Cooperating Through Leaders*

David K. Levine[1], Salvatore Modica[2], Aldo Rustichini[3]

Abstract

We model conflicts as a prisoners dilemma where two groups play the fight-or-cooperate game through leaders. Leaders make “recommendations and promises”, and each group complies with their most preferred recommendation, but punishes the chosen leaders if their promises are not fulfilled. Each group has a natural group leader who shares their preferences and addresses her group; and there is a third “common” leader who is concerned with general welfare and advances her proposals to both groups. We study the resulting game among the leaders.

Broadly speaking, the paper makes three points. First, the presence of a common leader is essential to reach a cooperative outcome - with only the two group leaders fighting is unavoidable. Second, a general condition favoring cooperation is that the group leaders can be adequately punished. The intuition is that punishment deters group leaders from making excessively optimistic promises, and this leaves room for a successful cooperation proposal by the common leader. Third, the order of moves matters, and roughly speaking the leaders who have “the last word” are better placed to be followed.

Keywords: Plural Societies, Polarization, Social Conflict, Accountability, Political Equilibria.

*First Version: January 6, 2022. We would like to thank Luigi Balletta, Rohan Dutta, and especially Sandeep Baliga for a short but illuminating exchange. DKL and SM gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL_01 and the Leverhulme Trust. AR thanks the U.S. Department of Defense, contract W911NF2010242

Email addresses: david@dklevine.com (David K. Levine), salvatore.modica@unipa.it (Salvatore Modica), aldo.rustichini@gmail.com (Aldo Rustichini)

[1]Department of Economics, RHUL and WUSTL

[2]UniversitÃ di Palermo, Dipartimento SEAS

[3]Department of Economics, University of Minnesota

1. Introduction

In 1905 the Russo-Japanese war was settled in the Treaty of Portsmouth with the mediation of US President Theodore Roosevelt, who won the Nobel Peace prize the year after for his efforts. But the intervention of a mediator is not always successful. Doyle and Sambanis (2006) analyze 121 civil wars between 1945 and 1999, and of these 99 ended with a military victory or a truce - that is, there was not a successful third parties intervention. Still, 14 ended with a negotiated settlement mediated by the UN.

In this paper we model situations like these, where the intervention of an external agent concerned with the well-being of both parties in a conflict - which we shall call "common leader" - may under some circumstances help reaching a cooperative outcome. Note that such an external figure enters the picture as a third player, the first two being the natural group leaders of the conflicting parties. In fact, since in practice it is leaders who conduct negotiations or take fighting decisions, we make the stark assumption that group members just evaluate the leaders' proposals and act accordingly, effectively playing the game through the leaders. The strategic interaction we study is then the game played by the leaders. We shall presently discuss and motivate the main features of the model introduced in the paper, but let us see first what it yields. In a word, the result is a characterization of the conditions under which we can expect a common leader to succeed in making the parties reach a cooperative outcome, and of course those where we should expect failure.

To set the stage we start with a symmetric two-by-two game between two groups, in which a cooperative action C taken by both groups gives a higher utility than when at least one player takes the alternative action F of fighting. We focus on the prisoners dilemma, where the groups would unavoidably fight. To this "underlying game" we add leaders who offer suggestions on the best course of actions to the groups and promise particular outcomes. Incidentally, leaders act to influence the outcome of the game because their own utility depends on that outcome. Group members act as followers, choosing the most promising proposal and complying with it; but they can and possibly will punish leaders who fail to deliver on their promises.

In the game leaders play they advance proposals to the groups in the form of action profiles in the underlying game (like CC, or FC). Such proposals may be interpreted "recommendations and promises". The leaders suggest how their followers should behave and what expectations they should have about the behavior of others if they do behave in that way. For example proposing FC to group 1 means "Fight, the others will play C"; the promise is that if the group follow they will get the FC payoff. The leaders' payoffs depend upon the outcome of that game and the punishments issued by their followers. The possibility of being punished restrains the leaders' temptation to make non-credible promises. In accordance with the opening examples we assume that there are two types of leaders: *group* leaders, whose utility from the outcomes in the underlying game is identical to that of their group (the leaders of the two countries or groups in our example); and a *common* leader (Roosevelt or the UN), whose utility we take to be the average utility of the two groups. In the leaders game there are therefore three players: two group leaders and a common leader. Group leaders address only their own group; the common leader talks to both groups.

To see how the leaders game is played suppose that the common leader proposes CC, that is "Cooperate, the other group will too". The groups evaluate the proposal at face value, that is assuming that if they accept the proposal and cooperate the other group will cooperate too. Suppose that group leader 2 proposes FF, so that CC is the best outstanding offer for group 2; they will then accept the proposal and play C. Now suppose that group leader 1 proposes to her group FC , "Let us fight, they will submit". Since group 1 is better off at FC than at CC they will comply with their group leader's proposal and play F. The implemented profile will then be FC. Group members and leaders receive the FC payoffs, but the common leader is also punished by group 2, because the realized payoff of group 2 from FC is lower than that implicitly promised by the common leader's proposal CC. For another example, suppose on the other hand that both group leaders play FF and the common leader plays CC: then the outcome is CC and no leader is punished.

At this point we can go back to the results of the analysis. First, we show that in the absence of a common leader the resulting game between the two group leaders has fighting, FF, as the only equilibrium outcome. So only if a common leader is present can conflict be avoided. Second, a general condition favoring cooperation is that the groups can adequately punish their group leaders. The intuition here is that punishment deters group leaders from making excessively optimistic promises - think for example of the first group leader proposing FC - and this leaves room for a successful CC proposal by the common leader. Indeed formally as the value of the group leaders' punishment goes to infinity the equilibrium probability of cooperation goes to 1. Thirdly, we find that the order of moves matters, and roughly speaking the leaders who have "the last word" are better placed to be followed. Thus in the opening example President Roosevelt was successful intervening after the conflicting parties had been at war for a year. The same can be said of the pacifying policy of Nelson Mandela in the 1990's regarding the apartheid system in South Africa, which had been in place with all its polarizing effects for thirty years (as all know Mandela won the Peace Nobel Prize too).

Regarding the role of the group members as passive followers, the point is that their expectations are *shaped* by the leaders, even in a full information context: they think the leaders are better at predicting behavior than they are, hence that on average they are better off following the advice of a leader. In the case of particularly charismatic "spiritual" leaders this happens independently of the reasoning capabilities of their followers. More generally, a group can be thought as a large set of players who individually may find it costly to acquire the information needed to fully understand the consequences of actions taken in the game, hence delegate to the leaders the assessment of the strategic situation, while maintaining the capability of punishing ex-post non delivering leaders.

Related Literature

The first paper we must mention is Baliga et al. (2011), who also study 2×2 games - two groups, two actions - with leaders who can be punished. Leaders are the two group leaders, whose purpose is to remain in power. They simultaneously choose an action - like our F or C - in the first place. Groups are heterogeneous, in that in each group each member has different payoffs from the four possible profiles. In a given group, an individual member then punishes the leader by not granting her support if from her point of view the leader's action is not a best response to the opposing leader's choice. Leaders do not make promises; they are punished for taking a wrong decision, not for failing to deliver on their promises. The paper studies political systems as defined by the fraction of supporters in the population a leader needs to remain in power. Despite the apparent similarities it is really a different setting. We stress promises and competition among leaders in a model where the relevant decisions are ultimately taken by the citizens, with a different research purpose.

Although our model is one of full information and multiple groups, the specification of the leaders' proposals arises from the same idea of expectation shaping as in Hermalin (1998), where a single leader of a single group is the only one who comes to learn a payoff relevant signal, and acting on the basis of the signal shapes the followers' expectations (in such a way that leader imitation by followers is an equilibrium).

We are not the first to point out the possible merits of third-party intermediation in conflicts. Meirowitz et al. (2019) is a remarkable paper where a "neutral broker who does not favor either of the players" increases the chance that the conflicting parties achieve desirable outcomes. But in that case the third party is really just a mediator with no interest in the outcome of the groups game, to whom the parties are somehow willing to reveal their private information. The behavior of our common leader is driven directly by her involvement in the game.

In the political economy literature it is typical to have competing leaders, see for example Dewan and Squintani (2018) and the literature there cited. But the general concern is on how a set of groups with differing preferences on the available alternatives chooses a leader who then *decides for all* - quite naturally, the standard model of electoral competition. We stress that ours is not a model of electoral competition. There is no "winning leader" whom all must follow; different groups may follow different leaders, as it is natural in conflict situations.

