Cooperating Through Leaders*

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

Abstract

We study games of conflict (among groups, or countries) where players can choose to fight or cooperate. We consider games where conflict is detrimental: the average welfare from outcomes in which conflict occurs (that is in which at least one of the players chooses fight) is smaller than the cooperation outcome. Depending on parameters, this game can be one of four types, called here Mutual interest, Stag hunt, Prisoner's dilemma and Chicken.

The novelty of our approach is that group choices are made under guidance of leaders who offer proposals to passive followers on the best course of action. *Accountability* of leaders is possible because of ex-post punishment when the realized utility is smaller than that implicitly promised. *Competition* among leaders is possible if groups are willing to listen to more than one leader. We prove that in all games the limit outcome is efficient under competition if accountability is sufficiently large.

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

*First Version: January 6, 2022. We would like to thank Rohan Dutta, and Sandeep Baliga for a short but illuminating exchange. DKL and SM gratefully acknowledge support from the MIUR PRIN 2017 n. 2017H5KPLL_01. 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, EUI and WUSTL

[2]Università di Palermo, Dipartimento SEAS

[3]Department of Economics, University of Minnesota

1. Introduction

We model groups interacting through the mediation of leaders, who offer guidance by proposing a course of action. Our setting is extremely simple: symmetric two-by-two games, with two players. We label cooperation the action that, if taken by both players, gives a utility higher than if both players choose the other action.

The novelty of our approach is that a different set of players, called leaders, offer proposals on the best course of actions to the groups, who then act as followers and take the action proposed by the chosen leader. Leaders act to influence the outcome of the game among the groups because their own utility depends on that outcome. The presence of leaders induces a new game between the leaders. A leader's utility from action profiles in the underlying game may be identical with that of a particular group - in which case we call the leader a "group leader" - or it may be an average of the utilities of the two groups, in which case the leader is a "common leader". The overall payoff of a leader is given by this direct utility plus a possible punishment inflicted by her followers; and the followers inflict their chosen leader a punishment when their realized utility is smaller than that implicitly promised by the leader. Punishments insure that leaders are accountable to their promises. Competition among leaders arises if, in addition to group leaders, a common leader is also present, and followers compare the proposal of their own group leaders with those of the common leader. This abstract structure provides a model of the role of political mediation in group conflict in polarized societies, an element that seems essential and so far not well studied.

The issue of polarization and potential conflict among groups has acquired particular relevance in the period following the second world conflict, as the new post-colonial order emerged. This development was anticipated in the farsighted book by Furnivall (2014) on the development of Burmese society after independence, and the conflicts potentially arising in a multi-ethnic society. Furnivall introduced the key concept of *plural society*, defined as "comprising two or more elements or social orders which live side by side, yet without mingling in one political unit". The concept was further elaborated by Rabushka and Shepsle (1971): "in the plural society - but not in the pluralistic society - the overwhelming preponderance of political conflicts is perceived in ethnic terms." The authors note that this definition "does not explain why some culturally diverse societies are plural and others are not. Typically, however, definitions are not called upon to perform such tasks. What is needed is a theory - a theory, we argue, of political entrepreneurship." Building this theory is the main purpose of this paper. Related ideas on polarized society were discussed in Lijphart (1977) and Fearon and Laitin (1996). Papers providing analytical foundations to this idea include Esteban and Ray (1994), Esteban and Ray (2011) and Duclos et al. (2004), who construct a general, well founded measure of polarization. The *salience* of ethnic conflict, which was the main parameter marking the transformation from pluralistic to plural society in Rabushka and Shepsle (1971), 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), which provide support for the theory). In the context of provision of public goods, a related issue is explored in Alesina et al. (1999); here individuals live in the same city but have different ethnicity and thus heterogeneous preferences on public good;

this fragmentation induces inefficiently low provision of public goods.

In the literature on polarized societies we have reviewed so far, groups act directly, with no intervention of political mediators. But large groups usually interact, particularly in the political arena, through leaders, so that in large part games between groups are really played by leaders. Our main contribution here is to provide a simple theoretical framework to analyze how group interaction is mediated by the rational and self-interested interventions of leaders.

As indicated, in this paper we study a family of underlying two-group games arising from conflict situations; but we emphasize that the construction of leaders' games from underlying followers' games can be implemented for any game, and is of more general interest. In our approach, leaders have preferences over outcomes (that is action profiles in the underlying group game). A leader is linked to a group because the leader shares, fully or partially, the utility of the group. This assumption is related to the idea of citizen-candidates introduced in the context of voting models, (Osborne and Slivinski (1996), Besley and Coate (1997)). Leaders compete to lead groups with the purpose of influencing outcomes, by proposing an action profile in the game between groups. Each group chooses the profile which, if realized, gives them the highest utility, and then implements the corresponding action of the group. The idea is the same as in Eliaz and Spiegler (2020), where a representative agent chooses among policy proposals and then selects and implements the one with the highest expected payoff. The difference here is that the proposers (which we call leaders) are modeled explicitly and that each of them may address several representative followers at the same time - just as many as there are groups. Moreover, crucially, the leaders can be made accountable for their actions. Indeed, if groups end up getting a lower utility than that implicitly promised, they inflict the leader (or leaders) responsible for the proposal some form of punishment. This punishment is a simplified form of accountability that binds leaders or politicians (see Ferejohn (1986), Maskin and Tirole (2004), Besley and Case (1995), Besley (2006)). Thus the interaction among leaders becomes central to the unfolding of the group conflict, and the underlying games between groups result in games between leaders. To briefly sum up, in these games leaders make policy proposals in the form of action profiles, then each group acts according to the proposal of their choice, and the leaders' payoff is the utility of the realized play plus the punishments inflicted for not delivering. We study equilibria of these games, and show that the outcomes of these games differ substantially from the equilibria of the underlying game.

In summary, we find that the presence of leaders transforms the nature of the underlying game among groups. The fact that the groups delegate decisions to the chosen leaders implies that the game that matters is played by leaders. In this model of games played through leaders groups can achieve cooperative outcomes in games, like the prisoners dilemma, where this is not possible if they play directly as groups. For cooperation to occur, two conditions must be met. First, there must be *competition* among leaders, that is, followers must be able to listen to many sides, and not just to what their group leaders proposes. Second, there must be *accountability*: bad proposals of leaders must be punished by followers, when the realized outcomes are worse than the promised ones. In a word, the insight offered by this paper is that competing, accountable leaders enable

groups to achieve cooperation with surprisingly high probability, tending to full probability as the accountability becomes sufficiently large.

There are other studies where delegation and/or leadership has a role. In the tradition of Barro (1973) (and Miquel (2007) for an application to divided societies), Baliga et al. (2011) develop a model of conflict between countries related to our conflict game, (see also Baliga and Sjöström (2004) and Baliga and Sjöström (2020)). Individuals in the countries (groups) have different payoffs, and may be hawkish (the aggressive action is dominant) or dovish (the accommodating strategy is dominant). There are leaders who choose strategies, and citizens retrospectively support or not the leader, depending on whether the action of the leader was a best response to that of the opponent from their point of view. The main difference with our approach is that the choice in our model is made by the citizens, not by the leader; the latter can only influence the choice of the citizens with their proposals. Also in Dutta et al. (2018) leaders choose strategies, and moreover their utility is not linked to that of the groups - which is the central feature of the present paper. In the tradition of games with common agency (Dixit et al. (1997)), 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. The setup is different from the one used here, where the direct utility of leaders and followers may be the same, and defined on outcomes, with no transfers; though leaders can be punished so their overall payoff may differ from that of the followers.

The sequel of the paper proceeds as follows. The underlying games of interest are introduced in Section 2, and it is seen there that they are of four types: Mutual Interest, Stag Hunt, Chicken and Prisoners Dilemma. The leaders game is defined and illustrated Section 3, and in the subsequent sections its equilibria are analyzed. We start in section 4 with the case where only group leaders are active, and in section 5 we introduce the active common leader. Section 6 deals with Mutual Interest and Stag Hunt, where cooperation is an equilibrium in the underlying game. Sections 7 and 8 concern Prisoners Dilemma and Chicken. The main properties of the leaders equilibria are described in Section 9. Section 10 examines the relation between these equilibria and the correlated equilibria of the underlying games. In Section 11 we look at the possibility that group and common leaders may be punished to different extent. And Section 12 contains concluding comments. Most proofs are in Appendix.

2. The Underlying Games

We study symmetric games with two players $k = 1, 2$, interpreted as large homogeneous groups. Each player has two possible actions, C (cooperation) and F (fight). We assume that if both play C they get a higher von Neumann-Morgenstern utility than if they both play F. This is just a labeling convention: if two groups enjoy war more than peace, say in pursuit of honor in battle, then that is their way of cooperating and get higher utility. Thus we convene that

$$\text{for } k = 1, 2, \ u_k(C, C) > u_k(F, F) \tag{1}$$

(where u_k is player k's utility). Using invariance under monotonic linear transformations and (1), we assume

$$\text{for both } k \ , u_k(C,C) = 1, u_k(F,F) = 0 \tag{2}$$

so, with $\lambda, \xi \in \mathbb{R}$, the family of games becomes the following:

	C	F
C	$1,1$	ξ, λ
F	λ, ξ	$0,0$

In summary, we will be considering two-by-two two players games with a specific label attached to the actions; the label cooperation is chosen to indicate a desirable social outcome, because of (1). We are examining the conditions on the political structure that, when cooperation is desirable, make it an equilibrium.

Considering the combination of the two possible inequalities between λ and 1 on the one hand and ξ and 0 on the other, we have two sets of possible games. One has with $\lambda > 1$, so the choice of F against C of the opponent is better than the choice of C: these are Prisoner's Dilemma if $\xi < 0$ and Chicken if $\xi > 0$. We call these *conflict games*, because (C, C) is not a Nash equilibrium of the game. The other set of possible games has $\lambda < 1$, so the choice of C against C of the opponent is better than the choice of F: they are Stag Hunt if $\xi < 0$ and Mutual Interest if $\xi > 0$. We call them *cooperation games*, because (C, C) is a Nash equilibrium of the game.

