Collusion Constrained Equilibrium☆

Rohan Dutta[1], David K. Levine[2], Salvatore Modica[3]

Abstract

We study collusion within groups in non-cooperative games. The primitives are the preferences of the players, their assignment to non-overlapping groups and the goals of the groups. Our notion of collusion is that a group coordinates the play of its members among different incentive compatible plans to best achieve its goals. Unfortunately, equilibria that meet this requirement need not exist. We instead introduce the weaker notion of *collusion constrained equilibrium*. This allows groups to put positive probability on suboptimal alternatives in certain razor's edge cases where the set of incentive compatible plans changes discontinuously. These collusion constrained equilibria exist and are a subset of the correlated equilibria of the underlying game. We examine four perturbations of the underlying game. In each case we show that equilibria in which groups choose the best alternative exist and that limits of these equilibria lead to collusion constrained equilibria. We also show that for a sufficiently broad class of perturbations every collusion constrained equilibrium arises as such a limit. We give an application to a voter participation game showing how collusion constraints may be socially costly.

Keywords: Collusion, Organization, Group

☆First Version: June 2, 2013. We are especially grateful to Michele Boldrin, Michael Chwe, Drew Fudenberg, Zacharias Maniadis, Andrea Mattozzi, Joel Sobel, seminar participants at Bocconi, Glasgow, Naples and Padua and a number of anonymous referees. We are grateful to the EIEF, to NSF Grant SES-08-51315 and to the MIUR PRIN 20103S5RN3 for financial support.

*Corresponding author David K. Levine, 1 Brooking Dr., St. Louis, MO, USA 63130

Email addresses: rohan.dutta@mcgill.ca (Rohan Dutta), david@dklevine.com (David K. Levine), salvatore.modica@unipa.it (Salvatore Modica)

[1]Department of Economics, McGill University

[2]Department of Economics, EUI and WUSTL

[3]Università di Palermo

1. Introduction

As the literature on collective action (for example Olson (1965)) has emphasized, groups often behave collusively while the preferences of individual group members limit the possible collusive arrangements that a group can enter into. Neither individual rationality - ignoring collusion - nor group rationality - ignoring individual incentives - provides a satisfactory theory of interaction between groups. We study what happens when collusive groups face internal incentive constraints. Our starting point is that of a standard finite simultaneous move non-cooperative game. We suppose that players are exogenously partitioned into groups and that these groups have well-defined objectives. Given the play of the other groups there may be several Nash equilibria within a particular group. We model collusion within that group by supposing that the group will agree to choose the equilibrium that best satisfies its objectives. This idea is not new. It has been used in the study of trading economics, for example, by Hu, Kennan and Wallace (2009). It is closely connected to the idea in mechanism design that within a mechanism a particular group must not wish to recontract in an incentive compatible way. In this setting the group could be a group of bidders in an auction as in McAfee and McMillan (1992) and Caillaud and Jéhiel (1998), or it might consist of a supervisor and agent in the Principal/Supervisor/Agent model of Tirole (1986).[4] In political economy Levine and Modica (2016)'s model of peer pressure and its application to the role of political parties in elections by Levine and Mattozzi (2016) use the same notion of collusion.

The key problem that we address is that *strict collusion constrained equilibria* in which groups simultaneously try to satisfy their goals subject to incentive constraints do not generally exist. We show that this is due to the discontinuity of the equilibrium correspondence and show how it can be overcome by allowing, under certain razor's edge conditions, randomizations by groups between alternatives to which they are not indifferent. This leads to what we call *collusion constrained equilibrium*. These are a special type of correlated equilibrium of the underlying non-cooperative game. To motivate the definition we first consider three perturbations of the underlying model. We consider models in which there is slight randomness in group beliefs. We consider models in which groups may overcome incentive constraints at a substantial enforcement cost. For both of these perturbations strict collusion constrained equilibria exist and as the perturbation vanishes the equilibria of the perturbed game converge to collusion constrained equilibria of the underlying game.

[4] See also the more general literature on hierarchical models discussed in Tirole (1992) or Celik (2009). For other types of mechanisms see Laffont and Martimort (1997). Most of these papers study a single collusive group. One exception is Che and Kim (2009) who allow multiple groups they refer to as cartels. In the theory of clubs, such as Cole and Prescott (1997) and Ellickson et al (2001), implicitly collusion takes place within (many) clubs - but the clubs interact in a market rather than a game environment.

Finally, we consider a model in which there is a non-cooperative meta-game played between "leaders" and "evaluators" of groups and in which leaders have a slight valence. In this game Bayesian perfect equilibria exist and as the valence approaches zero the equilibrium play path converges to a collusion constrained equilibria of the underlying game.

These upper hemicontinuity results with respect to three perturbations show that the set of collusion constrained equilibria is "big enough" in the sense of containing the limits of equilibria of perturbed models. It leaves open the question of whether the set is "too big" in the sense that perhaps not all collusion constrained equilibria arise as such limits. Indeed, we show in a simple example that limits from perturbed games lead to strict refinements of collusion constrained equilibria - albeit different refinements depending on which perturbation we consider. The example also raises the possibility that the set of collusion constrained equilibria is too big because some collusion constrained equilibria do not arise as any limit from perturbed games. The example, however, is degenerate. In our final theoretical result we consider a combination of belief and enforcement cost perturbation, and to eliminate degeneracy allow also perturbations to the group objective. Once again in these perturbed games strict collusion constrained equilibria exist and converge to collusion constrained equilibria of the underlying game. However, for this broader class of perturbations we have the converse as well: all collusion constrained equilibria of the underlying game arise as such limits. Hence the set of collusion constrained equilibria is "exactly the right size," being characterized as the set of limit points of strict collusion constrained equilibria for this broader class of perturbations.

In our theory incentive constraints play a key role. In applied work the presence of incentive constraints within groups has often been ignored. For example political economists and economic historians often treat competing groups as single individuals: it is as if the group has an unaccountable leader who makes binding decisions for the group. In Acemoglu and Robinson (2000)'s theory of the extension of the franchise there are two groups, the elites and the masses, who act without incentive constraints. Similarly in the current literature on the role of taxation by the monarchy leading to more democratic institutions the game typically involves a monarch and a group (the elite).[5] A related class of models contains leadership models in which a group benefits from its members coordinating their actions in the presence of imperfect information about the environment.[6] In this literature, however, there is no game between groups - the problem is how to exploit the information being acquired by leader and group members in the group interest.[7] In our leader/evaluator

[5]Hoffman and Rosenthal (2000) explicitly assume that the monarch and the elite act as single agents, and this assumption seems to be accepted by later writers such as Dincecco, Federico and Vindigni (2011).

[6]For example Hermalin (1998), Dewan and Myatt (2008) and Bolton, Brunnermeier and Veldkamp (2013).

[7]For example, Bolton, Brunnermeier and Veldkamp (2013) find that the leader should not put too much

perturbation we also assume that the group decision is made by a single leader, but we add to the game evaluators who punish the leader for violating incentive constraints. We focus on strategic interaction between groups and a central element of our model is accountability, in that a leader whose recommendations are not endorsed by the group will be punished.

The branch of the game theory literature that is most closely connected to the ideas we develop here is the literature that uses non-cooperative methods to analyze cooperative games. There, however, the emphasis has been on the endogenous formation of coalitions. One example is Ray and Vohra (1999) who introduce a game in which players bargain over the formation of coalitions by making proposals to coalitions and accepting or rejecting those proposals within coalitions. This literature generally describes the game by means of a characteristic function and involves proposals and bargaining. We work in a framework of implicit or explicit coordination among group members in a non-cooperative game among groups. This is similar in spirit to Bernheim, Peleg and Whinston (1987)'s variation on strong Nash equilibrium, that they call coalition-proof Nash equilibrium, although the details of our model are rather different.

To make the theory more concrete we study an example based on the voter participation model of Palfrey and Rosenthal (1985) and Levine and Mattozzi (2016). We consider two parties voting over a transfer payment and we depart slightly from the standard model by assuming that ties are costly. In this setting we find all the Nash equilibria, all the collusion constrained equilibria, and all the equilibria in which the groups have a costless enforcement technology. We study how the equilibria compare as the stakes are increased. The main findings for this game are the following. For small stakes nobody votes. For larger stakes in Nash equilibrium it is always possible for the small party to win. If the stakes are large enough in collusion constrained and costless enforcement equilibrium the large party preempts the small and wins the election. For intermediate stakes strict collusion constrained equilibria do not exist, but collusion constrained equilibria do. For most parameter configurations the collusion constrained equilibria are more favorable for the large party than Nash equilibrium, less favorable than costless enforcement equilibrium, and less efficient than either.

