The Politics of Punishment:
Why Dictators Join the International Criminal Court*

Leslie Johns† and Francesca Parente‡

6 February 2022

Online Appendix

Contents

---

*For helpful feedback, we thank Peter Rosendorff and Jim Vreeland. We also thank participants in the 2021 annual meeting of the American Political Science Association, and a seminar at New York University. Rebecca Pol provided excellent research assistance for this project. This research was funded by a grant from the Burkle Center for International Relations at UCLA and the Niehaus Center for Globalization and Governance at Princeton University.

†Professor, UCLA

‡Assistant Professor, Christopher Newport University

1 Theory

1.1 Main Model

1.1.1 Assumptions

We focus on strategic interactions between two actors: the government, G, and a rebel group, R. We assume that these two actors are competing for power over the state, which yields a payoff of $W > 0$. To isolate our causal mechanism, we assume that both players have complete information all aspects of the game.

The game begins when the government decides whether to join the ICC. We assume that the only impact of this publicly-observed decision is to raise the cost of violence to both players. After the government makes its decision, each player simultaneously chooses a level of effort to invest in violence, $e_i \geq 0$. These effort levels in turn determine the likelihood that each player wins power over the state. Namely, we assume that the probability that the government survives in power is:

$$\pi\left(e_G, e_R\right) = \frac{e_G}{e_G + e_R}$$

The probability that the rebel group gains power over the state is therefore: $1 - \pi$.

The payoffs of the players from the various outcomes depend on whether the government previously joined the ICC. If the government has not previously joined the ICC, we assume that the government's unit cost of violence is $c > 0$. We interpret this parameter as the strength of the government, relative to the rebel group. We assume that when the government is relatively strong, it can easily deploy violence at little cost. The corresponding government cost of violence, c, is therefore small. In contrast, we assume that when the government is relatively weak, violence is more costly in economic and political terms. The corresponding government cost of violence, c, is therefore large. We normalize the cost of violence for the rebel group to be equal to 1.[1]

In contrast, if the government has previously joined the ICC, we assume that the costs of violence are higher for both the government and the rebel group. Namely, we let parameter $\rho_H > 0$ denote the added unit cost of violence to the rebel group, and $\rho_L > 0$ denote the added unit cost of violence to the government. Higher values of ρ therefore correspond to situations in which players expect for the ICC to have more ability and willingness to prosecute international crimes. We additionally assume that $\rho_L < \rho_H$.

[1]This normalizing assumption does not affect any of our results. It simply reduces the number of parameters in the model.

The expected payoffs for both players from the various outcomes of the game are shown in Table A1.

Table A1: Expected Utility from Outcomes in the Theoretical Model

Actor	Don't Join	Join
Government	$W\pi - ce_G$	$W(1-\pi) - (c+\rho_L)\, e_G$
Rebel group	$W\pi - e_R$	$W(1-\pi) - (1+\rho_H)\, e_R$

1.1.2 Results

<u>**Selection**</u>

Proposition 1. *In equilibrium, weaker governments (with high values of c) join the ICC, while stronger governments (with lower values of c) do not join.*

Proof of Proposition 1. To solve the model generally, let c_i denote the cost of effort to actor i, and $\pi_i = \frac{e_i}{e_i+e_j}$ denote the probability that actor i gains the payoff from winning political power. Then the generic utility function for player i is:

$$U_i\left(e_i, e_j\right) = W\pi_i - c_i e_i$$

Maximizing the utility function with respect to the choice variable e_i yields the best response function:

$$e_i\left(e_j\right) = \left(\frac{e_j W}{c_i}\right)^{\frac{1}{2}} - e_j$$

Taking the intersection of the best response functions yields the equilibrium effort level:

$$e_i^* = \frac{W c_j}{\left(c_i + c_j\right)^2}$$

For the two different subgames, this yields the following behavior and outcomes:

	Don't Join	Join
Government effort (e_G)	$\frac{W}{(1+c)^2}$	$\frac{W(1+\rho_H)}{(1+c+\rho_L+\rho_H)^2}$
Rebel effort (e_R)	$\frac{Wc}{(1+c)^2}$	$\frac{W(c+\rho_L)}{(1+c+\rho_L+\rho_H)^2}$
Government survival probability (π)	$\frac{1}{1+c}$	$\frac{1+\rho_H}{1+c+\rho_L+\rho_H}$

So the government's expected utility from its two choices is:

$$\begin{aligned}
EU_G\,(\text{Don't Join}) &= W\left[\frac{1}{1+c}\right] - c\left[\frac{W}{(1+c)^2}\right] = \frac{W}{(1+c)^2} \\
EU_G\,(\text{Join}) &= W\left[\frac{1+\rho_H}{1+c+\rho_L+\rho_H}\right] - (c+\rho_L)\left[\frac{W\,(1+\rho_H)}{(1+c+\rho_L+\rho_H)^2}\right] = \frac{W(1+\rho_H)^2}{(1+c+\rho_L+\rho_H)^2}
\end{aligned}$$

