of Figure 9, $d=-0.5$. In both cases I assumed m is 0.5 .¹³

There are fewer periods of atrocities on average in both scenarios. The effect is larger in the first scenario than in the second scenario. This result is illustrated in part C of Figure 8, which shows the change in the average number of periods of atrocities in each of the two scenarios. The reduction in the number of periods of atrocities is greater in the first scenario ($d = -0.2$) than in the second ($d = -0.5$). This difference converges as r grows, because for values of $r \geq 4$ periods of atrocities go to zero in the first post-regime scenario and can go no lower.

¹³ The other parameters used in the simulations are the same as in the previous example— $F(a)$ is uniform of $[0,1]$ and $\delta=0.9$. The number of periods of atrocities is out of a twenty year time period. Simulations were repeated 1000 times.

Why the difference in the regime's effect in the two scenarios? The payoff from being deposed not having committed atrocities is closer to the payoff from being deposed having committed atrocities in the second scenario. Therefore leaders have a greater incentive to commit atrocities in the second scenario. Since they are going to be punished severely whether they commit atrocities or not they have less incentive to refrain from doing so. In such a situation the threat of punishment by P is less daunting given the relatively severe punishment meted out by domestic political rivals. To the extent that this model captures the interaction between human rights abusers and the states that offer them asylum, this result suggests that the ICC will have its greatest impact in states where the problem of atrocities was somewhat lessened to begin with. Put another way this model suggests the ICC will fail to deter atrocities in the most atrocity prone states but may have more of an impact in more marginal states.

Discussion and Extensions of the Model

Before turning to a discussion of the legal literature on the ICC I would like to highlight a few points that might help clarify the model and its results. First, *the number of periods in which atrocities are committed will not increase as a result of the creation of this institution*. In the introduction I mentioned that some scholars are concerned that fear of prosecution will induce some atrocity-committing leaders to hold on to office longer than they otherwise would. I will return to this issue later in the paper when I discuss criticism of the court by legal scholars. For now I will only point out that the reign of a leader who has committed atrocities will not be prolonged as a result of the regime modeled in this paper. The international institution modeled in this paper potentially reduces leaders'

EPV of committing atrocities which in turn means that leaders are less likely to do so in the first place (N will have to draw a higher a_t higher for them to choose to do so). In doing so, however it does not prolong the reign of leaders who have committed atrocities. Prior to the creation of the institution, L_i would leave power whenever $a_t < \alpha$ and after the creation of the institution L_i will leave office whenever $a_t < \alpha'$. Since $\tilde{U} \leq U$, α' cannot be lower as a result of the regime. In short, we can be certain that the institution modeled in this paper will *not increase* the number of periods of atrocities and may reduce them. I will return to this issue in the subsequent section.

Second is the issue of *commitment problems*. Persuading a leader who has committed atrocities to step down voluntarily can be difficult even when his grip on power is tenuous because of the commitment problems inherent in the agreement between the outgoing leader and the party that offers him asylum. Leaders will be unwilling to step down voluntarily if they think they may be punished for their crimes at a later date. At the time the leader negotiates an agreement for his ouster he still has bargaining leverage because of his hold on state power (however tenuous). Once he steps down however he has no such bargaining power.¹⁴ What is to stop his successor from renegeing on any agreement and holding him accountable for his past crimes? Concerns such as this may lead the dictator to attempt to hold on to power even when in the absence of these problems there would be a mutually beneficial agreement between him and his successor that could lead to his stepping down. The intervention of some relatively impartial foreign state, of the type that I have modeled in this paper, may be helpful in alleviating some of these commitment problems because such a state may not

¹⁴ There are clear exceptions to this general characterization. For example, Chile's outgoing dictator Augusto Pinochet maintained such bargaining power through his continued control over the military.

have the desire to exact retribution like a domestic political rival might. The cases described in the first paragraph of this paper were examples of precisely these kinds of agreements. Even in such cases these sorts of commitment problems are perhaps not completely absent, however.

Currently commitment problems play no role in the model in this paper. I chose to ignore commitment problems to focus on the effects of the new regime on leaders' incentives to commit atrocities. If all of the S_i 's promises of asylum are not credible then the kinds of agreements I have modeled in this paper are impossible. The cases mentioned in the opening paragraph of the paper clearly suggest that, as an empirical matter, human-rights-abusing dictators and the states that harbor them have managed to overcome this problem somehow. The motivation of this paper is to see if the creation of an international institution can help countries refrain from granting atrocity-committing dictators asylum and immunity. That question would be moot if commitment problems were such that asylum and immunity were impossible in the first place. Thus in the interests of focusing on the issue at hand—the effect of an international criminal regime in a world where such agreements *are* possible—I have not included commitment problems in the model. What *is* relevant to the issue at hand is whether commitment problems are exacerbated by the creation of the ICC, since if they are it would have implications for the desirability of the new regime. I will return to this issue below and argue that commitment problems in this sense are not likely to be a problem because the ICC lacks enforcement power and because many countries are not members of the regime.

Third, this analysis has indicated that, to the extent that it conforms to the assumptions of the model in this paper, the ICC may lead to the deterrence of some potential human rights abusers at the margin, but that analysis begs the question: *why did the world need an ICC to accomplish this goal?* In this model the ICC does not really do anything. The prosecutor in this model merely waits for human-rights-abusing dictators to surrender to him. It is true that what the institution modeled in this paper accomplishes could in large part be accomplished simply by states pledging not to harbor human rights abusers unless they pay some agreed-upon punishment (recognizing that in expectation states would fail to comply with those pledges $F(\alpha) - F(\tilde{\alpha})$ percent of the time). However such a treaty would still leave the practical problem of whom L_i would surrender to and who would extract the “payment” of $-m$ from L_i . These tasks had to be delegated to some entity. In theory they 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.

Fourth, the model has implications for a potential problem with the ICC that I raised in the introduction, namely 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?* The model in this paper 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 and it provides a coordination mechanism for states. A country that is not a member of the regime modeled in this paper has no less incentive to refrain from offering a leader asylum than does a member of the regime.

A fifth 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 m at a very high level, which will produce a large reduction in \tilde{U} and atrocities will be deterred. However that interpretation of the role of m in this model would be incorrect. Recall that m is the cost to *the outgoing leader* of the maximum punishment that P can hand down. P does not really have that much control over that parameter of the model. P has limitations on how harsh a punishment it can set, and, more importantly, P cannot alter L_i '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.