2. A Motivating Example

The simplest - and as indicated in the introduction a widely used - theory of collusion is one in which players are exogenously divided into groups subject to incentive constraints. The basic idea we explore in this paper is that if, given the play of other groups, there

weight on the information coming from followers (what they call "resoluteness" of the leader).

is more than one in-group equilibrium, then a collusive group should be able to agree or coordinate on their "most desired" equilibrium.

Example 1. We start with an example with three players. The first two players form a collusive group while the third acts independently. The obvious condition to impose in this setting is that given the play of player 3, players 1 and 2 should agree on the incentive compatible (mixed) action profile that gives them the most utility. However, in the following game there is no equilibrium that satisfies this prescription.

Each player chooses one of two actions, C or D and the payoffs can be written in bi-matrix form. If player 3 plays C the payoff matrix for the actions of players 1 and 2 is a symmetric Prisoner's Dilemma game in which player 3 prefers that 1 and 2 both cooperate (play C)

	C	D
C	$6, 6, 5$	$0, 8, 0$
D	$8, 0, 0$	$2, 2, 0$

Think here of player 3 as an altruist who would like to see players 1 and 2 cooperate. This is worth 5 to him, but of course they will not cooperate given these payoffs. In order to induce cooperation player 3 has the option of paying 4 each to players 1 and 2 if they cooperate. Call D this strategy of player 3 and suppose the resulting payoffs are as follows, where notice that players 1 and 2 are then in a coordination game:

	C	D
C	$10, 10, 0$	$0, 8, 5$
D	$8, 0, 5$	$2, 2, 5$

The interpretation of player 3's payoff is this. By subsiding players 1 and 2 conditional on cooperation, player 3 feels she has "done her best" and this gives her utility of 5. If they do not cooperate this is what she gets. If players 1 and 2 do cooperate then player 3 gets 5 from having done her best, plus the benefit of 5 from seeing them both cooperate, minus the utility cost of the subsidy of 8, which we assume to be 10; she thus receives a net utility of 0. In this game player 3 prefers that players 1 and 2 fail to cooperate so she can have the pleasure of knowing she "did her best" without the cost of actually doing it.

Let α^i denote the probability with which player i plays C. We examine the set of equilibria for players 1 and 2 given the strategy α^3 of player 3. The payoff matrix for those two players is

	C	D
C	$6 + 4(1 - \alpha^3), 6 + 4(1 - \alpha^3)$	$0, 8$
D	$8, 0$	$2, 2$

so that as α^3 starts at 1 the two players face a prisoner's dilemma game with a unique Nash equilibrium at D, D, and as α^3 decreases the payoff to cooperation is increasing until at $\alpha^3 = 1/2$ the game becomes a coordination game and the set of equilibria changes discontinuously with a second pure strategy equilibrium at C, C; for $\alpha^3 < 1/2$ there is an additional symmetric strictly mixed equilibrium in which $\alpha^1 = \alpha^2 = 1/2(1 - \alpha^3)$.

How should the group of player 1 and player 2 collude given the play of player 3? If $\alpha^3 > 1/2$ they have no choice: there is only one in-group equilibrium at D, D. For $\alpha^3 \leq 1/2$ they each get $6 + 4(1 - \alpha^3)$ at the C, C equilibrium, 2 at the D, D equilibrium, and strictly less than $6 + 4(1 - \alpha^3)$ at the strictly mixed equilibrium. So if $\alpha^3 \leq 1/2$ they should choose C, C. Notice that in this example there is no ambiguity about the preferences of the group: they unanimously agree which is the best equilibrium.

We may summarize the play of the group by the "group best response." If $\alpha^3 > 1/2$ then the group plays D, D while if $\alpha^3 \leq 1/2$ the group plays C, C. What is the best response of player 3 to the play of the group? When the group plays D, D player 3 should play D and so $\alpha^3 = 0$ which is not larger than 1/2; when the group plays C, C player 3 should play C and so $\alpha^3 = 1$ which is not less than or equal to 1/2. Hence there is no equilibrium of the game in which the group of player 1 and player 2 chooses the best in-group equilibrium given the play of player 3.

In this example, the non-existence of an equilibrium in which player 1 and player 2 collude is driven by the discontinuity in the group best response: a small change in the probability of α^3 leads to an abrupt change in the behavior of the group, for as α^3 is increased slightly above .5 the C, C equilibrium for the group abruptly vanishes. The key idea of this paper is that this discontinuity is a shortcoming of the model rather than an intrinsic feature of the underlying group behavior. To motivate our proposed alternative let us step back for a moment to consider mixed strategy equilibria in ordinary finite games. There also the best response changes abruptly as beliefs pass through the critical point of indifference, albeit with the key difference that at the critical point randomization is allowed. But the abrupt change in the best response function still does not make sense from an economic point of view. A standard perspective on this is that of Harsanyi (1973) purification, or more concretely the limit of McKelvey and Palfrey (1995)'s Quantal Response Equilibria: the underlying model is perturbed in such a way that as indifference is approached players begin to randomize and the probability with which each action is taken is a smooth function of beliefs; in the limit as the perturbation becomes small, like the Cheshire cat, only the randomization remains. Similarly, in the context of group behavior, it makes sense that as the beliefs of a group change the probability with which they play different equilibria varies continuously. Consider, for example, $\alpha^3 = 0.499$ versus $\alpha^3 = 0.501$. In a practical setting where nobody actually knows α^3 does it make sense to assert that in the former case player 1 and 2 with probability 1 agree that $\alpha^3 \leq 0.5$ and in the latter case that $\alpha^3 > 0.5$? We think it makes more sense that they might in the first case agree that $\alpha^3 \leq 0.5$ with 90% probability and mistakenly agree that $\alpha^3 > 0.5$ with 10% probability and conversely in the second case. Consequently when $\alpha^3 = 0.499$ there would never-the-less be a 10% chance that the group would choose to play D, D not realizing that C, C is incentive compatible, while when $\alpha^3 = 0.501$ there would be a 10% chance that they would choose to play C, C incorrectly thinking that it is incentive compatible. We will develop below a formal model in which groups have beliefs that are a random function of the true play of the other groups and are only approximately correct. For the moment we expect, as in Harsanyi (1973), that in that limit only the randomization will remain. Our first step is to introduce a model that captures the grin of the Cheshire cat: we will simply assume that randomization is possible at the critical point. In the example we assert that when $\alpha^3 = 0.5$ and the incentive constraint exactly binds, the equilibrium "assigns" a probability to C, C being the equilibrium that is played by the group.[8] That is, when the incentive constraint holds exactly we do not assume that the group can choose their most preferred equilibrium, but instead we assume that there is an endogenously determined probability that they will choose that equilibrium.

Remark. Discontinuity and non-existence is not an artifact of restricting attention to Nash equilibrium. The same issue arises if we assume that players 1 and 2 can use correlated strategies. When the game is a prisoner's dilemma, that is, $\alpha^3 > 1/2$ then strict dominance implies that the unique Nash equilibrium is also the unique correlated equilibrium. When $\alpha^3 \leq 1/2$ the correlated equilibrium set is indeed larger than the Nash equilibrium set (containing at the very least the public randomizations over the Nash equilibria), but these correlated equilibria are all inferior for players 1 and 2 to C, C and so will never be chosen. While it is true that the correlated equilibrium correspondence is better behaved than the Nash equilibrium correspondence - it is convex valued and upper hemicontinuous - this example shows that the selection from that correspondence that chooses the best equilibrium for the group is never-the-less badly behaved - it is discontinuous.

This bad behavior of the best-equilibrium correspondence is related to some of the earliest work on competitive equilibrium. Arrow and Debreu (1954) showed that the best choice from a constraint set is well-behaved when the constraint set is lower hemicontinuous. If it is, then the maximum theorem can be applied to show that the argmax is a continuous correspondence. However, neither the Nash nor correlated equilibrium correspondence used as a constraint set is lower hemicontinuous, and - as we have seen - the best-equilibrium correspondence can then fail to be continuous.

[8] This is similar to Simon and Zame (1990)'s endogenous choice of sharing rules.

3. Collusion Constrained Equilibrium

3.1. The Environment

We now introduce our formal model of collusive groups that pursue their own interest subject to within-group individual incentive constraints. The membership in these groups is exogenously given and the ability of a group to collude is independent of actions taken by players outside of the group. We emphasize that we use the word collusion in the limited meaning that the group can choose an equilibrium to its liking. The goals of the group - like those of individuals - are exogenously specified: we do not consider the possibility of conflict within the group over goals.