2.1. Average Welfare Restriction

The real restriction on symmetric games we introduce is that the unilateral deviation from the best common action profile reduces average welfare, where we take simple average assuming that the groups have equal size:

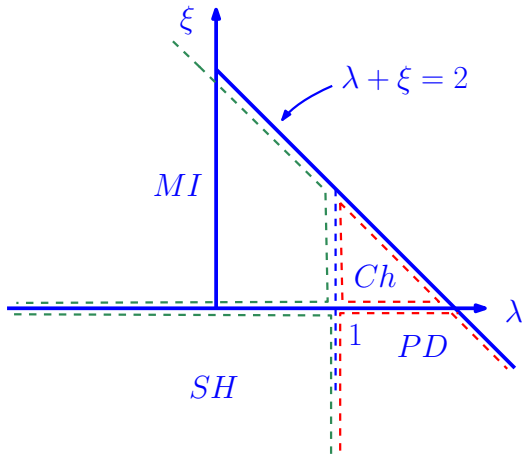
$$\text{for both } k, \ u_k(C,C) > \frac{1}{2}u_1(F,C) + \frac{1}{2}u_2(F,C) \tag{3}$$

that is $\lambda + \xi < 2$. Together with $u_i(C,C) > u_i(F,F)$ for both players it characterizes games where average players' payoff is highest at outcome CC. In this sense these are the games where conflict is detrimental. With this restriction the games we are studying can be visualized in (λ, ξ) space as in Figure 1.

2.2. Examples

We now consider some common examples of game families that have been considered in the literature. These examples, imposing a specific technology used to produce the utility values, carve out subsets of the space of the two parameters (λ, ξ). In this literature the analysis is usually much richer and complex than what may appear from the simple form we use here; considering these classic examples is useful however to put our approach in the perspective of a well known tradition.

Figure 1: **The family of games.** Everything is below the $\lambda + \xi = 2$ line. To the left of the $\lambda = 1$ line: above the horizontal axis ($\xi > 0$) there is Mutual Interest and below it there is Stag Hunt; these are the cooperation games. To the right of $\lambda = 1$: above the axis we have Chicken, below it is Prisoners Dilemma; these are conflict games.



Conflict over a Public Good

The first example is in the spirit of Esteban and Ray (2011) (see also Esteban and Ray (1994) and Esteban et al. (2012)) whose focus is on the issue of polarization. Consider a simple model of conflict between two large identical groups who compete for a public good, which is worth $v > 0$. If both compromise each gets $v/2$. If one group compromises and the other fights the latter wins $(1/2+a)v-c$ and the loser is left with $(1/2-a)v$ where $0 \leq a \leq 1/2$ is the *degree of polarization; c* is the cost of fighting, which includes both direct costs of effort and monitoring costs associated with peer pressure and discouragement of free-riding. If both groups fight each has an equal probability of winning but there is also battle damage bc to each group where $b \geq 0$ is the *intensity of conflict*, so both get $v/2 - (1+b)c$.

After normalization this model results in the family of games defined by

$$\lambda = 1 + \frac{av/c - 1}{(1+b)} \quad \textit{and} \quad \xi = 1 - \frac{av/c}{(1+b)}.$$

Here $\lambda > 0$ and $\xi < 1$. Note that $\lambda + \xi = 2 - 1/(1+b)$, therefore the constraint (3) is satisfied for all values of the parameters.

The game will be one of mutual interest when the relative mobilization cost c/v is large and a is small so that $c/v > a$; in this case compromise is strictly dominant for each group ($\lambda < 1$ and $\xi > 0$). If $c/v < a$ but the intensity of conflict is large enough that $c/v \cdot (1+b) > a$ the game is one of *chicken* ($\lambda > 1$ and $\xi > 0$). If both c/v and the intensity of conflict b are not too large relative to a so that $c/v < a$ and $(1+b) \cdot c/v < a$ the game is a *prisoners dilemma* (PD, with $\lambda > 1$ and $\xi < 0$).

Strategic Complements versus Strategic Substitutes

Baliga and Sjöström (2020), see also Baliga et al. (2011), concentrate on strategic complements

versus strategic substitutes. In Baliga et al. (2011), using our labels C and F, with our interpretation, a player receives a payoff of 0 if they both cooperate, but $-d$ if he cooperates and the other fights. If a player fights, he pays a cost c for both actions of the other, but receives an additional utility μ if the other cooperates. Adding c to all entries and then dividing by c the utility matrix in is our general format, with:

$$\lambda = \frac{\mu}{c} \quad \textit{and} \quad \xi = 1 - \frac{d}{c}. \tag{4}$$

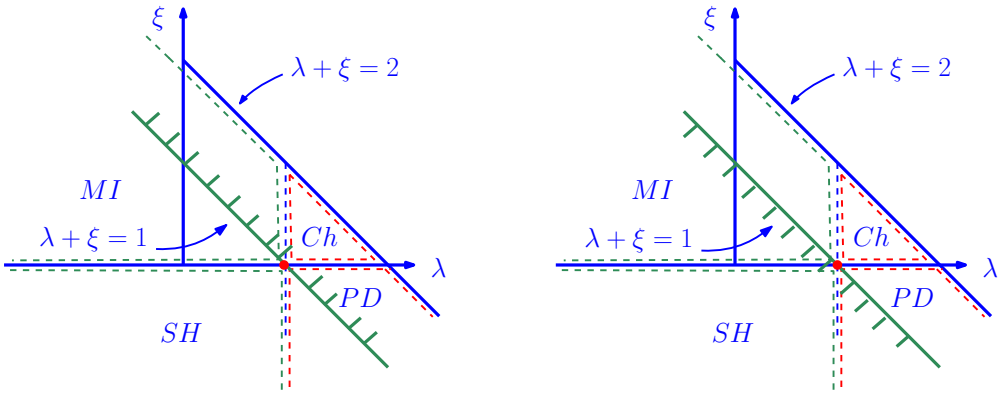
They assume $\mu < d$, so (3) holds. Also, if $\mu/c > 1$, then $d/c > 1$ and so $\mu/c > 1$ implies $\lambda > 1$ and $\xi < 0$. This is the *Prisoner's Dilemma* game. On the other hand, if $\mu/c < 1$ (that is, $\lambda < 1$) then

1. if $d/c < 1$ then $\xi > 0$ which together with $\lambda < 1$ gives the *Mutual Interest* game;
2. if $d/c > 1$ then $\xi < 0$ which together with $\lambda < 1$ gives the *Stag Hunt* game

Remarks

Note that in both examples some values of the pairs (λ, ξ) for each possible type of game are missing; that is the chosen functional form selects subsets of the possible values. The key difference between the two is the comparison of the total payoff from on occurrence of fight, $u_1(F,C)+u_2(C,F)$ and the joint cooperation $u_1(C,C)$: in the first example the first is larger, in the second the opposite holds. For example, in the case of conflict over a public good we have $\lambda + \xi \geq 2 - 1/(1+b) \geq 1$ and so the Stag Hunt game is excluded. On the other hand since in the second example $\lambda + \xi \leq 1$, the Chicken game is excluded. The Prisoner's Dilemma game and Mutual Interest are common to both, but some values of the parameters are excluded in both cases. The situation is visualized in Figure 2.

Figure 2: **Regions covered in the two examples.** The left panel describes the region covered by the ER parametrization; in that case $\lambda + \xi \geq 1$ so Stag Hunt and parts of PD and Mutual Interest are excluded. The right panel depicts the Baliga formulation; in that case $\lambda + \xi \leq 1$ so Chicken and parts of PD and Mutual Interest are excluded.



3. The Game Between Leaders

Interpreting players as large homogeneous groups we focus on the role of leaders in the collective decision making process. We take the view that, because groups are large, individual members have

little incentive to invest informing themselves about the consequences of collective actions, and with rational ignorance or limited knowledge of causality they instead listen to leaders and follow the leaders who make them the best offer. We consider two types of leaders: group leaders who have the same interest as that of a specific group and common leaders who care about both groups.

We now turn to the formal model. There is an underlying game between two groups, as described in the previous section. There are two groups $k \in \{1,2\}$, where each group has a representative follower. The followers choose actions $a_k \in \{F,C\}$, action profiles being denoted by $a \in A$, and all group k members receive utility $u_k(a_k, a_{-k})$ where $-k$ denote the other group. We assume that payoffs are distinct:[4]

$$\text{for all } k, a \neq a' \text{ implies } u_k(a) \neq u_k(a') \tag{5}$$

These utility functions give rise to the *underlying game*.

We now describe the leaders' game. This game is played by three leaders $\ell \in \{0,1,2\}$. Each leader has the same strategy set, which is equal to the set of action profiles, so he can choose $s^\ell \in A$. A profile of leaders' strategies is $s \equiv (s^\ell)_{\ell \in \{0,1,2\}} \in A^3$. The strategy of a leader is interpreted as his proposal to the society of a profile of behavior of groups, that may be examined by the followers.

There are two types of leaders. The leaders $\ell \in \{1,2\}$ are the leaders of groups 1 and 2 respectively and receive direct utility the same as the group: $U^\ell(a) = u_\ell(a)$. Leader $\ell = 0$ is a "common" leader who shares the preferences of both groups, with direct utility $U^0(a) = (u_1(a) + u_2(a))/2$. The interpretation is that the group leaders spend all their time with their group, while the common leader spends half his time with each group.

The followers in group k have the ability to impose a utility penalty that is proportional to the amount of time the leader spends with that group: for group leaders this is P, while for the common leader it is $P/2$. We suppose that each group k considers the proposals from the leaders who they are credible to them, that is the ones they can punish: both their own leader and the common leader, but not the leader of the other group. Among the proposals they consider, the followers choose the one promising them the highest utility. That is, given a strategy profile s of the leaders, follower k choose the strategy that maximize $u_k(s^\ell)$ over $\ell \in \{0,k\}$; this is unique by assumption (5), although it may be proposed by more than one leader. Denote this by $g^k(s) \in A$. Utility $u_k(g^k(s))$ is the implied promise to group k. Group k then implement their part in the chosen strategy, that is they play $g^k(s)_k$.

