Cooperating Through Leaders*

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

Abstract

We model conflict as a prisoners dilemma where two groups play the game through leaders: two group leaders who share their preferences and a "common" leader who is concerned with general welfare. Leaders make recommendations and promises to the groups.

An accountable common leader will implement cooperation if she has "the last word". She will preempt aggressive play by the group leaders, who will comply with her cooperation proposal. If the common leader does not move last cooperation can be achieved with high probability if the group leaders suffer large penalties for not delivering on their promises.

Keywords: Social Conflict, Polarization, 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]Università di Palermo, Dipartimento di Matematica e Informatica (corresponding author)

[3]Department of Economics, University of Minnesota

1. Introduction

In 1905 the Russo-Japanese war was settled with the Treaty of Portsmouth mediated by the 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, in the absence of a successful third party intervention. Still, 14 ended with a negotiated settlement mediated by the UN. Under which circumstances can mediators succeed in fostering cooperation between the members of two in-conflict groups? This is the issue this paper investigates.

To this end we model situations where the intervention of an external agent concerned with the well-being of both parties in a conflict - which we call the "common leader" - might help in reaching a cooperative outcome. Such an external figure enters the picture as a third player, the first two being the natural group leaders of the conflicting parties. 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 the game played by the leaders.

We start with an underlying two-by-two game between two groups with actions C, cooperate or F, fight, where the cooperative profile CC gives the highest payoff. We focus on the prisoners dilemma, where the groups if unassisted would unavoidably fight. To this game we add leaders, who offer suggestions on the best course of actions to the groups and promise particular outcomes. 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 punish leaders who fail to deliver on their promises.

The leaders 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 they follow the recommendation they will get the FC payoff. The leaders' payoffs depend upon the outcome of the underlying 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; and a *common* leader, whose utility we take to be the average utility of the two groups. Thus in the leaders game there are three players: two group leaders and a common leader. Group leaders make proposals only to their own group; the common leader makes a proposal 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 cooperate 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 to her group, so that CC is the best outstanding offer for group 2; they will then follow the common leader proposal and play C. Now suppose that group leader 1 proposes FC - "Let us fight, they will submit" - to her group. 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 which they followed.

We now describe the main results of the analysis. Firstly, in the absence of a common leader, in the resulting game between the two group leaders the only equilibrium outcome is fighting, FF. So only if a common leader is present can conflict be avoided.

Assume a common leader is there. Then the order of moves, that is how the game evolves, matters. In the opening example President Roosevelt was successful intervening after the conflicting parties had been at war for a year. Indeed if the common leader moves last she may be able to implement cooperation. The crucial condition is that she be held accountable for her promises, in other words that she can be adequately punished. That an accountable common leader should be given the right to have the last word might be regarded as the main message of the paper. The condition is described more concisely at the beginning of Section 4 after the model is introduced, and some intuition on why it is so is given in Section 4.1.

If the common leader does not move last cooperation is never an equilibrium. A general condition favoring cooperation is that the groups can inflict large punishments to their group leaders. Formally, as the level of the group leaders' punishment goes to infinity there is an equilibrium where the probability of cooperation goes to 1. The intuition here is that the threat of punishment tends to deter group leaders from making excessively optimistic promises - think of the first group leader proposing FC - and this leaves room for a successful CC proposal by the common leader.

Comment on the role of groups. Regarding the role of the group members as passive followers, the point is that their expectations can be *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 the 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, not by the leaders, with a different research purpose. A similar analysis where groups may interact to a greater or lesser extent can be found in Block et al. (2025).

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 cited there. 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 on *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 in Appendix D.

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. Section 4 is the core of the paper: main result, intuition and discussion are there. The following Section 5 contains the detailed analysis the models studied, and Section 6 concludes. Proofs are mostly in appendices.

2. The Model

2.1. The Underlying Game

The are two homogeneous 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 the 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:

[4] Action profiles are always written so that 1 proceeds 2, and we are omitting commas when possible.