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 α enter into an asylum agreement and receive an EPV 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 in this paper 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 EPV of committing atrocities and thereby deters some marginal human rights abusers. Granting P some ability to make

“plea bargains” could actually enhance the deterrent effect of the institution in this model. If P 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 $a_i < \alpha'$. Via these plea-bargained punishments each leader would be required to “pay” so that the leader obtained no surplus from the plea agreement. P would be playing a role akin to that of a price-discriminating monopolist. All leaders who drew an $a_i < \alpha'$ would be punished to the maximum extent possible while still inducing them to step down, so \tilde{U} would be even lower in a regime that allowed such plea bargains than in the one modeled in this paper. As such the deterrent effect would be greater.

Implications for Legal Criticism of the Court

The model developed above offers some interesting insights for the current debate surrounding the ICC. As mentioned in the introduction the ICC has been a very controversial international institution and the legal literature regarding it is voluminous. In this section I will focus on four main criticisms: (1) the ICC will fail to deter atrocity committers because it lacks the power to enforce its indictments (2) the creation of the ICC will prolong the reign of leaders who commit atrocities because the amnesty agreements that would induce them to step down will become illegal (3) the ICC will create an unjust two-tiered international judicial system in which state leaders are tried by the ICC according to due process of law and receive relatively light punishments while low-level atrocity perpetrators will be tried by less equitable domestic courts and receive harsher punishments (4) perversely the ICC will actually encourage atrocities because it

will deter humanitarian military intervention out of fear that the interveners' service members may be subject to politically motivated ICC prosecution.

1. *The ICC will fail to deter because it has no enforcement power.*

Goldsmith (2003) offers a forceful example of this 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”¹⁵

The main problem with this criticism is that it 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 Payam Akhavam (2001), who served as legal council to the International Criminal Tribunals for the Former Yugoslavia and Rwanda, 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.” (p.7)

In the model in this paper there is absolutely no outside enforcement. The ICC has no means of sanctioning either atrocity-committing leaders or the states that offer them sanctuary. Furthermore the model assumes no action by outside sovereign governments to force atrocity-committing leaders out of office. Despite the extreme weakness of enforcement powers for both the ICC and foreign states in this model, the institution may, in the right circumstances, still have some deterrent effect because leaders may recognize

¹⁵ Goldsmith 2003, 92. See Smidt (2001) for the same argument.

that a time may come when their best option is surrendering to the ICC. Under those circumstances they may have little hope of obtaining a 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).¹⁶

2. The ICC will prolong the reign of leaders who commit atrocities.

This argument has been made in a variety of places but the following quote from an interview with Stephen Krasner (2003) is a particularly nice example of this line of thinking.

“...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.”¹⁷

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. In the model in this paper those incentives are illustrated by a comparison of equation (2.2) with equation (2.3). The former is the maximum value of a_i that L_i would

¹⁶ Of the remainder 24 were detained by international forces, 11 were arrested by Serbia and seven were turned over by other states. Eighteen indicted war criminals remain at large.

¹⁷ Krasner Stephen D. 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 same argument can be found in Goldsmith (2003) and Goldsmith and Krasner (2003).

need to have in order to be willing to surrender to P and the latter is the maximum value of a_t for which L_i would be willing to step down if granted immunity. The former is unambiguously lower than the latter by the amount $\frac{m}{1+r+\delta\tilde{U}}$. In other words N would have to draw a lower (and therefore less likely) value of a_t before L_i would be willing to step down and be punished for past crimes. Thus in this sense the criticism is correct—the probability that a leader will step down and be punished for his crimes *is* lower than the probability that he would be willing to step down if he knew he would not be punished. Of course 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. As mentioned above the outcome that concerns Krasner and the other scholars cited these scholars does not occur in the model in this paper. The criticism of Krasner and the others is based on one of two assumptions neither of which is particularly plausible.

First the criticism may be based on the assumption 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 the model in this paper states only comply with the institution's indictments when it is costless to do so, that is when doing so offers them a at least as high an expected present value as not doing so. On occasion (cases where $a_t < \tilde{\alpha}$) states may refuse to offer asylum but only because they know that the leader in question will surrender to the prosecutor anyway. The scenario in which the reign of atrocity-committing leaders is prolonged cannot occur in the model in this paper 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 it 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 in this paper would become impossible because they would no longer be credible.¹⁸ The human rights abuser would know that S_i would have no incentive to comply with its asylum agreement and therefore he 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. No legal analysis of which I am aware suggests that the ICC has the power to impose such costs. On the contrary the ICC's lack of enforcement power is one of the central criticisms of the regime as I 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 to 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

¹⁸ The timing of the costs is important. If the country 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. S_i 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.

ICC does make it any easier or less costly to avail themselves of that option. My point is not that commitment problems do not exist but that nothing I have seen suggests that the ICC will make them more severe.

Suppose I am wrong, though, and that the creation of the ICC does, despite all of the existing indications to the contrary, make punishing states that harbor atrocity committers easier—including states that did not even sign the Rome Statute. Such a situation would still not guarantee that there would be more periods in which atrocities were committed than if the ICC did not exist. 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. In such a case the criticism that the ICC would prolong the reign of atrocity committers would be correct, but ultimately what we care about are amount of atrocities committed and this criticism cannot claim that the ICC will increase the number of states and years in which atrocities are committed—the argument is simply ambiguous. Ultimately of course this is an empirical question that the ICC has not been around long enough to answer.

The ICC will create a system of “Cadillac justice” in which leaders are tried fairly by the ICC and low-level perpetrators are subjected to harsh and unfair trial at home.

The point of this criticism, which was raised by Roth (2002) among others, is mainly normative—a system of “Cadillac justice” is unfair on its face. However, the model in this paper suggests that if such a system of “Cadillac justice” emerges as a result of the creation of the ICC—that is if the ICC insulates state leaders from the harsh punishments meted out by their domestic rivals and replaces them with the relatively light

punishments of the ICC—it will be detrimental to the deterrence of atrocities. If for instance the ICC were able to compel the domestic political agents to turn over deposed atrocities perpetrators rather than disposing of them themselves—an assumption that is not a feature of the model discussed above—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 of the leaders to whom this model applies would be given by their domestic political/legal systems if they were deposed. If a leader knew that his domestic opposition would be forced to turn him over to the ICC if he were deposed, he would know that he would face a less severe punishment as a result of the creation of the Court than he would have without the creation of the institution. The effect on deterring atrocities by the leader and on the leader’s incentives to step down would be the same as reducing the value of r , the comparative statics on which I already explored in Figure 3.

The ICC does not have a sufficiently long track record to tell us whether this will actually occur in practice. Failure to adhere to the norm of complementarity has created problems with international criminal tribunals in the past, particularly the one in Rwanda. The ICTR created precisely such a system of “Cadillac justice,” and in fact the ICTR does not even conduct its proceeding in Rwanda but in neighboring Tanzania. Because of this and other problems with the Tribunal, the Rwandan government, which was on the Security Council at the time the ICTR was established, actually voted against it and has been hostile toward its proceedings (Akhavan 1996, Henquet 1999). In part to rectify these types of problems the subsequently established tribunal on atrocities committed in the civil war in Sierra Leone contained much more local involvement.