There are other studies where delegation and/or leadership has a role. In Eliaz and Spiegler (2020), as here, a representative agent chooses among policy proposals and then selects and implements the one with the highest expected payoff (we explicitly model the proposers and their incentives, and allow each of them to address several representative followers). Like us, Dutta et al. (2018) consider punishment of leaders, but their punishment is based *ex ante* considerations and there are no common leaders. Prat and Rustichini (2003) explore the idea that games among principals can be played through the mediation of agents who receive transfers conditional on the action chosen, to induce them to play one action rather than another.

Loosely related to the present context, Esteban and Ray (1994), Esteban et al. (2012) and Duclos et al. (2004), construct a general, well founded measure of polarization. The *salience* of ethnic conflict is analyzed in Esteban and Ray (2008). These models are tested against data in several follow up studies (for example in Esteban et al. (2012)).

The leaders game built over an underlying game shares important features with the correlated equilibria of that underlying game: in both cases, thanks to a form of mediation, better outcomes than Nash equilibria can obtain; and in both solution concepts, leaders or the mediator suggest to followers an action profile, and followers respond. But the differences are deeper than the similarities. We provide the relevant comparison at the end of the appendix.

Outline of the Paper

In the next section we set up the model. In section 3 we dispose of the case where only group leaders are present, and in section 4 we analyse the model on which we focus. Section 5 concludes. Proofs are mostly in appendix.

2. The Model

2.1. The Underlying Game

The are two groups denoted by $k \in \{1, 2\}$, and each group has a representative follower. Follower k chooses action $a_k \in \{C, F\} \equiv A_k$, where C means cooperation and F fight. Action profiles (a_1, a_2) are denoted by $a \in A$, and at a all members of group k receive utility $u_k(a)$. These utility functions give rise to the *underlying game*.

The underlying game on which this paper is focused is the prisoners dilemma. If both followers play C they get a higher utility than if they both play F, and we set $u_k(CC) = 1$ and $u_k(FF) = 0$ for both k. Also, $u_1(FC) = u_2(CF) = \lambda > 1$ and $u_1(CF) = u_2(FC) = \xi < 0$.[4] The game matrix is thus

	C	F
C	$1, 1$	ξ, λ
F	λ, ξ	$0, 0$

We are only interested in games in which conflict is detrimental, so we assume that the average group payoff is maximum at CC:

$$(\lambda + \xi)/2 < 1. \tag{1}$$

2.2. The Leaders' Games

We now describe the games played by the leaders, which are the object of our analysis. There are three leaders $\ell \in \{0, 1, 2\}$: two *group leaders* $\ell = 1, 2$ who have the same interest as group $k = \ell$, and a *common leader* $\ell = 0$ who cares about both groups. The payoff of the leaders is the sum of a direct component and a possible punishment imposed by the followers.

The direct utility depends on the action profile $a \in A$ played by the followers in the underlying game. Denoting by $U^\ell(a)$ the utility leader ℓ obtains from profile a, we take $U^\ell(a) = u_\ell(a)$ for $\ell = 1, 2$; and we assume that the common leader's preferences coincide with utilitarian welfare: $U^0(a) = (u_1(a) + u_2(a))/2$.[5] We now describe how the profile a played by the followers and the possible punishments to the leaders are determined.

[4] Action profiles are always written so that 1 proceeds 2, and we are omitting commas when possible.
[5] We use superscripts for leaders and subscripts for followers.

The three leaders make *recommendations and promises* to their potential followers. Specifically, each leader makes a *proposal* $s^\ell \in A$, that is, an action profile in the underlying game. Proposing s^ℓ to group k means recommending the group to play s^ℓ_k and suggesting that the other group plays s^ℓ_{-k}, thus promising utility $u_k(s^\ell)$; in words the leader's proposal is "Follow me: play s^ℓ_k, and you will get utility $u_k(s^\ell)$".

We will consider three different extensive form games, differing in the order of the leaders' moves. In the first one the two group leaders move first, simultaneously choosing their proposals; their choices are communicated to the common leader, who then chooses her proposal; the triple of the resulting proposals made by the leaders is finally communicated to the followers, who choose their actions in the underlying game - a_1 and a_2 respectively - and thus determine their own payoff and the leaders' direct utility. In the second version the three leaders move simultaneously, and then the game proceeds as in the previous case - the leaders' choices are communicated to the followers who then choose actions C or F in the underlying game. In the third version the common leader moves first, choosing a proposal in A; her choice is communicated to the group leaders, who then simultaneously choose their proposals.

We still have to specify how the followers choose actions given a profile of proposals by the three leaders and how the payoffs at final nodes are determined; to this we turn. Observe that in all the extensive forms games just introduced the strategies of the leaders result in a triple of proposals communicated to the followers, which we denote by $s \equiv (s^0, s^1, s^2) \in A \times A \times A$. We assume that the follower of group k considers the proposal of the corresponding group leader and the one by the common leader, in other words follower k considers s^k and s^0.[6] Among the proposals they consider, the followers choose the one promising them the highest utility; and if follower k chooses $s^\ell = a$ then group k plays a_k, expecting $u_k(a)$. More precisely, given a triple s follower k chooses the proposal that maximizes $u_k(s^\ell)$ over the proposals s^0 and s^k she considers. Denote the chosen proposal by $g^k(s) \in A$.[7] Having chosen $g^k(s)$ group k then play their part $g^k(s)_k$, expecting to get $u_k(g^k(s))$. Therefore, given a triple (s^0, s^1, s^2) the *implemented action profile* in the underlying game will be $g(s) \equiv \left(g^k(s)_k\right)_{k=1,2} \in A$. This determines the utility of the groups, $u_k(g(s))$, and the direct utility of the leaders $U^\ell(g(s))$. If for example $s^0 = FC, s^1 = FC, s^2 = CF$ then $g^1(s) = FC$ and $g^2(s) = CF$ so both groups will play F, and $g(s) = FF$. Note that follower 1 is complying with the recommendations of both $\ell = 0$ and $\ell = 1$. Of course in this case no leader fulfills her promise (because all have promised $\lambda > 1$ but the realized utility is 0).

As to punishments, if a leader's promise is not fulfilled she will be punished by the groups who have complied with their proposals. Precisely, each group has the ability to impose a utility penalty $P > 0$ on their group leader, and $Q/2 > 0$ on the common leader (who then loses Q if punished by both groups). And if $u_k(g^k(s)) < u_k(g(s))$ then group k punishes any leader $\ell \in \{0, k\}$ such that

[6] As a benchmark we will analyse in Section 3 the case in which followers ignore the proposal of the common leader; in this case each group just follows their own group leader.

[7] The maximizer $g^k(s)$ for group k is unique because $a \neq a'$ implies $u_k(a) \neq u_k(a')$ for both k, though it may be proposed by more than one leader.

$s^\ell = g^k(s)$, where the punishment is P if $\ell = k$ and $Q/2$ if $\ell = 0$. In the example above group 1 punishes leaders $\ell = 0$ and $\ell = 1$, and group 2 punishes $\ell = 2$.

Finally, the payoffs. In all the three extensive forms, given the leaders' strategies, their payoffs - direct utility and punishments - depend only on the triple $s = (s^0, s^1, s^2)$ of proposals that are communicated to the followers. Denoting by $V^\ell(s)$ the payoff of leader ℓ, and letting $\mathbf{1}\{\mathfrak{c}\} = 1$ if condition $\mathfrak{c}$ is true and zero otherwise, the payoff of a group leader $\ell = 1, 2$ is

$$V^\ell(s) = U^\ell(g(s)) - P \cdot \mathbf{1}\{\ell = k \ \& \ g^k(s) = s^\ell \ \& \ u_k(s^\ell) < u_k(g(s))\} \tag{2}$$

and the payoff of the common leader is

$$V^0(s) = U^0(g(s)) - (Q/2) \cdot \sum\nolimits_{k=1,2} \mathbf{1}\{g^k(s) = s^0 \ \& \ u_k(s^0) < u_k(g(s))\}. \tag{3}$$

The games played by the three leaders which we have defined will be referred to as *leaders games*. The solution concept we adopt is a strengthening of the subgame perfect equilibrium: we require that in each subgame the leaders play a Nash equilibrium in weakly undominated strategies. The games are finite, so subgame perfect equilibria in mixed strategies exist. We call these *leaders equilibria*. We are particularly interested in the conditions under which the implemented profile in the equilibria of the leaders game is the cooperative outcome.

3. Nothing Is Gained With Only Group Leaders

We first establish that if each group only considers proposals from their own group leader the outcomes of the leaders game are the same as in the underlying game. This is in fact true quite generally, that is for any game with any number of groups:

Proposition 1. *For any leaders game, if each group only considers the proposal of their own group leader then at the Nash equilibria of the leaders game the distributions of action profiles chosen by the groups are the same as those induced by the Nash equilibria of the corresponding underlying game.*

The proof is in Appendix A. Specification of the general games is a trivial extension of the one given in the previous section. In the prisoners dilemma, which is the game on which the paper is focused, the above result implies that with only group leaders conflict is unavoidable.