$U^0(a) = (u_1(a) + u_2(a))/2$.[5] We now describe how the profile a played by the followers and the possible punishments to the leaders are determined.

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. We present the games going from the case in which the common leader moves last, to the one in which she moves first. 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).

[5] We use superscripts for leaders and subscripts for followers.

[6] As a benchmark we will analyze 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.

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 $s^\ell = g^k(s)$, 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 interested in the conditions under which the implemented profile in the equilibria of the leaders game is the cooperative outcome, at least with positive probability.

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. 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. The General Picture with a Common Leader

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.

Recall that in the underlying game both groups get 1 at CC and zero at FF; excluding punishments, the utility of the common leader is the groups' average payoff, so she gets 1 in CC (her most preferred outcome) and zero in FF. It turns out that in the games we are studying the possibility of reaching a cooperative outcome for sure in equilibrium critically depends on whether Q is larger or smaller than $\lambda + \xi$.

If $Q < \lambda + \xi$, which we interpret as the common leader not being sufficiently accountable, the common leader's incentives are not conducive to cooperation. If on the other hand $Q > \lambda + \xi$ then cooperation emerges as the sole equilibrium outcome, provided that the common leader has the final decision-making authority. Some intuition for this is provided in the next subsection. Note that in particular this condition is satisfied for all Q in games where $\lambda + \xi < 0$. The case where the common leader does not move last is commented after the next statement.

To emphasize the importance of the common leader's opportunity to speak last consider an augmented game where, in a preliminary stage, the common leader chooses which of the three games will be played. After this selection, the chosen game proceeds as described. In this extended game, a unique subgame perfect equilibrium emerges where the common leader selects the sequential game that enables her to move last, leading to cooperation being the only outcome in that subgame.

The main points this paper makes are contained the following:

Theorem (Main Result)**.** *In a game where $Q > \lambda + \xi$, if the common leader moves last the only equilibrium outcome of the game is cooperation, CC; and a common leader who can choose the order of moves chooses to move last, and implements cooperation.*

If on the other hand $Q < \lambda + \xi$, for any order of moves CC is never an equilibrium outcome. In particular, if the common leader moves last the equilibrium outcomes are FF if $P < -\xi$ and FC and CF if $P > -\xi$.

If the common leader does not move last an equilibrium exists with outcome a mixture of CC and FF, with the probability of CC going to 1 as P goes to infinity.

Proof. The first assertion is proved in Proposition 2, Section 5.1. A common leader with these incentives will choose to move last because this choice is the only one for which in the ensuing game the equilibrium outcome is cooperation for sure. That cooperation is not a sure equilibrium outcome in any game with other order of moves follows from the next results.

Proposition 2 also shows that if $\lambda + \xi - Q > 0$ and the common leader moves last then the only equilibrium outcomes are FC, CF and FF.

That in the games where the common leader does not move last there are no equilibria with sure CC as outcome follows from Propositions 3 and 4, Sections 5.2 and 5.3 respectively. Existence of the mixed equilibria where the probability of CC goes to 1 is proved there as well. □

Thus cooperation is attained if the common leader moves last when she is accountable in the sense that $Q > \lambda + \xi$. In this case she would preempt aggressive play by the group leaders, who will then comply with her cooperation proposal.

An unaccountable common leader moving last will not implement cooperation. If the common leader does not move last, then the outcome depends on whether the group leaders are accountable in the sense that $P > -\xi$. As we shall see, if $P < -\xi$ then the only equilibrium outcome is FF: group leaders who cannot be punished tend to make war, and a common leader cannot prevent that. On the other hand, for large P an equilibrium where cooperation obtains with high probability exists. If P is large the group leaders will seldom make aggressive proposals for fear of punishment and most often accept the common leader cooperation proposal.