The ICC, too, was created with the norm of complementarity in mind. If the Chief Prosecutor gives deference to state's domestic political and judicial systems then the domestic political agents who overthrew the previous leader will not have to turn deposited leaders over to the ICC and will be able to exact their regular retribution, which I assume is worse for the dictator than is the punishment that the ICC hand out. If instead domestic political agents are required or pressured to turn leaders over to the ICC then those leaders will probably be able to expect lighter punishments than they would have received prior to the creation of the ICC, and the ICC could actually lead to *more* atrocities according to the model in this paper.

The ICC may actually have the perverse effect of deterring intervention on behalf of human rights.

According to this argument the ICC will “deter the wrong parties.” (Smidt 2001, 160). States which would otherwise be willing to intervene to liberate the oppressed will now be deterred from doing so out of fear that their troops may be prosecuted by the ICC (for example, for the incidental deaths of civilians during the humanitarian intervention). The claim is that “the provisions of the ICC simply place too many risks on nations and their armed forces that are determined to rid the world of oppression.” (Smidt 2001, 158). This will then lead to more atrocities rather than less.

“Political prosecutions before the ICC are so probable that the forces of good may be deterred from taking on the forces of evil. Since the forces of evil will recognize the deterrent influence of such politically based prosecution on potential responders the leaders of these regimes may make the entirely rational decisions to commit acts of aggression, knowing they can act without fear of military intervention from foreign forces.”¹⁹

¹⁹ Smidt *op. cit.* 158. See Goldsmith, *op. cit.* 95-99 for the same argument.

The model in this paper is not directly relevant to this criticism because military intervention has no role in it either before or after the creation of the ICC. However I would like to point out two shortcomings inherent in this criticism. First the assumption that “the forces of good” are *not* deterred from sending troops into the heat of battle against a ruthless dictator but they *are* deterred by the prospect that some of those troops may face politically motivated prosecutions by the ICC is somewhat implausible. Second, this criticism fails to recognize that Article 98 of the Rome Statute allows states to create side agreements that address this very criticism. In these so-called “Article 98” or “nonsurrender” agreements states pledge to refrain from surrendering each other’s citizens to the ICC. As far as the United States is concerned “The American Service Members’ Protection Act of 2001” essentially requires such “Article 98” agreements before member states of the ICC are eligible for US military aid, including participation in peacekeeping missions. The first such agreement was concluded with Romania, only one month after the entry into force of the ICC. As of this writing the United States has concluded 92 of these agreements, the latest with Eritrea on July 8, 2004.²⁰ Once such an agreement is secured, Article 98 of the very treaty that created the Court dictates that the Court has no standing to try US service personnel.

Conclusion

This paper began with a puzzle: why would a group of states create an international institution, the International Criminal Court, that is, *prima facie*, so futile? The answer provided in this paper is that perhaps the ICC is not quite as futile as it first appears. The model in this paper demonstrates that in the right circumstances, the ICC could yield

²⁰ Information about these agreements can be found at <http://www.state.gov/t/pm/art98/>.

some reductions in the number of leaders who commit human rights atrocities. The importance of this finding is that it occurs in a model in which there is no international intervention to depose leaders who commit atrocities nor is there punishment of states that harbor such criminals after they are deposed. In short in a model with absolutely *no* punishments and outside enforcement the ICC can still have a deterrent effect on marginal human rights abuser.

The reason for this deterrent effect is that under the right conditions, when a leader has a high probability of being deposed leaders are actually willing to surrender to the institution modeled in this paper rather than face retribution from their domestic political opposition, which is assumed to be worse than the punishment handed out by the institution in this model. Knowing that the leader is willing to turn himself in, outside states have no incentive to offer him asylum. In such a situation, the best case scenario for the leader is to surrender to the international authorities. This outcome is worse than the best case scenario in an identical situation in the absence of the institution, which accounts for why the expected present value of committing atrocities is lower after the regime is created, and thus so is the number of atrocities committers at the margin. In this model the ICC has no enforcement capability and is understood merely as a coordination mechanism whereby states do not offer outgoing leaders asylum with impunity when those leaders would have been willing to surrender to the ICC.

A natural extension of the argument in this paper 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. That question is an empirical one and can only be answered prospectively by parameterizing the model from real-world cases—a project which is the object of ongoing empirical work. In the meantime I hope that this paper will inform the continuing legal debate in such a way that the ICC is no longer characterized as either a cure-all or a catastrophe-in-waiting. If the model in this paper is any guide, the ICC will be neither of those things. As long as the prosecutor respects the complementarity norm and does not institute a system of “Cadillac justice,” 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.

References

Akhavam, Payam. 1996. The International Criminal Tribunal for Rwanda: The Politics and Pragmatics of Punishment. *American Journal of International Law*. 90(3):501-510.

_____. 2001. Beyond Impunity: Can International Criminal Justice Prevent Future Atrocities? *American Journal of international Law*. 95(1): 7-31

Bueno de Mesquita, Bruce, Alastair Smith, Randolph Siverson and James D. Morrow. *The Logic of Political Survival*. Cambridge, MA: The MIT Press.

“Charles Taylor: A wanted man.”

<http://www.cnn.com/2003/WORLD/africa/06/10/liberia.taylor/>. Posted Thursday, December 4, 2003. Accessed October 1, 2004.

“Charles Taylor - preacher, warlord and president.”

<http://news.bbc.co.uk/2/hi/africa/2963086.stm>. Accessed October 1, 2004.

Fudenberg, Drew and Jean Tirole. *Game Theory*. Cambridge MA: The MIT Press.

Goldsmith Jack. 2003. The Self-Defeating International Criminal Court. *The University of Chicago Law Review*. 70(1): 89-104.

Goldsmith Jack and Stephen D. Krasner. 2003. The limits of idealism. *Daedalus*. 132(1): 47-63.

Henquet, Thomas. 1999. Mandatory Compliance powers vis-à-vis States by the Ad Hoc Tribunals and the International Criminal Court: A Comparative Analysis. *Leiden Journal of International Law*. 12(4): 969-99.

International Criminal Tribunal for the former Yugoslavia. 2003. *Tenth annual report of the International Criminal Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia since 1991*. Available at: <http://www.un.org/icty/rappannu-e/2003/AR03-e.pdf>

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>.

Morrow, James D. 2001. The Institutional Features of Prisoner of War Treaties. *International Organization*. 55(4): 971-91.