Therefore, given a profile of leaders' strategies s, the *implemented action profile* 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 $U^\ell(g(s))$.

After actions are implemented and direct utility accrue, followers of group k impose a punishment to the followed leaders when the obtained utility is strictly less than the one promised implicitly with the suggested action profile. Precisely, if $u_k(g^k(s)) < u_k(g(s))$ then group k punishes $\ell \in \{0,k\}$ such that $s^\ell = g^k(s)$, where the punishment is P if $\ell = k$ and $P/2$ if $\ell = 0$.

[4] In terms of our parameters this says $\lambda, \xi \notin \{0,1\}$

The sum of the direct utility and the punishments obtained as we have just described determine the payoff of leader ℓ, $V^\ell(s)$, for any strategy profile s. We let $1\{\mathfrak{c}\} = 1$ if condition $\mathfrak{c}$ is true and zero otherwise. Then the payoff of a group leader $\ell = 1, 2$ is

$$V^\ell(s) = U^\ell(g(s)) - P \cdot 1\{\ell = k \ \& \ g^k(s) = s^\ell \ \& \ u_k(s^\ell) < u_k(g(s))\} \tag{6}$$

and of the common leader

$$V^0(s) = U^0(g(s)) - (P/2) \cdot \sum\nolimits_{k=1,2} 1\{g^k(s) = s^0 \ \& \ u_k(s^0) < u_k(g(s))\}. \tag{7}$$

We call the game played by the set of leaders indexed by $\ell \in \{0, 1, 2\}$, with $S^\ell = A$ and the utilities V^ℓ just defined, a *leaders game*. It is a finite game, hence an equilibrium in mixed strategies exists. We are interested in Nash equilibria in weakly undominated strategies of the leaders game. We call this a *leaders equilibrium*.

In Section 4 we start by considering the benchmark case where only the two group leaders are present, and each group only considers the proposal of their own group leader. In this case there is no competition among leaders, each group just follows their group leader's proposal. Not surprisingly, the resulting leaders' game turns then out to be essentially the same as the underlying game.

3.1. Illustration of underlying and leaders' games

To illustrate the leaders' game we take the Prisoner's Dilemma as underlying game. The 2×2 payoff matrix of the underlying game is the one on page 4, with $\lambda > 1$ and $\xi < 0$.

Consider the case where the followers compare proposals from own group leader and the common leader. The 4×4 matrix resulting from the common leader playing CC is in Table 1. The three payoffs in each entry are naturally ordered with the leaders' index (first common then the other two).

Table 1: **From Underlying to Leaders Game: Common Leader and Group Leader.** Prisoners Dilemma is the underlying game. The table contains the leaders' payoffs when the common leader plays CC, ordered according to the the leaders' index (first common then the other two).

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

Look for example at the entry (CF, FC), corresponding to leaders' strategy profile (CC, CF, FC): all leaders get 1 because for both groups the best proposal comes from the common leader, so both groups play C, the implemented action is CC, both groups get 1 which was the promised utility, hence the chosen common leader is not punished and all leaders get 1. Or consider the payoffs when

the two group leaders play (FC, CC): the first group choose their group leader and play F; group 2 receives proposal CC from both leaders they listen to and they play C; so the implemented action profile is FC; group leader 1 gets the promised λ; the common leader and leader 2 are followed by group 2 and are punished (since the group gets ξ against a higher promise of 1). Note that the common leader gets direct utility of $(\lambda + \xi)/2$ and a punishment of $P/2$ (inflicted by group 2).

3.2. Informal Description of the Equilibria

We continue our illustration using the Prisoner's Dilemma game to provide some intuition for the structure of the equilibria, and show how cooperation may arise in equilibrium when competition among leaders and accountability exist.

We begin with the case in which the only leaders are the two group leaders, and each group only consider proposals by their own leader. The game among leaders is a four-by-four game, with each action profile of the underlying game a strategy in the leaders' game; for example, a strategy for a leader in the leaders' game is (F, F), that is "fight on both sides". The only equilibrium in this leaders' game is the strategy profile in which both leaders propose to fight to both groups, and the outcome is the bad equilibrium of the underlying prisoner's dilemma.[5] It is important to note that group leaders have to propose, at equilibrium, (F, F). They cannot, for example, propose fight for their group and cooperation for the other (that is, (F, C) for the first leader), which would produce the same outcome, because this would entail ex-post punishment by the followers that would compare the implicit rosy promise with the realized bad outcome: thus, the anticipation of future punishment prevents the first leader from sweetening the pill and proposing (F, C). This truthfulness condition opens the way for the intervention of the common leader, when one is active.

In fact, equilibrium outcomes change if in addition to group leaders there is also a common leader whose preferences are average of those of the two groups, and whose proposals are considered by both groups. In this case each group considers the proposals of their leader and those of the common leader. Clearly, the proposal of "fighting on both sides" by both group leaders is beaten by the proposal of the common leader of cooperation of both groups, (C, C). But this proposal of the common leader is in turn easily beaten, for example, by the (F, C) proposal of the first leader (and (C, F) by the second). However, as we have already noted, the two group leaders cannot both play (F, C) and (C, F) respectively for sure, because they anticipate that these proposal would produce the bad outcome (with low utility for both groups, a utility they share) and the consequent punishment imposed by followers. As will be shown, the only equilibrium is then a mixed strategy one, in which group leaders randomize between "aggressive" play (that is (F, C) for the first leader and (C, F) for the second) and conservative play (F, F); the common leader will mix too, between proposing cooperation (C, C) and effectively opting out by playing (F, F).

The probabilities at equilibrium of these various action profiles proposed by the leaders depend on the parameters, and vary across different equilibria. But when the cost of punishment is large

[5]It will be shown below that it is true in general that the equilibrium outcomes of the leaders game with only group leaders are the same as those of the underlying game between the groups.

group leaders will want to limit the probability of the aggressive proposals, which may imply costly punishment, thus leaving room for a winning cooperation proposal (C, C) of the common leader.

4. No Competition among Leaders

We start by studying the case where only group leaders are present, that is where $\ell \in \{1, 2\}$, and each group only considers proposals from their own group leader.[6] Without a common leader there is no competition among leaders - follower k just plays what her group leader recommends. Our first result says that in this case the outcomes of the leaders game are the same as in the underlying game. This is actually true for any leaders game, with any number of groups, and even without the assumption (5). Proving the statement for this more general case requires no additional effort, so we state it for this case:

Theorem 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 groups are the same as those induced by the Nash equilibria of the corresponding underlying game.*

The proof is in Appendix. Thus, without competition among leaders there are no improvements over the outcomes of the underlying game.[7]

5. Games with Common Leader

We now introduce competition among leaders, considering the case in which the group leaders compete with a common leader. This means that followers consider the proposals of their own group leader and of the common leader. Having disposed of the no-competition case in the previous section we examine this case in detail in the sequel of the paper.

Notation. The proposal by group leader k "we play F and they play C" will be denoted by $F^k C^{-k}$. This is FC for leader 1 and CF for leader 2.

6. The Cooperation Games

We begin with the cooperation games (Mutual Interest and Stag Hunt). From theorem 1 we know that in the game with only group leaders the equilibrium outcomes are those of the underlying game, so in the mutual interest game we have efficiency already without a common leader. The next theorem shows that with a common leader efficiency obtains also in Stag Hunt.

Theorem 2. *With a common leader, in the Mutual Interest and Stag Hunt games for any value of P there is a unique leadership equilibrium, with implemented action profile (C, C).*

[6] The model trivially extends to the case of K groups: just take $k, \ell \in \{1, 2, \ldots, K\}$ instead of $k, \ell \in \{1, 2\}$.

[7] As the proof shows, the equilibrium strategy profile in the leaders game implementing a Nash equilibrium of the underlying game is not necessarily unique, but for any equilibrium in the leaders game the induced mixed action profile in the underlying game is unique.

Proof. The CC outcome is the most preferred by the common leader and she can guarantee that outcome by proposing it, because $u_k(F^k, C^{-k}), u_k(F^k, F^{-k}) < 1$ so the group leaders best response to CC by the common leader is to propose C to their group. □

7. The Prisoners Dilemma

In this case the leaders' game can be considerably simplified. For the group leaders, the strategies CC and C^kF^{-k} are weakly dominated by FF. For the common leader, the strategies CF and FC are then weakly dominated by FF for all $P > 0$. So the analysis is reduced to the game where the group leaders only play F^kC^{-k} or FF and the common leader plays only CC or FF. In summary, the game is reduced to a simpler game with three players, each player with two actions. This simplified game is presented in table 2.

Table 2: The game after elimination of weakly dominated strategies. Left panel: utilities for the choices of the common leader equal to CC. Right panel: choice of FF.

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

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

The proof of the above statements is in the appendix, lemmas 9 and 10. Given this it can be shown that in equilibrium the common leader must play CC with positive probability, and in symmetric equilibrium the group leaders must play FC with positive probability.

The next theorem states what the equilibria of the leaders' game are. For P large, equilibrium is unique and the implemented action profile is either full cooperation or tends to full cooperation. For small P, the equilibrium is again unique and both groups fight. The proof is in Appendix B.2.

Theorem 3. *For P sufficiently low (more precisely $P < \min\{-\xi, \lambda + \xi\}$) there is a unique equilibrium outcome, in which both groups fight, and both groups get zero utility. For P sufficiently large (more precisely $P > \min\{-\xi, \lambda + \xi\}$) there is a unique equilibrium outcome, where the probability of the cooperation outcome is either equal to 1 (if $\lambda < 2$) or tends to 1 (if $\lambda > 2$).*

8. The Chicken Game

The symmetric equilibria of the underlying chicken game survive as leadership equilibria when there is no common leader (this follows from Theorem 1). And the presence of the common leader is not sufficient to change this fact:

Theorem 4. *The outcomes FC and CF of the underlying game are equilibrium outcomes of the leaders' game for all (λ, ξ, P).*

This is proved in Appendix C, Lemma 19. But interesting new possibilities emerge in the symmetric mixed equilibrium we consider next.