We may observe that there is a difference in the order of magnitude of the leaders' punishments required for cooperation. For the common leader it is enough that $Q > \lambda + \xi$, so any $Q > 2$ will do. In the cases where the probability of equilibrium cooperation depends on P, to obtain high probability P must go to infinity.

4.1. Some intuition

Assume that $Q > \lambda + \xi$ and that the common leader moves last. First, in this case a group leader gets at least zero by playing FF. Indeed, suppose leader 1 plays FF; if 2 plays CF then the common leader will play FF and get zero in the resulting FF outcome, since her alternative is to play CC to implement CF in which case she would get $(\lambda + \xi - Q)/2 < 0$; hence leader 1 will get zero. In this case the common leader effectively blocks the aggressive CF attempt by leader 2. If on the other hand 2 does not play CF then the common leader will play CC and implement cooperation, where leader 1 gets 1.

Given this, neither group leader will play aggressively. Suppose leader 1 plays FC. If 2 plays FC as well then the implemented action is either FC, in which case 2 gets $\xi < 0$, or FF, in which case 1 gets $\xi - P < 0$ - and neither outcome is possible in equilibrium. Similar arguments show that whatever 2 does at least one group leader gets a strictly negative payoff, contradicting the initial claim. Excluding aggressive play the group leaders are promising at most 1 to their respective groups, hence the common leader will implement cooperation by playing CC.

In the game where the three leaders move simultaneously mixed equilibria arise naturally. Suppose for example that the common leader plays CC for sure. If leader 1 does not play FC then 2 would play CF, in which case 1's best response is FC. This is the kind of circularity that leads to mixing. It is proved in the analysis of the game that leader 1 will randomize only between FC and FF, and similarly 2 between CF and FF. Still assuming that the common leader plays CC for sure, leader 1 playing FC get punished if 2 plays CF; and the higher the level of P the higher the risk of doing it. Thus for high P both group leaders tend to refrain from playing aggressively (playing FF instead), and cooperation obtains with high probability.

4.2. Discussion

Recall the main results. Without a common leader the sure outcome is fight: the presence of an external social welfare maximizer is essential for existence of a cooperative equilibrium. The common leader should be given the last word and she will then implement cooperation, provided she is accountable ($Q > \lambda + \xi$). An unaccountable common leader cannot do much good. The possibility of reaching a cooperative outcome remains open when the common leader does not move last if the group leaders are accountable ($P > -\xi$), and chances improve as the level of their punishment increases. If the group leaders do not suffer significant penalties for lying they will produce fighting.

It is not easy to single out a specific game timing as being more realistic than the others. Of course real conflicts and negotiations are far more complex than the simple model we have studied, and hardly ever do some parts have a "last word". Most often, especially in international affairs, one observes many rounds of negotiations, with no player really moving last. In these sometimes rather dramatic situations we tend to view the simultaneous moves game as the most appropriate. It is also conceivable that the common leader has little authority and after she offers her plans the group leaders go back to their groups and make "final" proposals ignoring her recommendations. On the other hand it is also possible that the common leader steps in later in the conflict, when presumably the group leaders have already given some advice to the groups, and it may be hard for them to change it, either because they are locked in or because *they* lost some authority due to the prolonged conflict, and in those cases it makes sense for the common leader to be modeled as moving last.

The successful mediation of President Roosevelt seems to fit the last picture reasonably well. He intervened after the war had been going on for more than a year, and his strong interest in maintaining good relations with both parties might suggest that the necessary incentives condition was satisfied.

However, cases of individual interventions in international affairs are not common. Especially in contexts of civil wars within nations the obvious candidate to play the role of common leader is the United Nations. Now: can the factions involved in local conflicts punish the UN? The sensible answer seems that no, they cannot: Q here is essentially zero. The incentives of the common leader are then entirely determined by the parameters λ and ξ of the underlying game, and as we know they are conducive to cooperation in games where $\lambda + \xi < 0$. This condition may or may not be satisfied in particular cases. If it is not, remember that the probability of cooperation increases with the level of P, that is with the strength of the pressure groups can exert on their leaders.