“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>

Penrose, Mary Margaret. 2000. Lest We Fail: The Importance of Enforcement in International Criminal Law. *American University International Law Review*. 15(2): 321-94.

Roth Kenneth. 2002. Introduction. *Social Research*. 69(4):

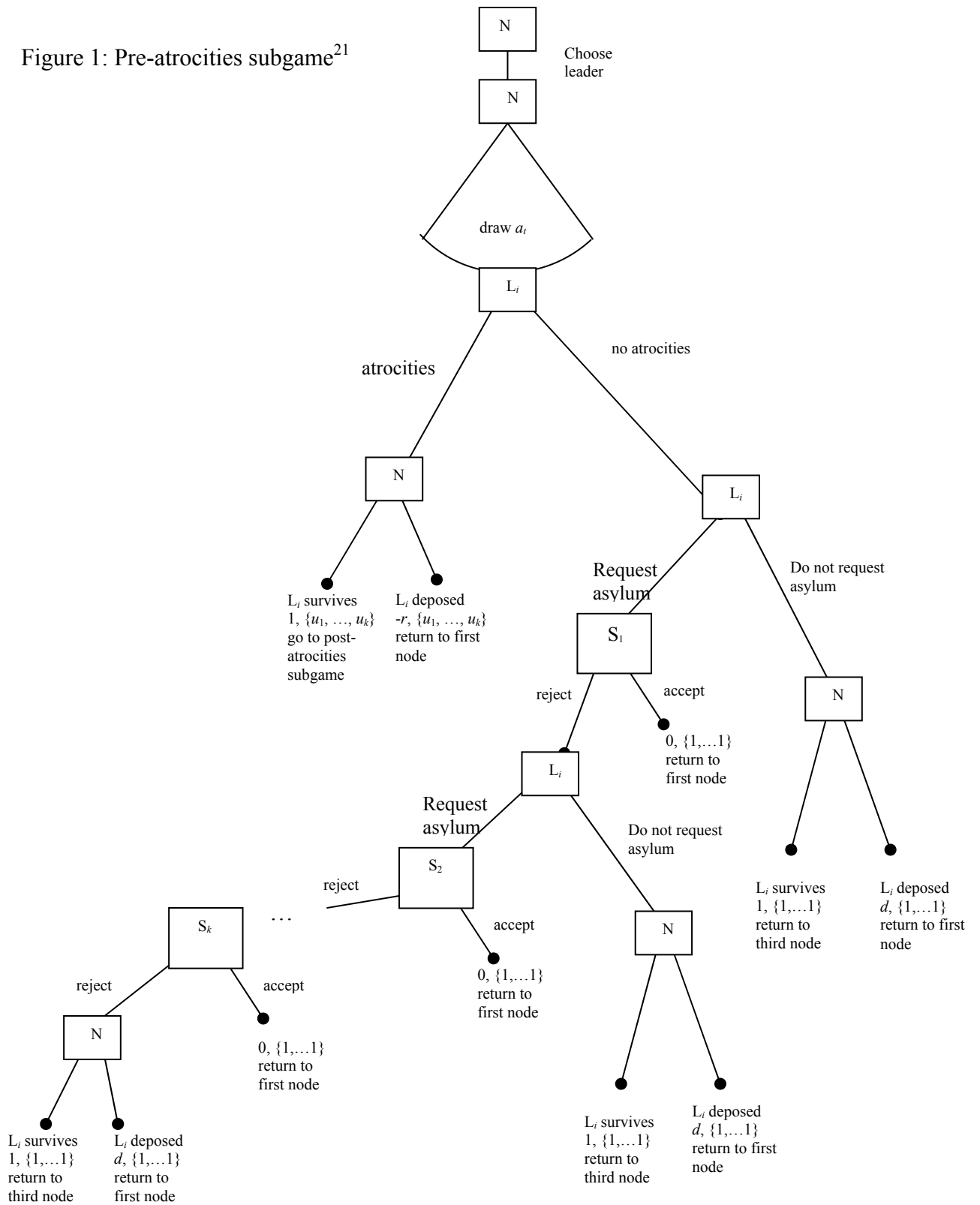
Rudolph, Christopher. 2001. Constituting an Atrocities Regime: The Politics of War Crimes Tribunals. *International Organization*. 53(3) 655-91.

Smidt, Michael L. 2001. The International Criminal Court: An Effective Means of Deterrence? *Military Law Review*. 167: 156-240.

“Taylor's new Nigerian home” <http://news.bbc.co.uk/2/hi/africa/3142101.stm>. Accessed October 1, 2004.

Wippman, David. 2000. Atrocities, Deterrence and the Limits of International Justice. *Fordham International Law Journal*. 23(3): 473-88.

Figure 1: Pre-atrocities subgame²¹



²¹ The first payoff listed is L_i 's. The vector refers to the payoffs of $\{S_1, \dots, S_k\}$

Figure 2: Post-atrocities subgame (see footnote to Figure 1)