The government will thus be willing to join iff:

$$\Psi \equiv (1+\rho_H)^2\,(1+c)^2 - (1+c+\rho_L+\rho_H)^2 \geq 0$$

Note that:

$$\begin{aligned}
\Psi\,(c=0) &= (1+\rho_H)^2 - (1+\rho_H+\rho_L)^2 < 0 \\
\frac{\partial \Psi}{\partial c} &= 2\,(1+\rho_H)^2\,(1+c) - 2\,(c+\rho_L+\rho_H+1) \\
&= 2\left[\rho_H - \rho_L + \rho_H^2 + c\left(2\rho_H + \rho_H^2\right)\right] > 0 \\
\lim_{c\to\infty} \Psi &= \lim_{c\to\infty} \left\{\left(1+2\rho_H+\rho_H^2\right)(1+c)^2 - \left[(1+c)^2 + 2\,(1+c)\,(\rho_L+\rho_H) + (\rho_L+\rho_H)^2\right]\right\} \\
&= \lim_{c\to\infty} \left\{\left(2\rho_H+\rho_H^2\right)(1+c)^2 - 2\,(1+c)\,(\rho_L+\rho_H) - (\rho_L+\rho_H)^2\right\} \\
&= \lim_{c\to\infty} (1+c) \times \lim_{c\to\infty} \left[\left(2\rho_H+\rho_H^2\right)(1+c) - 2\,(\rho_L+\rho_H)\right] - (\rho_L+\rho_H)^2 = \infty
\end{aligned}$$

So by the intermediate value theorem, there exists a unique $\bar{c} > 0$ such that $\Psi(c) < 0$ for all $c < \bar{c}$ and $\Psi(c) > 0$ for all $c > \bar{c}$. So higher values of c are (weakly) more likely to join than lower values of c. □

Violence

Proposition 2. *Joining the ICC always lowers the total violence in the state.*

Proof of Proposition 2. Total violence if the government does not join the ICC is:

$$\frac{W}{(1+c)^2} + \frac{Wc}{(1+c)^2} = \frac{W}{1+c}$$

Total violence if the government joins the ICC is:

$$\frac{W(1+\rho_H)}{(1+c+\rho_L+\rho_H)^2} + \frac{W(c+\rho_L)}{(1+c+\rho_L+\rho_H)^2} = \frac{W}{1+c+\rho_L+\rho_H}$$

Note that the latter quantity is always less than the former quantity. □

<u>**Survival**</u>

Proposition 3. *For governments that select into the ICC, joining the ICC increases the probability of surviving in office.*

Proof of Proposition 3. The probability that the government survives in office if it joins increases (relative to not joining) iff:

$$\frac{1}{1+c} < \frac{1+\rho_H}{1+c+\rho_L+\rho_H} \quad \Leftrightarrow \quad 1+c+\rho_L+\rho_H < (1+\rho_H)(1+c) \quad \Leftrightarrow \quad \frac{\rho_L}{\rho_H} < c$$

Note that:

$$\begin{aligned} \Psi\left(c = \frac{\rho_L}{\rho_H}\right) &= (1+\rho_H)^2\left(1+\frac{\rho_L}{\rho_H}\right)^2 - \left(1+\frac{\rho_L}{\rho_H}+\rho_L+\rho_H\right)^2 \\ &= \frac{(1+\rho_H)^2(\rho_L+\rho_H)^2}{\rho_H^2} - \left[\frac{(1+\rho_H)(\rho_L+\rho_H)}{\rho_H}\right]^2 = 0 \end{aligned}$$

So for governments that select into the ICC ($c > \bar{c}$), joining the ICC increases the probability of surviving in office. □

1.2 Model Extension: Punish the Winner Less than the Loser

1.2.1 Assumptions

Continue to assume the same players, actions, and preferences if the government does not join the Rome Statute

However, assume that if the government joins the Rome Statute, then the ICC costs are imposed after the conflict ends. Namely, suppose that the group that hold power pays a low unit cost, ρ_L, while the group that does not hold power pays a high unit cost, ρ_H, where $0 < \rho_L < \rho_H$.

Then expected utility for the subgame in which the government signs the Rome Statute is:

$$\begin{aligned} U_i(e) &= W\pi_i - c_i e_i - \rho_L e_i \pi_i - \rho_H e_i (1-\pi_i) \\ &= W\left(\frac{e_i}{e_i+e_j}\right) - c_i e_i - \rho_L\left(\frac{e_i^2}{e_i+e_j}\right) - \rho_H\left(\frac{e_i e_j}{e_i+e_j}\right) \end{aligned}$$

1.2.2 Results

Selection

Proposition 4. *In equilibrium, weaker governments (with high values of c) join the ICC, while stronger governments (with lower values of c) do not join.*

Proof of Proposition 4. If the government does not join the Rome Statute, then subgame behavior matches that in Proposition 1.

However, if the government joins the Rome Statute, then the following optimization process applies. We begin with the first- and second-order conditions:

$$\begin{aligned} \frac{\partial U_i(e)}{\partial e_i} &= W\left[\frac{e_j}{(e_i+e_j)^2}\right] - c_i - \rho_L\left[\frac{e_i^2+2e_ie_j}{(e_i+e_j)^2}\right] - \rho_H\left[\frac{e_j^2}{(e_i+e_j)^2}\right] \\ \frac{\partial^2 U_i(e)}{\partial e_i^2} &= \frac{-2We_j}{(e_i+e_j)^3} - \rho_L\left[\frac{(e_i+e_j)^2(2e_i+2e_j) - \left(e_i^2+2e_ie_j\right)2(e_i+e_j)}{(e_i+e_j)^4}\right] + \frac{2\rho_H e_j^2}{(e_i+e_j)^3} \\ &= \frac{2e_j\left[(\rho_H-\rho_L)e_j - W\right]}{(e_i+e_j)^3} \end{aligned}$$

Note that:

$$\begin{aligned} \frac{\partial U_i(e)}{\partial e_i}(e_i=0) &= W\left(\frac{1}{e_j}\right) - c_i - \rho_H \\ &\Rightarrow \quad \frac{\partial U_i(e)}{\partial e_i}(e_i=0|e_j>0) > 0 \quad \Leftrightarrow \quad \frac{W}{c_i+\rho_H} > e_j \\ \frac{\partial^2 U_i(e)}{\partial e_i^2} &< 0 \Leftrightarrow e_j \in \left(0,\ \frac{W}{\rho_H-\rho_L}\right) \end{aligned}$$