Theorem 5. *There is a mixed strategy equilibrium of the leaders' game in which the common leader plays CC with probability tending to 1, as P becomes large, and the group leaders play F^kC^{-k} with probability p and FF with probability $1-p$; the value of p tends to 0 as P becomes large.*

This is proven in two theorems in Appendix C. The first, Theorem 21, describes an equilibrium that exists for small P ($P < \lambda+\xi$) in which the common leader plays CC for sure and group leaders randomize between F^kC^{-k} and FF. The probability of F^kC^{-k} in this equilibrium is

$$\tilde{p} = \frac{\lambda - 1}{P + \lambda + \xi - 1}.$$

The second, Theorem 22, describes an equilibrium that exists for larger P ($P > \lambda+\xi$) in which the common leader randomizes between CC, FC and CF, with the probability of CC tending to 1 as P becomes large, and the group leaders randomize between F^kC^{-k} and FF, with the probability of F^kC^{-k} tending to zero as P grows large. In the limit the equilibrium implemented action profile is CC with probability 1.

9. Properties of the Equilibria for Small and Large P

We summarize here the payoff relevant properties of the leaders equilibria considered so far, as the punishment size becomes small or large. This concerns the accountability of the leaders, so it is a central issue in this paper. The statement below follows directly from the various results proved in appendix on the equilibria.

Theorem 6. *For all the leaders equilibria of the games with a common leader considered in the paper the following holds:*

(1) As $P \to 0$ the limit equilibria replicate outcome distributions of equilibria of the corresponding underlying games.

(2) With the exception of the asymmetric pure equilibria in the Chicken game (see Theorem 4), as $P \to \infty$ the equilibrium probability of cooperation and average group payoff tend to 1.

Proof. Part (1). In the cooperation games the CC outcome is common to the leaders and the underlying games equilibria. For the chicken game this is a corollary to Theorem 21. For the prisoners dilemma this is part 1(a) of Theorem 11.

Part (2). Again in the case of the cooperation games efficiency holds for any P. For the chicken game this is the last statement of Theorem 22. For the prisoners dilemma the claim follows from part 2(b) of Theorem 11 because if $P \to \infty$ then $\tilde{q}$ and $\hat{q}$ tend to 1 and $\tilde{p}$ and $\hat{p}$ tend to zero. □

The content of the result is clear: with competition among leaders brought about by the presence of a common leader, adequate accountability is necessary and sufficient for efficiency (we discuss the Chicken exception below). Without punishment, the leaders participation adds nothing to the underlying game. On the other hand with sufficiently large P - in fact not so large as we shall see - the outcome becomes not only better, but reaches full efficiency. This is the main message of the

present paper: in games where the group conflict is detrimental and the unmediated Nash equilibria are undesirable, the interaction through competing, accountable leaders enable groups to achieve cooperation with surprisingly high probability.

The two asymmetric equilibria outcomes in the underlying chicken game survive as leaders equilibrium implemented action profile for all P. It may come as a surprise that the inefficiency arising in the chicken game is harder to overcome than the prisoners dilemma. But the fact is that in the PD equilibrium both groups are badly worse off than in the cooperative outcome, and then a common leader may come to the rescue; in the chicken pure equilibria on the other hand one party is relatively well off (possibly better off than in the CC outcome), and when a group acts aggressively, with or without the mediation of a group leader, neither the other group leader nor a common leader can do anything to dissuade them.

As we know, at least in the Chicken game, better outcomes than in Nash equilibria may be reached also through the intervention of an external, uninterested mediator - in the correlated equilibria of the game. We turn to comparison of leaders and correlated equilibria next. Of course fixing the (λ, ξ) parameters we already know what happens asymptotically. But as we shall see the mixed leaders equilibria fare better than the correlated equilibria of the underlying game already for moderate values of P. What makes the difference is that on the one hand the common leader is interested in cooperation (her most preferred outcome) and this is therefore what she tends to propose; and that on the other hand the group leaders are discouraged to make aggressive proposals by the threat of the punishment that may come as a consequence.

10. Comparison with Correlated Equilibria

The leaders' game shares some important features with the canonical correlated equilibrium: 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 actually 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 by the players of the underlying games.

Nonetheless both solution concepts produce sets of equilibrium action profiles, so the comparison from the point of view of welfare is of some interest. We take as measurement of welfare the average utility of players in the underlying game: so we sum the utility of the two groups and ignore the

welfare of the leaders (which may include punishments). Still we are comparing two sets, and in the case of the leaders' game we have an additional parameter to take into account, which is the punishment. Thus the comparison changes for different values of P. We will call efficient the outcome where both players in the underlying game get a utility of one.

In this section we show that outcomes leaders game typically dominate correlated equilibria in average welfare. More precisely, we show that in all games the largest average utility at outcomes of equilibria of the leaders' game is larger than the largest utility at correlated equilibrium outcomes. Call $\Delta(A)$ the set of correlated strategies (A is the set of action profiles of the underlying game), with generic element μ.

The comparison is trivial in the case of the mutual interest game: both solution concepts predict a unique outcome, and the outcome is efficient. The comparison is also easy for the prisoners dilemma and the stag hunt game; but the two sets do not coincide, so the comparison is meaningful.

In the prisoner's dilemma, the outcome predicted by the leaders' game is unique: it is the efficient outcome in the limit as P tends to $+\infty$, and the zero utility outcome when $P = 0$. There is a unique correlated equilibrium of the underlying game, which is the zero utility outcome. Thus in this case the leaders' equilibrium dominates the correlated.

In the stag hunt, the outcome of the leaders' game is unique, and it is the efficient outcome for any value of P. The set of correlated equilibria in not a singleton, so we consider the best and worst possible outcomes. The best outcome for correlated equilibria is the efficient one. Since the set of utilities in a correlated equilibria is convex, all the values between the best and worst utility are correlated equilibrium outcomes. The worst correlated equilibrium outcome is the one induced by the mixed strategies of the underlying game. Thus in this case too the leaders' equilibrium weakly dominates the correlated.

In the Chicken game the comparison is more complex, and we turn to it now. For fixed (λ, ξ) the comparison is straightforward:

Theorem 7. *In the Chicken game, given (λ, ξ), the average payoff in any correlated equilibrium of the underlying game is bounded away from 1, so for large enough P it is lower than the average payoff in the mixed equilibrium of the leaders game (which goes to 1 as $P \to \infty$, see Theorem 22).*

The proof of this is in Appendix D. Before considering the situation with varying parameters we compare "worst against worst" equilibria. The lowest correlated equilibrium payoff is computed in Appendix D. As shown in appendix, for small P the Chicken game has a mixed leaders equilibrium whose payoff is increasing in P, so the lowest occurs for $P = 0$ where its outcome distribution is the same as in the mixed equilibrium of the underlying game. The asymmetric outcomes of the underlying game are leaders equilibrium outcomes as well, so the worst leaders equilibrium is either the mixed or one of the pure equilibria of the underlying game (whichever is worse). It is proved in Appendix D that both yield higher payoff than the worst correlated equilibrium.

10.1. The Size of P

We go back to the "best against best" comparison in the Chicken game removing the restriction of fixed (λ, ξ), and ask how large P must actually be for the mixed leaders equilibrium to beat the best correlated equilibrium of the underlying game.

As we mentioned above, for $P < \lambda + \xi$ there is a mixed leaders equilibrium where the common leader plays CC for sure and the group leaders mix between $F^k C^{-k}$ and FF with probability $\tilde{p}$ on the former - we shall refer to it as "the common leader equilibrium" in the sequel. As P crosses a threshold a little above $\lambda+\xi$ the mixed leaders equilibrium is the one described in Theorem 22, where the common leader mixes between CC, FC and CF and the group leaders mix between $F^k C^{-k}$ and FF with a probability $p < \tilde{p}$ on $F^k C^{-k}$. The average group payoff in this equilibrium is higher than in the common leader equilibrium.[8] Unfortunately, the mixing probability p is the root of a cumbersome equation, and it is most easily computed numerically for each set of parameters value. This makes comparisons over varying parameters not convenient. So in the following estimates we use the common leader equilibrium, which is easier although to our disadvantage.

As shown in Appendix D the highest payoff in the correlated equilibrium is

$$\overline{\pi}^{corr}(\lambda, \xi) \equiv \frac{\xi + (\lambda + \xi)(\lambda - 1)}{\xi + 2(\lambda - 1)}$$

which of course depends on (λ, ξ); notice that it goes to 1 if $\lambda + \xi \to 2$ or $\lambda \to 1$.