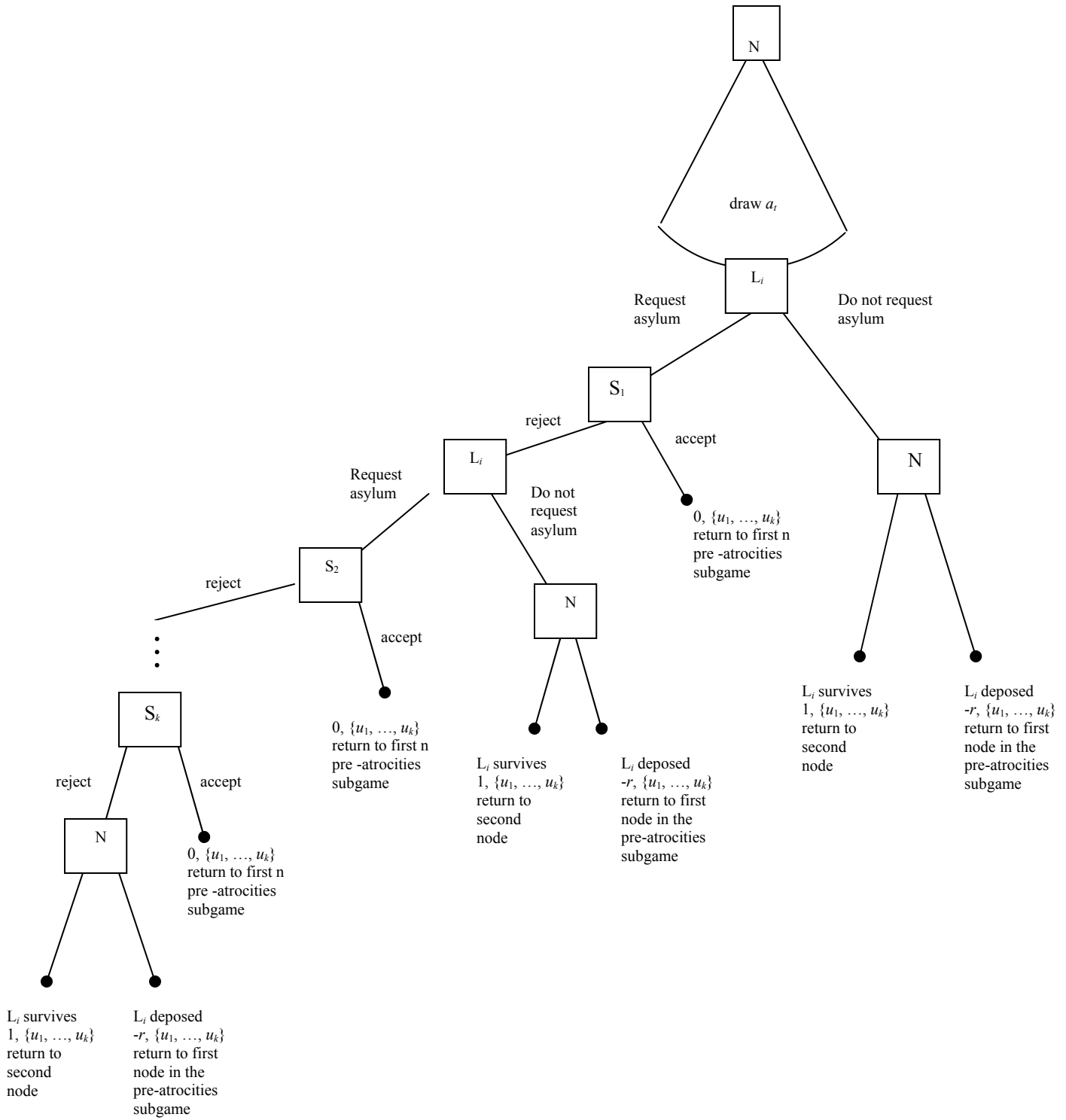


Figure 3: Comparative statics on a_t and r

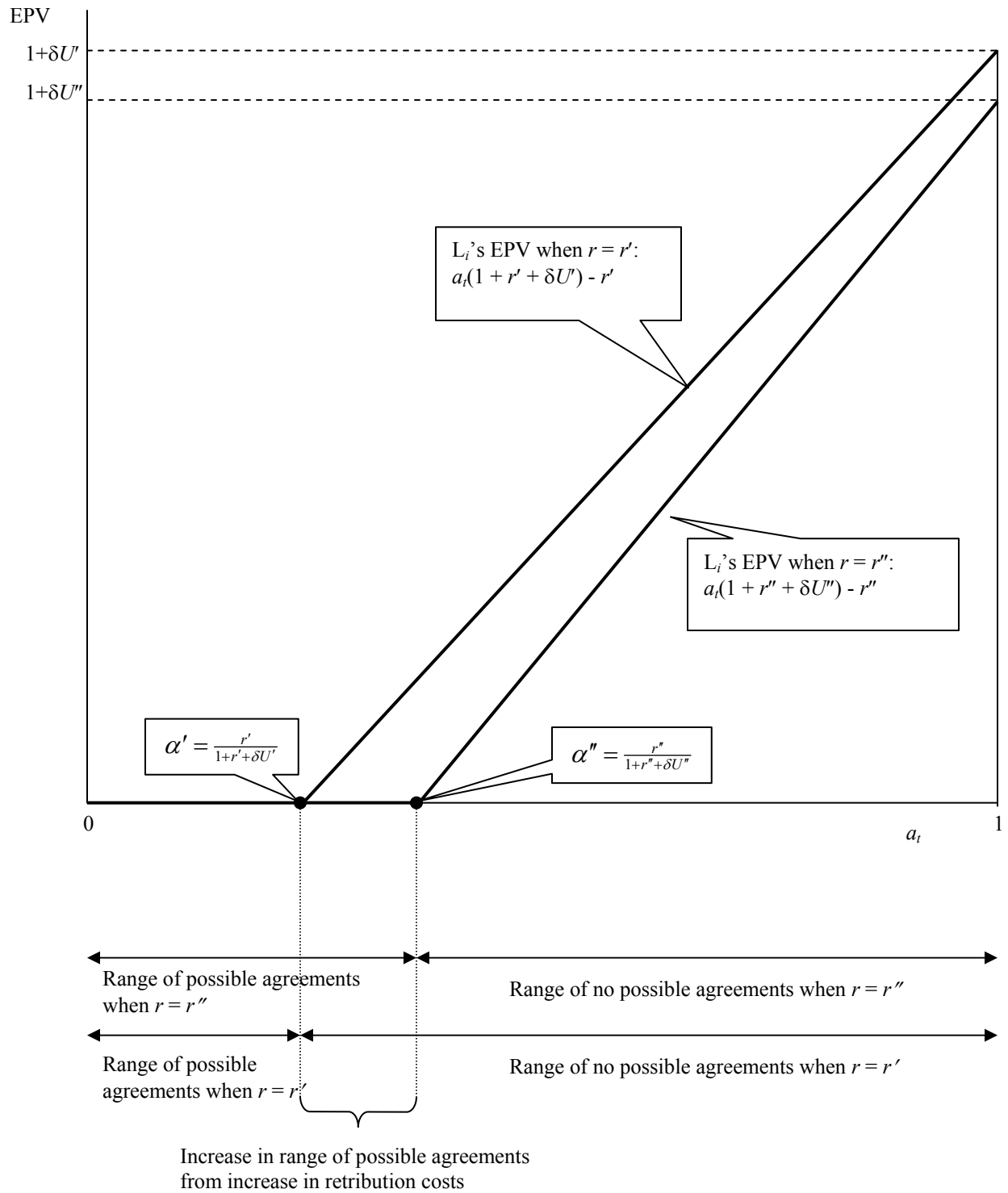


Figure 4: Differences in scale arising from different norms of succession in two different states