4. Analysis of the Leaders Games

From now on the presence of a common leader is always assumed, so each group considers the proposals by their group leader and by the common leader. We study the three extensive form games defined in Section 2 in the following order: (1) the group leaders move first, then the common leader; (2) all leaders move simultaneously; and finally (3) the common leader moves first, then the group leaders. For ease of exposition we start working under the assumption that the common leader's punishment is not too low, specifically that $Q > |\lambda+\xi|$. The case of low Q is analyzed and discussed in Section 4.5.

We summarize the main findings here for convenience; details are spelled out in the next three subsections. The reader wishing to skip the formalities may refer to the present statement and go directly to the discussion section 4.4.

Summary of Results. *The main results of the analysis ($Q > |\lambda + \xi|$ assumed) are the following:*

If the common leader moves after the group leaders the implemented profile in any equilibrium of the leaders game is CC: cooperation obtains for sure.

In the simultaneous moves game:

If $P < -\xi$ the only equilibrium outcome is FF, fighting.

If $P > -\xi + (\lambda - 1)(Q - (\lambda + \xi))$ in the unique equilibrium the common leader plays CC and the group leaders mix, assigning to FF a probability which goes to zero as $P \to \infty$. Hence as P grows large the equilibrium probability of cooperation goes to 1.

If the common leader moves first, then both for $P < -\xi$ and $P > -\xi + (\lambda - 1)(Q - (\lambda + \xi))$ there are equilibria with same outcomes as in the previous case; but for any P there are also equilibria with outcome FF.

We now turn to the detailed study of the equilibria in the three games.

4.1. The Group Leaders Move First, Then the Common Leader

We start with the sequential game where the group leaders simultaneously move first, and the common leader moves after them. The strategy set of each group leader is the set A of action profiles in the underlying game (her possible proposals). The common leader has 16 information sets - one for each of the possible 4×4 choices of the group leaders - and at each one she can choose a proposal in the set A; thus the strategy set of the common leader is the 16-fold Cartesian product of A.

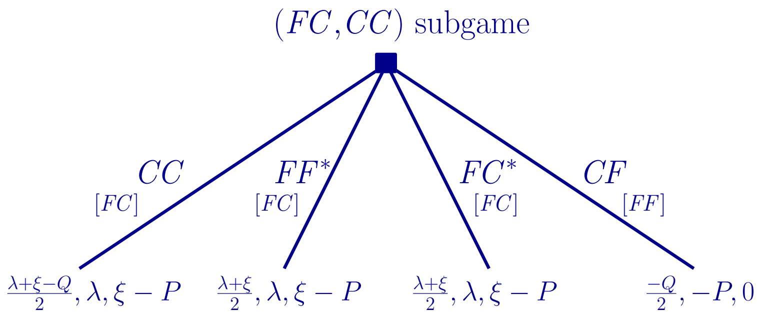
Proposition 2. *In the sequential case where the group leaders move first, in* all *subgame perfect equilibria of the leaders game (no dominance restriction) the only equilibrium outcome is CC. All leaders and both groups get 1.*

Proof. Since the action set of each leader is $\{FC, FF, CC, CF\}$ there are 16 possible subgames corresponding to each pair of group leaders choices; and in each subgame the common leader has 4 choices.

In Figure 1 we illustrate the working of the model in the subgame corresponding to the profile FC,CC of the two group leaders. The other cases are similar. If the common leader plays CC then the implemented action is FC, and the second group punish the two leaders they follow, namely their group leader and the common leader (who have promised 1 against a realized group payoff of ξ). Whence the leftmost payoff $(\lambda + \xi - Q)/2, \lambda, \xi - P$. The other payoffs are obtained similarly. Obviously the best replies of the common leader are FF and FC - by which she induces the FC

outcome which gives her $(\lambda+\xi)/2$. Note that both choices yield the same payoffs to the group leaders. This latter fact is always true when the common leader has multiple best responses.

Figure 1: Shown is the FC, CC-subgame. In brackets the implemented profile.



Observe that whenever the common leader can induce the CC outcome without being punished she will do it because CC is her most preferred outcome. Recall that we are assuming $Q > |\lambda+\xi|$ (no restrictions on P). For each pair of group leaders actions the common leader best response(s) determine their payoffs; in the case just seen for example the group leaders payoff is $\lambda, \xi - P$. Table 1 displays the 16 possibilities in the case $\lambda+\xi > 0$, the row and column players being respectively $\ell = 1$ and $\ell = 2$. The corresponding common leader best responses are in brackets. The stars indicate the two players best responses. The unique pure equilibrium of the game clearly has both group leaders playing FF and the common leader playing CC.

Table 1: Group leaders' payoffs in the 16 subgames, with common leader's best responses in square brackets

	CF	FF	CC	FC
FC	$-P,-P$ $[CC, FF]$	$-P, 0^*$ $[FF]$	$\lambda^*, \xi-P$ $]FF, FC]$	λ^*, ξ $[FC]$
FF	$0^*, -P$ $[FF]$	$\mathbf{1}^*, \mathbf{1}^*$ $[CC]$	$1, 1^*$ $[CC]$	$1, 1^*$ $[CC]$
CC	$\xi-P, \lambda^*$ $[FF, CF]$	$1^*, 1$ $[CC]$	$1, 1$ $[CC]$	$1, 1$ $[CC]$
CF	ξ, λ^* $[CF]$	$1^*, 1$ $[CC]$	$1, 1$ $[CC]$	$1, 1$ $[CC]$

If $\lambda+\xi < 0$ the only difference in the resulting matrix is that when the group leaders play (FC, FC) or (CF, CF) the common leader's best response will be FF so the leader playing aggressively is punished. The equilibria, as is elementary checked from the resulting matrix, have group leaders play (FF, FF), (CF, FF), (FF, FC) or (CF, FC). The common leader always plays CC and establishes cooperation.

Observe that in this case all the subgame perfect equilibria have cooperation as outcome: no domination restriction is required. $\square$

In this game anticipation of the common leader's move will deter group leaders from playing aggressively. Indeed if say leader 1 plays FC she will get $-P$. The reason is that if group 2 play C they get at most $\xi < 0$, while if they play FF the common leader will play FF as well (with CC she would get punished) so group 2 can ensure zero payoff by playing FF; thus if leader 1 plays FC the outcome will be FF and she will get $-P$. But she can ensure a zero payoff by playing FF, hence she will not play FC. Similarly leader 2 will not play CF. Barred FC and CF, whatever pair of strategies the group leaders play the common leader can play CC and induce cooperation, which is her most preferred outcome.

4.2. All Leaders move simultaneously

Of course here the only subgame is the whole game, so we just have to find the Nash equilibria in weakly undominated strategies. And recall that with simultaneous moves the strategy set of all leaders is the set A of profiles of the underlying game, hence the game is $4 \times 4 \times 4$. However this case is a little more involved to analyse because mixed equilibria naturally arise. Consider for instance the profile $(s^0, s^1, s^2) = (CC, FC, CC)$. Here follower 1 complies with leader 1 because $u_1(FC) > u_1(CC)$ so group 1 plays F; and group 2 receives the proposal CC from both $\ell = 0, 2$; so group 2 plays C; therefore the implemented action profile is FC; leader 1 gets λ; the common leader and leader 2 are followed and punished by group 2 (since the group gets ξ against a promise of 1); so the common leader gets utility of $(\lambda + \xi)/2 - Q/2$ and leader 2 gets $\xi - P$. Of course the strategies of leaders $\ell = 0, 2$ are not best response - they would be better off playing FF for example; but then leader 1's strategy becomes suboptimal; and so on.

Elimination of weakly dominated strategies considerably simplifies this game. Indeed, for a group leader $\ell = k \in \{1, 2\}$, a proposal $s^\ell = (a_1, a_2)$ is weakly undominated if and only if $a_\ell = F$. So leader $\ell = 1$ will only play FC or FF and $\ell = 2$ will only play CF or FF. This is proved in Lemma 1 in Appendix A. Given this, for the common leader the strategies CF and FC are (strictly) dominated by FF, so the common leader will only play CC or FF. This is Lemma 2 in Appendix A.

Therefore the analysis is reduced to the $2 \times 2 \times 2$ game presented in Table 2, where the three payoffs in each entry are naturally ordered with the leaders' index (first common then the other two).

Table 2: The reduced game. The left panel shows utilities when the common leader plays CC; in the right panel are the payoffs when the common leader plays FF.