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; 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 successful when it was able to advance a cooperative plan; but the prevalent outcome was conflict. These events could be interpreted as the outcome of conflicts without a common leaders, or with group leaders superseding the common leader's efforts, with low levels of punishments for the group leaders. Indeed in the case of 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 cases this is what the model predicts: high frequency of conflict, and sporadic occurrences of the cooperative outcome proposed by the common leader.

In the context of internal politics the typical situation is different: conflicting political parties are locked in stalemate or have provoked some sort of crisis, and they resort to asking an external figure for help. In this cases Q may stand for reputational loss, and it can be relevant.

A case in point may be the 1952 US Presidential election. It was a contest between two very distant group leaders, and Dwight Eisenhower was asked to be a candidate by both parties.[8] Prior to the election he had no political affiliation, and a failed presidency would tarnish his reputation as a war hero. In office Eisenhower did act for the public good, and his success at bringing the nation together is indicated by the fact that he has been the most popular president of the US post-war history.

The crisis of 2011 in Italy is another episode we can discuss. In the midst of a serious financial crisis, in November professor Monti - then Rector of the Bocconi University - succeeded Berlusconi as prime minister with the largest majority ever recorded in the Italian republican history. In the preceding non cooperative phase the conflict was taking place within the ruling coalition: Berlusconi wanted to spur the economy, while the minister of the economy Tremonti (representing the Lega and the productive North) would instead focus on countering possible adverse repercussions of the subprime crisis. This conflict upset markets, the situation became unsustainable and Berlusconi resigned. On the other hand the Italian economy was on solid ground. The cooperative, first-best outcome was to look beyond the two short-term alternatives over which the parties were fighting and focus on the structural pro-growth long overdue reforms. Monti swiftly approved a stability budget which reassured markets, and then began acting on Italy's structural bottlenecks, notably on its rigid labour market and its unsustainable pension system.

5. Analysis of the Games

We first summarize the formal results; details are offered in the next three subsections.

Summary of Results. (1) In the game where the common leader moves after the group leaders: If $Q > \lambda + \xi$, the only equilibrium outcome is cooperation. If $Q < \lambda + \xi$ the only equilibrium outcomes are FC and CF if $P > -\xi$ and FF if $P < -\xi$.

(2) In the simultaneous moves game: If $P < -\xi$ the only equilibrium outcome is FF. If $P > -\xi$ the unique equilibrium outcome is a mixture of CC and FF, where the probability of CC goes to 1 as $P \to \infty$.

(3) In the game where the group leaders move after the common leader: Both for $P < -\xi$ and $P > -\xi$ there are equilibria as in the previous case; but equilibria with outcome FF also exist for any P.

[8] The Republican Robert A. Taft was an arch-conservative who opposed the New Deal, opposed US entry into World War II, opposed the Nuremberg trials, opposed NATO, the UN, and labor unions. The Democrat, Adlai Stevenson, along with Eleanor Roosevelt was the leader of the progressive movement: for labor unions and for expanding the New Deal, and a leading internationalist who was instrumental in the founding of the UN. See Cotter (1983).

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

5.1. 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.[9]

Proposition 2. *If the common leader moves after the group leaders:*

If $Q > \lambda + \xi$ the only equilibrium outcome is cooperation, CC;

If $Q < \lambda + \xi$ the only equilibrium outcomes are FC and CF if $P > -\xi$ and FF if $P > -\xi$.

