Moderating Conflicts with Radical Hardliners Hardliners*

June 15, 2020

Gabriele Gratton†
UNSW

Bettina Klose‡
Purdue

Abstract

We study a model of civil conflict in which players may be "hardliners" who strongly dislike any position that differs from theirs, but barely distinguish between positions closer and farther from it. In our model, for a pre-determined set of ideological positions, each faction chooses whether to exert effort to coerce other factions. Full-scale conflict arises whenever two or more factions exert such effort. We show that if extremist factions are sufficiently more hardliners than the moderate one, there exist circumstances under which more radical conflicts induce moderate and peaceful outcomes. We discuss how a third party can induce moderate and peaceful outcomes by means of favoring more radical leaderships in extremist factions. Such an intervention can be successful only if the cost of fighting is sufficiently large. Otherwise, it induces more conflict and more radical outcomes. Interventions that reduce the cost of conflict increase the likelihood of both full-scale conflict and radical outcomes.

Keywords: civil conflict; extremism; radicalization.

*We are grateful to Peter Buisseret, Maria Cubel, Sambuddha Ghosh, and Sarah Walker for helpful comments.

†School of Economics, UNSW Business School, UNSW Sydney. Email: g.gratton@unsw.edu.au.

‡Krannert School of Management, Purdue University. Email: bklose@purdue.edu.

1 Introduction

The rise of fundamentalist Salafi movements in the Middle East poses the question of how to manage conflicts in the presence of extremist factions who seem uninterested in any form of compromise. In fact, extremist factions are often *hardliners*: they strongly dislike any position that differs from their objective, but barely distinguish between positions closer and farther from it or positions that differ on dimensions that are not in the faction's agenda.

This characterization of extremist factions is at odds with the rationalist workhorse model of preferences, which assumes that even radical fundamentalists strongly prefer policies closer to their bliss point to others farther from it. However, this assumption is far from being unanimously accepted. For example, Osborne (1995) is "uncomfortable with the implication of concavity that extremists are highly sensitive to differences between moderate candidates" (see also Eguia, 2013). Similarly, Kamada and Kojima (2014) argue that non-concave utility functions are a better representation of ideological preferences, especially when motivated by moral or religious values (see also Michaeli and Spiro, 2015). In this paper, we study a stylized model of conflict in which different factions have arbitrarily convex or concave preferences over the policy space. The analysis of the model allows us to show how policy implications depend on the relative convexity of the factions' preferences.

In our model, there is a moderate and two extremist factions. Each faction strategically chooses whether to exert costly effort to coerce other factions. When two or more factions exert effort, we say that they *fight* a full-scale conflict. The value of coercing other factions depends on how radical the conflict is, which can be interpreted as the distance in policy bliss-points between extremist factions and the moderate faction. We say that a faction is more hardliner if its value of cohering the other factions is less elastic with respect to how radical the conflict is. We characterize the set of equilibria of and show that, perhaps surprisingly, when extremist factions are sufficiently more hardliners than the moderate faction, radicalizing the conflict may induce a moderate peace. Intuitively, radicalizing the extremists yields a steeper increase in the moderate faction's incentives to exert effort.[1]

This result has potentially key policy implications for a foreign intervention wishing to induce moderate and peaceful outcomes. Civil conflicts are typically dominated by radical actors while we rarely see armies of moderate, democratic rebels. Klose and Kovenock (2015) offer a simple theoretical justification: as extremists are ideologically further away from each other, they have large incentives to fight. On the contrary, moderates that lie between the two extreme factions are closer to each extreme and

[1]Our model also captures the fact that more radical leaders might also be less likely to receive sufficient support to organize the fight.

thus have weaker incentives to fight. Lacking a moderate leadership to deal with, foreign intervention can aim to influence the leadership of extremist factions,[2] favoring leaders with relatively more moderate positions so as to moderate the ultimate outcome of the conflict and reduce the likelihood of a full-scale conflict. In contrast, our result shows that there are circumstances under which more radical extremist leaderships lead to an uprising of moderates and ultimately to a moderate and peaceful outcome.