So the binding constraint to identify an optimal e_i is:

$$e_j \in \left(0,\ \frac{W}{c_i+\rho_H}\right)$$

This translates to the system of constraints:

$$e_G \in \left(0,\ \frac{W}{1+\rho_H}\right)$$
$$e_R \in \left(0,\ \frac{W}{c+\rho_H}\right)$$

We can now identify the best response function for each player:

$$\frac{\partial U_i(e)}{\partial e_i} = 0 \Leftrightarrow We_j - c_i\left(e_i+e_j\right)^2 - \rho_L\left(e_i^2+2e_ie_j\right) - \rho_H e_j^2 = 0$$
$$\Leftrightarrow \left(c_i+\rho_L\right)e_i^2 + 2\left(c_i+\rho_L\right)e_ie_j + \left[\left(c_i-\rho_H\right)e_j^2 - We_j\right] = 0$$

$$\begin{aligned}
e_i\left(e_j\right) &= \frac{-2\left(c_i+\rho_L\right)e_j \pm \sqrt{4\left(c_i+\rho_L\right)^2 e_j^2 - 4\left(c_i+\rho_L\right)\left[\left(c_i+\rho_H\right)e_j^2 - We_j\right]}}{2\left(c_i+\rho_L\right)} \\
&= \frac{\left(c_i+\rho_L\right)^{\frac{1}{2}}\left[We_j+\left(\rho_L-\rho_H\right)e_j^2\right]^{\frac{1}{2}} - \left(c_i+\rho_L\right)e_j}{\left(c_i+\rho_L\right)} \\
&= \left[\frac{We_j+\left(\rho_L-\rho_H\right)e_j^2}{\left(c_i+\rho_L\right)}\right]^{\frac{1}{2}} - e_j
\end{aligned}$$

This translates to the best response functions:

$$\begin{aligned}
e_G\left(e_R\right) &= \left(\frac{We_R+\left(\rho_L-\rho_H\right)e_R^2}{c+\rho_L}\right)^{\frac{1}{2}} - e_R \\
e_R\left(e_G\right) &= \left(\frac{We_G+\left(\rho_L-\rho_H\right)e_G^2}{1+\rho_L}\right)^{\frac{1}{2}} - e_G
\end{aligned}$$

One possible solution to this system is: $e_G = e_R = 0$. However, these values do not maximize the utility function because of the second-order condition. Alternatively, substitution yields:

$$
\begin{aligned}
e_G &= \left(\frac{W\left[\left(\frac{We_G+(\rho_L-\rho_H)e_G^2}{1+\rho_L}\right)^{\frac{1}{2}} - e_G\right] + (\rho_L-\rho_H)\left[\left(\frac{We_G+(\rho_L-\rho_H)e_G^2}{1+\rho_L}\right)^{\frac{1}{2}} - e_G\right]^2}{c+\rho_L} \right)^{\frac{1}{2}} \\
&\quad - \left[\left(\frac{We_G+(\rho_L-\rho_H)e_G^2}{1+\rho_L}\right)^{\frac{1}{2}} - e_G\right] \\
\Leftrightarrow & \quad \frac{W\left[\left(\frac{We_G+(\rho_L-\rho_H)e_G^2}{1+\rho_L}\right)^{\frac{1}{2}} - e_G\right] + (\rho_L-\rho_H)\left[\left(\frac{We_G+(\rho_L-\rho_H)e_G^2}{1+\rho_L}\right)^{\frac{1}{2}} - e_G\right]^2}{c+\rho_L} \\
&\quad = \frac{We_G+(\rho_L-\rho_H)e_G^2}{1+\rho_L} \\
\Leftrightarrow & \quad 0 = W\left[\left(\frac{W+(\rho_L-\rho_H)e_G}{1+\rho_L}\right)^{\frac{1}{2}} - e_G^{\frac{1}{2}}\right] + (\rho_L-\rho_H)e_G^{\frac{1}{2}}\left[\left(\frac{W+(\rho_L-\rho_H)e_G}{1+\rho_L}\right)^{\frac{1}{2}} - e_G^{\frac{1}{2}}\right]^2 \\
&\quad - (c+\rho_L)e_G^{\frac{1}{2}}\left(\frac{W+(\rho_L-\rho_H)e_G}{1+\rho_L}\right)
\end{aligned}
$$

Define

$$
\begin{aligned}
\Delta(x) &\equiv W\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right] + (\rho_L-\rho_H)x^{\frac{1}{2}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]^2 - (c+\rho_L)x^{\frac{1}{2}}\Upsilon(x) \\
&\text{where} \quad \Upsilon(x) \equiv \frac{W+(\rho_L-\rho_H)x}{1+\rho_L}
\end{aligned}
$$

Note the following property of the Δ function:

$$
\Delta(0) = W\Upsilon(0)^{\frac{1}{2}} = \frac{W^{\frac{3}{2}}}{(1+\rho_L)^{\frac{1}{2}}} > 0
$$

Also note that:

$$
\Upsilon\left(\frac{W}{1+\rho_H}\right)^{\frac{1}{2}} - \left(\frac{W}{1+\rho_H}\right)^{\frac{1}{2}} = \left(\frac{W+(\rho_L-\rho_H)\left(\frac{W}{1+\rho_H}\right)}{1+\rho_L}\right)^{\frac{1}{2}} - \left(\frac{W}{1+\rho_H}\right)^{\frac{1}{2}} = 0
$$