CC	CF	FF
FC	$0, -P, -P$	$\frac{\lambda+\xi-Q}{2}, \lambda, \xi$
FF	$\frac{\lambda+\xi-Q}{2}, \xi, \lambda$	$1, 1, 1$

FF	CF	FF
FC	$0, -P, -P$	$0, -P, 0$
FF	$0, 0, -P$	$0, 0, 0$

In the reduced game a strategy profile may be written as a vector of the form (q, p_1, p_2), q being the probability that the common leader plays CC, p_1 the probability that leader $\ell = 1$ plays FC and p_2 the probability that $\ell = 2$ plays CF.

The next result characterizes the equilibria of the leaders' game. All the equilibria are at least partially mixed, and the mixing probabilities are given in the next two displayed equations. Equation (4) below describes the profile where the common leader plays CC for sure and $p_1 = p_2 = \tilde{p}$:

$$\tilde{q} = 1\,, \qquad \tilde{p} \equiv \frac{\lambda - 1}{\lambda - 1 + P + \xi}. \tag{4}$$

Note that $\tilde{p}$ converges to 0 as P becomes large so the induced outcome converges to cooperation as P becomes large. The pair $(\hat{q}, \hat{p})$ in (5) below describes a fully mixed profile, where $p_1 = p_2 = \hat{p}$.

$$\hat{q} \equiv \frac{P}{P + \lambda - 1 - \hat{p}(P + \lambda + \xi - 1)}\,, \qquad \hat{p} \equiv \frac{1}{1 + Q - (\lambda + \xi)} \tag{5}$$

Precisely, the equilibria are the following (the result is proved in Appendix A, where all the equilibria are identified in a sequence of lemmas):

Proposition 3. *In the simultaneous moves game:*
If $P < -\xi$ the equilibria are all $(q, 1, 1)$ for $-P/\xi < q \leq 1$, with outcome FF.
If $P > -\xi$:
If $P + \xi < (\lambda - 1)(Q - (\lambda + \xi))$ the unique equilibrium is $(\hat{q}, \hat{p}, \hat{p})$
If $P + \xi > (\lambda - 1)(Q - (\lambda + \xi))$ the unique equilibrium is $(1, \tilde{p}, \tilde{p})$.

If the punishment of the group leaders is small, the only possible equilibrium outcome is conflict, just as in the underlying game. Consider now what happens for $P > -\xi$. The $(1, \tilde{p}, \tilde{p})$ equilibrium obtains uniquely if P is large enough, that is, if the group leaders bear adequate responsibility for their actions. In this equilibrium the common leader plays CC for sure, and as $P \to \infty$ the probability of aggressive play by the group leaders goes to zero, so the cooperative outcome obtains with probability 1 in the limit. Punishing the common leader harshly and leaving the group leaders relatively free - $Q > \lambda + \xi$ and $0 < P + \xi < (\lambda - 1)(Q - (\lambda + \xi))$ - is not as good. In the $(\hat{q}, \hat{p}, \hat{p})$ equilibrium the group leaders will play *FF* with high probability, but the common leader will also play *FF* with positive probability.

4.3. The Common Leader Moves First, Then the Group Leaders

We lastly consider the sequential game where the common leader moves first. In this case a group leader has four information sets, each corresponding to a proposal chosen by the common leader; her strategy set is thus the 4-fold Cartesian product of A. We shall see that in this case, unlike in the simultaneous moves version, there are equilibria where the common leader plays *FF* for sure; however in all of them all leaders are worse off than in the corresponding equilibria of the simultaneous moves game.

The weak dominance arguments in the reduction lemmas 1 and 2 still apply to this extensive form. For the sake of completeness the relative statements appear as Lemmas 9 and 10 in the Appendix (see page 21). We then conclude as before that the proposals CF and FC of the common leader are dominated. Therefore the group leaders have only two relevant information sets (corresponding to choices CC and FF of the common leader); and since their undominated proposals are only FC and FF for leader 1 and CF and FF for leader 2 the game is that of Table 1, reproduced below for convenience, where the two matrices report payoffs to the three leaders in the two subgames.

By backward induction we restrict attention to Nash equilibria in each subgame, where each such equilibrium is a pair of proposals, one by each group leader. It is easiest to look at these equilibria

Table 3: The reduced game. The left panel shows utilities when the common leader plays CC; in the right panel are payoffs when the common leader plays FF.

CC	CF	FF
FC	$0, -P, -P$	$\frac{\lambda+\xi-Q}{2}, \lambda, \xi$
FF	$\frac{\lambda+\xi-Q}{2}, \xi, \lambda$	$1, 1, 1$

FF	CF	FF
FC	$0, -P, -P$	$0, -P, 0$
FF	$0, 0, -P$	$0, 0, 0$

right away. The FF-subgame has the unique equilibrium $\phi \equiv (FF, FF)$, with implemented profile FF, for any value of P. Equilibria in the CC-subgame depend on the value of P.

If $P < -\xi$ then in the CC-subgame between the group leaders the unique equilibrium is the aggressive play $\alpha \equiv (FC, CF)$, with implemented profile FF. Therefore the only equilibrium pair in the two subgames is (α, ϕ) - α in the CC-subgame and ϕ in the FF-subgame. If $P > -\xi$ the CC-subgame has three equilibria: two in pure strategies, $\eta^1 \equiv (FC, FF)$ and $\eta^2 \equiv (FF, CF)$ with outcomes respectively FC and CF; and a mixed equilibrium with $p_1 = p_2 = \frac{\lambda-1}{\lambda-1+P+\xi} \equiv \tilde{p}$ (where p_1 is the probability that 1 plays FC, p_2 the probability that 2 plays CF). Possible equilibrium pairs are then $(\eta^1, \phi), (\eta^2, \phi)$ and $(\tilde{p}, \phi)$.

Equilibria of the leaders game where the common leader uses a pure strategy will be written as for example $(FF, (\eta^1, \phi))$ - meaning that the common leader plays FF and the equilibria played in the two subgames are respectively η^1 and ϕ. We are now ready to state

Proposition 4. *In the sequential case where the common leader moves first:*

If $P < -\xi$ the equilibria are $(CC, (\alpha, \phi))$, $(FF, (\alpha, \phi))$ and all those where the common leader mixes between CC and FF. In all of them the outcome is FF and all get zero, as in the $(q, 1, 1)$ equilibria of the simultaneous moves version.

For all $P > -\xi$ the profiles $(FF, (\eta^1, \phi))$ and $(FF, (\eta^2, \phi))$ with outcome FF are equilibria; in addition:

- If $P + \xi < (\lambda - 1)(Q - (\lambda + \xi))$ the only other equilibrium is $(FF, (\tilde{p}, \phi))$ also with outcome FF (in the simultaneous version where the unique equilibrium was $(\hat{q}, \hat{p}, \hat{p})$).

- If $P + \xi > (\lambda - 1)(Q - (\lambda + \xi))$ the only other equilibrium is $(CC, (\tilde{p}, \phi))$ (the latter corresponding to the unique $(1, \tilde{p}, \tilde{p})$ equilibrium of the simultaneous moves game).

Proof. If $P < -\xi$ the only equilibrium pair in the subgames is (α, ϕ), and in either of them the common leader gets zero; so she is indifferent between CC and FF.

Suppose now $P > -\xi$. If in the CC-subgame the group leaders play η^1 or η^2 then by playing CC the common leader gets $\frac{\lambda+\xi-Q}{2} < 0$ (punished by the group whose leader plays FF), so she will play FF.

Consider lastly the third possible equilibrium in the CC-subgame, which we denoted by $\tilde{p}$. If this is played then the common leader playing CC gets $(1-p)^2 + 2p(1-p)\frac{\lambda+\xi-Q}{2}$, which is positive at $\tilde{p}$ if and only if $P + \xi > (\lambda - 1)(Q - (\lambda + \xi))$. Recalling that if she plays FF she gets zero, she will then play CC if this condition is true and FF if it is false. □

Compared to the simultaneous moves version here there are two new equilibria with outcome FF for however large $P > -\xi$, namely $(FF, (\eta^1, \phi))$ and $(FF, (\eta^2, \phi))$. Consider $\eta^1 \equiv (FC, FF)$ for illustration. The point is that if leader 1 plays FC and the common leader plays CC leader 2 is better off playing FF getting ξ than fighting back with CF which yields $-P$. But at (FC, FF) the common leader playing CC would get punished by group 2 and get a negative payoff; her best response is FF which guarantees zero. So if one of the group leaders intends to play aggressively a cooperative proposal by the common leader is not viable, and all leaders end up playing FF.

4.4. Discussion of the general picture

Recall first Proposition 1: without a common leader the sure outcome is fight. In other words to open the possibility of a cooperative outcome in the leaders game the presence of a common leader is essential. The following discussion concerns the case where all leaders are present (and $Q > |\lambda+\xi|$ is assumed); the summary of the relevant results is given at the beginning of the section, on page 6.