While in our model the moderates know the ideological position of extremist leaders, in reality their incentive to fight only depends on their perception of how radical extremist factions are. Thus, the same result could also be achieved by an intervention that manipulates moderates' belief, convincing them that extremist leaders are more radicalized.[3]

The 1970s in Italy offer a possible example of our mechanism at work. The decade was characterized by frequent acts of terrorism and political violence from right and left-wing extremists. Violence began to decline in the early 1980s when the increasingly violent tactics of extremists alienated popular support (Bull and Cooke, 2013), leading to coalition governments including parties previously excluded from the executive, and bringing workers unions firmly on the side of democratic legality. Bull (2007) suggests that the escalation of violence was deliberately encouraged by US and European governments with the objective of forcing moderate forces to work together against extremist factions. This "strategy of tension" closely resembles our radicalizing interventions. Jenkins (1990) argues that a similar strategy was employed in Belgium between 1982 and 1986. More recently, Satter (2003) and Felshtinsky and Pribylovskyr (2008) argue that the apartment bombings in 1999 Russia were in fact perpetuated by the Russian security forces and falsely attributed to the Chechen independence movement. The attacks then signaled to moderate Russians that more was at stake for them than the Chechen independence itself, increasing the popular support for the resumption of military operations in Chechnya.

Our model captures several patterns observed in civil conflicts. While only interventions that radicalize extremist leaderships can induce moderate and peaceful outcomes, this result can be achieved only if the cost of fighting is sufficiently large. Otherwise, the same intervention induces more conflict and more radical outcomes. We also discuss the effects of interventions targeting the cost of conflict, for example by lifting an arms embargo or smuggling weapons into the country. Regan (1996,

[2]Tiernay (2015) shows evidence that leadership changes have large impacts on the termination of civil conflicts (see also Fearon and Laitin, 2007). "Leadership changes are a factor in the termination of between 25% and 40% of civil wars" (James Fearon, cited in *How to Stop the Fighting, Sometimes*, The Economist, November 9th, 2013). Hamlin and Jennings (2007) consider the endogenous choice of leadership within factions.

[3]Baliga and Sjöstrom (2012) study how a third-party can induce full-scale conflict by manipulating information.

2000) and Elbadawi (1999) put forward evidence that the major effect of interventions is prolonging the conflict by reducing the of cost of fighting.[4] In our model, the combination of low fighting costs and radical extremist positions results in full-scale conflict with radical outcomes. This pattern fits the development of the Afghan civil war after the US smuggled weapons into the country to help Mujaheddin forces fight the Soviet occupation. Similarly, after the arrival of the US-led Coalition troops in 2003, the dismantling of the Iraqi Army dramatically lowered the cost of organizing armed militias, fostering the chances of a civil war. In our model, when the cost of fighting is sufficiently low, conflict cannot be avoided, but an intervention that favors leaders with more moderate positions induces a more moderate outcome.

2 A Model of Civil Conflict

There are three factions: a moderate M and two extremists, E_1 and E_2. Each faction i chooses whether to fight or not. Choosing to fight entails a cost $\gamma > 0$. If $n \geq 1$ factions fight, then each fighting faction wins with probability $1/n$. If no faction fights, then each faction wins with probability $1/3$.

For faction $i \in \{M, E_1, E_2\}$, the value of winning v^i is a continuous and increasing function of how *radical* the conflict is, which we measure by $R > 0$. We assume that $v^i(0) = 0$ for all $i \in \{M, E_1, E_2\}$ and we focus on a symmetric case such that $v^{E_1} = v^{E_2} = v^E$.

We represent a mixed strategy for faction i by its probability of fighting $\sigma^i \in [0, 1]$. Our solution concept is Nash equilibrium. Following (loosely) Esteban and Ray (1999) we say that an equilibrium is *extremist* if only extremist players fight with positive probability. We say that an equilibrium is *moderate* if only moderate players fight with positive probability. Furthermore, an equilibrium is a *dictatorship* if only one faction fights with positive probability; otherwise it is a *conflict*. Note that there might exist both *extremist dictatorships* or an *extremist conflict*; in contrast, a moderate equilibrium is always a dictatorship.

The key to the analysis is that a faction i will fight only if the opponent factions fight with sufficiently low probability.

Lemma 1. *Fighting is a best response for faction $i \in \{M, E_1, E_2\}$ if $\sum_{j \neq i} \sigma^j \leq 2B^i(R)$, where*

$$B^i(R) \equiv 2 - \frac{3\gamma}{v^i(R)} \tag{1}$$

is greater when fighting is costlier or the conflict is more radical.