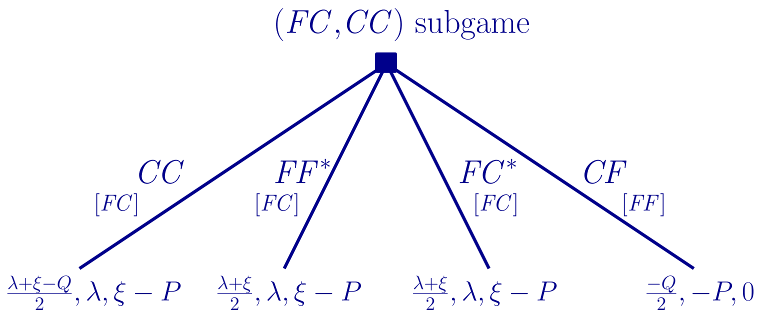
Proof. Since the action set of each leader is $\{FC, FF, CC, CF\}$ there are 16 possible subgames corresponding to each pair of group leaders choices (and in each subgame the common leader has 4 choices). In each subgame, given the common leader's optimal choice the implemented action is determined, and this determines the group leaders' payoffs. These can therefore be displayed in a 4×4 matrix (as in Table 1 below), the row and column players being respectively $\ell = 1$ and $\ell = 2$, each entry containing their payoffs in the corresponding subgame.

Start with the case $0 < \lambda + \xi < Q$. 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.

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. 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 (recall that we are assuming $Q > \lambda + \xi$). The corresponding common leader best responses are in brackets. The stars indicate

[9] In the following proposition no weak dominance restriction is required: the results hold for *all* subgame perfect equilibria.

Figure 1: Shown is the (FC,CC)-subgame. Best responses of the common leader are stared. In brackets the implemented profile is shown.



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. For example the (FC,CC) entry $\lambda, \xi - P$ comes from the (FC,CC)-subgame shown in Figure 1, where the common leader's best responses are FF and FC.

	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^*	$[CF]$	$1^*, 1$	$[CC]$	$1, 1$	$[CC]$	$1, 1$	$[CC]$

Consider next the case $\lambda + \xi < 0$. Accept for a moment that there is no equilibrium where group leader 1 plays FC or leader 2 plays CF. Barring those plays the 4×4 payoff matrix becomes 3×3, and it is immediate to verify that in all cases the best response of the common leader is CC and cooperation obtains. It remains to exclude aggressive play by the group leaders in equilibrium. Suppose 1 plays FC. If 2 plays FC, FF or CF then since $\lambda + \xi < 0$ clearly the common leader plays FF, implemented profile is FF and 1 gets $-P$. If 2 plays CC the common leader's best response now depends on Q (see Figure 1): if $Q < |\lambda + \xi|$ she will play CF (getting $-Q/2$ in the resulting outcome FF) and 1 will get $-P$; if $Q > |\lambda + \xi|$ the common leader will play FF or FC (getting $(\lambda + \xi)/2$ in the outcome FC), and 2 will get $\xi - P$. In both cases a group leader would get less than zero. But any group leader can guarantee herself zero by playing FF. Indeed, considering leader 1, if 2 plays CF the common leader will play FF, outcome will be FF and 1 will get zero; if 2 plays otherwise the common leaders will play CC and 1 will get 1. Obviously the argument holds for both group leaders.

Lastly, suppose $0 < Q < \lambda + \xi$. 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. Table 1 now becomes Table 2.

All pure strategy equilibria of the leaders game have now outcomes 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.

Table 2: Case $Q < \lambda + \xi$ and $P > -\xi$

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

To finish consider $P < -\xi$. The matrix is the same as in the case $P > -\xi$, what change are the group leaders' best responses: if 1 plays FC then 2's best response is CF, and vice versa. In the only equilibrium both group leaders play aggressively and get punished, and the common leader opts out by playing FF or CC. □

5.2. *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 $4 \times 4 \times 4$. This case is a little more involved to analyze because mixed equilibria naturally arise.

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 5 in Appendix B. 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 6 in Appendix B. Therefore the analysis is reduced to the $2 \times 2 \times 2$ game presented in Table 3, where the three payoffs in each entry are naturally ordered with the leaders' index (first common then the other two).

Table 3: 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 $p_1 = p_2 = \hat{p}$.