Given that a common leader's preferences coincide with social welfare and that cooperation is socially optimal, her natural role in a conflict situation is to try to make the parties cooperate. We find that her effort are likely to succeed if P is large, except that when the group leaders have the last word and one of them tends to be aggressive then the common leader has no way to avoid conflict. Indeed the situation is as follows: if common leader has the last word cooperation obtains for sure. Otherwise cooperation obtains with high probability if P is large enough, except that if group leaders move last an equilibrium with outcome FF where the common leader essentially stays out is also possible.

The first thing to observe is that it is crucial that the groups can inflict significant losses to partisan leaders who fail to deliver on their promises. This obviously depends on the institutional context and the circumstances, but the point is that if the group leaders do not suffer significant penalties for lying they will tend to produce fighting and leave little room for the intermediation activity of external mediators.

The other fact that the model uncovers is that the order of moves matters. Of course real conflicts and negotiations are far more complex than the little model we have studied, and hardly ever do some parts have a "last word". Still, the model suggests that when a common leader does then she is likely to succeed in achieving a cooperative solution. In practice this may happen when the conflicting parties have been involved in a conflict for a while and at some point an external mediator starts to be part of the game.

Some real life cases seem to fit this picture reasonably well, such as the successful mediation of President Roosevelt. Note that he had a strong interest in maintaining good relations with both parties, so we are well into the case of adequately large Q. Another, more recent renowned figure that may be regarded as a successful common leader is Nelson Mandela. He had been jailed from 1964 to 1982 for opposing the apartheid system in South Africa. The groups, both on the white and black side, were radical and had been fighting for three decades. Mandela was elected President in 1994, and in his inaugural speech he declared: "The time for the healing of the wounds has come.

[...] The moment to bridge the chasms that divide us has come." His presidency gave a serious blow to the apartheid.[8]

One can also think of examples in more day-to-day internal politics. In this context it is usually the case that renowned figures from outside politics are asked to step in after political or financial crises. A case in point is Italy, where common leaders emerged in several episodes of severe national emergencies, especially in 2011 when professor Monti, then Rector of the Bocconi University, became prime minister in the midst of a financial crisis, where - with Berlusconi prime minister for the fourth time - the spread between interest rate on Italian debt and German bonds reached the alarming figure of 600 base points during the summer. In this case the preceding non cooperative phase may well be thought of as due to the low pressure - weak feasible punishment in our terms - by the constituency on the incumbent. Professor Monti was appointed by the Italian president "in the interest of the common good" on November 16th, and none of his government included politicians.[9] His government swiftly approved stability budget law which included pension reform and real estate tax, and this reassured markets: in less than a month that spread value nearly halved.

More often however, especially in international affairs, one observes many rounds of negotiations, with no player really "moving last". In our view in these more common, sometimes rather dramatic, situations the more appropriate model is that with all leaders moving simultaneously. The corresponding results apply therefore typically to civil wars within nations, and in these contexts the obvious candidate to play the role of common leader is the United Nations (UN). Doyle and Sambanis (2000) and Doyle and Sambanis (2006) conduct a detailed analysis of the UN operations since its onset. They list 121 civil wars between 1945 and 1999; of these, 99 ended with a military victory or a truce, that is without successful third parties interventions; and of the remaining 22, 14 ended with a negotiated settlement mediated by the UN, and in 12 of these cases there was no recurrence within 2 years from the settlement. So the UN was rather successful when it was able to advance a cooperative plan. But the prevalent outcome was conflict. This could be interpreted as the outcome of conflicts without a common leaders, or with group leaders superseding the common leader's efforts; but one might also argue that these are cases where the punishments for the group leaders are too low. Indeed in civil wars one may think that the punishment, seen as cost of failure, is particularly high (in the limit, death); but in war life is at risk whether you have promised victory or not, so that the additional punishment inflicted by followers is actually small. And in this case this is what the model predicts: high frequency of conflict, and sporadic occurrences of the cooperative outcome proposed by the common leader.

[8] The quotations are from Mandela's Reuters obituary, taken from reuters.com/article/uk-mandela-obituary-idUKBRE9B417G20131206. Incidentally, Mandela government's results were positive on the economic side as well: per capita GDP fell at the rate of 1.35% per year in the decade preceding 1994, and rose by 1.4% per year in the following decade (data from FRED).

[9] Quotation from https://presidenti.quirinale.it/Elementi/207444

4.5. The Case of Small Punishment for the Common Leader

We now turn to the case $Q < |\lambda + \xi|$. In all three games, if the common leader cannot be effectively punished she will induce equilibria with outcomes FF, FC or CF. This happens even in the most favorable case when she has the last word, where if $Q > \lambda + \xi$ the equilibrium with outcome CC obtains without strings attached. We comment on this case here. The complete result is spelled out in Proposition 6 at the end of the Appendix. It will be clear that it suffices to look at the case $\lambda + \xi > 0$, that is when $0 < Q < \lambda + \xi$.

Consider the game where the common leader moves after the group leaders, and to make the point even more starkly we also assume $P > -\xi$ which should curb the groups' incentive to play aggressively. It is easily verified (see the proof of Proposition 6) that all equilibria have outcomes FC or CF. The problem is that the common leader does not mind leaving a winner and a loser on the field, even if the loser punishes her, because the punishment is effectively irrelevant. For example, the profile where all leaders play FC is an equilibrium because even if leader 2 switched to FF the common leader would force the FC outcome by playing CC - she would be punished by group 2, but she prefers being punished at the asymmetric outcome - getting $(\lambda + \xi - Q)/2$ - than playing FF, inducing outcome FF and getting zero.

The bottom line is that a common leader who is listened to but not punished cannot induce cooperation. At best she can mislead one group into following her. This is an *external* figure who can make false promises at no cost, and in pursuing her interest will then not hesitate to impose a loss to a group by deliberately lying to them. It is doubtful whether a group may ever choose to follow her recommendations. The model just says that if they do they will have to regret it.

5. Conclusions

We have studied how political leadership can fundamentally alter outcomes in societies with group conflict when leaders are accountable to groups. We rely on a model of leadership which may be useful in general environments: given an underlying game among players, we construct a game among leaders in which the leaders' strategies are action profiles proposed by each leader to the society of players-followers, which can be interpreted as "recommendations and promises". Followers choose among the proposals to maximize their utility.

The main insight derived from the analysis of our model is that conflict in polarized societies can be substantially reduced, *under appropriate conditions*, thanks to the mediation of interested leaders. The existence of leaders by itself cannot accomplish anything useful: with only group leaders the equilibrium outcomes are the same as in the game with no leaders (Proposition 1). With common leaders, our analysis has identified a main driving force of the cooperation results: accountability of group leaders. In general cooperation and good outcomes are possible when the accountability of leaders is sufficiently large. In the limit of high accountability, cooperation may be realized with probability 1. On the other hand, as our introductory example shows, even common leader such as the UN may fail to induce cooperation. Lack of sufficient accountability of group leaders (a small P in our notation) is the most likely of such failures explanation within our framework.

Our setup relies on simplifying assumptions, and some of these assumptions may be in contrast with important real world regularities. In the model, leaders share precisely the utility of their constituencies, so their incentives are perfectly in line with those of the groups. Leaders do not have a political career to pursue, nor derive utility from being leaders. Leaders cannot profit directly or indirectly on their position. The common leader in particular is assumed to share the interests of society as a whole. Followers, on their part, make the task of the leaders as easy as possible: they hear what the leaders say, and take their promises at face value, with the understanding that punishment will follow if the leader does not deliver. Finally, punishment must be sufficiently high for cooperation to arise. Fortunately, our analysis makes clear the leaders' role, so it can be taken to provide the best case scenario for possible positive effects of mediation in group conflict. Systematic empirical research will have to decide which are the realistic ranges of the losses groups can impose on leaders.

The behavior of followers in our model is extremely simplified, but it is not completely unrealistic: in large and complex societies, understanding the structure of payoffs from social actions is at the same time very hard (because societies are complex) and unrewarding (because the action of each player - even when he has acquired enough information to evaluate the best choice - is in itself irrelevant). Thus a first simple approximation is to assume, as we do, that followers just consider the promised utility, and choose the highest.

Appendix A. Omitted Proofs

We collect proofs omitted from the text, including the relative statement.