[4] In this context, Balch-Lindsay et al. (2008) and Regan (2002) suggest that only biased interventions might reduce the length of the conflict by inducing a military victory of the favored faction.

Proof. All proofs are in Appendix A. □

If faction i has a higher $B^i(R)$, then it is more willing to fight even if it knows that other factions are likely to fight. That is, it is more willing to participate in a full-scale conflict. Therefore, we call $B^i(R)$ faction i's *bellicosity*. Obviously, since $v^{E_1} = v^{E_2}$, then also $B^{E_1}(R) = B^{E_2}(R) = B^E(R)$.

For the remaining analysis, we rule out uninteresting cases in which factions have a strictly dominant strategy. Assumption 1 says that (i) it is worth fighting for a victory if all other factions do not fight, and (ii) fighting is not a best response when all other factions fight with probability 1.[5]

Assumption 1. $0 < B^i(R) < 1$ *for all factions* $i \in \{M, E_1, E_2\}$.

Interpreting the model. A conflict may be more radical for various reasons and our model accomodates several interpretations. For example, a conflict is more radical if the extremist factions promote policies that are farther from the moderates' preferred policies. Similarly, a conflict is more radical if the issue at stake is more important. For example, because the three factions fight over an indivisible resource of value R.

Bellicosity is naturally increasing in how radical the conflict is, R, for all factions. Yet, one faction's bellicosity may be more or less sensitive to R. From (1), we see that faction i's bellicosity is more sensitive to R when the value of winning v^i is more sensitive to to changes in R. In particular, it is useful to define the elasticity of faction i's value of winning with respect to R:

$$\epsilon^i(R) = R\frac{d}{dR}\ln\left(v^i(R)\right).$$

If the value of winning v^i is a (weakly) convex function of R, then the faction is not very sensitive to changes in R for low values of R, but becomes more sensitive to changes in R for larger values. If R, as in our spatial politics model below, represents the distance between moderate policies and the extreme policies of the extremist factions, then the standard assumption of convexity in preferences implies that v^i is (weakly) convex. This is a reasonable assumption for our moderate faction, M. Yet, radical hardliners are hardly represented by convex preferences: they pay little attention to the difference between two positions that are both far from their bliss point and rarely prefer a compromise to a lottery between extreme outcomes. Therefore, if we wish to model radical hardliners, we may need to consider the case in which v^i is a convex function.

To make this point clearer, we also study one example of our model in a more standard spatial model of politics. Let $P \subseteq \mathbb{R}^2$ be a two-dimensional policy space. The

[5]For given functions v^i, this restricts the range of admissible values for R to $R \in (\underline{R}, \bar{R})$ with $\min\left\{v^M(\underline{R}), v^E(\underline{R})\right\} = \frac{3}{2}\gamma$ and $\max\left\{v^M(\bar{R}), v^E(\bar{R})\right\} = 3\gamma$.

moderate faction M has utility given by $u^M\left(b^W\right) = -\alpha\left[\left(b_1^W - b_1^M\right)^m + \left(b_2^W - b_2^M\right)^m\right]$, where $\alpha > 0$, $b^M = \left(b_1^M, b_2^M\right) \in P$ is the bliss point of the moderate faction and $b^W = \left(b_1^W, b_2^W\right)$ is the policy set by the winner. We assume that if the moderate faction wins, then $b^W = b^M$. Notice that the parameter $m > 0$ represents how sensitive the moderate faction is to marginal changes in policy.

Extremist faction E_k, $k \in \{1,2\}$ only cares about the i-th policy dimension: it has utility given by $u^{E_k}\left(b^W\right) = -\beta\left(b_k^W - b_k^{E_k}\right)^e$, where $\beta > 0$, $b_k^{E_k}$ is the k-th dimension faction $E_k's$bliss point, and $e > 0$ represents how sensitive extremist factions are to marginal changes in policy. Furthermore, we assume that, if faction E_k wins, then $b_k^W = b_k^{E_k}$ and $b_{j\neq k}^W = b_{j\neq k}^M$.