Our basic setting is that of a standard normal form game. There are players $i = 1, 2, \ldots I$; player i chooses actions from a finite set $a^i \in A^i$ and receives utility $u^i(a^i, a^{-i})$. On top of this standard normal form game we have the structure of groups $k = \mathbf{1}, \mathbf{2}, \ldots \mathbf{K}$. There is a fixed assignment of players to groups $i \mapsto k(i)$. Notice that each player is assigned to exactly one group and that the assignment is fixed and exogenous. We use $a^k \in A^k$ to denote (pure) profiles of actions within group k and a to denote the profile of actions over all players. Like individuals, groups have well-defined objectives given by a payoff function $v^k(a^k, a^{-k}) = \sum_{i|k(i)=k} \beta^i u^i(a^k, a^{-k})$ for some positive utility weights $\beta^i > 0$. This implies on the one hand a preference for Pareto efficient plans, but also agreement on the welfare weights.

We assume that groups can make plans independently from other groups. We take this to mean that each group k has an independent *group randomizing device* the realization of which is known to all group members but not to players who are not group members. One implication of this is that the play of group k appears from the perspective of other groups to be a correlated strategy - a probability distribution $\rho^k \in R^k$ over pure action profiles A^k. In addition to the group randomizing device the individual players in group can randomize, so that by using the group randomizing device the group can randomly choose a profile of mixed strategies for group members. We let $\alpha^k \in \mathcal{A}^k$ represent such a profile, albeit we take $\mathcal{A}^k \subseteq R^k$ so that rather than regarding α^k as a profile of mixed strategies we choose to regard it as the generated distribution over pure strategy profiles A^k. Hence if the group mixes over a subset $B^k \subseteq \mathcal{A}^k$ using the group randomizing device the result is in the convex hull of B^k which we write as $H(B^k)$.

Players choose deviations $d^i \in D^i = A^i \cup \{0\}$ where the deviation $d^i = 0$ means "mix according to the group plan." Individual utility functions then give rise to a function

$$U^i(d^i, \alpha^k, a^{-k}) = \begin{cases} \sum_{a^k} u^i(a^i, a^{k-i}, a^{-k})\alpha^k[a^k] & d^i = 0 \\ \sum_{a^k} u^i(d^i, a^{k-i}, a^{-k})\alpha^k[a^k] & d^i \neq 0 \end{cases}.$$

It is convenient also to have a function that summarizes the degree of incentive incompatibility of a group plan. Noting that the randomizations of groups are independent of one another, for $\alpha^k \in \mathcal{A}^k, \rho^{-k} \in R^{-k}$ we define

$$G^k(\alpha^k, \rho^{-k}) = \max_{i|k(i)=k, d^i \in D^i} \sum_{a^{-k}} \left(U^i(d^i, \alpha^k, a^{-k}) - U^i(0, \alpha^k, a^{-k}) \right) \Pi_{j \neq k} \rho^j[a^j] \geq 0$$

which represents the greatest expected gain to any member of group k from deviating from the plan α^k given the play of the other groups. The condition for group incentive compatibility is simply $G^k(\alpha^k, \rho^{-k}) = 0$.

The key properties of the model are embodied in $G^k(\alpha^k, \rho^{-k})$ and

$$v^k(\alpha^k, \rho^{-k}) = \sum_a v^k(a^k, a^{-k}) \alpha^k[a^k] \Pi_{j \neq k} \rho^j[a^j]$$

Both functions are continuous in (α^k, ρ^{-k}) and it follows from the standard existence theorem for Nash equilibrium in finite games that for every ρ^{-k} there exists an α^k such that $G^k(\alpha^k, \rho^{-k}) = 0$. These properties together with $\mathcal{A}^k$ being a closed subset of R^k are the properties that are used in the remainder of the paper. For example, we could take $\mathcal{A}^k$ to be all correlated strategies by group k if we thought they had access to arbitrary correlating devices, or we could take $\mathcal{A}^k$ to be the mixed strategy of a representative individual in a homogeneous group if we thought such a group was restricted to anonymous play.[9]

3.2. Equilibrium

We now give our definition of *collusion constrained equilibrium*. As motivation, let us try to capture the idea that groups may randomize because their beliefs about the play of other groups are random. One possibility is to introduce an explicit random belief model and work directly with equilibria in that model. We shall do this subsequently, but we take a more direct approach here: we observe that when the degree of randomness is small the set of equilibria does not depend very much on the randomization, and formalize this by looking at the limit as randomness vanishes - a kind of Chesire cat model in which the cat

[9]It would not be appropriate to assume $\mathcal{A}^k$ convex for the following reason. We want public randomizations over incentive compatible plays. But a distribution over profiles which is a correlated equilibrium (hence incentive compatible) with respect to some correlating device is not necessarily generated by public randomization over incentive compatible profiles. For example, a group which has no correlating devices available and can only mix among action profiles cannot achieve the usual $(1/3, 1/3, 1/3)$ distribution in the game of chicken without violating incentive compatibility, because that distribution is obtainable only through the public randomization that puts weight $1/3$ on the three pure strategy profiles - which are not all incentive compatible. However a convex $\mathcal{A}^k$ containing the pure profiles would also contain the above distribution even if no correlating devices are available to the group. We must thus dispense with a convexity assumption on $\mathcal{A}^k$ to properly account for incentive compatibility within groups.

(the randomization) disappears and only the grin (equilibrium) is left. In doing so, our goal will be to provide a parsimonious model - one in which it is reasonably easy to say what equilibria are, in which the set of equilibria is reasonably small, one which puts weight only on the best equilibria at non-critical points. We shall subsequently show that our notion of equilibrium captures the limits of a variety of perturbed models and that in a precise sense all of these equilibria arise as limits of perturbed models.

Our preliminary goal, then, is to capture the idea that at critical points - where the group best equilibrium correspondence according to v^k is discontinuous - small random errors in beliefs may lead to playing equilibria that are not the best at the critical point. This will also "fill in" the discontinuity, which will lead to an existence theorem. To implement this program we must specify which randomizations over equilibria are allowed. Our approach is to consider which is the worst of the best equilibria for nearby beliefs. No slightly random belief would lead to a choice of an equilibrium that gives a smaller group utility than this. Hence we consider the equilibria that give at least this reservation value of group utility. This approach has the advantage that at critical points we need to compute just a single number for each group, and then we allow arbitrary randomization over the equilibria that give at least this group utility. As we shall see, this definition is relatively easy to work with.

Recall that $G^k(\alpha^k, \rho^{-k})$ measures the greatest gain in utility to any group member of deviating from the plan α^k. The greatest incentive compatible group utility is given by

$$V^k(\rho^{-k}) = \max_{\alpha^k \in \mathcal{A}^k \mid G^k(\alpha^k, \rho^{-k})=0} v^k(\alpha^k, \rho^{-k})$$

For the solutions to the maximization problem we say:

Definition 1. The *group best response set* $B^k(\rho^{-k})$ is the set of plans α^k satisfying $G^k(\alpha^k, \rho^{-k}) = 0$ *and* $v^k(\alpha^k, \rho^{-k}) = V^k(\rho^{-k})$.

Here $V^k(\rho^{-k})$ has the role of a group reservation utility. Note that $B^k(\rho^{-k})$ is closed. We can then define

Definition 2. $\rho \in R$ is a *strict collusion constrained equilibrium* if $\rho^k \in H[B^k(\rho^{-k})]$ for all k.

As we have seen in the example above these do not generally exist. Hence we now allow perturbations of beliefs about the other groups plans, and define the worst best utility for group k for beliefs near ρ^{-k} as

$$V^k_\epsilon(\rho^{-k}) = \inf_{|\sigma^{-k} - \rho^{-k}| < \epsilon} V^k(\sigma^{-k}).$$

Observing that this is non-increasing in ϵ so we may take the limit to get just the grin and define $V_S^k(\rho^{-k}) = \lim_{\epsilon \to 0} V_\epsilon^k(\rho^{-k})$ as the group *shadow* reservation utility. Further observing that as $\sigma^{-k} \to \rho^{-k}$ the incentive compatible plans at σ^{-k} should converge to those at ρ^{-k} we are led to the following definition:

Definition 3. The *shadow response set* $B_S^k(\rho^{-k})$ is the set of plans α^k that satisfy $G^k(\alpha^k, \rho^{-k}) = 0$ and $v^k(\alpha^k, \rho^{-k}) \geq V_S^k(\rho^{-k})$. [10]

Like $B^k(\rho^{-k})$ we have $B_S^k(\rho^{-k})$ closed. Note that since $V_S^k(\rho^{-k}) \leq V^k(\rho^{-k})$ we have $B_S^k(\rho^{-k}) \supseteq B^k(\rho^{-k})$. We know from example 1 that $B^k(\rho^{-k})$ may fail to be upper hemi-continuous. We show in the Appendix that by contrast the correspondence $B_S^k(\rho^{-k})$ must be upper hemicontinuous. Consequently $B_S^k(\rho^{-k}) = B^k(\rho^{-k})$ implies that $B^k(\rho^{-k})$ is also upper hemicontinuous at ρ^{-k} and we say that ρ^{-k} is a *regular point* for group k. Otherwise we say that ρ^{-k} is a *critical point* for group k.