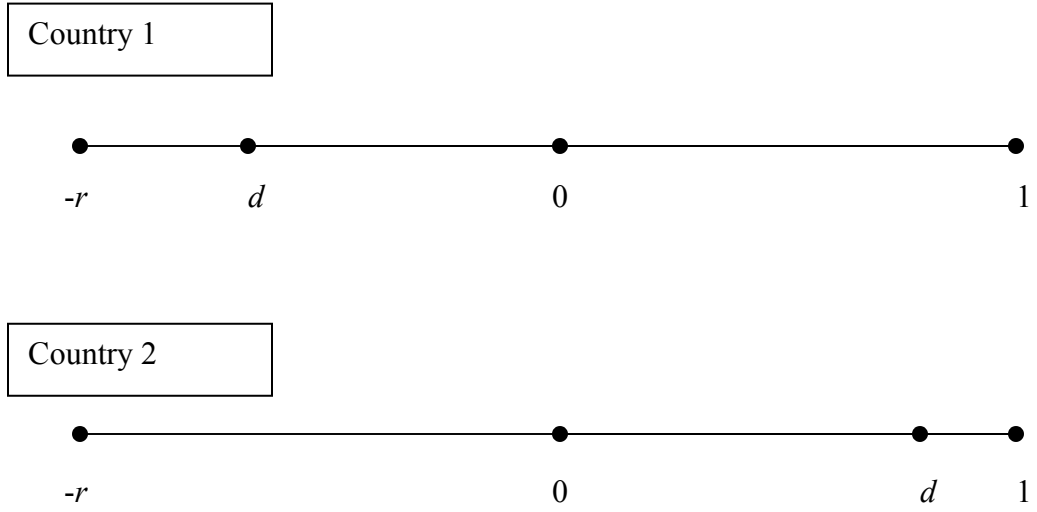


Figure 5: Periods of atrocities for various values of r and d .

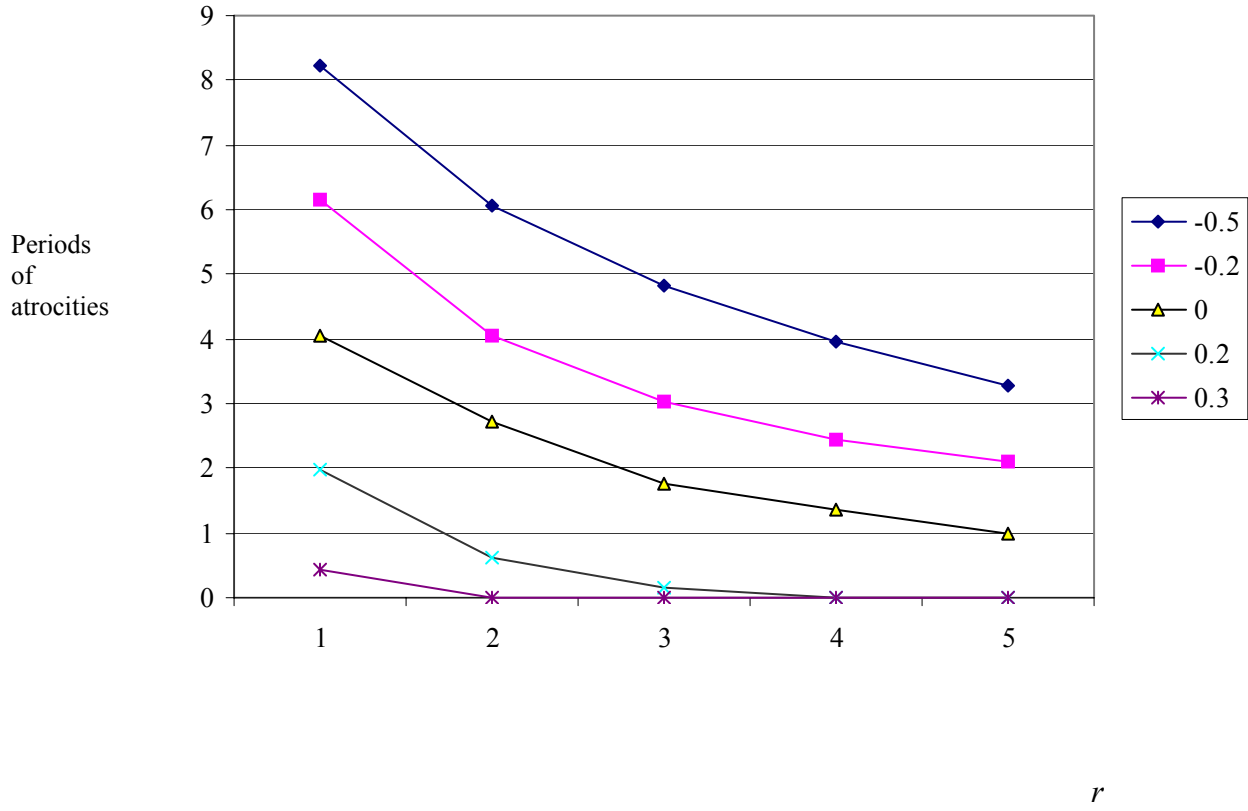


Figure 6: Periods of atrocities for various values of r and q .