So:

$$
\Delta\left(\frac{W}{1+\rho_H}\right) = -(c+\rho_L)\left(\frac{W}{1+\rho_H}\right)^{\frac{1}{2}}\Upsilon\left(\frac{W}{1+\rho_H}\right) = -(c+\rho_L)\left(\frac{W}{1+\rho_H}\right)^{\frac{3}{2}} < 0
$$

Finally, note that:

$$\Delta'(x) = W\left[\frac{\Upsilon'(x)}{2\Upsilon(x)^{\frac{1}{2}}} - \frac{1}{2x^{\frac{1}{2}}}\right] + (\rho_L - \rho_H)\left\{x^{\frac{1}{2}}2\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]\left[\frac{\Upsilon'(x)}{2\Upsilon(x)^{\frac{1}{2}}} - \frac{1}{2x^{\frac{1}{2}}}\right] + \frac{1}{2x^{\frac{1}{2}}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]^2\right\}$$
$$- (c+\rho_L)\left[x^{\frac{1}{2}}\Upsilon'(x) + \frac{\Upsilon(x)}{2x^{\frac{1}{2}}}\right]$$
$$= \frac{W}{2}\left[\frac{\left(\frac{\rho_L-\rho_H}{1+\rho_L}\right)}{\left(\frac{W+(\rho_L-\rho_H)x}{1+\rho_L}\right)^{\frac{1}{2}}} - \frac{1}{x^{\frac{1}{2}}}\right]$$
$$+ (\rho_L - \rho_H)\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]\left[\frac{\left(\frac{\rho_L-\rho_H}{1+\rho_L}\right)x^{\frac{1}{2}}}{\left(\frac{W+(\rho_L-\rho_H)x}{1+\rho_L}\right)^{\frac{1}{2}}} + \frac{\Upsilon(x)^{\frac{1}{2}}}{2x^{\frac{1}{2}}} - \frac{3}{2}\right]$$
$$- \left(\frac{c+\rho_L}{2x^{\frac{1}{2}}}\right)\left[2x\left(\frac{\rho_L-\rho_H}{1+\rho_L}\right) + \left(\frac{W+(\rho_L-\rho_H)x}{1+\rho_L}\right)\right]$$
$$= \frac{W}{2}\left[\frac{\rho_L-\rho_H}{[W+(\rho_L-\rho_H)x]^{\frac{1}{2}}(1+\rho_L)^{\frac{1}{2}}} - \frac{1}{x^{\frac{1}{2}}}\right]$$
$$+ (\rho_L - \rho_H)\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]\left[\frac{(\rho_L-\rho_H)x^{\frac{1}{2}}}{[W+(\rho_L-\rho_H)x]^{\frac{1}{2}}(1+\rho_L)^{\frac{1}{2}}} + \frac{\Upsilon(x)^{\frac{1}{2}}}{2x^{\frac{1}{2}}} - \frac{3}{2}\right]$$
$$- \left(\frac{c+\rho_L}{2x^{\frac{1}{2}}}\right)\left(\frac{W-3(\rho_H-\rho_L)x}{1+\rho_L}\right)$$

Note that $\lim_{W\to\infty}\Upsilon(x) = \infty$. So:

$$\lim_{W\to\infty}\Delta'(x) = -\frac{1}{2x^{\frac{1}{2}}}\lim_{W\to\infty}[W] - (\rho_H - \rho_L)\left[\lim_{W\to\infty}\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]\left[\lim_{W\to\infty}\frac{\Upsilon(x)^{\frac{1}{2}}}{2x^{\frac{1}{2}}} - \frac{3}{2}\right]$$
$$- \left(\frac{c+\rho_L}{2x^{\frac{1}{2}}}\right)\lim_{W\to\infty}\left(\frac{W-3(\rho_H-\rho_L)x}{1+\rho_L}\right) = -\infty < 0$$

By the IVT, for large W there exists a unique $x \in \left(0, \frac{W}{1+\rho_H}\right)$ that solves $\Delta(x) = 0$. This is the government's equilibrium level of effort.

Note that:

$$\frac{\partial x}{\partial c} = \frac{-\Delta_c}{\Delta_x} = \frac{x^{\frac{1}{2}}\Upsilon(x)}{\Delta_x} < 0 \quad \text{for large } W$$

The total equilibrium violence for this subgame is:

$$e_G^* + e_R^* = \left[\frac{Wx + (\rho_L+\rho_H)x^2}{(1+\rho_L)}\right]^{\frac{1}{2}}$$

The government probability of survival for this subgame is:

$$\pi_G = \frac{e_G}{e_G + e_R} = \frac{e_G}{\left[\frac{We_G + (\rho_L+\rho_H)e_G^2}{(1+\rho_L)}\right]^{\frac{1}{2}}} = \frac{e_G^{\frac{1}{2}}(1+\rho_L)^{\frac{1}{2}}}{\left[W + (\rho_L+\rho_H)\, e_G\right]^{\frac{1}{2}}} = \left[\frac{x}{\Upsilon(x)}\right]^{\frac{1}{2}}$$

The government's expected utility from this subgame is:

$$\begin{aligned} U_G(x) &= W\pi_G - (c+\rho_H)\,x + (\rho_H - \rho_L)\,x\pi_G \\ &= W\left[\frac{x}{\Upsilon(x)}\right]^{\frac{1}{2}} + (\rho_H - \rho_L)\left[\frac{x^{\frac{3}{2}}}{\Upsilon(x)^{\frac{1}{2}}}\right] - (c+\rho_H)\,x \end{aligned}$$