We expect that limits of equilibria with random beliefs should place weight only on $B_S^k(\rho^{-k})$ (which we will subsequently demonstrate) so we adopt the following definition.

Definition 4. $\rho \in R$ is a *collusion constrained equilibrium* if $\rho^k \in H[B_S^k(\rho^{-k})]$ for all k.

The key to collusion constrained equilibrium is that we allow plans in $B_S^k(\rho^{-k})$ not merely in $B^k(\rho^{-k})$. If in a collusion constrained equilibrium $\rho^k \notin H[B^k(\rho^{-k})]$ we say that group k engages in *shadow mixing.* This means that the group puts positive probability on equilibria in $B_S^k(\rho^{-k}) \backslash B^k(\rho^{-k})$, that are not the best possible.

Our example above shows that shadow mixing may be necessary in equilibrium, as we spell out next.

Example. [*Example 1 revisited*] In the example we take $k(1) = k(2) = \mathbf{1}, k(3) = \mathbf{2}$. We take group utility to be defined by equal welfare weights $\beta^1 = \beta^2 = 1$.

To apply the definition of collusion constrained equilibrium we first compute for group $k = \mathbf{1}$ the best utility $V^{\mathbf{1}}(\rho^{\mathbf{2}})$ where since there is one player $\rho^{\mathbf{2}}$ may be identified with α^3. For $\alpha^3 \leq 1/2$ we know that the best equilibrium for group $k = \mathbf{1}$ is (C, C) with corresponding group utility $V^{\mathbf{1}}(\rho^{\mathbf{2}}) = 12 + 8(1 - \alpha^3)$, while for $\alpha^3 > 1/2$ the only equilibrium is (D, D) with group utility $V^{\mathbf{1}}(\rho^{\mathbf{2}}) = 4$. For $\alpha^3 \neq 1/2$ we have $V_S^{\mathbf{1}}(\rho^{\mathbf{2}}) = V^{\mathbf{1}}(\rho^{\mathbf{2}})$, and the shadow response and best response sets are the same: (C, C) for $\alpha^3 < 1/2$ and (D, D) for $\alpha^3 > 1/2$. At $\alpha^3 = 1/2$ the worst best utility for nearby beliefs are those for $\alpha^3 > 1/2$ giving a group utility of 4, whence the set of incentive compatible plans that give at least this utility are the equilibria (C, C) and (D, D), that is $B_S^{\mathbf{1}}(\rho^{\mathbf{2}}) = \{(C, C), (D, D)\}$. For the group $k = \mathbf{2}$

[10] The set $B_S^k(\rho^{-k})$ is a kind of shadow of nearby best equilibria - hence we refer to mixing over this shadow as shadow mixing.

consisting solely of individual 3 the shadow best response set is just the usual best response set.

Clearly there is no equilibrium with $\alpha^3 \neq 1/2$. On the other hand when $\alpha^3 = 1/2$ the group can shadow-mix 50-50 between (C, C) and (D, D), leaving player 3 indifferent between C and D; so this is a collusion constrained equilibrium. We conclude that there is a unique collusion constrained equilibria with $\boldsymbol{\rho^1}$ a 50-50 mixture over $\{(C, C), (D, D)\}$ and $\boldsymbol{\rho^2}$ a 50-50 mixture over $\{C, D\}$.

It should be apparent that collusion constrained equilibria use as correlating devices only the private randomization device available to each player and the group randomization device. We refer to correlated equilibria of the underlying game that use only these randomizing devices as *group correlated equilibria.*

Theorem 1. *Collusion constrained equilibria exist and are a subset of the group correlated equilibria of the underlying game.*

The theorem is proved in the Appendix.

4. Three Model Perturbations

We now study how collusion constrained equilibrium arises as a limit of equilibria in perturbed models. This is useful in the same way that the non-cooperative Nash demand game is useful in understanding the cooperative Nash bargaining solution. Equilibria in the perturbed models are standard in the sense that groups make best choices and there is no shadow mixing. In each case such equilibria are shown to exist. We consider three different types of perturbations. First, as discussed above, we consider the possibility that group beliefs are random. We will then consider the possibility that incentive constraints can be overcome by a costly enforcement technology. Finally, we suppose that group decisions are taken by a leader who has valence in the sense of being able to persuade group members to do as he wishes, but that if he issues orders that are not followed he is punished. In each case we take a limit: as beliefs become less random, enforcement becomes more costly, or valence shrinks; and in each case we show that the limit of equilibria of the perturbed games are collusion constrained equilibria in the unperturbed game. We emphasize that these are upper hemicontinuity results, that do not show that every collusion constrained equilibrium arises this way. The issue of lower hemicontinuity is considered subsequently.

4.1. Random Belief Equilibrium

We now show that collusion constrained equilibria are limit points of strict collusion constrained equilibria when beliefs of each group about behavior of the other groups are random and the randomness tends to vanish. We start by describing a random belief model. The idea is that given the true play ρ^{-k} of the other groups, there is a common belief σ^{-k} by group k that is a random function of that true play. Notice that these random beliefs are shared by the entire group - we could also consider individual belief perturbations, but it is the common component that is of interest to us, because it is this that coordinates group play. Conceptually if we think that a group colludes through some sort of discussions that gives rise to common knowledge - looking each other in the eye, a handshake and so forth - then it makes sense that during these discussions a consensus emerges not just on what action to take, but underlying that choice, a consensus on what the other groups are thought to be doing. We must emphasize: our model is a model of the consequences of groups successfully colluding - we do not attempt to model the underlying processes of communication, negotiation and consensus that leads to their successful collusion.

Definition 5. A density function $f^k(\sigma^{-k}|\rho^{-k})$ is called a *random group belief model* if it is continuous as a function of (σ^{-k}, ρ^{-k}); for $\epsilon > 0$ we say that the random group belief model is *only ϵ-wrong* if it satisfies $\int_{|\sigma^{-k}-\rho^{-k}|\leq\epsilon} f^k(\sigma^{-k}|\rho^{-k})d\sigma^{-k} \geq 1-\epsilon$.

In other words if the model is only ϵ-wrong then it places a low probability on being far from the truth. In Web Appendix 2 we give for every positive ϵ an example based on the Dirichlet distribution of a random group belief model that is only ϵ-wrong. We also define

Definition 6. A *group decision rule* is a function $b^k(\rho^{-k}) \in H[B^k(\rho^{-k})]$, measurable as a function of ρ^{-k}.

Notice that for given beliefs ρ^{-k} we are assuming that the group colludes on a response in $B^k(\rho^{-k})$ which is the set of the best choices for the group that satisfy the incentive constraints, and does not choose points in $B^k_S(\rho^{-k})\backslash B^k(\rho^{-k})$ as would be permitted by shadow mixing.

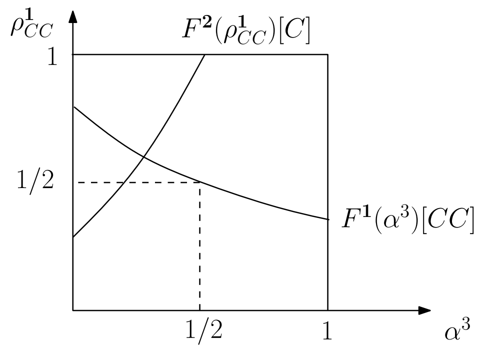
Definition 7. For a group decision rule b^k and random group belief model f^k *the group response function* is the distribution $F^k(\rho^{-k})[a^k] = \int b^k(\sigma^{-k})[a^k]f^k(\sigma^{-k}|\rho^{-k})d\sigma^{-k}$. If we have rules and belief models for all groups then a $\rho \in R$ that satisfies $\rho^k = F^k(\rho^{-k})$ for all k is called a *random belief equilibrium* with respect to b^k and f^k.