In this parametric example, $1/e$ $(1/m)$ naturally measures how hardliner an extremist (moderate) faction is. Furthermore, we can appropriately normalize the policy space so that $\left|b_1^{E_1} - b_1^M\right| = \left|b_2^{E_2} - b_2^M\right| = R$, yielding

$$v^M(R) = \alpha R^m;$$
$$v^E(R) = \beta R^e,$$

and

$$\epsilon^M(R) = m/R$$
$$\epsilon^{E_k}(R) = e/R, \text{ all } k \in \{1,2\}.$$

We will return to this parametrization in interpreting some of our results.

3 Conflict and Radicalization

We now partially characterize the set of equilibria. Proposition 0 says that if the moderate faction is sufficiently more bellicose than extremist factions then the unique equilibrium is moderate. Otherwise, either an extremist conflict or an extremist dictatorship are equilibria.

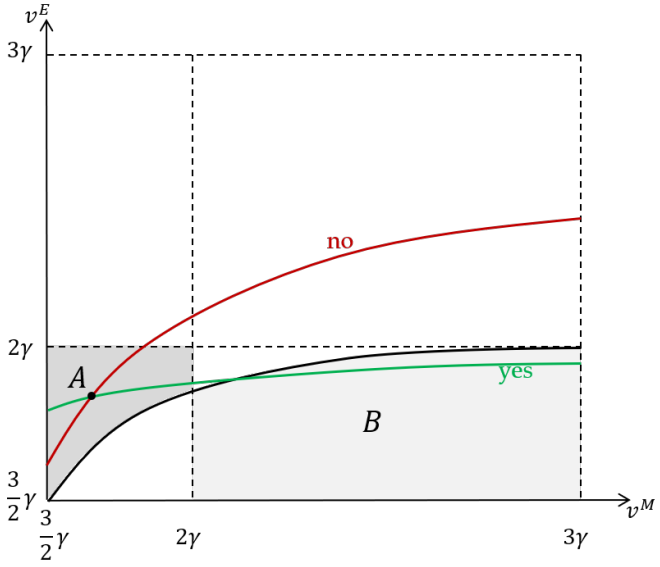
Proposition 0. *The unique equilibrium is moderate if and only if*

$$B^M(R) > \max\left\{1/2, 2B^E(R)\right\}.$$

Otherwise,

1. *if and only if $B^M(R) \leq 2B^E(R)$, there exist extremist conflict equilibria. In any ex-*

Figure 1: Regions of Conflict (A) and Moderation (B)



tremist conflict equilibrium, each extremist faction fights with probability

$$\sigma^{E_1} = \sigma^{E_2} = \bar{\sigma} \equiv \min\left\{2B^E(R), 1\right\}$$

and the moderate faction does not fight, $\sigma^M = 0$.

2. *if and only if* $B^i(R) \leq 1/2$ *for all* $i \in \{M, E_1, E_2\}$, *there exist extremist dictatorship equilibria.*

In an extremist conflict equilibrium, a full-scale conflict arises with probability $\bar{\sigma}^2$. Hence, the probability of a moderate victory in such an equilibrium is $(1-\bar{\sigma})^2/3$.

Radicalizing to Moderate. We can now derive our main results regarding comparative statics on R. We say that an intervention that increases R is *radicalizing;* one that reduces R is *de-radicalizing.* For example, a foreign power can favor more radical extremist leaderships, therefore making the policy distance between moderate and extremist factions wider and the conflict more radical. Similarly, international politics may increase the importance of the issue at stake or even focus the conflict on a more important issue, again radicalizing the conflict. In what follows, we consider the following scenario. An extremist dictatorship has switched to an extremist conflict. Thus, we require that the pre-intervention parameters are in the region where both extremist dictatorships and extremist conflicts are equilibria. The shaded region A in Figure 1 represents the space of pre-intervention scenarios.

Assumption 2. *The pre-intervention* R^P *is such that* $B^M(R^P) \leq 2B^E(R^P) \leq 1/2$.

Furthermore, an intervention cannot switch the situation from one type of equilibrium to another, unless the original type of equilibrium ceases to exist. Therefore,

full moderation can be achieved only by reaching values of (v^M, v^E) for which only moderate equilibria exist. The shaded area B represents this region in Figure 1. An intervention changes the values of v^M and v^E along a path determined by the relative curvature of the utility functions of the three factions. Figure 1 depicts two such patterns—labeled "yes" and "no"—starting from a point in region A. Since v^M is increasing in R, a necessary condition for a successful intervention is to radicalize the conflict: increase R.