Note that:

$$\begin{aligned} \Delta(x) = W\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right] + (\rho_L+\rho_H)\,x^{\frac{1}{2}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]^2 - (c+\rho_L)\,x^{\frac{1}{2}}\Upsilon(x) = 0 \\ \Leftrightarrow (c+\rho_L)\,x^{\frac{1}{2}}\Upsilon(x) = W\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right] + (\rho_L - \rho_H)\,x^{\frac{1}{2}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]^2 \\ \Leftrightarrow c = W\left[\frac{\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}}{x^{\frac{1}{2}}\Upsilon(x)}\right] + (\rho_L - \rho_H)\frac{\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]^2}{\Upsilon(x)} - \rho_L \\ \Leftrightarrow (c+\rho_H)\,x = Wx^{\frac{1}{2}}\left[\frac{\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}}{\Upsilon(x)}\right] + (\rho_H - \rho_L)\,x\left[\frac{\Upsilon(x) - \left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]^2}{\Upsilon(x)}\right] \end{aligned}$$

So the government's expected utility from this subgame–given equilibrium behavior–is equivalent to:

$$\begin{aligned} U_G(x) &= W\left[\frac{x^{\frac{1}{2}}}{\Upsilon(x)^{\frac{1}{2}}}\right] + (\rho_H - \rho_L)\left[\frac{x^{\frac{3}{2}}}{\Upsilon(x)^{\frac{1}{2}}}\right] \\ &\quad - \left\{Wx^{\frac{1}{2}}\left[\frac{\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}}{\Upsilon(x)}\right] + (\rho_H - \rho_L)\,x\left[\frac{\Upsilon(x) - \left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]^2}{\Upsilon(x)}\right]\right\} \\ &= W\left[\frac{x}{\Upsilon(x)}\right] - (\rho_H - \rho_L)\left[\frac{x^{\frac{3}{2}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]}{\Upsilon(x)}\right] \end{aligned}$$

The government wants to join the ICC iff:

$$\theta(c) \equiv W\left[\frac{x}{\Upsilon(x)}\right] - (\rho_H - \rho_L)\left[\frac{x^{\frac{3}{2}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]}{\Upsilon(x)}\right] - \frac{W}{(1+c)^2} > 0$$

Define $x_0 \equiv \lim_{c\to 0} x$. Recall that x_0 is finite because it must fall within the interval $\left(0, \frac{W}{1+\rho_H}\right)$.

So:

$$\theta(0) = W\left[\frac{x_0}{\Upsilon(x_0)}\right] - (\rho_H - \rho_L)\left[\frac{x_0^{\frac{3}{2}}\left[\Upsilon(x_0)^{\frac{1}{2}} - x_0^{\frac{1}{2}}\right]}{\Upsilon(x_0)}\right] - W < 0$$

$$\Leftrightarrow -(\rho_H - \rho_L)\, x_0^{\frac{3}{2}}\left[\Upsilon(x_0)^{\frac{1}{2}} - x_0^{\frac{1}{2}}\right] < W\left[\Upsilon(x_0) - x_0\right]$$

This always holds because $x < \Upsilon(x)$ and $x^{\frac{1}{2}} < \Upsilon(x)^{\frac{1}{2}}$ for all $x \in \left(0, \frac{W}{1+\rho_H}\right)$.

Next note that:

$$\begin{aligned}
\frac{d\theta(c)}{dc} &= \frac{\partial\theta(c)}{\partial c} + \frac{\partial\theta(c)}{\partial x}\left(\frac{\partial x}{\partial c}\right) \\
&= \frac{2W}{(1+c)^3} + \left(\frac{\partial x}{\partial c}\right)\frac{\partial}{\partial x}\left\{W\left[\frac{x}{\Upsilon(x)}\right] - (\rho_H - \rho_L)\left[\frac{x^{\frac{3}{2}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]}{\Upsilon(x)}\right]\right\} \\
&= \frac{2W}{(1+c)^3} + \frac{W^2 x^{\frac{1}{2}}}{(1+\rho_L)\,\Delta_x \Upsilon(x)} \\
&\quad - \left[\frac{(\rho_H - \rho_L)\,x}{2\Delta_x}\right]\left\{\frac{x\Upsilon'(x)}{\Upsilon(x)^{\frac{1}{2}}} - x^{\frac{1}{2}} + 3\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right] - 2x\left[\frac{\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}}{\Upsilon(x)}\right]\Upsilon'(x)\right\}
\end{aligned}$$

Note that:

$$\begin{aligned}
0 < \frac{d\theta(c)}{dc} \Leftrightarrow 0 < {} & \frac{2}{(1+c)^3} + \frac{W x^{\frac{1}{2}}}{(1+\rho_L)\,\Delta_x \Upsilon(x)} \\
& - \left[\frac{(\rho_H - \rho_L)\,x^{\frac{3}{2}}}{2\Delta_x \Upsilon(x)}\right]\left\{\frac{x\Upsilon'(x)}{W\Upsilon(x)^{\frac{1}{2}}} - \frac{x^{\frac{1}{2}}}{W} + \frac{3\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]}{W} - 2x\left[\frac{\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}}{W\Upsilon(x)}\right]\Upsilon'(x)\right\} \equiv M
\end{aligned}$$

Recall that $\lim_{W\to\infty} \Delta_x = -\infty$. Define $\hat{x} \equiv \lim_{W\to\infty} x$. Recall that $\hat{x}$ is finite because it must fall within the interval $\left(0, \frac{W}{1+\rho_H}\right)$.