In the Appendix the following is proved:

Theorem 2. *If for each k we have group decision rules b^k and for each k and n we have random group belief models $f^k_{\epsilon_n}$ that are only ϵ_n-wrong then there exist random belief equilibria ρ_n with respect to b^k* and $f^k_{\epsilon_n}$. *Moreover if $\epsilon_n \to 0$ and $\rho_n \to \rho$ then ρ is a collusion constrained equilibrium.*

Example (Random belief equilibrium in example 1)**.** In Web Appendix 1 we analyze the random belief model corresponding to the Dirichlet belief model defined in Web Appendix 2. The figure below shows what the group response functions look like in our three player example. The key point is that the random belief equilibrium value of α^3 lies below 1/2, that is, as $\epsilon \to 0$ the collusion constrained equilibrium is approached from the left and above.

Figure 4.1: Beliefs equilibrium



4.2. Costly Enforcement Equilibrium

We now assume that each group k has a costly enforcement technology that it can use to overcome incentive constraints. In particular, we assume that every plan α^k is incentive compatible provided that the group pays a cost $C(\alpha^k, \rho^{-k})$ of carrying out the monitoring and punishment needed to prevent deviation. Levine and Modica (2016) show how cost of this type arise from peer discipline systems and Levine and Mattozzi (2016) study these systems in the context of voting by collusive parties: we give an example below. We assume $C^k(\alpha^k, \rho^{-k})$ to be non-negative and continuous in α^k, ρ^{-k} and adopt the following

Definition 8. A function $C^k(\alpha^k, \rho^{-k})$ is an *enforcement cost* if $C^k(\alpha^k, \rho^{-k}) = 0$ whenever $G^k(\alpha^k, \rho^{-k}) = 0$.

In other words enforcement is costly only if there is a deviation that needs to be deterred and nearby plans have similar enforcement costs. A particular example of such a cost function would be $C^k(\alpha^k, \rho^{-k}) = G^k(\alpha^k, \rho^{-k})$, that is, the cost of deterring a deviation is equal to the biggest benefit any player receives by deviating. Notice that we allow the possibility that incentive incompatible plans have zero cost.

With this technology we define

Definition 9. The *enforced group best response set* $B^k_C(\rho^{-k})$ is the set of plans α^k such that $v^k(\alpha^k, \rho^{-k}) - C^k(\alpha^k, \rho^{-k}) = \max_{\tilde{\alpha}^k \in \mathcal{A}^k} v^k(\tilde{\alpha}^k, \rho^{-k}) - C^k(\tilde{\alpha}^k, \rho^{-k})$.

Notice that again there is no shadow mixing here, just a choice of the group's best plan. Then we have the usual definition of equilibrium

Definition 10. $\rho \in R$ is a *costly enforcement equilibrium* if $\rho^k \in H[B_C^k(\rho^{-k})]$.

Notice that if the cost of enforcement is zero then the group can achieve the best outcome ignoring incentive constraints, an assumption, as we indicated in the introduction, often used by political economists and economic historians. We are interested in the opposite case in which enforcing non-incentive compatible plans is very costly. We then define

Definition 11. A sequence $C_n^k(\alpha^k, \rho^{-k})$ of cost functions is *high cost* if there are sequences $\gamma_n^k \to 0$ and $\Gamma_n^k \to \infty$ such that $G^k(\alpha^k, \rho^{-k}) > \gamma_n^k$ implies $C_n^k(\alpha^k, \rho^{-k}) \geq \Gamma_n^k$.

In the Appendix we prove[11]

Theorem 3. *Suppose $C_n^k(\alpha^k, \rho^{-k})$ is a high cost sequence. Then for each n a costly enforcement equilibrium ρ_n exists, and if $\lim_{n\to\infty} \rho_n \to \rho$ then ρ is a collusion constrained equilibrium.*

Example 2. We give a simple example of a costly enforcement technology and a high cost sequence based on Levine and Modica (2016). Specifically, we view the choice of α^k by group k as a social norm and assume that the group has a monitoring technology which generates a noisy signal of whether or not an individual member i complies with the norm. The signal is $z^i \in \{0, 1\}$ where 0 means "good, followed the social norm" and 1 means "bad, did not follow the social norm." Suppose further that if member i violates the social norm by choosing $\alpha^i \neq \alpha^k$ then the signal is 1 for sure while if he adhered to the social norm so that $\alpha^i = \alpha^k$ then the signal is 1 with probability π_n. When the bad signal is received the group member receives a punishment of size P^i.[12]

It is convenient to define the individual version of the gain to deviating

$$G^i(\alpha^k, \rho^{-k}) = \max_{d^i \in D^i} \sum_{a^{-k}} \left(U^i(d^i, \alpha^k, a^{-k}) - U^i(0, \alpha^k, a^{-k}) \right) \Pi_{j \neq k} \rho^j[a^j] \geq 0.$$

For the social norm α^k to be incentive compatible we need $P^i - \pi_n P^i \geq G^i(\alpha^k, \rho^{-k})$ which is to say $P^i \geq G^i(\alpha^k, \rho^{-k})/(1 - \pi_n)$. If the social norm is adhered to, the social cost of the punishment is $\pi_n P^i$, and the group will collude to minimize this cost so that it will choose $P^i = G^i(\alpha^k, \rho^{-k})/(1 - \pi_n)$. The resulting cost is then $(\pi_n/(1 - \pi_n))\, G^i(\alpha^k, \rho^{-k})$. Hence in this model $C_n^k(\alpha^k, \rho^{-k}) = (\pi_n/(1 - \pi_n)) \sum_{k(i)=k} G^i(\alpha^k, \rho^{-k})$.

[11] Actually it is not essential that $\Gamma_n^k \to \infty$, just that it be "big enough" that it would never be worth paying such a high cost.

[12] Here the coercion takes the form of punishment - but it could equally well be the withholding of a reward.

Since $C_n^k(\alpha^k, \rho^{-k}) = 0$ if and only if $G^k(\alpha^k, \rho^{-k}) = \max_{i|k(i)=k} G^i(\alpha^k, \rho^{-k}) = 0$ it follows that $C_n^k(\alpha^k, \rho^{-k})$ is an enforcement cost. We claim that as $\pi_n \to 1$, that is, as the signal quality deteriorates, this is in fact a high cost sequence. Certainly $C_n^k(\alpha^k, \rho^{-k}) \geq (\pi_n/(1-\pi_n))\, G^k(\alpha^k, \rho^{-k})$. Choose $\gamma_n^k \to 0$ such that $\Gamma_n^k \equiv (\pi_n/(1-\pi_n))\, \gamma_n^k \to \infty$. Then for $G^k(\alpha^k, \rho^{-k}) > \gamma_n^k$ we have $C_n^k(\alpha^k, \rho^{-k}) \geq \Gamma_n^k$ as required by the definition.

Example (Costly enforcement equilibrium in example 1). We use the high cost sequence just defined. In Web Appendix 1 we show that the costly enforcement equilibrium of our three-player game consists of the group randomizing half half between CC and DD while player 3 plays $\alpha^3 = (4-3\pi_n)/2$ for all $\pi_n > 4/5$. This equilibrium converges to the collusion constrained equilibrium as $\pi_n \to 1$. Notice that the collusion constrained equilibrium value of $\alpha^3 = 1/2$ is approached from the right while the group randomization in the costly enforcement equilibrium is constant and equal to the limiting constrained equilibrium value. This is the opposite of what we have seen in the random belief model where the approach is from the left and above.

4.3. Leader/Evaluator Equilibrium

In this section we study a model of leadership and define a non-cooperative game that respects incentive constraints. Leaders give their followers instructions - they tell them things such as "let's go on strike" or "let's vote against that candidate." The idea is that group leaders serve as explicit coordinating devices for groups. Each group will have a leader who tells group members what to do, and if he is to serve as an effective coordinating device these instructions cannot be optional. However, we do not want leaders to issue instructions that members would not wish to follow - that is, that are not incentive compatible. Hence we give them incentives to issue instructions that are incentive compatible by allowing group members to "punish" their leader. Indeed, we do observe in practice that it is often the case that groups follow orders given by a leader but engage in *ex post* evaluation of the leader's performance.

The leader/evaluator game is governed by two positive parameters ν, P. The parameter v measures the "valence" of a leader: this has a concrete interpretation as the amount of utility that group members are ready to give up to follow the leader.[13] Alternatively, ν can be thought of as measuring group loyalty. The parameter P represents a punishment that can be levied by a group member against the leader.[14]

[13] It is convenient notationally and for the statement of results that all leaders have the same valence; this also implicitly assumes that followers of a leader are equally willing to sacrifice. This entails no loss of generality since as long as the willingness to sacrifice is positive we can linearly rescale u^i to units in which willingness to sacrifice is the same.

[14] Again this might depend on k but we can rescale ν^k so that punishment is the same for all leaders.

Our non-cooperative game goes as follows:

Stage 1: Each leader privately chooses a plan $\alpha^k \in \mathcal{A}^k$: conceptually these are orders given to the members who must obey the orders.