On the other hand, the average group payoff in the common leader equilibrium can be computed to be

$$\tilde{\pi}(\lambda, \xi, P) \equiv (1 - \tilde{p})(1 + \tilde{p}(\lambda + \xi - 1)) = \frac{P + \xi}{(P + \xi + \lambda - 1)^2} \cdot (P + \lambda(\lambda + \xi - 1))$$

so that fixing $\alpha \leq 1$, for each (λ, ξ) there is a threshold that P must reach so that $\tilde{\pi}(\lambda, \xi, P) = \alpha$ - in particular for each (λ, ξ) in the set $\overline{\pi}^{corr}(\lambda, \xi) = \alpha$. To put ourselves in the most unfavorable position, for each α we pick the *highest* P-threshold in the set $\overline{\pi}^{corr}(\lambda, \xi) = \alpha$. Denote this by $P(\alpha)$.[9] By construction, for $P > P(\alpha)$ the average group payoff in the leaders equilibrium is higher than in any correlated equilibrium with average payoff α. The graph of $P(\alpha)$ is in Figure 3. $P(\alpha) \leq 1$ for $\alpha \lessapprox 0.77$; For $\alpha = 0.9$ this is $P(\alpha) = 2.4$; for $\alpha = 0.99$ it is $P(\alpha) = 9.75$. [10]

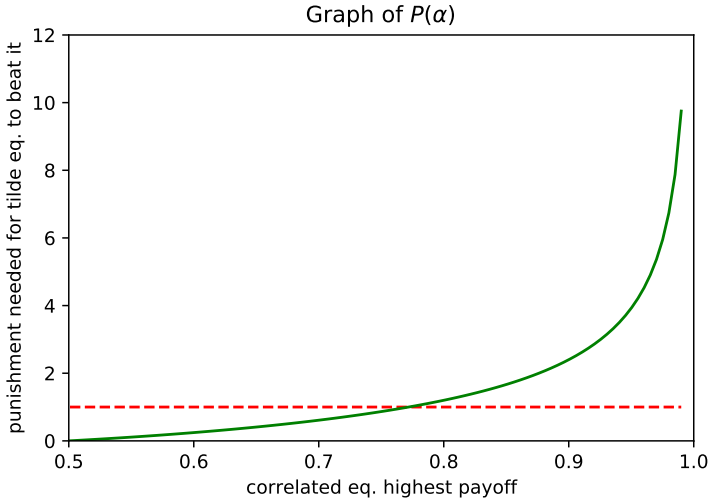
In conclusion, for parameters in the interior of the chicken region what our computations show is that typically, for values of P in the same range as the players' payoffs the leaders equilibrium yields higher payoff than any correlated equilibrium of the underlying game.

[8] The reason is that at $\tilde{p}$ the common leader prefers FC and CF to CC; the group leaders raise the probability of FF, to the extent that the common leader's payoff from CC goes up and reaches that from FC and CF (which go down); in the end the group leader's payoff is higher, and so the average group payoff.

[9] The procedure is spelled out in Appendix D.

[10] The last value is somewhat high, but consider that the correlated payoff is high when the game approaches a cooperation game ($\lambda \to 1$) or a game where cooperation is not detrimental ($\lambda + \xi \to 2$), in which case conflict is not too pronounced.

Figure 3: **Graph of $P(\alpha)$.** This is the P value above which the mixed leaders equilibrium is higher than *any* correlated equilibrium which gives average group payoff equal to α. Higher values of α are harder to beat. $P(\alpha) \leq 1$ for $\alpha \leq 0.77$; For $\alpha = 0.95$ this is $P(\alpha) = 3.92$; for $\alpha = 0.99$ it is $P(\alpha) = 9.75$. The dashed horizontal line at height 1 is displayed for convenience.



11. The Role of Differential Punishment

We have considered so far the hypothesis that the punishment for leaders is the same for group leaders and for common leaders. In the Prisoners Dilemma, for large enough P the equilibria are the common leader equilibrium mentioned above and another one where the group leaders again mix between F^kC^{-k} and FF - but with probability $\hat{p}$ displayed below on F^kC^{-k} - and common leader mixes between CC and FF, with probability $\hat{q}$ on CC. The probabilities in the latter equilibrium are

$$\hat{p} = \frac{1}{P - (\lambda + \xi - 1)}, \qquad \hat{q} = \frac{P}{P + \lambda - 1 - \frac{P+\lambda+\xi-1}{P-(\lambda+\xi-1)}}$$

In the Chicken game for large P the equilibrium is close to the common leader equilibrium so we consider this one for simplicity in the following discussion.

We now ask how outcomes would differ if the punishments were allowed to be different across leaders. This thought experiment may clarify which one of the leaders has to be given the appropriate incentive. These will turn out to be the group leaders. Specifically, we allow for different punishments for the group leaders and the common leader, letting P^c and P^g denote punishments for common and group leaders respectively.

Theorem 8. *Assume that the conditions on P in Theorems 3 and 5, respectively $P \lesseqgtr \min\{-\xi, \lambda+\xi\}$ and $P \lesseqgtr \lambda + \xi$, hold for both P^c and P^g. Then the structure of the mixed equilibria remains the same, with tilde and hat probabilities given by*

$$\tilde{p} = \frac{\lambda - 1}{P^g + \lambda + \xi - 1}, \qquad \tilde{q} = 1, \qquad \hat{p} = \frac{1}{P^c - (\lambda + \xi - 1)}, \qquad \hat{q} = \frac{P^g}{P^g + \lambda - 1 - \frac{P^g+\lambda+\xi-1}{P^c-(\lambda+\xi-1)}}$$

To prove Theorem 8 only minor modifications are needed of the given arguments: in the various incentive constraints one has to specify which P is involved.[11]

In both equilibria $(\tilde{q}, \tilde{p}, \tilde{p})$ and $(\hat{q}, \hat{p}, \hat{p})$ the average group payoff is given by $q(1-p)\,(1+p(1-\delta))$. It is apparent that in the former only the group leaders punishment matters: $\tilde{p}$ tends to 0 as P^g tends to $+\infty$. In the latter both P^g and P^c are involved; but it is found by numerical computations that average group payoff is again increasing in P^g, and as P^g becomes large it is *decreasing* in P^c.

12. Conclusions and Discussion

We gather here our conclusions, remarks on how we should evaluate our results within the broader research agenda on diverse societies, and more in general on the study of the relationship between ruling classes and citizens.

In this paper we have examined how political entrepreneurship can fundamentally alter outcomes in societies with group conflict. We rely on a model of leadership which may be useful in more 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. Followers choose among the proposals to maximize their utility.

The main insight derived from our model and analysis is that conflict in polarized societies can be substantially reduced, *under appropriate conditions*, thanks to the mediation of interested leaders. The simple existence of leaders by itself cannot accomplish anything useful: the equilibrium outcomes are the same as in the game with no leaders, unless appropriate conditions are met. Our analysis has identified two main forces: competition among leaders and accountability. If there is competition among leaders, then in general cooperation and good outcomes are possible when the accountability of leaders is sufficiently large. In the limit of high accountability, full cooperation is realized. Our results seem to temper the bleak picture that may emerges from the literature on group conflict: a truce among groups in conflict is possible, under appropriate conditions. However, one has to put this conclusion in the appropriate perspective. Our setup relies on simplifying assumptions, and some of these assumptions are in contrast with important regularities in political life.

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 built to share the interests of the 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.

[11] For the *PD* for example we only need to rewrite the preference conditions $CC \succcurlyeq_c FF$ and $FC \succcurlyeq_k FF$. The former was $(1-p)\,(1-p(1+P-(\lambda+\xi))) \geq 0$, and becomes $(1-p)\,(1-p(1+P^c-(\lambda+\xi))) \geq 0$. And $FC \succcurlyeq_k FF$ was $P \leq q\,(P+\lambda-1-p(P+\lambda+\xi-1))$ and is now $P^g \leq q\,(P^g+\lambda-1-p(P^g+\lambda+\xi-1))$ Then the resulting modifications of the tilde and hat probabilities follow.

Fortunately, our analysis makes clear the leaders' role, so it can be taken to provide the best case scenario for possible positive effects of political mediation in group conflict. Systematic empirical research will have to decide which are the realistic ranges of the losses voters can impose on leaders.

The behavior of followers in our model is extremely simplified. On the other hand the assumption of unsophisticated behavior is not so unrealistic: in large and complex societies, understanding the structure of the payoff 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 consider the promised utility, and choose the highest.

A natural extension of the model presented here, in the direction of a more realistic behavior of followers, is a foundation of their behavior based on a model of information acquisition on relevant parameters affecting the utility of players. This information is hard to gather, so it is delegated to leaders or parties, which can do that through costly effort, and then send messages (for example, political programs) to the entire society. Followers may then interpret the signals sent in the light of what they know and choose rationally the best action.[12]

[12]In a different context, a similar idea is presented in Matějka and Tabellini (2021).

Appendix A. Proof of Theorem 1

The statement of the theorem is given here in the more general case in which there are $K \geq 2$ groups. The definition of the leaders' game is a natural extension of the one provided for the case $K = 2$.

Theorem (Theorem 1 in the text)**.** *For any leadership game the outcomes in the underlying game induced by the the Nash equilibrium of the leadership game are the same induced by the Nash equilibria of the 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. □

Appendix B. Analysis of the Prisoner's Dilemma

Appendix B.1. Elimination of Weakly Dominated Strategies

We begin with some preliminary Lemmas to eliminate weakly dominated strategies.

Lemma 9. *For group-k leader the strategies CC and $C^k F^{-k}$ are weakly dominated by FF.*

Proof. We let $k = 1$. Fix any profile s^{-k} of the other leaders.

Consider CF first. Suppose that $g(CF, s^{-k})_1 = F$; then group 1 must have accepted a proposal FF or FC by the common leader, so that by playing CF or FF group-1 leader gets the same payoff (λ or 0, no punishment). Suppose $g(CF, s^{-k})_1 = C$; then the common leader must have proposed CF as well and group-1 leader gets $\xi < 0$, while in this case by proposing FF she gets 0 and no punishment.

Now consider CC and suppose first $g(CC, s^{-k})_1 = F$; then group 1 must have accepted a proposal FC by the common leader, and therefore CC and FF yield the leader the same payoff. Suppose $g(CC, s^{-k})_1 = C$ so that her proposal is accepted; the competing offers may have been CC, CF or FF; if all other proposals are CC then her payoff does not change if she plays FF; if there is a CF or an FF by some $\ell \neq 1$ then group-1 leader is strictly better off by playing FF (she gets zero, while with CC she gets $\xi - P$). □

In view of this lemma we may assume that group leader k plays only $F^k C^{-k}$ or FF; we let p_k denote the probability of $F^k C^{-k}$.