Consider the components of function M:

$$\lim_{W\to\infty}\left\{\frac{2}{(1+c)^3}\right\} = \frac{2}{(1+c)^3} > 0$$

$$\lim_{W\to\infty}\left\{\frac{Wx^{\frac{1}{2}}}{(1+\rho_L)\,\Delta_x\Upsilon(x)}\right\} = \frac{\hat{x}^{\frac{1}{2}}}{(1+\rho_L)}\lim_{W\to\infty}\left\{\frac{1}{\Delta_x}\right\}\lim_{W\to\infty}\left\{\frac{W}{\Upsilon(x)}\right\}$$

$$= \frac{\hat{x}^{\frac{1}{2}}}{(1+\rho_L)}\lim_{W\to\infty}\left\{\frac{1}{\Delta_x}\right\}\lim_{W\to\infty}\left\{\frac{1+\rho_L}{1+\frac{(\rho_L-\rho_H)\hat{x}}{W}}\right\} = \hat{x}^{\frac{1}{2}}\lim_{W\to\infty}\left\{\frac{1}{\Delta_x}\right\} = 0$$

$$\lim_{W\to\infty}\left\{\frac{(\rho_H-\rho_L)\,x^{\frac{3}{2}}}{2\Delta_x\Upsilon(x)}\right\} = \frac{(\rho_H-\rho_L)\,\hat{x}^{\frac{3}{2}}}{2}\lim_{W\to\infty}\left\{\frac{1}{\Delta_x\Upsilon(x)}\right\} = 0$$

$$\lim_{W\to\infty}\left\{\frac{x\Upsilon'(x)}{W\Upsilon(x)^{\frac{1}{2}}} - \frac{x^{\frac{1}{2}}}{W} + \frac{3\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right]}{W} - 2x\left[\frac{\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}}{W\Upsilon(x)}\right]\Upsilon'(x)\right\}$$

$$= \hat{x}\Upsilon'(\hat{x})\lim_{W\to\infty}\left\{\frac{1}{W\Upsilon(\hat{x})^{\frac{1}{2}}}\right\} - \hat{x}^{\frac{1}{2}}\lim_{W\to\infty}\left\{\frac{1}{W}\right\} + 3\lim_{W\to\infty}\left\{\frac{\Upsilon(\hat{x})^{\frac{1}{2}} - \hat{x}^{\frac{1}{2}}}{W}\right\} - 2\hat{x}\Upsilon'(\hat{x})\lim_{W\to\infty}\left\{\frac{\Upsilon(\hat{x})^{\frac{1}{2}} - \hat{x}^{\frac{1}{2}}}{W\Upsilon(\hat{x})}\right\}$$

$$= 3\left[\lim_{W\to\infty}\left\{\frac{\Upsilon(\hat{x})^{\frac{1}{2}}}{W}\right\} - \hat{x}^{\frac{1}{2}}\lim_{W\to\infty}\left\{\frac{1}{W}\right\}\right] - 2\hat{x}\Upsilon'(\hat{x})\left[\lim_{W\to\infty}\left\{\frac{1}{W\Upsilon(\hat{x})^{\frac{1}{2}}}\right\} - \hat{x}^{\frac{1}{2}}\lim_{W\to\infty}\left\{\frac{1}{W\Upsilon(\hat{x})}\right\}\right]$$

$$= 3\lim_{W\to\infty}\left\{\frac{\left(\frac{W+(\rho_L-\rho_H)\hat{x}}{1+\rho_L}\right)^{\frac{1}{2}}}{W}\right\} = \frac{3}{(1+\rho_L)^{\frac{1}{2}}}\lim_{W\to\infty}\left\{\frac{\left[W+(\rho_L-\rho_H)\,\hat{x}\right]^{\frac{1}{2}}}{W}\right\}$$

$$= \frac{3}{(1+\rho_L)^{\frac{1}{2}}}\lim_{W\to\infty}\left\{\frac{1}{2\left[W+(\rho_L-\rho_H)\,\hat{x}\right]^{\frac{1}{2}}}\right\} = 0$$

So:

$$\lim_{W\to\infty} M = \frac{2}{(1+c)^3} > 0$$

Finally, note that:

$$0 < \theta(c) \quad\Leftrightarrow\quad 0 < Wx - (\rho_H-\rho_L)\,x^{\frac{3}{2}}\left[\Upsilon(x)^{\frac{1}{2}} - x^{\frac{1}{2}}\right] - \frac{\Upsilon(x)\,W}{(1+c)^2} \equiv N(c)$$

Define $\bar{x} \equiv \lim_{c\to\infty} x(c)$. Note that $0 \leq \bar{x}$ and $\bar{x}$ is finite. Then:

$$\lim_{c\to\infty} N(c) = W\bar{x} - (\rho_H-\rho_L)\,\bar{x}^{\frac{3}{2}}\left[\Upsilon(\bar{x})^{\frac{1}{2}} - \bar{x}^{\frac{1}{2}}\right]$$

So:

$$0 < \lim_{c \to \infty} N(c) \quad \Leftrightarrow \quad (\rho_H - \rho_L)\, \bar{x}^{\frac{1}{2}} \left[\Upsilon(\bar{x})^{\frac{1}{2}} - \bar{x}^{\frac{1}{2}} \right] < W$$

This holds for large W. So by the IVT, there is a unique crossing for large W at which $\theta(c) = 0$. The government joins iff c is sufficiently large. □

<u>**Violence**</u>

Proposition 5. *Joining the ICC always lowers the total violence in the state.*

Proof of Proposition 5. Total violence if the government does not join the ICC is:

$$\frac{W}{1+c}$$

Total violence if the government joins the ICC is:

$$\left[\frac{Wx + (\rho_L + \rho_H)\, x^2}{(1+\rho_L)} \right]^{\frac{1}{2}}$$

where x is implicitly defined by $\Delta(x) = 0$ from the Proof of Proposition 4. □

2 Empirics

2.1 H1: Political Competition Increases Joining

These descriptive statistics describe the data in our main analysis.