Stage 2: Each player i with $k(i) = k$ serves as an evaluator and observing the plan α^k of the leader chooses a response $d^i \in A^k \cup \{0\}$.

Payoffs: The evaluator receives utility $U^i(d^i, \alpha^k, \alpha^{-k})$ if $d^i \neq 0$ and $U^i(0, \alpha^k, \alpha^{-k}) + \nu$ if $d^i = 0$, that is, he takes as given the other players in the group have followed orders and gets a bonus of ν for acquiescing to the leader's plan. Let Q^k denote the number of evaluators who chose $d^i \neq 0$. The leader receives $v^k(\alpha^k, \alpha^{-k}) - PQ^k$, that is, for each evaluator who disagrees with his decision he is penalized by P. Note that the leader and evaluator do not learn what the other groups did until the game is over.

Definition 12. We say that ρ is a *Bayesian perfect equilibrium* of the leader/evaluator game if there are probability distributions μ^k over $\mathcal{A}^k$ for the leaders and measures $\eta^i(\alpha^k)$ over $A^k \cup \{0\}$ measurable as a function of α^k for the evaluators such that

(i) $\rho^k = \int \sigma \mu^k(d\sigma)$

(ii) μ^k is optimal for the leader given ρ^{-k} and η^i

(iii) for all $\alpha^k \in \mathcal{A}^k$ and evaluators i the measure $\eta^i(\alpha^k)$ is optimal for the evaluator given α^k and ρ^{-k}.

Note that (iii) embodies the idea of "no signaling what you do not know" [15] that beliefs about the play of leaders of other groups is independent of the plan chosen by the leader of the own group.[16]

For this game to have an interesting relation to collusion constrained equilibrium, two things should be true.

- The evaluators must be able to punish the leader enough to prevent him from choosing incentive incompatible plans. A sufficient condition is that the punishment is greater than any possible gain in the game, that is, $P > \max v^k(\alpha^k, \alpha^{-k}) - \min v^k(\alpha^k, \alpha^{-k})$.

- The leader should be able to avoid punishment by choosing an incentive compatible plan. However the leader can only guarantee avoiding punishment if the evaluators strictly prefer not to deviate from his plan. If $\nu = 0$ this is true only for plans that are strictly incentive compatible and such plans may not exist. Hence the assumption

[15]It is known for finite games that this is an implication of sequentiality and Fudenberg and Tirole (1991) use this condition to define Bayesian perfect equilibrium for a class of games. Since the leader/evaluator game is not finite sequentiality is complicated. Hence it seems most straightforward to follow Fudenberg and Tirole (1991) and define Bayesian perfect directly with the "no signaling what you do not know" condition.

[16]Since the leader has no way of knowing if other leaders have deviated he should not be able to signal this through his own choice of action.

$\nu > 0$ is crucial: it assures that the leader can always avoid punishment by choosing an incentive compatible plan.

The following result is proved in the Appendix.

Theorem 4. *Suppose $\nu_n \to 0$ and $P_n > \max v^k(\alpha^k, \alpha^{-k}) - \min v^k(\alpha^k, \alpha^{-k})$. Then for each n a Bayesian perfect equilibrium ρ_n of the leader/evaluator game exists, and if $\lim_{n\to\infty} \rho_n = \rho$ then ρ is a collusion constrained equilibrium.*

Example (Leader/Evaluator equilibrium in example 1). For $\alpha^3 < 1/2$, playing CC is incentive compatible for the group, the question is how much can they mix out of the unique bad equilibrium DD when $\alpha^3 > 1/2$ given that they are willing to forgo gains no larger than ν. Web Appendix 1 shows that the equilibrium is $\hat{\alpha}^3 = (2+\nu)/4 > 1/2$ and that the group shadow mixes between the unique mixture $\hat{\alpha}^1 = \hat{\alpha}^2$ that is the smallest solution of $-4(\hat{\alpha}^1)^2(1-\hat{\alpha}^3) + 2\hat{\alpha}^1 = \nu$ and CC with probability

$$\frac{0.5 - (\hat{\alpha}^1)^2}{1 - (\hat{\alpha}^1)^2}$$

on CC. Note that as $\nu \to 0$ we have $\hat{\alpha}^1 \to 0$ so that in the limit the group shadow mixes between CC and DD as expected. Notice also that $\alpha^3 > 1/2$ so that the solution is on "the same side" of 1/2 as the costly enforcement equilibrium, but the opposite side of the belief equilibrium. The solution differs from both, however, in that the group does not randomize between CC and DD, but rather between CC and a mixed strategy.

5. Refinements of Collusion Constrained Equilibrium

In the perturbations we have considered the result is always that the limit of the perturbation is a collusion constrained equilibria. If there are several such equilibria, do the different limits converge to the same equilibrium? Not always. In this section we present an example with a continuum of collusion constrained equilibria each of which can be obtained as the limit of an appropriately chosen perturbation.

The example is a variation of Example 1, where player 3 gets zero for sure if he plays C, and the good equilibrium in the coordination game for the group which results if player 3 plays D is only weakly incentive compatible.

Example 3. The matrix on the left below contains the payoffs if player 3 plays C, the right one results if she plays D:

	C	D
C	6, 6, 0	0, 8, 0
D	8, 0, 0	2, 2, 0

	C	D
C	8, 8, 0	0, 8, 5
D	8, 0, 5	2, 2, 5

In this game clearly player 3 must play D with probability 1: if he plays C with any positive probability then it is strictly dominant for players 1 and 2 to play D in which case player 3 strictly prefers to play D. When player 3 plays D players 1 and 2 have exactly two equilibria: CC and DD; and any mixture between them is a collusion constrained equilibrium. To see this observe that independently of the utility weights entering $v^{\mathbf{1}}$, for any belief perturbation around $\alpha^3 = 0$ the worst equilibrium for the group is always DD so $V_S^{\mathbf{1}}(\alpha^3 = 0)$ is the utility the group obtains in that equilibrium. Thus any mixture between DD and CC satisfies the equilibrium condition, where of course in all strictly mixed equilibria the group gets utility higher than $V_S^{\mathbf{1}}$.

Now consider the perturbations. For any random beliefs C has positive probability so the group must play DD, so the only limit of random belief equilibria is DD. For costly collusion equilibrium on the other hand the better equilibrium CC for the group has zero cost so that will be chosen: the unique limit in this case is CC. Finally, for leadership equilibrium since the compliance bonus ν is positive again CC will be chosen, the unique limit is again CC. Notice that not only do the different perturbations sometimes pick different points out of the collusion constrained equilibrium set, but the collusion constrained equilibria involving strict mixtures do not arise as a limit from any of the perturbed models.

This example is non-generic because it depends heavily on the fact that when player 3 plays a pure strategy D players 1 and 2 are indifferent to deviating from CC. If we try to construct an example of this type in the interior then players 1 and 2 must shadow mix in the correct way to make player 3 indifferent and this should pin down what the shadow mixture must be. In the example we get around this by assuming that the pure strategy for player 3 is a strict best response so that there are a continuum of shadow mixtures by 1 and 2 that are consistent with player 3 playing D.

6. Lower Hemicontinuity

Roughly speaking, when we consider a perturbation such as random belief equilibrium, leadership equilibrium, or costly enforcement equilibrium we are exhibiting a degree of agnosticism about the model we have written down. That is we recognize that our model is an imperfect representation but hopefully reasonable approximation of a more complex reality and ask whether our equilibrium might be a good description of what happens in that more complex reality. This is the spirit behind refinements such as trembling hand perfection and concepts such as Harsanyi (1973)'s notion of purification of a mixed equilibrium. It is the question addressed by Fudenberg, Kreps and Levine (1988) who show how refinements do not capture the equilibria of all nearby games. We have shown that collusion constrained equilibrium does a good job of capturing random beliefs, costly enforcement and leadership equilibria. We know by example that there may be collusion constrained equilibria that do not arise as a limit of any of these. We now ask whether for a given collusion constrained equilibrium there is a story we can tell in the form of a perturbation representing a more complex reality that justifies the particular collusion constrained equilibrium.

Each of the perturbations we have considered has embodied a story or justification about why groups might be playing the way they are playing. We now consider a richer class of perturbations that combines elements of beliefs with costly enforcement and a perturbation of the group objective function. Specifically, we use the following:

Definition 13. A *perturbation* for each group k consists of a continuous belief perturbation $r_k^{-k}(\rho^{-k}) \in R^{-k}$, an enforcement cost function function $C^k(\alpha^k, \rho^{-k})$ and a continuous objective function $w^k(\alpha^k, \rho^{-k})$. A *perturbed equilibrium* ρ is defined by the condition $\rho^k \in H[\arg\max_{\alpha^k} w^k(\alpha^k, r_k^{-k}(\rho^{-k})) - C^k(\alpha^k, r_k^{-k}(\rho^{-k}))]$.