Proof of Proposition 1

This result is actually true for any leaders game, with any number of groups. Observe that the model trivially extends to the case of K groups: just take $k, \ell \in \{1, 2, \ldots, K\}$ instead of $k, \ell \in \{1, 2\}$. Proving the statement for this more general case requires no additional effort, so we state it for this case:

Statement. *For any leaders game, if each group only considers the proposal of their own group leader, then at the Nash equilibria of the leaders game the distributions of action profiles chosen by the groups are the same as those induced by the Nash equilibria of the corresponding underlying game.*

Proof. For a mixed strategy $\hat{\sigma}^k$ of leader k we let $\hat{\sigma}^k_{A_k}$ the induced distribution on A_k. Our first claim is that

$$\forall \hat{\alpha} \in NE(UG) \exists \hat{\sigma} \in NE(LG) : \forall k,\ \hat{\sigma}^k_{A_k} = \hat{\alpha}_k, \tag{A.1}$$

where $NE(UG)$ and $NE(LG)$ denote the sets of Nash equilibria of the underlying game and leaders' game respectively. Consider a mixed action profile $\hat{\alpha} \in NE(UG)$. For any action $b_k \in \text{supp}(\hat{\alpha}_k)$ choose

$$a_{-k}(b_k) \in \text{argmin}_{c_{-k} \in A_{-k}} u_k(b_k, c_{-k}). \tag{A.2}$$

Define now $\hat{\sigma}^k$ as:

$$\hat{\sigma}^k(a) \equiv \sum\nolimits_{a_k \in A_k} \hat{\alpha}(a_k) \delta_{(a_k, a_{-k}(b_k))}(a). \tag{A.3}$$

If all leaders j different form k follow the strategy defined in (A.3) then leader k is facing the probability on A^{-k} given by $\hat{\alpha}_{-k}$. Consider now a possible strictly profitable deviation $\hat{\tau}^k$ from $\hat{\sigma}^k$. Since by following $\hat{\sigma}^k$ the k leader incurs no punishment cost, the increase in net utility to leader k from $\hat{\tau}^k$ is at least as large as the increase in direct utility, and the direct utility is the utility of the followers. Thus $\hat{\tau}^k$ would have a marginal on A_k that is a profitable deviation for player k from $\hat{\alpha}_k$ against $\hat{\alpha}_{-k}$, a contradiction with $\hat{\alpha} \in NE(UG)$.

The second claim is:

$$\forall \hat{\sigma} \in NE(LG),\ \text{if}\ \hat{\alpha}_k \equiv \hat{\sigma}^k_{A_k},\ \text{then}\ \hat{\alpha} \in NE(UG). \tag{A.4}$$

Consider in fact a strictly profitable deviation β_k from $\hat{\alpha}_k$ of a player k in the underlying game. Extend β_k to a profitable deviation τ^k in the leaders game of the k^{th} group leader following the construction in equations (A.2) and (A.3). This deviation would insure for group leader k, the same utility as β_k, which would then be higher than $\hat{\sigma}^k$, since the direct utility of τ^k is higher than $\hat{\sigma}^k$, and its punishment cost is zero; a contradiction with the assumption that $\hat{\sigma}^k$ is a best response. $\square$

Lemmas for the Reduction

Lemma 1. *For group leader $\ell = k \in \{1, 2\}$ the strategy s^k is weakly dominated if and only if $s^k_k = C$.*

Proof. We first show that strategies with $s^k_k = C$ are dominated by FF. Let $a, b \in A$ denote the strategies chosen by the two groups, and $g = (a_1, b_2)$ the implemented profile. Take $k = 1$.

Consider $s^1 = CF$ first. If $s^0 \in \{FF, FC\}$ then $g = (F, b_2)$ and direct utility is $U^1(g) \geq u_1(FF)$, and the same occurs if $s^1 = FF$; the inequality implies that there is no punishment either way, so under both strategies $V^1(g) = U^1(g)$. Suppose now $s^0 \in \{CF, CC\}$. If $s^0 = CF$ then $g = CF$ and $V^1(g) = \xi < 0$; on the other hand if $s^1 = FF$ then $g = FF$ so $V^1(g) = 0$ (no punishment since $u_1(g) = u_1(s^1)$). If $s^0 = CC$ then $g = (C, b_2)$ and $V^1(g) = u_1(g)$ (no punishment because leader 1 is not followed by her group); under FF nothing changes.

Consider now $s^1 = CC$. If $s^0 = FC$ then $g = (F, b_2)$ and $V^1(g) = u_1(g)$ (no punishment since leader 1 is not followed), and the same holds if $s^1 = FF$. If $s^0 = CF$ then $g = CF$ therefore $s^1 = CC$ yields $V^1(g) = \xi - P < 0$, while under $s^1 = FF$ we would have $g = FF$ and $V^1(g) = 0$ (no punishment since $u_1(g) = u_1(s^1)$). If $s^0 = CC$ then $g = (C, b_2)$ and $V^1(g) \leq u_1(g)$ (if $b_2 = F$ the inequality is strict because leader 1 is punished); in this case $s^1 = FF$ yields the unfollowed leader 1 payoff $V^1(g) = u_1(g)$. Suppose finally that $s^0 = FF$; if $b_2 = F$ then $g = CF$ and $V^1(g) = \xi - P$ while if $s^1 = FF$ then $g = FF$ and $V^1(g) = 0$ (no punishment since $u_1(s^1) = u_1(g)$); if $b_2 = C$ then $g = CC$ and $V^1(g) = u_1(CC) = 1$; but if $s^1 = FF$ then $g = FC$ whence $V^1(g) = u_1(g) = \lambda > 1$ (no punishment since $u_1(g) > u_1(s^1)$).

To show that any strategy with $s^k_k = F$ is not weakly dominated note that $s^1 = FF$ is a unique best response to $s^0 = CF$, and $s^1 = FC$ is a unique best response to $s^0 = s^2 = CC$. □

From Lemma 1 follows

Lemma 2. *After eliminating the dominated strategies in Lemma 1, strategies CF and FC for the common leader are strictly dominated.*

Proof. We do it for CF. This proposal is rejected by group 1 who will play F (because $s^1 \in \{FC, FF\}$), and accepted for sure by group 2; so the implemented profile is FF and group 2 will punish the common leader. She is better off by playing FF (which yields zero), strictly for any $Q > 0$. □

Proof of Proposition 3

We restate Proposition 3. The analysis which is organized considering three possible cases for the value of q, namely $q = 0$, $q = 1$ and then $q \in (0, 1)$.

Statement. *In the prisoners dilemma leaders game:*

If $P < -\xi$ the equilibria are all $(q, 1, 1)$ for $-P/\xi < q \leq 1$, with outcome FF.

If $P > -\xi$:

If $Q < \lambda + \xi$ the equilibria are $(1, 1, 0)$ and $(1, 0, 1)$ with outcomes FC and CF, and $(1, \tilde{p}, \tilde{p})$

If $Q > \lambda + \xi$ and $P + \xi < (\lambda - 1)(Q - (\lambda + \xi))$ the unique equilibrium is $(\hat{q}, \hat{p}, \hat{p})$
If $Q > \lambda + \xi$ and and $P + \xi > (\lambda - 1)(Q - (\lambda + \xi))$ the unique equilibrium is $(1, \tilde{p}, \tilde{p})$.

Proof. There is no equilibrium with $q = 0$ for any $P > 0$, from Lemma 3. Consider $P < -\xi$. We have equilibrium $(1, 1, 1)$ from Lemma 4; from Lemma 7 we have $(q, 1, 1)$ for $\max\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\} < q < 1$; and from Lemma 8 the same equilibrium for $-\frac{P}{\xi} < q < \frac{P}{P+\lambda-1}$ if that interval is nonempty. The last two give $(q, 1, 1)$ for $-\frac{P}{\xi} < q < 1$. Therefore if $P < -\xi$ we have $(q, 1, 1)$ for $-\frac{P}{\xi} < q \leq 1$, as in the statement. Turn to $P > -\xi$. For $Q < \lambda + \xi$ Lemma 5 gives $(1, 0, 1)$, $(1, 1, 0)$ and $(1, \tilde{p}, \tilde{p})$; for $Q > \lambda + \xi$ Proposition 5 gives $(\hat{q}, \hat{p}, \hat{p})$ if $P + \xi < (\lambda - 1)(Q - (\lambda + \xi))$, and Lemma 5 gives $(1, \tilde{p}, \tilde{p})$ for the reverse inequality. □

Lemmas for Proposition 3

We concentrate on the interesting cases in which the relevant inequalities among combinations of parameters hold strictly.

Equilibria with $q = 0$. We start with the fact that there are no such equilibria.