Table A2: Descriptive Statistics for Models (1)–(3)

Variable	Mean	Minimum	Maximum	N
Year	2007	1998	2018	1,194
ICC join	0.03	0	1	1,194
Political competition	2.08	0	4	1,100
Log(GDP per capita)	7.90	5.23	11.15	1,107
Rule of law	−0.71	−2.61	1.84	1,194
Total violence	0.81	0	7	1,194
Inter-state violence	0.09	0	6	1,194
Intra-state violence	0.72	0	7	1,194
Foreign aid*	19.31	−21.41	24.66	1,095
Africa	0.46	0	1	1,194
Middle East/North Africa	0.23	0	1	1,194
Asia	0.26	0	1	1,194
Central & South America	0.05	0	1	1,194

*hyperbolic sine transformation

As a robustness check for our dependent variable, we coded whether/when each state accepted jurisdiction of the ICC. For most states, this is equivalent to ratification of the Rome Statute. However, a handful of states accepted jurisdiction using other rare procedures. These include Ivory Coast in 2003 and Ukraine in 2013. We then performed the statistical tests for H1 in our main analysis. Our results remained robust:

Table A3: Political Competition Increases Ratification of the Rome Statute

Dependent Variable: Years to Ratification

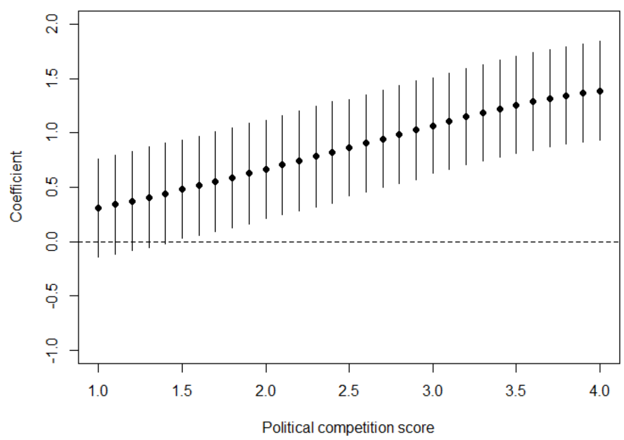
	(A1)	(A2)	(A3)
Explanatory Variable			
Political competition	0.84***	0.69**	0.63**
	(0.26)	(0.28)	(0.28)
Control Variables			
Rule of law		0.56	0.72
		(0.46)	(0.50)
Log (GDP per capita)		−0.30	−0.48
		(0.19)	(0.30)
Foreign aid		0.21	0.02
		(0.22)	(0.11)
Total violence		0.02	
		(0.17)	
Intra-state violence			0.03
			(0.18)
Inter-state violence			−0.29
			(0.61)
Region dummies	Yes	Yes	Yes
Events	24	23	23
States	90	85	85
Observations (state-year)	1,108	940	940

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

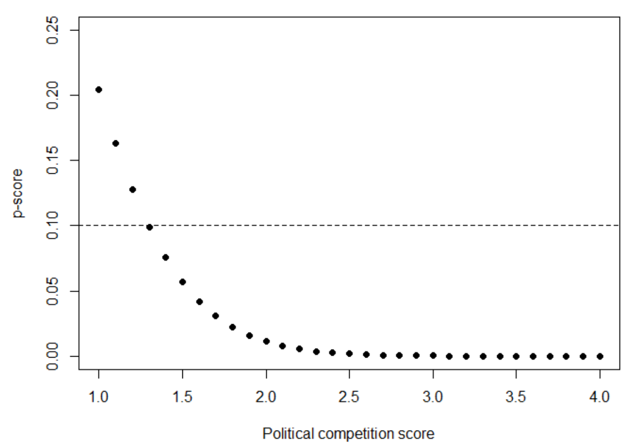
As a robustness check for our explanatory variable, we re-ran Model (1) using hypothetical political competition scores for the years in which they were missing. Note that this is an extremely conservative estimate of the imputed coefficients because an actual imputation model would use information from prior years available for the state to make its estimate. For example, Afghanistan has data for 1998-2001 but not 2002-2003; an imputation model might use information from 1998-2001 to create the scores for 2002-2003. Here I have assumed a completely naive model in which 2002-2003 scores are coded as any hypothetical political competition score, regardless of what the score was in previous years. Even under this extremely conservative and naive assumption, the coefficient would still be statistically significant at almost any hypothetically imputed political competition score value. The average political competition score in the dataset is 2.08; political competition scores below that are likely much more restrictive than what imputation software would provide for states on which there is zero data.

Figure A1: Results for Imputed Values of Political Competition

(a) Coefficients



(b) p-Values



Note: The x-axis indicates the hypothetical imputed political competition score for states that lack *Political competition* data in any year. The horizontal line indicates the point at which the result would lose statistical significance at the $p = 0.10$ level.