Proposition 1. *If an intervention induces moderation, then it is radicalizing. Such an intervention exists if and only if there exists R such that $B^M(R) > \max\{1/2, 2B^E(R)\}$. A sufficient condition for such an intervention to exists is $v^E(R^*) < 2\gamma$, where R^* is the ideological distance such that $v^M(R^*) = 3\gamma$.*

Such intervention is Pareto improving. In fact, an extremist's expected payoff in an extremist conflict in region A is 0. Intuitively, in an extremist conflict, all the expected value of a victory is dissipated into fighting. Thus, the intervention deters a conflict which brings no expected advantage to the fighters.

Hardliners and the cost of fighting Proposition 1 says that an intervention that induces moderation exists only if there exists $R > R^P$ such that $B^M(R^P) \leq 2B^E(R^P)$ and $B^M(R) > 2B^E(R)$. Because $v^i(0) = 0$ for all factions, a necessary condition is that v^E grows faster than v^M from $R = 0$ to $R = R^P$ but slower than v^M for values of R greater than R^P. Returning to our parametric example, it is easy to see that such a combination is only possible if e is sufficiently smaller than m. That is, the possibility of an intervention that induces moderation relies on the extremist factions to be hardliners relative to the moderate faction.[6] By Proposition 1, moderation may then be achieved only by making the bliss points of the extremist factions more extreme.

The necessary and sufficient condition in Proposition 1 is satisfied if and only if there exists R such that $v^M(R) \in (2\gamma, 3\gamma)$ and

$$v^E(R) < \left[\frac{1}{3\gamma} + \frac{1}{2}v^M(R)^{-1}\right]^{-1}. \tag{2}$$

Notice also that (2) can never be satisfied if the cost of fighting γ is sufficiently small.

Remark 1. A radicalizing intervention that achieves full moderation and avoids full-scale conflict exists only if the cost of fighting is sufficiently large.

[6]Notice that the moderate faction having convex preferences and the extremist factions having non-convex preferences is neither a sufficient nor a necessary condition for the existence of an intervention that induces moderation. Nevertheless, it is necessary that the extremist factions have preferences that are more hardliner: their value of winning is less elastic to changes in R.

Moderating Conflicts and the Anarchic Chaos. We now turn to the question of how interventions can moderate conflict when full moderation is not achievable. We impose a regularity condition: for very low levels of R, extremist factions are more bellicose than the moderate faction.

Assumption 3. *Let R^{**} be the ideological distance such that $B^M(R^{**}) = 0$. Then $B^E(R^{**}) > 0$.*

Whenever the condition in Proposition 1 is not met, an extremist conflict is an equilibrium for all R satisfying assumptions 1 (by Proposition 0). Therefore, an intervention that cannot ensure full moderation can only affect the likelihood of conflict and final policy outcomes.

Proposition 2. *If an intervention that induces moderation does not exist, then: (i) a de-radicalizing intervention reduces the probability of full-scale conflict and increases the chances of a moderate victory; (ii) a radicalizing intervention increases the probability of full-scale conflict and reduces the chances of a moderate victory.*

Intuitively, interventions that reduce R induce more peaceful and moderate outcomes because the probability of fighting of the extremist groups is increasing in R. When the conditions for moderation via a radicalizing intervention are not met, then such an intervention instead increases all factions' willingness to fight.

Returning to our spatial politics example, when the conditions for moderation via a radicalizing intervention are not met, a radicalizing intervention not only increases the chances of an extremist victory, but also induces more extremist policies. Instead, a moderating intervention induced both greater chances of a moderate victory *and* less extreme policies if moderates lose.

We finally consider the effect of reducing the cost of fighting. For example, a foreign intervention could establish or lift an embargo on armaments or introduce a different technology. Recall that full moderation can be achieved by increasing the moderates' incentive to fight. However, reducing the cost of fighting never achieves full moderation, because it increases the incentive to fight for all three factions.

Proposition 3. *Reducing the cost of fighting increases the probability of a full-scale conflict and reduces the chances of a moderate victory.*

We conclude by highlighting the peril of excessive interventions in increasing R or reducing γ. Both can in fact precipitate the conflict to a situation in which $1 < \min\left\{B^E(R), B^M(R)\right\}$. By Lemma 1, then the unique equilibrium is one in which all three factions fight with non-zero probability and the outcome is likely to be extreme. We think of this situation as an *anarchic chaos* similar to the Afghan civil war after the retreat of the Soviet Army in the late 1980's.