Lemma 10. *The probability that the common leader plays either CF or FC is zero.*

Proof. We do it for CF. This proposal is rejected by group 1 who will play F, and accepted for sure by group 2 who will play F and punish the common leader. She is better off by playing FF (strictly if $P > 0$). □

Appendix B.2. Nash Equilibria in Prisoners' Dilemma

In the previous section we have simplified the leaders' game when the underlying game is the prisoners' dilemma to a three players game, each player with two actions. This simplified game is reported in table 2 of the main text. Thanks to this simplification, we can describe a strategy profile of the three players with a vector of the form (q, p^1, p^2) where q is the probability that the common leader plays CC ($(1-q)$ that he plays FF), and p^k the probability that the k group leader plays FC ($(1-p^k)$ that he plays FF).

The next theorem characterizes the equilibria of the leaders' game when the underlying game is the Prisoners' Dilemma. We first introduce some notation. The pair $(\hat{q}, \hat{p})$ in (B.1) describes a pair of mixed strategies in the simplified game ($\hat{q}$ for the common leader and $\hat{p}$ for each of the group leaders). It does not give full cooperation, but the induced outcome tends to cooperation as P becomes large, because $\hat{q}$ tends to 1 and $\hat{p}$ tends to 0.

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

The equation (B.2) defines a different pair of mixed strategies (actually pure for the common leader); note that $\tilde{p}$ tends to 0 as P becomes large.

$$\tilde{q} = 1, \qquad \tilde{p} \equiv \frac{\lambda - 1}{\lambda - 1 + P + \xi} \tag{B.2}$$

Finally, the inequality (B.3) links the three parameters together, and decides (see the last point in theorem 11) whether the equilibrium as P becomes large is B.1 or B.2.

$$\xi + (\lambda - 1)(\lambda + \xi) > (\lambda - 2)P \tag{B.3}$$

We can now present the theorem:

Theorem 11. *In the leaders' game with prisoners' dilemma underlying game:*

1. *If $P < \lambda + \xi$:*
   (a) *If $P < -\xi$ the equilibria are all $(q, 1, 1)$ for any $q \in \left(\frac{P}{-\xi}, 1\right]$;*
   (b) *If $P > -\xi$ the equilibria are $(1, \tilde{p}, \tilde{p})$ and the set $\{(1, 1, 0), (1, 0, 1)\}$*
2. *If $P > \lambda + \xi$:*
   (a) *If $P < -\xi$ the equilibria are all $(q, 1, 1,)$ for any $q \in \left(\frac{P}{-\xi}, 1\right]$, and the set*

$$\{(q, \hat{p}, \hat{p}) : \min\{-\frac{P}{\xi}, \frac{P}{P + \lambda - 1}\} < q < \max\{-\frac{P}{\xi}, \frac{P}{P + \lambda - 1}\}\};$$

   (b) *If $P > -\xi$:*
      i. *if the inequality (B.3) holds, then the equilibria are $(\tilde{q}, \tilde{p}, \tilde{p})$;*
      ii. *if the inequality (B.3) does not hold, then the equilibria are $(\hat{q}, \hat{p}, \hat{p})$;*

Proof. The proof follows from consideration of the cases examined in section Appendix B.3. □

We turn to the proof of theorem 3:

Proof. The proof follows from theorem 11. As P becomes small, the only relevant case is 1.(a), in which both $P < -\xi$ and $P < \lambda + \xi$. In this case the two group leaders play FC and CF respectively for sure, so the outcome in the underlying game is (F, F) for sure.

As P becomes large, the only relevant case is 2.(b), in which both $P > \lambda + \xi$ and $P > -\xi$. In this case the nature of the equilibrium is decided by the inequality B.3. Note that whether this equality holds or not for large P depends on whether λ is smaller or larger than 2. □

Appendix B.3. Analysis of Equilibria in PD

We will identify all the equilibria in the game; the analysis is organized considering three possible cases for the value of q, namely $q = 0$, $q = 1$ and then $q \in (0, 1)$. We concentrate on the interesting cases in which the relevant inequalities among combinations of parameters hold strictly.

Equilibra with $q = 0$

Lemma 12. *If $P > 0$, there is no equilibrium with $q = 0$*

Proof. If the common leader sets $q = 0$ then the leaders' game is the bottom 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 acions of group leaders, CC yields 1, and FF yields 0, to the common leader, hence setting $q = 1$ is the best response. □

Equilibra with $q = 1$

In the first lemma we deal with the case of small P:

Lemma 13. *If $\xi < -P$ 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 14. *If $\xi > -P$:*

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 $\lambda + \xi - P > 0$.*
2. *There is an equilibrium where group leaders play a mixed strategy if and only if:*

$$\xi + (\lambda - 1)(\lambda + \xi) + (2 - \lambda)P > 0. \tag{B.4}$$

The mixed strategy is $\tilde{p}$ in equation (B.5).

Note that, for fixed λ and ξ as P becomes large the equilibria as in lemma 14 fail to exist, and also equilibria as in case (1) of lemma 13 fail to exist, and the same for the equilibria in case (2) of the same lemma when $\lambda > 2$. In summary equilibria with $q = 1$ exist for P large if and only if $\lambda < 2$.

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

$$p^1 = p^2 = \frac{\lambda - 1}{\lambda - 1 + P + \xi} \equiv \tilde{p} \tag{B.5}$$

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

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

1. If $\lambda + \xi - P > 0$ then there are two equilibria, $(q, p_1, p_2) = (1, 0, 1), (1, 1, 0)$. This follows because CC gives $\frac{\lambda + \xi - P}{2}$, while FF gives 0 to the common leader.
2. If $\lambda + \xi - P < 0$ 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 gets

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

which is larger than 0 (hence CC better than FF) if and only if (B.4) holds. □

Equilibra with $q \in (0,1)$

To set up the analysis we assume that the common leader is playing q and compare the corresponding expected payoff from FC and FF for group leader in the two cases: group leader plays CF and FF (thus, four comparisons overall). In the first case FC is better than CC if and only if

$$q > -P/\xi \tag{B.6}$$

In the second case FC is better than CC if and only if

$$q > \frac{P}{P+\lambda-1} \tag{B.7}$$

In lemmas 15 and 16 we consider the two extreme possible cases for q:

Lemma 15. *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, and so for any such q the payoff to the common leader at the best response of the group leaders from CC is 1, and from FF is zero, so no $q \in (0,1)$ can be part of an equilibrium. □

Lemma 16. *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 is large.

Proof. The condition on q implies that FC for group leader 1 (CF for 2) is dominant. At this best response (FC, CF) of the group leaders, both CC and FF give a payoff of 0, hence any q (in particular any satisfying that condition) is part of an equilibrium of the form described. □

Next we consider the intermediate cases for the values of q. At these values of q the game with $q$expected payoffs of group leaders has three equilibria, two pure strategies and one mixed. We deal with pure strategies in lemma 17.

Lemma 17.

1. $\frac{P}{P+\lambda-1} < q < -\frac{P}{\xi}$ *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, consider for example the profile (FF, CF) (the other is (FC, FF)). In this case CC gives $\frac{\lambda+\xi-P}{2}$, 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, note that with value of q in that range with the strategy profile (FC, CF), both CC and FF give zero to the common leader, hence $(q,1,1)$ with any q in the range is an equilibrium. Instead, the other possible equilibrium with q-expected payoffs has CC giving value 1

to the common leader, and FF giving 0, hence no equilibrium with the q component in the open interval $(0,1)$ can exist. □

Lemma 18. *An equilibrium with* $\min\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\} < q \leq \min\{\max\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\}, 1\}$ *exists, with a mixed strategy* $(\hat{q}, \hat{p}, \hat{p})$ *defined in equations B.8 and B.10 below.*

Proof. For q to be part of an equilibrium, the common leader has to be indifferent between CC and FF which is true if and only if:

$$p = \frac{1}{P+1-\lambda-\xi} \equiv \hat{p} \tag{B.8}$$

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:

$$p = \frac{P+\lambda-1-P/q}{P+\lambda+\xi-1} \equiv f(q) \tag{B.9}$$

Combining equations B.8 and B.9 we conclude that an equilibrium with q in the range exists if both $f(q) = \hat{p}$ and

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

and since $f(-\frac{P}{\xi}) = 1$ and $f(\frac{P}{P+\lambda-1}) = 0$ with f strictly increasing, there is unique $\hat{q}$ in the given range such that

$$f(\hat{q}) = \hat{p}. \tag{B.10}$$

it is easy to check that this $\hat{q}$ is indeed the value in equation (B.1).

Note that for P large, $\max\{-\frac{P}{\xi}, \frac{P}{P+\lambda-1}\} = -\frac{P}{\xi} > 1$, hence in this case we must check whether an equilibrium exists with $\frac{P}{P+\lambda-1} < q < 1$. Since $f(\frac{P}{P+\lambda-1}) = 0$, we now compare compare $f(1) = \frac{\lambda-1}{P+\lambda+\xi-1}$ and $\hat{p}$; so a solution exists if and only if $\frac{\lambda-1}{P+\lambda+\xi-1} > \frac{1}{P+1-\lambda-\xi}$; this in turn is equivalent to:

$$(\lambda-2)\,P > \lambda(\xi+\lambda-1) \tag{B.11}$$

.

So if $\lambda > 2$ we have an interior equilibrium for large values of P. In the other case (that is, $\lambda < 2$) we have $f(q) < \hat{p}$ for all $\frac{P}{P+\lambda-1} \leq q \leq 1$, and the equilibrium is $(1, \tilde{p}, \tilde{p})$ with $\tilde{p}$ introduced earlier in equation (B.5). □

Appendix C. Analysis of the Chicken Game

Appendix C.1. The Pure Strategy equilibria

We describe here symmetric equilibria, so we formulate the lemma focusing on one outcome, (F, C). The same statement holds for the outcome (C, F).

Lemma 19. *In the Chicken game the outcome (F, C) of the underlying game is an equilibrium outcome of the leaders' game for all (λ, ξ, P).*

Proof. We write $BR^\ell(a^0, a^1, a^2)$ the best response of leader ℓ to the profile (a^0, a^1, a^2). The proof has three parts, for $\ell \in \{0, 1, 2\}$:

$$FC \in BR^\ell(FC, FC, FC) \tag{C.1}$$

So in each step we proceed from the assumption that the other leaders are playing (F, C) and consider the best response of the leader under consideration. We then examine the expected utility from the different possible choices of the leader under consideration, and claim that conclude that his best response is (F, C).