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

In Appendix B it is proved that the equilibria are the following:

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* $(1, \tilde{p}, \tilde{p})$*.*

In this game, if P is small ($P < -\xi$) then the equilibrium outcome is FF, just as in the underlying game; for fixed Q, if P is large, that is if the group leaders bear adequate responsibility for their actions, then the relevant equilibrium is $(1, \tilde{p}, \tilde{p})$: the common leader plays CC for sure, the probability of cooperation increases with P, and as $P \to \infty$ the probability of aggressive play by the group leaders is small for obvious reasons and in fact it goes to zero, so the cooperative outcome obtains with probability 1 in the limit. This equilibrium is independent of Q.

Large Q alone is not sufficient for cooperation with high probability. The relevant equilibrium becomes $(\hat{q}, \hat{p}, \hat{p})$, and if $Q \to \infty$ then the probability of cooperation goes to 1 only if also $P \to \infty$. On the other hand, in this equilibrium when $P + \xi$ and $Q - (\lambda + \xi)$ are close to zero fighting emerges with probability close to 1.

5.3. 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 outcome is FF; 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 14 and 15 in Appendix C. 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 3, reproduced here for convenience.

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 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.

Table 3: 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	$0, 0, -P$	$0, 0, 0$

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. Recall that in the equilibrium $(CC, (\tilde{p}, \phi))$ the probability of cooperation depends on P, and it goes to 1 as $P \to \infty$. We are now ready to state

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

If $P < -\xi$ the only equilibrium outcome is FF; the equilibria are $(CC, (\alpha, \phi))$, $(FF, (\alpha, \phi))$ and those where the common leader mixes between CC and FF.

If $P > -\xi$ and $Q < \lambda + \xi$ the profile $(CC, (\tilde{p}, \phi))$ is an equilibrium; the other equilibria are $(CC, (\eta^1, \phi)), (CC, (\eta^2, \phi))$ with outcomes FC and CF respectively.

If $P > -\xi$ and $Q > \lambda + \xi$ the equilibria $(FF, (\eta^1, \phi))$ and $(FF, (\eta^2, \phi))$ with outcome FF exist, together with another equilibrium: $(FF, (\tilde{p}, \phi))$ also with outcome FF if $P < -\xi + (\lambda - 1)(Q - (\lambda + \xi))$; and $(CC, (\tilde{p}, \phi))$ if $P > -\xi + (\lambda - 1)(Q - (\lambda + \xi))$.

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 $P > -\xi$. Assume first $0 < Q < \lambda + \xi$. If the continuation equilibria are (η^1, ϕ) and (η^2, ϕ) obviously the common leader prefers CC to FF. Given continuation $(\tilde{p}, \phi))$ the same holds if and only if the CC payoff, which is $(1-p)^2 + 2p(1-p)\frac{\lambda+\xi-Q}{2}$, is larger than zero at $\tilde{p}$; this occurs if and only if $P > -\xi + (\lambda - 1)(Q - (\lambda + \xi))$, which is implied by $0 < Q < \lambda + \xi$.

Next consider $\lambda + \xi < 0$ or $0 < \lambda + \xi < Q$. 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. If on the other hand the continuation is $(\tilde{p}, \phi))$ the conclusion follows from the analysis of the previous case. □

Here the cooperative outcome CC may emerge only in some of the many equilibria, mainly if P is large. Compared to the simultaneous moves version, here there are two new equilibria with outcome FF for however large $P > -\xi$. 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 her 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.

Overall, having the common leader move first is less favorable to cooperation since it raises the possibility of unfavorable equilibria that do not exist in the simultaneous move case.

6. Conclusions

We have studied how political leadership can fundamentally alter outcomes in societies with group conflict. We rely on a model of leadership which may be useful in general environments: given an underlying game among groups, we construct a game among leaders in which the leaders' strategies are action profiles proposed by each leader to the groups, which can be interpreted as "recommendations and promises". Groups choose among the proposals to maximize their utility. Our particular attention, differently from existing literature on leadership, is on the role of common leaders acting as *super partes* mediators, who have regard for common interests to all groups and try to maximize aggregate welfare.