4 Conclusions

Rationalist models of civil conflict assume that both moderates and extremists have convex preferences over the policy space. Yet, radical fundamentalists are unlikely to exhibit such preferences. Instead, they are hardliners who barely distinguish policies close to their bliss point to others farther from it. In fact, fundamentalist Salafi jihadists often target moderate Muslims as much as (or more than) secularists. It is therefore important to understand the dynamics of conflict in a model that does not rely on the assumption of convex preferences.

Our model highlights how the presence of extremist hardliners impacts civil conflicts. We show that when extremists are sufficiently more hardliners than moderates, there exist circumstances in which a moderate faction rises to successfully defeat extremists. However, in order to induce such an outcome, interventions must aim to increase the moderates' value of victory—perhaps by informing moderate citizens of the consequences of an extremist victory—rather than simply reducing their cost of fighting. We highlighted how such an intervention may find historical parallels in the so-called strategies of tension in 1970s Italy, 1980s Belgium, and 1990s Russia.

While our results suggest that such an intervention is the only option to achieve a moderate and peaceful outcome, we have also pointed out the risks associated with it. Specifically, if the cost of fighting becomes sufficiently small, a radicalizing intervention increases the likelihood of full-scale conflict and induces more extreme outcomes, precipitating the polity in an anarchic chaos similar to the Afghan civil war after the retreat of the Soviet Army in the late 1980's.

References

Balch-Lindsay, Dylan, Andrew J. Enterline, and Kyle A. Joyce, "Third-party intervention and the civil war process," *Journal of Peace Research*, 2008, *45* (3), 345–363.

Baliga, Sandeep and Tomas Sjöstrom, "The Strategy of Manipulating Conflict," *American Economic Review*, October 2012, *102* (6), 2897–2922.

Bull, Anna Cento, *The Strategy of Tension and the Politics of Nonreconciliation*, Berghahn Books, 2007.

__ **and Philip Cooke**, *Ending Terrorism in Italy*, Rutledge, 2013.

Eguia, Jon X, "On the spatial representation of preference profiles," *Economic Theory*, 2013, *52* (1), 103–128.

Elbadawi, Ibrahim, "Civil wars and poverty: The role of external interventions, political rights and economic growth," *Working Paper, World Bank, Washington D.C.*, 1999.

Esteban, Joan and Debraj Ray, "Confict and Distribution," *Journal of Economic Theory*, 1999, *87* (2), 379–415.

Fearon, James and David D. Laitin, "Civil War Termination," *Presented at the 2007 Annual Meetings of the American Political Science Association*, 2007.

Felshtinsky, Yuri and Vladimir Pribylovskyr, *The age of assassins: the rise and rise of Vladimir Putin*, Gibson Square Books, 2008.

Hamlin, Alan and Colin Jennings, "Leadership and conflict," *Journal of Economic Behavior & Organization*, 2007, *64* (1), 49–68.

Jenkins, Philip, "Strategy of tension: The Belgian terrorist crisis 1982-1986," *Terrorism*, 1990, *13* (4-5), 299–309.

Kamada, Yuichiro and Fuhito Kojima, "Voter Preferences, Polarization, and Electoral Policies," *American Economic Journal: Microeconomics*, November 2014, *6* (4), 203–236.

Klose, Bettina and Dan Kovenock, "Extremism Drives Out Moderation," *Social Choice and Welfare*, 2015, *44* (4), 961–887.

Michaeli, Moti and Daniel Spiro, "Norm conformity across societies," *Journal of Public Economics*, 2015, *132*, 51–65.

Osborne, Martin J., "Spatial Models of Political Competition Under Plurality Rule: A Survey of Some Explanations of the Number of Candidates and the Positions They Take," *Canadian Journal of Economics*, 1995, *2*, 261–301.

Regan, Patrick M., "Conditions of successful third-party intervention in intrastate conflicts," *Journal of Conflict Resolution*, 1996, *40* (2), 336–359.

_ , *Civil wars and foreign powers: outside intervention in interstate conflict*, The University of Michigan Press, 2000.

_ , "Third-party interventions and the duration of intrastate conflicts," *Journal of Conflict Resolution*, 2002, *46* (1), 55–73.