Consider first $\ell = 2$. Given $a^1 = a^0 = (F, C)$, group 1 will choose F no matter what the other leader offers, because this is the largest utility it can receive, and group 2 has the proposal (F, C) of the common leader. Considering the possible choices of a^2: (C, C) gives a utility $\xi - P$ (because $\xi < 1$, so group 2 will follow leader 2 , but the outcome then will be (λ, ξ) rater than the implicit promise $(1, 1)$ of leader 2, and hence leader 2 will be punished. (C, F) gives a utility $-P$ (group 2 will follow leader 2 and play F but the outcome is then $(0, 0)$ and so leader 2 gets the 0 utility and the punishment because the realized 0 is smaller than the promised λ). (F, C) gives a utility ξ (because both common leader and leader 2 promise the same utility profile). Finally, (F, F) gives a utility ξ (because the associated utility vector is $(0, 0)$, and common leader is promising ξ). Our claim follows.

Consider next $\ell = 0$. We proceed noting that $a^1 = a^2 = (F, C)$, and thus group 1 is choosing F.

(C, C) gives a utility of $\frac{\lambda+\xi-P}{2}$ (because group 1 will choose F, following the group leader, while group 2 will choose C, following the common leader, expecting utility 1 rather than the ξ proposed by the group leader. Thus the outcome is (F, C), thus the common leader direct gets utility $\frac{\lambda+\xi}{2}$, and group 2 punishing the common leader). (C, F) gives a utility of $-P/2$ (because group 1 will follow leader 1, and group 2 will follow the common leader and play F expecting λ. Thus the outcome is (F, F) and average utility of groups equal to 0 and punishment of common leader by group 2. (F, C) gives a utility of $\frac{\lambda+\xi}{2}$ (because all leaders are proposing the same action profile). (F, F) gives a utility of $\frac{\lambda+\xi}{2}$ (because the proposal of the common leader will be ignored).

Consider finally $\ell = 1$. Assuming $a^0 = a^2 = (F, C)$, we note that group 1 is considering the utility λ from the common leader (with choice C), and group 2 is considering the utility ξ from both common leader and group leader 2. Group 1 is choosing F, following the common leader, no matter what group leader 1 is going to propose. The choice $a^1 = (F, C)$ gives leader 1 a utility of λ (group 1 is choosing F, because this is then the only proposal they receive, and group 2 is choosing C); but λ is the largest possible utility, hence (F, C) is a best response of group leader 1. □

The next lemma shows that different degrees of communications between groups and leaders does not alter this conclusion:

Lemma 20. *In the Chicken game the outcome (F, C) of the underlying game is an equilibrium outcome of the leaders' game for all (λ, ξ, P) and for all γ functions.*

Proof. We adopt the formulation in which all leaders formulate a proposal, and the γ function only selects a special subset of the proposal that each group observes. So there is in each case a vector (s^0, s^1, s^2) of proposals, one from each leader. In the case $\gamma^1 = \{0, 1\}$, group 1 observes (s^0, s^1), whereas in $\gamma^1 = \{0, 1, 2\}$ group 1 observes (s^0, s^1, s^2), and so on.

Consider the pure strategy equilibrium (FC, FC, FC) identified, in the case $\gamma(1) = \{0, 1\}$, in lemma (19). Consider now the same pure strategy profile, but in the game where $\gamma(1) = \{0, 1, 2\}$, and consider the best response of leader 2. We claim that his best response in the same in the games with the two different γ functions. In fact, the expected utility from the choice of each strategy profile leader 2 can take is the same in both games, since the other two leaders in the pure strategy profile are taking the same strategy. Hence the strategy profile (FC, FC, FC) is an equilibrium irrespective of the γ function. □

Theorem 21. *For $P \leq \lambda + \xi$ or $2\lambda + \xi \leq 3$ there is an equilibrium where the common leader plays CC for sure, and the group leaders play $F^k C^{-k}$ with probability $\tilde{p}$ and FF with probability $1 - \tilde{p}$, with $\tilde{p}$ as defined in B.2.*

Proof. The utility matrix when the common leader plays CC is the following:

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

Consider first a group leader, given the others' strategies: if she plays FF he gets

$$p\xi + 1 - p = 1 - p(1 - \xi)$$

while if she plays FC she gets

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

so indifference between FF and FC holds if and only if:

$$p = \frac{\lambda - 1}{\lambda - 1 + \xi + P} = \tilde{p}.$$

This is smaller than 1 because $\xi > 0$. For a group leader proposing C cannot improve utility, since C is proposed by the common leader already. And indeed as we see from the utility matrix CF yields the same utility as FF and CC is weakly worse.

Consider now the common leader. The reduced utility matrix when she plays CC is this

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

so by playing CC she gets

$$(1-p)\left(1+p(\lambda-1+\xi-P)\right).$$

This value is strictly positive because it is easily verified that at $p=\tilde{p}$ one has $1+p(\lambda-1+\xi-P)>0$. From the reduced utility matrix in the case in which the common leader plays FF:

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

we see that FF gives zero, less than CC.

Consider lastly the utility from playing FC. The utility matrix is

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

so her utility is

$$p^2(-P/2)-p(1-p)(P-(\lambda+\xi))/2+(1-p)^2(\lambda+\xi)/2$$

Thus the common leader prefers CC to FC if the following difference is positive:

$$p^2P+p(1-p)\left((\lambda+\xi)-P\right)+(1-p)^2(2-(\lambda+\xi))$$

so for $P\leq\lambda+\xi$ this is certainly positive for any (ξ,λ) pair in the chicken region. Consider next $P>\lambda+\xi$. As $P\to\infty$, since $\tilde{p}\to 0$ and $\tilde{p}P\to\lambda-1$ the limit of the above difference is easily computed to be $(1/2)\left(3-2\lambda-\xi\right)$, which is positive for $2\lambda+\xi\leq 3$. We now show that for $2\lambda+\xi\leq 3$ the above difference is strictly positive for all $P>\lambda+\xi$. Neglecting the $1/2$ factor we can re-write it as

$$2p^2\left(1+P-(\lambda+\xi)\right)-p(P+4-3(\lambda+\xi))+2-(\lambda+\xi).$$

We are assuming $P>\lambda+\xi$ so the first term is positive; and we now show that the remaining part is positive as well, which inserting $\tilde{p}$ becomes

$$\left[2-(\lambda+\xi)\right]\left[\lambda-1+\xi+P\right]>(\lambda-1)\left(P+4-3(\lambda+\xi)\right).$$

This is found to be equivalent to

$$P\left(3-2\lambda-\xi\right)>2(\lambda-\xi-1)-(\lambda+\xi)(2\lambda-\xi-2)$$

so since $3-2\lambda-\xi$ it suffices to show that the right member is negative, equivalently $(\lambda+\xi)(2\lambda-\xi-2)>2(\lambda-\xi-1)$; this in turn can be checked to simplify to $2(\lambda-1)^2>\xi(\xi-\lambda)$ which is true since $\xi<1<\lambda$ implies $\xi-\lambda<0$. □

It may be useful to state the following

Corollary. *For $P = 0$ the outcome distribution of the above equilibrium is the same as in the mixed equilibrium of the underlying game.*

Proof. For $P = 0$ we have $\tilde{p} = p(F)$ where $p(F)$ is the probability of F in the mixed equilibrium of the underlying game. Then the claim follows because in the leaders equilibrium: the probability of FF is $\tilde{p}^2$; outcomes FC and FF have probability $\tilde{p}(1-\tilde{p})$; and CC has probability $(1-\tilde{p})^2$. Given $\tilde{p} = p(F)$ this is as in the mixed equilibrium of the underlying game. □

We next state and prove the result concerning equilibrium in the case $P > \lambda + \xi$ and $2\lambda + \xi > 3$. Recall that the difference utility from CC minus utility from FC is

$$(1/2)\left[p^2 P - p(1-p)\left(P - (\lambda+\xi)\right) + (1-p)^2(2-(\lambda+\xi))\right]$$

We re-write this as

$$\left(P + 1 - (\lambda+\xi)\right)p^2 - \left[P/2 - 1 + (3/2)\left(2-(\lambda+\xi)\right)\right]p + \left(2-(\lambda+\xi)\right)/2 \tag{C.2}$$

Theorem 22. *For each pair (ξ, λ) with $2\lambda + \xi > 3$ there is a $\overline{P}(\xi,\lambda) > \lambda + \xi$ such that for $P \leq \overline{P}$ the equilibrium in the previous theorem still exists. For $2\lambda + \xi > 3$ and $P > \overline{P}$ the mixed leadership equilibrium can be described as follows. There is a $p(P)$, $0 < p(P) < \tilde{p}$ such that the group leaders play FC with probability $p(P)$ and FF with probability $1 - p(P)$; the common leader plays CC with probability q and FC and CF with probability $(1-q)/2$ each, with (writing p for $p(P)$)*

$$q = \frac{\xi + (1+p)P}{2\lambda + \xi - 2 - 2p\left(\lambda + \xi - 1\right) + P(1-p)} < 1.$$

As $P \to \infty$ we have $p(P) \to 0$ and $q \to 1$.