The belief perturbation is a simplification of the random belief model which assumed that beliefs were random but near correct most of the time. Now we are going to assume that they are deterministic and near correct. As in the random belief model we allow that beliefs are slightly wrong and do not require that two groups agree about the play of a third. The model of costly enforcement is exactly the same model we studied earlier. In addition we are now agnostic about the group objective and allow the possibility that the model may be slightly wrong in this respect. From a technical point of view it helps get rid of non-generic examples. As we are only interested in small perturbations we define

Definition 14. A sequence of perturbations $r_{kn}^{-k}, C_n^k, w_n^k$ is said to *converge* as $n \to \infty$ if $\max_{\rho^{-k}} |r_{kn}^{-k}(\rho^{-k}) - \rho^{-k}| \to 0$, if C_n^k is a high cost sequence, and if $\max_{\alpha^k, \rho^{-k}} |w_n^k(\alpha^k, \rho^{-k}) - v^k(\alpha^k, \rho^{-k})| \to 0$. We say that ρ is *justifiable* if there is a convergent sequence of perturbations together with perturbed equilibria $\rho_n \to \rho$.

Our main result, proven in the Appendix, is

Theorem 5. *A perturbed equilibrium exists for any perturbation, and ρ is justifiable if and only if it is a collusion constrained equilibrium.*

7. A Voting Participation Game

What difference do groups make? Collusion constrained equilibria are a subset of the set of group correlated equilibria, so we should expect that often the equilibria that are rejected are going to have better efficiency properties than those that are accepted. However, that comparison is not so interesting because it is the fact that the group is collusive that enables it to randomize privately from the other groups - that is, coordinate their play.[17] A more useful comparison is to ask what happens if the players play as individuals without correlating devices to coordinate their play, versus what happens if they are in collusive groups. In addition to static Nash equilibrium a second useful benchmark is to analyze the case in which there is free (costless) enforcement - so that incentive constraints do not matter.

Our setting for studying the economics of collusion is a voter participation game. We start with a relatively standard Palfrey and Rosenthal (1985)/Levine and Mattozzi (2016) framework: there are two parties, the "large" party has two voters, players 1 and 2, the "small" party one voter, player 3. Voters always vote for their own party, but is is costly to vote - a cost we normalize to 1 - and voters may choose whether or not to vote. The party that wins receives a transfer payment from the losing party. We assume that the cost or benefit to each member of the large party is $\tau > 0$ so that the cost or benefit to the small party member is 2τ. Usually it is assumed that a tie means that each party has a 50% chance of winning the prize, meaning that the election is a wash and no transfer payment is made. In case nobody votes we maintain this assumption that the status quo is unchanged and everyone gets 0. But when voting does take place it is often not the case in practice that a tie is innocuous - it may result in a deadlocked government or in conflict between the parties. So we we assume that a tie where each party casts one vote results in a deadlock that is - for simplicity - just as bad as a loss.

The payoffs can be written in bi-matrix form. If player 3 does not vote the payoff matrix for the actions of players 1 and 2 (where 0 represents do not vote and 1 represents vote) is

	1	0
1	$\tau - 1, \tau - 1, -2\tau$	$\tau - 1, \tau, -2\tau$
0	$\tau, \tau - 1, -2\tau$	$0, 0, 0$

This is may not be a prisoner's dilemma game between players 1 and 2, but it does have a unique dominant strategy equilibrium at which neither votes. If player 3 does vote the payoff matrix for the actions of players 1 and 2 becomes

	1	0
1	$\tau - 1, \tau - 1, -2\tau - 1$	$-\tau - 1, -\tau, -2\tau - 1$
0	$-\tau, -\tau - 1, -2\tau - 1$	$-\tau, -\tau, 2\tau - 1$

If $\tau > 1/2$ this is a coordination game for player 1 and 2 due to the fact that a tie is as bad

[17] The random belief model, in particular, only makes sense if the group is colluding, otherwise how can they agree on their beliefs?

as a loss: for a large party member not voting and having a tie results in $-\tau$ while voting and winning results in $\tau - 1 > -\tau$. Similarly voting and having a tie is as bad as a loss and it would be better to not vote and lose, suffering the same loss but not paying the cost of voting.

We now summarize the structure of collusion constrained, Nash and free enforcement equilibria in this model, with the full analysis in Web Appendix 3. There are a number of equilibria of different kinds in the various ranges of τ: (i) an equilibrium N where nobody votes (only for $\tau < 1/2$); (ii) an equilibrium L in which player 3 does not vote and the large group wins by casting a single vote. In the case of Nash there is also (iii) an equilibrium S in which only player 3 votes (and wins); (iv) equilibria L_2, L_3 where player 3 plays a pure strategy and the group members randomize with positive probability on both voting and not voting; (v) a fully mixed equilibrium M in which the large group members randomize as in the previous case; (v) two asymmetric mixed equilibria A in which only one of the group members votes with positive probability. In the case of collusion constrained equilibrium (CCE) and free enforcement equilibria (FEE) there are two types of equilibria with player 3 mixing, which in the CCE case involve shadow mixing: (vi) m_1 and M_1 in which the large group either stays out or casts a single vote; and (vii) m_2 and M_2 in which the large group either stays out or casts two votes. In all the equilibria where player 3 mixes the probability that neither group member votes is always $\rho^{\mathbf{1}}[0,0] = 1/2\tau$.

We define $\tilde{\tau} \equiv 1/(3-\sqrt{5}) \approx 1.31$. The entire set of equilibria is then given by the following table calculated in Web Appendix 3.

lower τ	upper τ	CCE	Nash	FEE
0	1/2	N	N	N
1/2	3/4	m_2	S	L, M_1, M_2
3/4	1	m_2	S	L
1	$\tilde{\tau}$	m_2, m_1, L	S, L, A	L
$\tilde{\tau}$	3/2	m_2, m_1, L	S, L, M, A, L_3	L
3/2	2	L	S, L, L_3	L
2	∞	L	S, L, L_2, L_3	L

The model has elements of both external and internal conflict. There is conflict between the groups as each hopes to get the transfer. There is also conflict within the large group as each prefers that the other votes. There are two sources of inefficiency in the model: total welfare (the sum of the utilities of the all three players) is reduced if players vote and is further reduced if there is a tie with one player from each group voting.

There are several basic points. If $\tau < 1/2$ then it is strictly dominant for player 3 not to vote: if the group casts no votes not voting gives 0 rather than $\tau - 1$, and if the group does cast votes then voting has no effect or results in an undesirable tie. Given that player 3 is not voting and $\tau < 1/2$ it is optimal both for player 1 and player 2 individually not to vote and for the group as a whole for neither of them to vote - there is no conflict here between individual incentives and group goals. Hence - in all types of equilibrium, CCE, Nash and FEE - when $\tau < 1/2$ the unique equilibrium involves no voting and this is efficient.

The interesting case is what happens when the stakes increase to $\tau > 1/2$. Here it cannot be an equilibrium for nobody to vote because in this case player 3 would prefer to vote. Of particular interest are the S and L equilibria: these are always the best for the small and large group respectively. To see this, observe that the best that can happen if nobody in a group votes is to get 0. On the other hand the best thing that can happen if a group casts at least one vote is that it casts only one vote and it wins, in which case the group gets $2\tau - 1$. When $\tau > 1/2$ this is better than not voting. In the equilibrium S and L in which exactly one player votes total welfare is always -1 reflecting the cost of the single vote that is cast.

Additional observations from Web Appendix 3 are the following. There are a few parameter ranges where there are equilibria giving higher welfare than the S, L value of -1: for FEE the M_1 when it exists gives higher welfare; for CCE m_1 gives higher welfare in the range $1 \leq \tau \leq 9/8$. All remaining equilibria give welfare less than -1. In the Nash case S is always an equilibrium and indeed for $\tau < 1$ this is the only equilibrium. By contrast in CCE and FEE the small player always gets a negative utility. Moreover in both cases when the stakes τ are high enough the only equilibrium is L - although this occurs for a smaller value of τ for FEE than CCE.

In the range $1/2 < \tau < 3/2$ shadow mixing is a possibility for CCE and for $1/2 < \tau < 1$ there is a unique CCE with shadow mixing m_2. In the shadow mixing equilibria the small group does better than at L while the large group does worse than L.