Satter, David, *Darkness at Dawn: The Rise of the Criminal Russian State*, New Haven: Yale Univ. Press, 2003.

Tiernay, Michael, "Killing Kony: Leadership change and civil war termination," *Journal of Conflict Resolution*, 2015, *59* (2), 175–206.

A Proofs

Proof of Lemma 1. Let j and k be the opponents of i. Then fighting is a best-response for faction i whenever

$$\underbrace{v^i(R)\left[\frac{\sigma^j\sigma^k}{3}+\frac{\sigma^j\left(1-\sigma^k\right)+\sigma^k\left(1-\sigma^j\right)}{2}+\left(1-\sigma^j\right)\left(1-\sigma^k\right)\right]-\gamma}_{\text{exp. payoff of fighting}}\geq$$

$$\underbrace{v^i(R)\frac{\left(1-\sigma^j\right)\left(1-\sigma^k\right)}{3}}_{\text{exp. payoff of not fighting}}$$

which reduces to $\sigma^j+\sigma^k\leq 2B^i(R)$. If the inequality is strict, fighting is the unique best-response for i. □

Proof of Proposition 0. **First statement.** In the unique equilibrium, the moderate faction fights with probability 1 and both extremists do not fight.

Existence. We begin by noticing that an extremist faction prefers not to fight whenever $\sigma^M=1$. To see why, recall that $B^M(R)<1$ by Assumption 1. Thus, the condition $B^M(R)>2B^E(R)$ implies that $B^E(R)<1/2$. Lemma 1 then implies that fighting is not a best-response for an extremist faction E_i, $i\in\{1,2\}$ because

$$\sigma^{E_j}+\sigma^M\geq\sigma^M=1>2B^E(R),\ j\neq i.$$

Hence $\sigma^E=0$. Existence then follows from the condition $B^M(R)>1/2$, which implies

$$\sigma^{E_1}+\sigma^{E_2}=0\leq 1<2B^M(R).$$

Hence $\sigma^M=1$.

Uniqueness. We have shown above that under the conditions of Proposition 0 an extremist faction fights only if $\sigma^M<1$. By Lemma 1 this requires $\sigma^{E_1}+\sigma^{E_2}\geq 2B^M(R)>1$, which implies that $\sigma^{E_1}>0$ and $\sigma^{E_2}>0$. By Lemma 1 $\sigma^{E_i}>0$ only if $\sigma^{E_j}\leq\sigma^{E_j}+\sigma^M\leq 2B^E(R)$, $i\in\{1,2\}$, $j\neq i$. Thus,

$$\sigma^{E_1}+\sigma^{E_2}<2\cdot 2B^E(R)<2B^M(R).$$

But then M fights with probability 1 by Lemma 1.

Second statement. Let $\left(\sigma^M,\sigma^{E_1},\sigma^{E_2}\right)=(0,\bar{\sigma},\bar{\sigma})$ be an equilibrium. Then by Lemma 1, $2B^M(R)\leq\sigma^{E_1}+\sigma^{E_2}=2\bar{\sigma}\leq 4B^E(R)\Leftrightarrow B^M(R)\leq 2B^E(R)$.

Let $B^M(R) \leq 2B^E(R)$. Then

$$\sigma^{E_1} + \sigma^{E_2} = 2\bar{\sigma} \geq 4B^E(R) \geq 2B^M(R).$$

Therefore, $\sigma^M = 0$ is a best-response for M.

If $\bar{\sigma} = 1$ then

$$\sigma^M + \sigma^{E_i} = \bar{\sigma} = 1 \geq 2B^E(R),\ i \in \{1,2\}$$

and $\sigma^{E_j} = 1 = \bar{\sigma}$ is a best-response for E_j, $j \neq i$. If $\bar{\sigma} = 2B^E(R) < 1$, then

$$\sigma^M + \sigma^{E_i} = \bar{\sigma} = 2B^E(R),\ i \in \{1,2\}$$

and $\sigma^{E_j} = 2B^E(R) = \bar{\sigma}$ is a best-response for E_j, $j \neq i$.