Proof. It is clear from the proof of the previous theorem that for each pair (ξ, λ) with $2\lambda + \xi > 3$ there is a $\overline{P}(\xi,\lambda) > \lambda + \xi$ such that for $P \leq \overline{P}$ that equilibrium still exists (because for $P \leq \lambda + \xi$ it is positive for any (ξ, λ) pair). Precisely, $\overline{P}(\Gamma)$ is the value at which for $p = \tilde{p}$ the function in (C.2) as a function of P is zero. Note that in this function, for fixed $P > \lambda + \xi$ the coefficient of p^2 is positive; the function is positive at $p = 0$, and the derivative there is

$$\begin{aligned} &2p(P + 1 - (\lambda+\xi)) - \left[P/2 - 1 + (3/2)\left(2-(\lambda+\xi)\right)\right]\big|_{p=0} \\ &= -(3/2)\left(2-(\lambda+\xi)\right) - P/2 + 1 = -(1/2)\left[3\left(2-(\lambda+\xi)\right) + P - 2\right] < 0 \end{aligned}$$

because $P > \lambda + \xi$ whence

$$3\left(2-(\lambda+\xi)\right) + P - 2 > 3\left(2-(\lambda+\xi)\right) + (\lambda+\xi) - 2 = 2\left(2-(\lambda+\xi)\right).$$

At $p = 1$ the value is $P/2 > 0$ so both roots are less than 1 (incidentally, the smaller one becomes

smaller as P grows larger, in fact it goes to zero). For each $P > \overline{P}(\Gamma)$ the function is negative at $p = \tilde{p}$ (by construction). Define $p(P)$ to be the root of (C.2) on the left of $\tilde{p}$; so $0 < p(P) < \tilde{p} < 1$. By construction for $p = p(P)$ we have $CC \sim_c FC \sim_c CF$.

We let p to be the $p(P)$ defined above. Consider a group leader. If she plays FF she gets (in square brackets what the common leader plays)

$$q\,(1 - p(1 - \xi)) + (1 - q)/2\,(\lambda + \xi - p\lambda)$$

while by playing FC she gets: (note that $q + (1 - q)/2 = (1 + q)/2$)

$$((1 + q)/2)\,[\lambda - p(\lambda + P)] - ((1 - q)/2)P$$

so she is indifferent if

$$((1 + q)/2)\,[\lambda - p(\lambda + P)] - (1 - q)/2)P.$$

This simplifies to

$$q = \frac{\xi + (1 + p)P}{2\lambda + \xi - 2 - 2p\,(\lambda + \xi - 1) + P(1 - p)},$$

and it can be checked that $q < 1$ if and only if $p < \tilde{p}$, which is true by construction. This ends the equilibrium argument, since in this case it is apparent that no leader has a profitable deviation.

Finally, as $P \to \infty$ we have $p(P) \to 0$ since $p(P) < \tilde{p}$ and $\tilde{p} \to 0$; and given this it is immediate that $q \to 1$. □

Appendix D. Proof of statements in Section 10

We consider first the case of the best correlated equilibrium. The incentive compatibility constraints in the definition of correlated equilibria have the value $\mu(FF)$ appearing in the two inequalities $\mu(a)(\lambda - 1) \geq \mu(FF)\xi$, with $a \in \{FC, CF\}$ (these are the inequalities corresponding to the communication of the action F). On the other hand, the value $\mu(FF)$ does not appear in the total welfare sum; thus in any solution $\hat{\mu}$ of the maximization of total welfare over the set of correlated strategies, necessarily $\hat{\mu}(FF) = 0$, that is:

$$\hat{\mu}(CC) + \hat{\mu}(CF) + \hat{\mu}(FC) = 1. \tag{D.1}$$

Adding the incentive compatibility constraint of the first and second player upon communication of the C action we obtain:

$$(\hat{\mu}(FC) + \hat{\mu}(CF))\xi \geq 2\hat{\mu}(CC)(\lambda - 1) \tag{D.2}$$

From (D.1) and (D.2) we conclude that the total probability on the two non cooperation action

profiles FC and CF is bounded below:

$$\hat{\mu}(FC) + \hat{\mu}(CF) \geq \frac{2(\lambda - 1)}{2(\lambda - 1) + \xi} \tag{D.3}$$

In summary, the best correlated equilibrium is:

$$\hat{\mu}(CC) = \frac{\xi}{\xi + 2(\lambda - 1)}, \hat{\mu}(FC) = \hat{\mu}(CF) = \frac{\lambda - 1}{\xi + 2(\lambda - 1)}, \hat{\mu}(FF) = 0 \tag{D.4}$$

with average utility:

$$\frac{\xi + (\lambda + \xi)(\lambda - 1)}{\xi + 2(\lambda - 1)} \tag{D.5}$$

Since $\lambda + \xi < 2$ this is clearly less than 1.

We next turn to "worst against worst" comparison. An argument analogous to the one above can be applied to determine the worst correlated equilibrium, $\underline{\mu}$, which is:

$$\underline{\mu}(CC) = 0, \underline{\mu}(FC) = \underline{\mu}(CF) = \frac{\xi}{2\xi + \lambda - 1}, \underline{\mu}(FF) = \frac{\lambda - 1}{2\xi + \lambda - 1} \tag{D.6}$$

with average utility:

$$\frac{(\lambda + \xi)\xi}{2\xi + \lambda - 1} \tag{D.7}$$

The mixed equilibrium in the underlying game is seen to yield payoff

$$\frac{\lambda\xi}{\lambda - 1 + \xi}$$

while the asymmetric pure equilibria give of course $(\lambda + \xi)/2$. Either of the two may yield lowest payoff, but both are easily verified to be higher than in the worst correlated equilibrium. Indeed: the mixed equilibrium is better than the worst CE if $\frac{\lambda\xi}{\lambda - 1 + \xi} > \frac{\xi(\lambda + \xi)}{\lambda - 1 + 2\xi}$, that is if $1 > \xi$. and the asymmetric beats it if $(\lambda + \xi)(\lambda - 1) > 0$. This proves the claim in the text.

For the sake of completeness we compare mixed and asymmetric equilibria of the underlying chicken game. Asymmetric better than mixed if

$$\frac{\lambda\xi}{\lambda - 1 + \xi} < \frac{\lambda + \xi}{2} \tag{D.8}$$

This is equivalent to $\lambda > \frac{1 + \sqrt{1 + 4\xi(1 - \xi)}}{2}$ but $\xi(1 - \xi) \leq 1/4$, so (D.8) is equivalent to:

$$\frac{1 + \sqrt{1 + 4\xi(1 - \xi)}}{2} \leq \frac{1 + \sqrt{2}}{2} \approx 1.207$$

so we conclude that for $\lambda \geq 1.207$ asymmetric beats mixed for all ξ; for $1 < \lambda < 1.207$ it depends on ξ; for $\lambda = 1$ mixed beats asymmetric for all $0 < \xi < 1$.

Lastly we spell out the procedure used to compute $P(\alpha)$ in Section 10.1. The level set $\overline{\pi}^{corr}(\lambda, \xi) = \alpha$ describes a curve

$$\xi(\lambda, \alpha) = \frac{(2\alpha - \lambda)(\lambda - 1)}{\lambda - \alpha}.$$

We insert the function $\xi = \xi(\lambda, \alpha)$ describing the α level set of the correlated utility, so that the equation $\tilde{\pi}(\lambda, \xi(\lambda, \alpha), P) = \alpha$ implicitly defines a $P(\lambda, \alpha)$; then for each given α we compute the *highest* such P over λ in the level set $\overline{\pi}^{corr}(\lambda, \xi(\lambda, \alpha)) = \alpha$. It is seen that in fact $P(\alpha)$ corresponds to the point $\lambda = 2\alpha, \xi = 0$ in the correlated payoff α-level set. Then the inequality $\tilde{\pi}(\lambda, \xi, P) > \alpha$ becomes

$$\frac{P}{(P + 2\alpha - 1)^2} \cdot (P + 2\alpha(2\alpha - 1)) > \alpha$$

which is seen to be equivalent to

$$P > (2\alpha - 1)\sqrt{\frac{\alpha}{1 - \alpha}}.$$

References

Alesina, A., R. Baqir, and W. Easterly (1999): "Public goods and ethnic divisions," *The Quarterly journal of economics*, 114, 1243–1284.

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.

Baliga, S. and T. Sjöström (2004): "Arms races and negotiations," *The Review of Economic Studies*, 71, 351–369.

——— (2020): "The strategy and technology of conflict," *Journal of Political Economy*, 128, 3186–3219.

Barro, R. J. (1973): "The control of politicians: an economic model," *Public choice*, 19–42.

Besley, T. (2006): *Principled agents?: The political economy of good government*, Oxford University Press on Demand.

Besley, T. and A. Case (1995): "Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits," *The Quarterly Journal of Economics*, 110, 769–798.

Besley, T. and S. Coate (1997): "An economic model of representative democracy," *The quarterly journal of economics*, 112, 85–114.

Dixit, A., G. M. Grossman, and E. Helpman (1997): "Common agency and coordination: General theory and application to government policy making," *Journal of political economy*, 105, 752–769.

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.

——— (2011): "Linking conflict to inequality and polarization," *American Economic Review*, 101, 1345–74.

Fearon, J. D. and D. D. Laitin (1996): "Explaining interethnic cooperation," *American political science review*, 90, 715–735.

Ferejohn, J. (1986): "Incumbent performance and electoral control," *Public choice*, 5–25.

Furnivall, J. S. (2014): *Colonial policy and practice*, Cambridge University Press.

Lijphart, A. (1977): *Democracy in plural societies: A comparative exploration*, Yale University Press.

Maskin, E. and J. Tirole (2004): "The politician and the judge: Accountability in government," *American Economic Review*, 94, 1034–1054.

Matějka, F. and G. Tabellini (2021): "Electoral competition with rationally inattentive voters," *Journal of the European Economic Association*, 19, 1899–1935.

Miquel, G. P. (2007): "The control of politicians in divided societies: The politics of fear," *Review of Economic studies*, 74, 1259–1274.

Osborne, M. J. and A. Slivinski (1996): "A model of political competition with citizen-candidates," *The Quarterly Journal of Economics*, 111, 65–96.

Prat, A. and A. Rustichini (2003): "Games played through agents," *Econometrica*, 71, 989–1026.
Rabushka, A. and K. A. Shepsle (1971): "Political entrepreneurship and patterns of democratic instability in plural societies," *Race*, 12, 461–476.