As a robustness check for creating our sample, we utilized the "Electoral Democracy Index" (polyarchy) from the V-Dem data set. We first used a cutoff value of 0.5, and then used a cutoff value of 0.4. We then continued to use the "competitiveness of participation" variable from Policy dataset. Our results remained robust:

Table A4: Political Competition Increases Probability of Joining the ICC (Dictatorship Coded Using V-Dem)

Dependent Variable: Years to Join

	(A4)	(A5)	(A6)	(A7)
Cutpoint	0.50	0.50	0.40	0.40
Explanatory Variable				
Political competition	0.64^{**}	0.53^{*}	0.86^{**}	0.74^{*}
	(0.27)	(0.29)	(0.35)	(0.41)
Control Variables				
Rule of law		0.46		0.34
		(0.52)		(0.63)
Log (GDP per capita)		−0.43		−0.62
		(0.29)		(0.39)
Foreign aid		−0.03		−0.06
		(0.07)		(0.08)
Intra-state violence		−0.00		−0.08
		(0.19)		(0.27)
Inter-state violence		−0.17		−0.19
		(0.61)		(0.60)
Region dummies	Yes	Yes	Yes	Yes
Events	20	19	13	12
States	97	92	80	75
Observations (state-year)	1,186	1,015	973	809

$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

As a robustness check for our explanatory variable of political competition, we utilized the "Elections Multiparty" variable from the V-Dem data set. This variable is coded from 0 to 4, where 0 indicates "No-party or single-party and there is no meaningful competition" and 4 indicates that elections are multi-party with restrictions only on extremist parties. This variable is only coded in the year of an election, but we filled in non-election years using the most recent measurement. Thus, the variable should be interpreted as "level of competition in the most recent election". The sample here consists of dictatorships, as coded by V-Dem's "Electoral Democracy Index", using a cutoff value of 0.5. Our results remained robust:

Table A5: Political Competition (as Multiparty Elections) Increases Probability of Joining the ICC

Dependent Variable: Years to Join

	(A8)	(A9)
Explanatory Variable		
Political competition	0.62^{***}	0.52^{**}
	(0.22)	(0.23)
Control Variables		
Rule of law		0.45
		(0.49)
Log (GDP per capita)		-0.47^{*}
		(0.26)
Foreign aid		−0.02
		(0.07)
Intra-state violence		0.09
		(0.14)
Inter-state violence		−0.04
		(0.51)
Region dummies	Yes	Yes
Events	26	25
States	93	88
Observations (state-year)	1,098	969

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

2.2 H2: Joining the ICC Decreases Violence

These descriptive statistics describe the data in our main analysis.

Table A6: Descriptive Statistics for Models (4)–(9)

Variable	Mean	Minimum	Maximum	N
Post-ratification	0.77	0	1	651
Total violence	0.69	0	7	651
Intra-state violence	0.68	0	7	651
Log (GDP per capita)	6.95	5.35	9.39	625
Foreign aid*	20.71	−17.05	23.89	651
Polity	1.89	−7	8	617
Rule of law	−0.75	−2.13	0.46	651
Africa	0.74	0	1	651

*hyperbolic sine transformation

As a robustness check on our measure of violence, we constructed a measure called PRIO violence using the UCDP/PRIO Armed Conflict Dataset. The measure consists of the number of episodes of violence that occurred within a given state-year, multiplied by the intensity of those episodes. Intensity is coded as either 1 (25-999 battle deaths) or 2 (1,000+ battle deaths). If a state experiences multiple violence episodes within a given year of different intensities, we take the average intensity level of those episodes. The results are robust to this alternative specification of the violence measure.

Table A7: Joining the ICC Decreases Violence (Measured from PRIO Data)

Dependent Variable: PRIO violence

	(A10)	(A11)	(A12)
Post-join	-0.41^{**}	-0.60^{**}	-0.55^{**}
	(0.21)	(0.25)	(0.26)
Rule of law		-1.96^{***}	-1.66^{***}
		(0.23)	(0.24)
Log (GDP per capita)		-0.28	-0.31^{*}
		(0.17)	(0.17)
Foreign aid		0.77^{***}	0.70^{***}
		(0.12)	(0.13)
Polity			-0.06^{**}
			(0.03)
Africa	0.36	0.78^{***}	0.95^{***}
	(0.22)	(0.28)	(0.31)
States	31	31	30
Observations (state-year)	651	625	592

$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

2.3 H3: Joining Increases Leader Survival

These descriptive statistics describe the data in our main analysis.

Table A8: Descriptive Statistics for Models (10)–(12)

Variable	Mean	Minimum	Maximum	N
Post-join	0.77	0	1	758
Tenure year	7.46	0	46	758
Out of office	0.13	0	1	758
Polity	2.09	−7	8	691
Log(GDP per capita)	7.00	5.35	9.39	730
Foreign aid*	20.74	−17.05	23.89	758
Total violence	0.79	0	7	758
Africa	0.79	0	1	758

*hyperbolic sine transformation

As a robustness check, we also ran the models using all dictatorships as the sample. Our main models only consider survival among dictatorships that actually ratify, which we expect will bias us against finding any statistically significant results. However, we check this by considering the sample of all dictatorships below. We find that the results hold:

Table A9: Joining the ICC Increases Leader Survival in Office

Dependent Variable: Years to Losing Office
(Event: Removal from Office)

	(A13)	(A14)
Explanatory Variable		
Post-join	0.08	-0.69^{**}
	(0.20)	(0.28)
Control Variables		
Log (GDP per capita)		-0.20
		(0.12)
Total violence		0.05
		(0.07)
Polity		0.07^{***}
		(0.02)
DAC aid		0.00
		(0.02)
Region dummies	Yes	Yes
Never-ratifiers	Yes	Yes
N events	146	118
N states	94	74
N observations (leader-year)	1,701	1,176

$p < 0.1^{*}, p < 0.05^{**}, p < 0.01^{***}$

Note: Models (A13)–(A14) show results including only those states that eventually join the ICC. Here there are 86 states, representing 272 different leaders, of whom 184 lost office at some point (118 lose office in Model (A14) because of missing data).