Lemma 3. *If $P > 0$, there is no equilibrium with $q = 0$*

Proof. If the common leader sets $q = 0$ then the leaders' game is the right panel of table 2 (ignoring the common leader's utility). This game has a unique Nash Equilibrium in dominant strategies in which both group leaders play FF. At this profile of actions of group leaders, CC yields 1, and FF yields 0, to the common leader, hence setting $q = 1$ is the best response. □

Equilibria with $q = 1$. We deal in turn with small P and larger P:

Lemma 4. *If $P < -\xi$ then there is a unique equilibrium with $q = 1$, with $(q, p_1, p_2) = (1, 1, 1)$.*

Proof. Since $\lambda > 1$ and $\xi < -P$, if $q = 1$ we see from table 2 that the action FC is dominant for the first group leader CF for the second). When group leaders play the action profile (FC, CF) then both CC and FF give utility 0 to the common leader, hence $(1, 1, 1)$ is the only equilibrium with $q = 1$. □

Lemma 5. *If $P > -\xi$:*

1. *There are two equilibria where group leaders play pure strategies: $(q, p_1, p_2) \in \{(1, 0, 1), (1, 1, 0)\}$ if and only if $Q < \lambda + \xi$. In these equilibria the outcome is FC or CF.*
2. *There is an equilibrium where group leaders play a mixed strategy if and only if:*

$$P + \xi > (\lambda - 1)(Q - (\lambda + \xi)) \tag{A.5}$$

The mixed strategy is $\tilde{p}$ in equation (A.6).

Proof. If $P > -\xi$ then at $q = 1$ the game among group leaders has three equilibria, the two pure profiles (FF, CF), (FC, FF) and a mixed one with:

$$p_1 = p_2 = \frac{\lambda - 1}{\lambda - 1 + P + \xi} \equiv \tilde{p} \tag{A.6}$$

Note that $\lambda > 1$ and our assumption that $P > -\xi$ insure that $\tilde{p} \in (0, 1)$.

We first consider the possible equilibria where group leaders play pure strategies:

1. If $\lambda + \xi - Q > 0$ then there are two equilibria, $(q, p_1, p_2) = (1, 0, 1), (1, 1, 0)$. This follows because CC gives $(\lambda + \xi - Q)/2$, while FF gives 0 to the common leader.
2. If $\lambda + \xi - Q < 0$ then there are no equilibria $(1, p_1, p_2)$ with $p_i \in \{0, 1\}$, because in this case the utility to the common leader from CC is lower than the one from FF.

We then consider the the possible equilibria where group leaders play a mixed strategy. At any mixed strategy profile (p, p), with $p \in (0, 1)$ of the group leaders the common leader playing CC gets

$$(1 - p)^2 + 2p(1 - p)\frac{\lambda + \xi - Q}{2}$$

and at $\tilde{p}$ this is larger than 0 (hence CC better than FF) if and only if (A.5) holds. □

Equilibria with $q \in (0, 1)$. To set up the analysis we assume that the common leader is playing q and compare a group leader's payoffs from FC and FF for each of the two possible strategies CF and FF of the other group leader. From Table 2 we see that in the first case FC is better than FF if and only if

$$q > -P/\xi \tag{A.7}$$

while in the second case FC is better than FF if and only if

$$q > \frac{P}{P + \lambda - 1} \tag{A.8}$$

In lemmas 6 and 7 we consider the two extreme possible cases for q:

Lemma 6. *There is no equilibrium with* $0 < q < \min\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\}$.

Proof. The condition on q implies that the action FF is dominant for both group leaders, but the common leader's best response to (FF, FF) is $q = 1$. □

Lemma 7. *There is an equilibrium with any q such that* $\max\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\} < q < 1$, *of the form* $(q, 1, 1)$.

Of course the set of such q's may be empty; this is the case when $P > -\xi$.

Proof. The condition on q implies that (FC, CF) is dominant for the group leaders, and at this profile the common leader gets zero both from CC and FF; the conclusion follows. □

Next we consider the intermediate cases for the values of q. At these values of q the game between the group leaders has three equilibria, two pure strategies and one mixed. We deal with pure strategies of group leaders in lemma 8. Observe that $\frac{P}{P+\lambda-1} < -\frac{P}{\xi}$ iff $P > 1-(\lambda+\xi)$.

Lemma 8. 1. *If* $\frac{P}{P+\lambda-1} < q < -\frac{P}{\xi}$ *then there is no equilibrium with* $p_i \in \{0,1\}$ *(that is, with group leaders playing pure strategies)*

2. *For any value* $-\frac{P}{\xi} < q < \frac{P}{P+\lambda-1}$*, there is an equilibrium in pure strategies for group leaders of the form* $(q,1,1)$.

Proof. For the first case, the two pure strategy equilibria in the resulting group leaders game are (FF,CF) and (FC,FF); consider the first (the second is analogous). In this case CC gives $\frac{\lambda+\xi-Q}{2}$, and FF gives 0. Considering only the cases in which the inequalities holds strictly, it follows that the best response of the common leader to this strategy profile of the group leaders is either $q=0$ or $q=1$, hence not in the open interval $(0,1)$.

For the second case, with q in that range the two pure strategy equilibria in the group leaders' game are (FC,CF) and (FF,FF). At the first profile the common leader gets zero from either CC or FF whence the equilibria; at the second one the common leader gets 1 from CC and zero from FF, hence the best reply is not interior. □

We lastly deal with the case of fully mixed equilibrium.

Proposition 5. *An equilibrium with* $\min\{-\frac{P}{\xi},\frac{P}{P+\lambda-1}\} < q < \min\{\max\{-\frac{P}{\xi},\frac{P}{P+\lambda-1}\},1\}$ *exists, with the mixed strategy* $(\hat{q},\hat{p},\hat{p})$ *defined in equation* (5)*, if and only if* $Q > \lambda+\xi$ *and* $P+\xi < (\lambda-1)(Q-(\lambda+\xi))$.

Proof. For q to be part of a fully mixed equilibrium the common leader has to be indifferent between CC and FF, which is true if and only if $(1-p)^2 + 2p(1-p)\frac{\lambda+\xi-Q}{2} = 0$ that is if

$$p = \frac{1}{1+Q-(\lambda+\xi)} \equiv \hat{p} \tag{A.9}$$

Note that $0 \le \hat{p} \le 1$ if and only if $Q \ge \lambda+\xi$. On the other hand the indifference for group leader 1 (for example) between FC and FF requires:

$$-pP + (1-p)(q\lambda - (1-q)P) = pq\xi + (1-p)q$$

which is rewritten as:

$$p = \frac{P+\lambda-1-P/q}{P+\lambda+\xi-1} \equiv f(q) \tag{A.10}$$

Combining equations A.9 and A.10 we conclude that an equilibrium with q in the range exists if $0<q<1$, $f(q)=\hat{p}$ and

$$\min\{-\frac{P}{\xi},\frac{P}{P+\lambda-1}\} < q < \max\{-\frac{P}{\xi},\frac{P}{P+\lambda-1}\}.$$

Observe that $\frac{P}{P+\lambda-1} < -\frac{P}{\xi}$ if and only if $P+\lambda+\xi-1>0$, in which case f is strictly increasing; and f is strictly decreasing if the inequalities are reversed. Since $f(\frac{P}{P+\lambda-1}) = 0$ and $f(-\frac{P}{\xi}) = 1$, there is unique $\hat{q}$ in the given range such that

$$f(\hat{q}) = \hat{p}. \tag{A.11}$$

it is easy to check that this $\hat{q}$ is indeed the value in equation (5).

If $\max\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\} < 1$ - that is if $P < -\xi$ - we are done. For $P > -\xi$ we have $\max\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\} = -\frac{P}{\xi} > 1$, hence in this case we must check whether an equilibrium exists with $\frac{P}{P+\lambda-1} < q < 1$. Since $f(\frac{P}{P+\lambda-1}) = 0$ and in this case f is increasing, the equilibrium exists if $f(1) > \hat{p}$, that is if

$$\frac{\lambda-1}{P+\lambda+\xi-1} > \frac{1}{Q-(\lambda+\xi-1)}$$
$$P+\xi < (\lambda-1)(Q-(\lambda+\xi))$$

where we have used that (since $P > -\xi$) $P+\lambda+\xi-1>0$ and that $Q-(\lambda+\xi-1)>0$ because $\hat{p} \in (0,1)$. If the inequality is false then the conditions for the equilibrium $(1, \tilde{p}, \tilde{p})$ are met. □

Reduction Lemmas for the Case Where the Common Leader Moves First

We are considering the extensive form game in which the common leader chooses a proposal in the set A. Differently from the simultaneous move game, we proceed to the elimination of the strategies of the group leaders in each subgame induced by the choice of action of the common leader.

Lemma 9. *For any proposal by the common leader, for group leader $\ell = k \in \{1,2\}$ the proposal s^k is weakly dominated in the corresponding subgame if and only if $s^k_k = C$.*

Proof. The argument consists in showing that FF is always at least as good. The step is similar to the one given in the simultaneous move case. We spell it out for the CC-subgame.

As before let $a, b \in A$ denote the strategies chosen by the two groups, and $g = (a_1, b_2)$ the implemented profile, and take $k = 1$.