The main insight derived from the analysis of the model is that conflict in polarized societies can be substantially reduced thanks to the intervention of leaders who share the interests of the groups in conflict, which we call *common leaders*. Partisan leaders alone cannot accomplish anything useful, without a common leader the equilibrium outcomes are the same as in the game with no leaders.

More specifically, a common leader should be given the right to act after the group leaders have advanced their proposals, provided she is accountable; and this is the case if the punishment she may have to suffer is higher than her payoff from a winner-loser outcome. If this condition is satisfied the common leader will implement cooperation, by preempting aggressive behavior of the group leaders.

In the games where the common leader cannot bring up cooperation the main driving force of cooperation is accountability of group leaders: cooperation (in equilibrium) is possible with positive probability when the group leaders are accountable in the sense that their punishment is higher than the disutility suffered as loser in a winner-loser outcome. In the limit as the value of punishment goes to infinity cooperation may be realized with probability 1.

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. They 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 assuming, as we do, that followers just consider the promised utility and choose the highest seems a reasonable approximation.

Appendix A. Proof of Proposition 1

This result is in fact 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. $\square$

Appendix B. Proofs for the Simultaneous Moves Game

Lemmas for the Reduction

Lemma 5. *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 5 follows

Lemma 6. *After eliminating the dominated strategies in Lemma 5, 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 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 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 7. Consider $P < -\xi$. We have equilibrium $(1,1,1)$ from Lemma 8; from Lemma 11 we have $(q,1,1)$ for $\max\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\} < q < 1$; and from Lemma 12 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 9 gives $(1,0,1)$, $(1,1,0)$ and $(1,\tilde{p},\tilde{p})$; for $Q > \lambda + \xi$ Proposition 13 gives $(\hat{q},\hat{p},\hat{p})$ if $P + \xi < (\lambda - 1)(Q - (\lambda + \xi))$, and Lemma 9 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 7. *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 3 (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 8. *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 3 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 9. *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:*

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

The mixed strategy is $\tilde{p}$ *in equation (B.2).*

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:

$$p_1 = p_2 = \frac{\lambda - 1}{\lambda - 1 + P + \xi} \equiv \tilde{p} \tag{B.2}$$

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 (B.1) 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 3 we see that in the first case FC is better than FF if and only if

$$q > -P/\xi \tag{B.3}$$

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

$$q > \frac{P}{P + \lambda - 1} \tag{B.4}$$

In lemmas 10 and 11 we consider the two extreme possible cases for q:

Lemma 10. *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 11. *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 12. Observe that $\frac{P}{P+\lambda-1} < -\frac{P}{\xi}$ iff $P > 1-(\lambda+\xi)$.

Lemma 12. 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 13. *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

$$p = \frac{1}{1+Q-(\lambda+\xi)} \equiv \hat{p} \tag{B.5}$$

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:

$$p = \frac{P+\lambda-1-P/q}{P+\lambda+\xi-1} \equiv f(q) \tag{B.6}$$

Combining equations B.5 and B.6 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

$$f(\hat{q}) = \hat{p}. \tag{B.7}$$

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. □

Appendix C. Case Where the Common Leader Moves First

We prove here the reduction lemmas stated in the text. 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 14. *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 14 follows

Lemma 15. *After eliminating the dominated proposals in Lemma 14, 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. □

Appendix D. 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 16. *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. □

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.

BLOCK, J., R. DUTTA, AND D. K. LEVINE (2025): "Leaders and Social Norms: On the Emergence of Consensus or Conflict," *Journal of Economic Behavior and Organization,* forthcoming.

COTTER, C. P. (1983): "Eisenhower as party leader," *Political Science Quarterly*, 98, 255–283.

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.