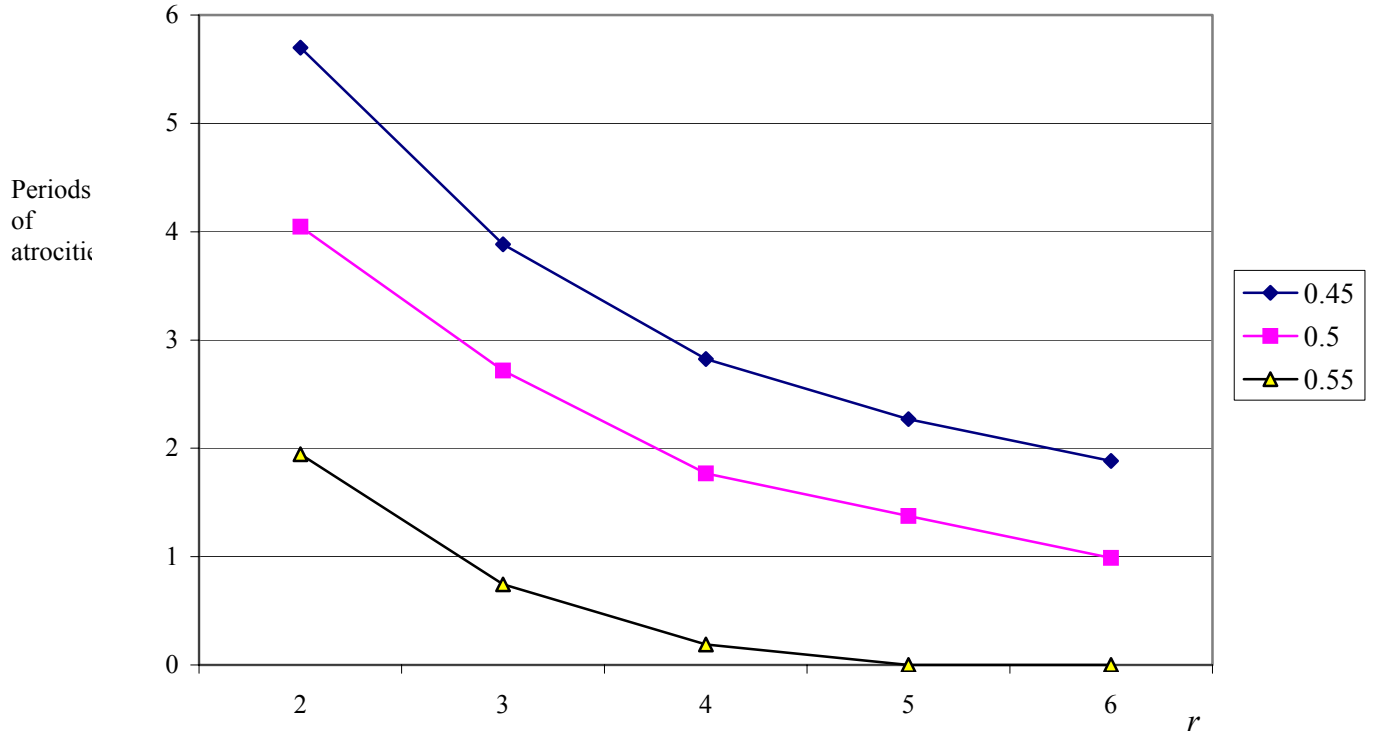


Figure 7: Post-indictment subgame (see footnote to Figure 1)

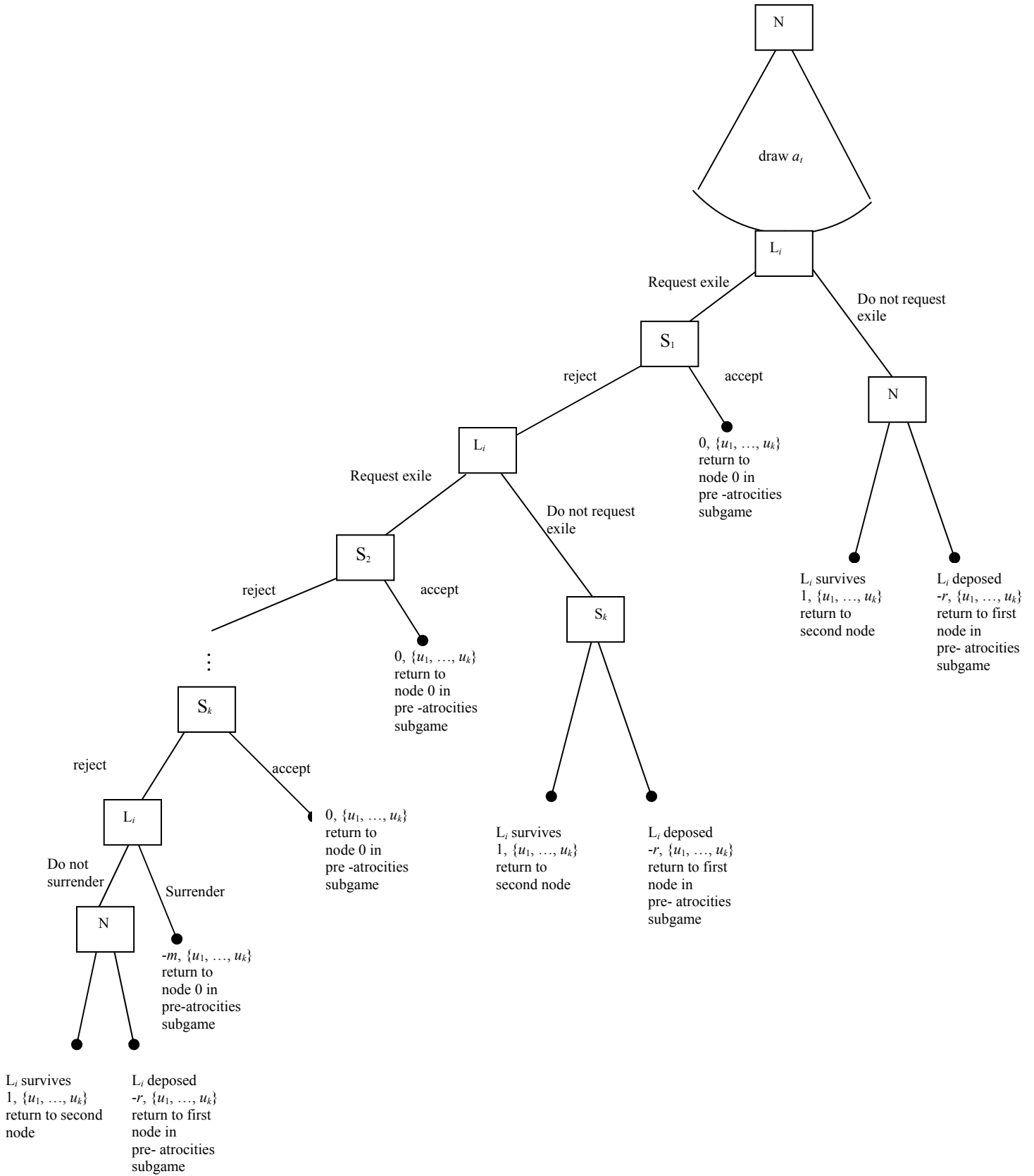
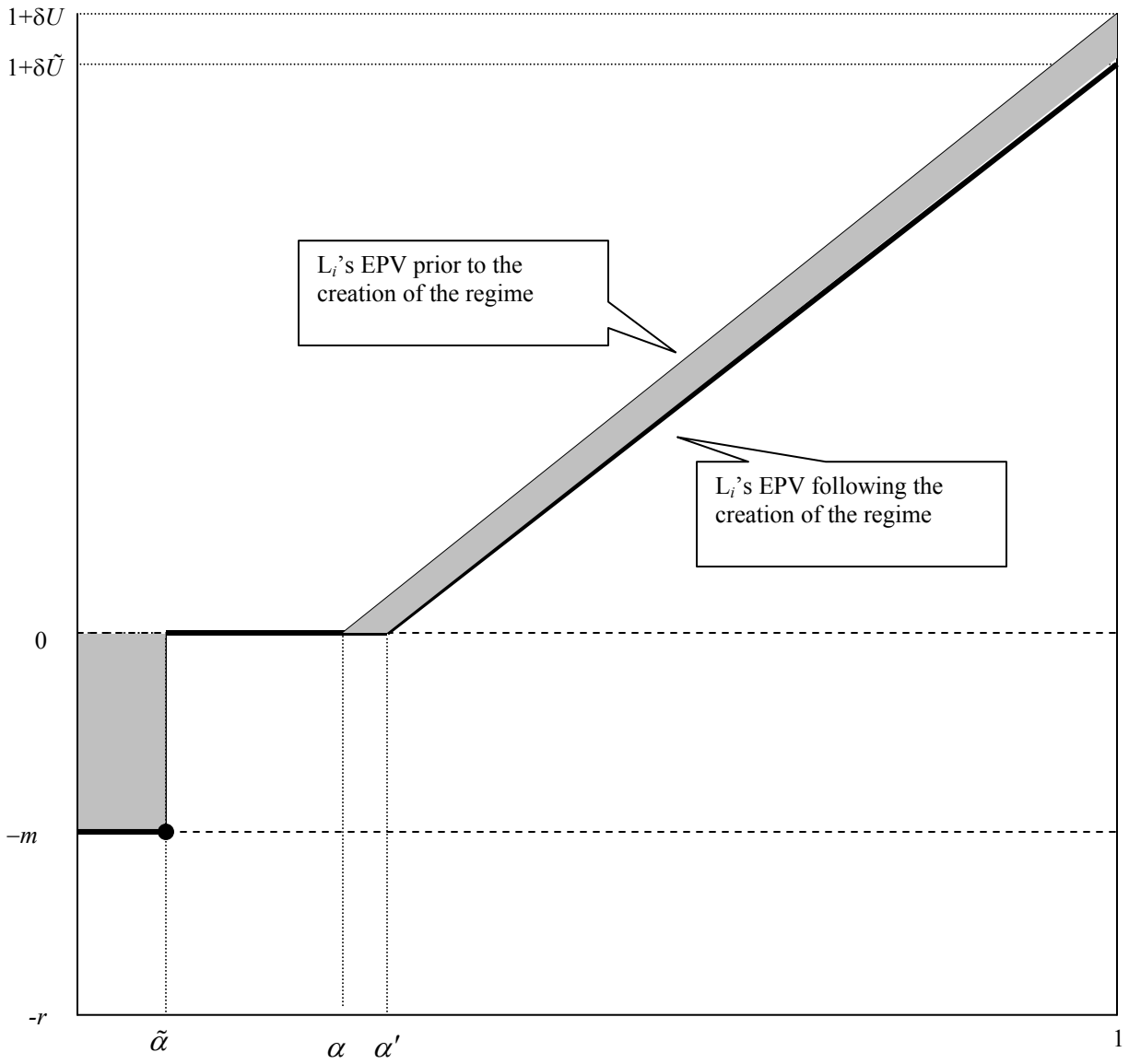