If leader 1 recommends C to her group when the common leader plays CC then $g = (C, b_2)$, whence $V^1(g) = u_1(g)$ (no punishment because leader 1 is not followed by her group); but under FF the implemented action and hence her payoff do not change. The other cases are analogous.

The proof for the other subgames is similar, and follows the pattern we have seen for the simultaneous move game.

To show that any strategy with $s^k_k = F$ is not weakly dominated note that FF is a unique best response in the CF-subgame, and FC is a unique best response in the CC-subgame if also leader 2 proposes CC. □

From Lemma 9 follows

Lemma 10. *After eliminating the dominated proposals in Lemma 9, proposals CF and FC for the common leader are strictly dominated.*

Proof. We do it for FC. Since in the FC-subgame leader 1 will play either FC or FF group 1 will play F. And however leader 2 plays in the subgame follower 2 will comply with te common leader recommendation; so the implemented profile will be FF and group 2 will punish the common leader who thus gets $-Q < 0$. She is strictly better off by playing FF which yields zero. □

Details for the case of small Q

To avoid being too lengthy we focus on the range $\lambda + \xi > 0$.

Proposition 6. *Assume* $Q < \lambda + \xi$.

If the group leaders move first:

- If $P < -\xi$ *the equilibrium outcome is FF; both group leaders play aggressively - FC and CF respectively - and get punished, and the common leader opts out by playing FF or CC.*

- If $P > -\xi$ *in all equilibria the outcome is FC or CF; either all play FC or all play CF and no leader gets punished; or the two group leaders play* (FC, FF) *or* (FF, CF) *and the common leader plays CC and gets punished.*

In the simultaneous moves game:

- If $P < -\xi$ *the equilibria are all* $(q, 1, 1)$ *for* $-P/\xi < q \leq 1$, *with outcome FF.*

- If $P > -\xi$ *the equilibria are* $(1, 1, 0)$ *and* $(1, 0, 1)$ *with outcomes FC and CF, and* $(1, \tilde{p}, \tilde{p})$

If the common leader moves first:

- If $P < -\xi$ *the equilibria are* $(CC, (\alpha, \phi))$ *and* $(FF, (\alpha, \phi))$, *with outcome FF (as in the* $(q, 1, 1)$ *equilibrium of the simultaneous moves version).*

- If $P > -\xi$ *the equilibria are* $(CC, (\eta^1, \phi)), (CC, (\eta^2, \phi))$ *and* $(CC, (\tilde{p}, \phi))$ *(corresponding to* $(1, 1, 0)$, $(1, 0, 1)$ *and* $(1, \tilde{p}, \tilde{p})$ *of the simultaneous moves version).*

Proof. Assume $P > -\xi$. Low punishment $Q < \lambda + \xi$ for the common leader implies that she is better off at FC and CF, even if punished by one group, than at FF. The 4×4 matrix above now becomes Table A.4.

Table A.4: Case $Q < \lambda + \xi$ and $P > -\xi$

	CF	FF	CC	FC
FC	$-P, -P\ [CC, FF]$	$\lambda^*, \xi^*\ [CC]$	$\lambda^*, \xi - P\ ]FF, FC]$	$\lambda^*, \xi^*\ [FC]$
FF	$\xi^*, \lambda^*\ [CC]$	$1, 1\ [CC]$	$1, 1\ [CC]$	$1, 1\ [CC]$
CC	$\xi - P, \lambda^*\ [FF, CF]$	$1, 1\ [CC]$	$1, 1\ [CC]$	$1, 1\ [CC]$
CF	$\xi^*, \lambda^*\ [CF]$	$1, 1\ [CC]$	$1, 1\ [CC]$	$1, 1\ [CC]$

All pure strategy equilibria of the leaders game have now outcomes FC or CF. The problem here is that the common leader does not mind leaving a winner and a loser on the field, even if the loser punishes her. For example, the group leaders profile FC, FC is an equilibrium because even if leader 2 switches to FF the common leader will force the FC outcome by playing CC.

Still with $Q < \lambda + \xi$ assume now $P < -\xi$. The matrix is the same as in the case $P > -\xi$, what change are the best responses. Here both group leaders will play aggressively - FC and CF respectively - because they are better off being punished at FF than being the losers at FC or CF. This ends the first case.

The proof for the simultaneous moves case is contained in the proof presented in the appendix.

For the case where the common leader moves first the considerations made in the text for the case $Q > \lambda + \xi$ are sufficient to establish the present assertions. $\square$

Comparison with Correlated Equilibria

The leaders game built over an underlying game shares important features with the correlated equilibria of that underlying game: in both cases, thanks to a form of mediation, better outcomes than Nash equilibria can obtain; and in both solution concepts, leaders or the mediator suggest to followers an action profile, and followers respond. But the differences are deeper than the similarities.

In correlated equilibria the single mediator has no direct interest in the outcome; followers respond strategically to the action suggested privately to each, by updating the posterior on the action profile played by others, and would never want to punish the mediator. In the leaders game there are competing leaders with a direct interest in the outcome, so that their utility is affected by the action of the followers; the latter respond to the leaders' suggestions by choosing the best action profile from their point of view, and typically punish the chosen leaders with positive probability in equilibrium. Most importantly, although action profiles are implemented by the groups, the strategic interaction is among the leaders, not between the players of the underlying games.

We compare the sets of equilibrium action profiles taking as measurement of welfare the average utility of players in the underlying game (ignoring the welfare of the leaders which may include punishments).

Proposition 7. *If $P < -\xi$ the leaders equilibrium payoff is the same as the correlated payoff; otherwise it is strictly higher.*

Proof. The correlated payoff is zero. In the case where the group leaders move first we the group average payoff is 1. Consider next the simultaneous moves game. For $P < -\xi$ the leaders equilibrium payoff is zero. Turn to $P > -\xi$. The condition for average payoff in $(1, \tilde{p}, \tilde{p})$ to be positive is $\tilde{p}^2 * 0 + \tilde{p}(1-\tilde{p})(\lambda + \xi) + (1-\tilde{p})^2 * 1 > 0$, equivalently $P + \xi > -(\lambda - 1)(\lambda + \xi)$. But the $(1, \tilde{p}, \tilde{p})$ equilibrium obtains in the range $P + \xi > (\lambda - 1)(Q - (\lambda + \xi))$ which implies the condition above. Consider lastly $(\hat{q}, \hat{p}, \hat{p})$. If the common leader plays FF the possible outcomes are FC and FF both with positive probability hence average payoff is positive. If the common leader plays CC then the condition becomes as above $\hat{p}(\lambda + \xi) + (1 - \hat{p}) > 0$ which is $Q > 0$. In the case where the common leader moves first, in the only equilibria not corresponding to those of the simultaneous case both groups get zero. $\square$

References

BALIGA, S., D. O. LUCCA, AND T. SJÖSTRÖM (2011): "Domestic political survival and international conflict: is democracy good for peace?" *The Review of Economic Studies*, 78, 458–486.

DEWAN, T. AND F. SQUINTANI (2018): "Leadership with Trustworthy Associates," *American Political Science Review*, 112, 844–859.

DOYLE, M. AND N. SAMBANIS (2000): "International Peacebuilding: a Theoretical and Quantitative Analysis," *The American Political Science Review*, 94, 779–801.

——— (2006): *Making war and building peace: United Nations peace operations*, Princeton University Press.

DUCLOS, J.-Y., J. ESTEBAN, AND D. RAY (2004): "Polarization: concepts, measurement, estimation," *Econometrica*, 72, 1737–1772.

DUTTA, R., D. K. LEVINE, AND S. MODICA (2018): "Collusion constrained equilibrium," *Theoretical Economics*, 13, 307–340.

ELIAZ, K. AND R. SPIEGLER (2020): "A Model of Competing Narratives," *American Economic Review*, 110, 3786–3816.

ESTEBAN, J., L. MAYORAL, AND D. RAY (2012): "Ethnicity and conflict: An empirical study," *American Economic Review*, 102, 1310–42.

ESTEBAN, J.-M. AND D. RAY (1994): "On the measurement of polarization," *Econometrica: Journal of the Econometric Society*, 819–851.

——— (2008): "On the salience of ethnic conflict," *American Economic Review*, 98, 2185–2202.

HERMALIN, B. E. (1998): "Toward an Economic Theory of Leadership: Leading by Example," *The American Economic Review*, 88, 1188–1206.

MEIROWITZ, A., M. MORELLI, R. K. W., AND F. SQUINTANI (2019): "Dispute Resolution Institutions and Strategic Militarization," *Journal of Political Economy*, 127, 378–418.

PRAT, A. AND A. RUSTICHINI (2003): "Games played through agents," *Econometrica*, 71, 989–1026.