Third statement. Let $\left(\sigma^M, \sigma^{E_1}, \sigma^{E_2}\right) = \left(0, 0, \sigma^{E_2}\right)$ with $\sigma^{E_2} \in (0,1]$ be an equilibrium. Then for $i, j \in \{M, E_1\}, i \neq j$ we must have $2B^i(R) \leq \sigma^j + \sigma^{E_2} \leq 1$ by Lemma 1. Thus by symmetry of the extremists, $B^k(R) \leq \frac{1}{2}$ for all factions $k \in \{L, M, R\}$

Let $B^i(R) \leq \frac{1}{2}$ for all factions $i \in \{L, M, R\}$. Then $2B^i(R) \leq 1$ for all factions i and by Lemma 1, each faction i prefers not to fight whenever there exists a faction j that fights with probability 1. In particular, $\left(\sigma^M, \sigma^{E_1}, \sigma^{E_2}\right) = (0,0,1)$ and $\left(\sigma^M, \sigma^{E_1}, \sigma^{E_2}\right) = (0,1,0)$ are extremist dictatorship equilibria. □

Proof of Proposition 1. **First statement.**By Assumption 2, the pre-intervention R^P is such that $v^M\left(R^P\right) \leq 2\gamma$. By Proposition 0, full moderation requires $v^M(R) > 2\gamma$. Since v^M is an increasing function of R, an intervention might induce moderation only if it increases R.

Second statement. By Proposition 0, for a given R, a moderate (dictatorship) equilibrium exists if and only if $B^M(R) > \max\left\{1/2, 2B^E(R)\right\}$. Notice that if such an R exists, then an intervention that radicalizes the conflict from R^P to R induces a moderate equilibrium. Otherwise, there is no intervention that can induce a moderate equilibrium.

Second statement. Let $v^E(R^*) < 2\gamma = \frac{6\gamma v^M(R^*)}{2v^M(R^*)+3\gamma}$. Notice that

$$B^M(R) > 2B^E(R) \Leftrightarrow v^E(R) < \frac{6\gamma v^M(R)}{2v^M(R)+3\gamma}$$

Recall that v^E and v^M are continuous and increasing in R. Therefore, there exists $\epsilon > 0$ such that for all $R : 0 < R^* - R < \epsilon$, $v^E(R) < 2\gamma$ and $2\gamma < v^M(R) < 3\gamma$. Then, by Proposition 0, the unique equilibrium at R is moderate. □

Proof of Proposition 2. Assume that an intervention that induces moderation does not exist: there exists no R such that $B^M(R) > \max\left\{1/2, 2B^E(R)\right\}$. By Assumptions 1 and 3, the post-intervention equilibrium is an extremist conflict for all R. Since $B^E(R)$ is increasing in R, in such an equilibrium, reducing R decreases the probability of a full-scale conflict

$$\bar{\sigma}^2 = \min\left\{2B^E(R)^2, 1\right\}$$

and increases the probability of a moderate victory $(1-\bar{\sigma})^2/3$. □

Proof of Proposition 3. By Assumption 2, the initial condition is an extremist conflict and

$$\begin{aligned} v^M(R) &\leq 2\gamma \\ v^E(R) &\geq \frac{6\gamma v^M(R)}{2v^M(R)+3\gamma}. \end{aligned} \tag{3}$$

We first show that for any γ', an extremist conflict continues to exist. That is, there exists no γ' such that

$$\begin{aligned} v^M(R) &> 2\gamma' \\ v^E(R) &< \frac{6\gamma' v^M(R)}{2v^M(R)+3\gamma'}. \end{aligned} \tag{4}$$

Notice that this requires

$$\gamma' < \frac{v^M(R)}{2} \leq \gamma.$$

Also,

$$\frac{6\gamma v^M(R)}{2v^M(R)+3\gamma}$$

is increasing in γ. Therefore, using (3),

$$v^E(R) \geq \frac{6\gamma v^M(R)}{2v^M(R)+3\gamma} > \frac{6\gamma' v^M(R)}{2v^M(R)+3\gamma'}$$

which contradicts (4).

We now show that reducing γ increases the probability of a full-scale conflict and decreases the probability of a moderate victory. Recall that in an extremist conflict the probability of a full-scale conflict is

$$\bar{\sigma}^2 = \min\left\{2B^E(R)^2, 1\right\}$$

and the probability of a moderate victory is given by $(1-\bar{\sigma})^2/3$. Noticing that

$$\frac{dB^E(R)}{d\gamma} = -\frac{3}{v^E(R)} < 0$$

concludes the proof. □