Figure 8: Total effect of the regime when L_i believes outcome 3i will occur when $a_i < \tilde{\alpha}$



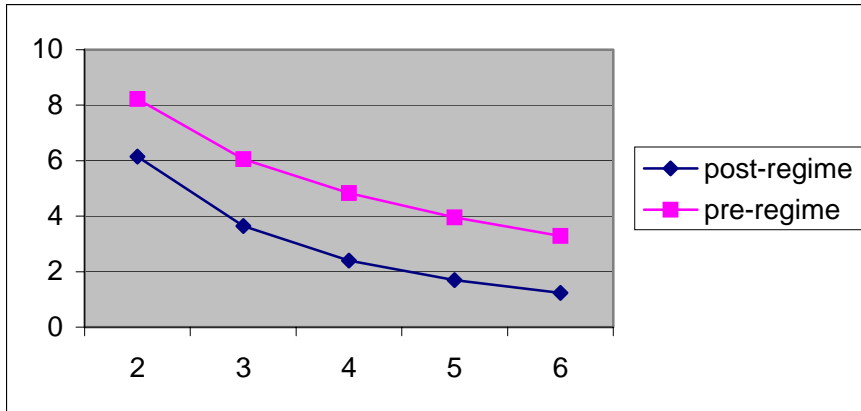
F will not accept L_i 's request for asylum for values of a_i in this range.

F will accept L_i 's request for asylum for values of a_i in this range.

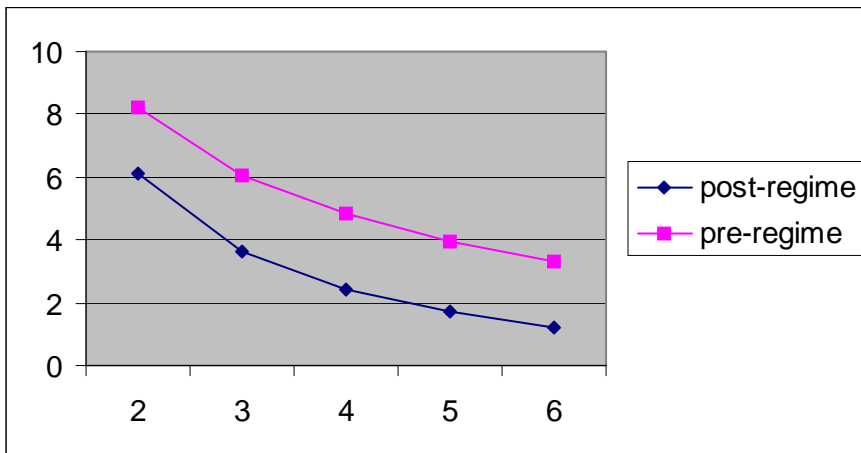
L_i will not request exile for a_i in this range.

Figure 9: Periods of atrocities before and after the creation of the regime

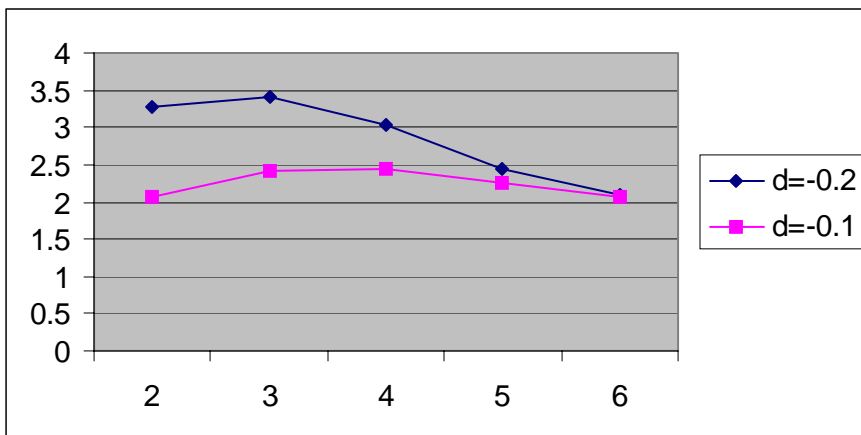
Part A: $d = -0.2$



Part B: $d = -0.5$



Part C: Reduction in Periods of atrocities



WORD COUNT: 13,026