It is interesting to compare m_2 and M_2 in the range $1/2 < \tau < 3/4$, the former for CCE and the latter for FEE. In both equilibria the group mixes the same way, but the third player must vote more frequently in CCE than in FEE. The reason is that if the third player votes too infrequently then the incentive constraint fails when both members of the group vote.

Another observation of interest is that there are CCE and FEE that give the large group more utility but a lower probability of winning. Specifically in $1/2 < \tau < 3/4$ for FEE we have that M_1 is better for the large group than M_2 but gives them a smaller probability of winning, and the same is true for CCE in the range $1 < \tau < \tilde{\tau}$ for the shadow mixing equilibria $m_1$and m_2.

In the range $3/4 < \tau < 1$ equilibrium of all types are unique, which allows for sharp equilibrium comparison. The Nash equilibrium is S, and the FEE is L. The CCE is less efficient than either, but the large group does better than S and does worse than L. In this range as the stakes τ increase the probability of both members of the large group voting, the probability of everyone voting and the probability of the large group winning all increase, while total welfare decreases.

In a rough sense Nash is best for the small group, FEE is best for the large group and CCE is in between. This rough "in between" picture also emerges in the sense that CCE changes more gradually in favor of the large group as τ increases than does FEE.

Remark. With respect to welfare of the large group we have computed it in the obvious way as expected utility. For shadow mixing whether or not this is correct depends on the underlying model - with random beliefs it is correct. However in costly enforcement equilibrium shadow mixing appears as actual mixing, meaning that the group must be indifferent between the alternatives. In m_1 and m_2 staying out is strictly worse than casting either one or two votes. Hence in the costly enforcement equilibrium the cost of overcoming the incentive constraints to allow the casting of votes must exactly equal the difference in utility between casting the votes and staying out: that is to say, all the gain from vote casting must be dissipated in enforcement cost. Hence, in the limit, we should evaluate the utility of the group as the least utility of profiles over which shadow mixing occurs - that is to say, the utility from staying out. From Web Appendix 3 we know that the expected utility to the large group in m_1, m_2 is $3 - 2\tau - \frac{1}{2\tau}$ and $-3 + 2\tau + \frac{1}{2\tau}$ respectively while the probability of player 3 not voting is $\frac{1}{\tau}$ and $1 - \frac{1}{\tau}$ respectively. Hence the utility of staying out is $2(1 - \tau)$ and -2 respectively and this is the appropriate utility for the large group. In particular in the range $1 \leq \tau \leq 9/8$ it is no longer true that m_1 does better from an overall welfare perspective than L and S.

In the leadership case the utility assigned to a group when shadow mixing is ambiguous. From the perspective of the followers the correct calculation is expected utility. From the perspective of the leader the correct calculation is the least utility of profiles over which shadow mixing occurs - from the leader's point of view the punishment needed to make him indifferent dissipates the benefit of the better profiles. One may wonder why anyone would agree to be leader given that they get less utility than the followers. Although a discussion of who leaders are and why they are leaders is beyond the scope of this paper it is natural to imagine they get some additional compensation from the group for agreeing to be leader. In this case the follower utility seems the most relevant.

8. Conclusion

We study exogenously specified collusive groups and argue that the "right" notion of equilibrium is that of collusion constrained equilibrium. We start from the observation that groups such as political, ethnic, business or religious groups often collude. We adopt the simple assumption that a group will collude on the within-group equilibrium that best satisfies group objectives. We find that this seemingly innocuous assumption disrupts existence of equilibrium in simple games. We show that the existence problem is due to a discontinuity of the equilibrium set, and propose a "fix" which builds on the presumption that a group cannot be assumed to be able to play a particular within-group equilibrium with certainty when at that equilibrium the incentive constraints are satisfied with equality. This "tremble" implies that the group may put positive probability on actions which give group members lower utility but are strictly incentive compatible. We show that the resulting equilibrium notion has strong robustness properties and indeed is both upper and lower hemicontinuous with respect to a class of perturbations. This makes collusion constrained equilibrium a strong foundation for analyzing exogenous groups (including dynamic models where people flow between exogenous groups based on economic incentives as in the Acemoglu (2001) farm lobby model), which in some sense is the case that Olson (1965) had in mind and is of key importance in much of the political economy literature. This is not to argue that endogenous group formation is not of interest - but it is important to understand what happens as a consequence of group formation before building models of group formation and collusion constrained equilibrium is step in that direction.

References

Acemoglu, Daron (2001): "Inefficient Redistribution," *American Political Science Association*: 649-661.

Acemoglu, Daron and James A. Robinson (2000): "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective, *Quarterly Journal of Economics* 115: 1167-1199.

Aliprantis, Charalambos and Kim Border (2007): *Infinite Dimensional Analysis: A Hitchhiker's Guide (3rd Edition),* Springer

Arrow, K. J. and G. Debreu (1954): "Existence of an equilibrium for a competitive economy," *Econometrica.*

Bernheim, B.Douglas, Bezalel Peleg, Michael D Whinston (1987): "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory* 42: 1-12.

Bolton, Patrick, Markus Brunnermeier and Laura Veldkamp (2013): "Leadership, Coordination and Corporate Culture", *Review of Economic Studies* 80: 512-537

Caillaud, Bernard and Philippe Jéhiel (1998): "Collusion in Auctions with Externalities," *RAND Journal of Economics* 29: 680-702.

Celik, G. (2009): "Mechanism design with collusive supervision," *Journal of Economic Theory.*

Che, Yeon-Koo and Jinwoo Kim (2006): "Robustly Collusion-Proof Implementation," *Econometrica* 74: 1063–1107.

Che, Yeon-Koo and Jinwoo Kim (2009): "Optimal Collusion-Proof Auctions," *Journal of Economic Theory.*

Cole, H. and E. C. Prescott (1997): "Valuation Equilibrium with Clubs," *Journal of Economic Theory* 74: 19-39.

Dewan, Torun and David Myatt (2008): "The Qualities of Leadership: Direction, Communication, and Obfuscation", *American Political Science Review* 102: 351-368

Dincecco, Mark, Giovanni Federico and Andrea Vindigni (2011): "Warfare, Taxation, and Political Change: Evidence from the Italian Risorgimento", *The Journal of Economic History* 71: 887-914

Ellickson, B., B. Grodal, S. Scotchmer and W. R. Zame (2001): "Clubs and the Market: Large Finite Economies," *Journal of Economic Theory* 101: 40-77.

Fudenberg, Drew, David Kreps and David Levine (1988): "On the robustness of Equilibrium Refinements", *Journal of Economic Theory* 44: 354-380.

Fudenberg, Drew and Jean Tirole (1991): Perfect Bayesian equilibrium and sequential equilibrium, *Journal of Economic Theory* 53: 236-260

Harsanyi, J. C. (1973): "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-strategy Equilibrium Points," *International Journal of Game Theory* 2: 1-23.

Hermalin, Benjamin E. (1998): "Toward an Economic Theory of Leadership: Leading by Example", *American Economic Review* 88: 1188-1206

Hoffman, Philip T. and Jean-Laurent Rosenthal (2000): "Divided We Fall: The Political Economy of Warfare and Taxation", Mimeo, California Institute of Technology.

T. Hu, J. Kennan and N. Wallace (2009): "Coalition-Proof Trade and the Friedman Rule in the Lagos-Wright Model," *Journal of Political Economy.*

Laffont, Jean-Jacques and David Martimort (1997): "Collusion Under Asymmetric Information," *Econometrica* 65: 875-911.

Levine, David and Salvatore Modica (2016): "Peer Discipline and Incentives within Groups", *Journal of Economic Behavior and Organization,* in press.

Levine, David and Andrea Mattozzi (2016): "Voter Participation with Collusive Parties," http://www.dklevine.com/archive/refs4786969000000001234.pdf.

McAfee, R. Preston and John McMillan (1992): "Bidding Rings," *American Economic Review* 82: 579-599.

McKelvey, R. D. and T. R.Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior* 10: 6-38.

Olson, Mancur (1965): *The Logic of Collective Action: Public Goods and the Theory of Groups.*

Palfrey, Thomas R., and Howard Rosenthal (1985): "Voter participation and strategic uncertainty," *American Political Science Review* 79: 62-78.

Ray, Debraj and Rajiv Vohra (1999): "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior* 26: 286-336.

Simon, L.K. and W.R. Zame (1990): "Discontinuous Games and Endogenous Sharing Rules," *Econometrica* 58: 861-872.

Tirole, J. (1986): "Hierarchies and Bureaucracies: on the Role of Collusion in Organizations" *Journal of Law, Economics and Organization* 2: 181-214.

Tirole, J. (1992): "Collusion and the Theory of Organizations," in J. J. Laffont, ed., *Advances in Economic Theory: Sixth World Congress*, Vol. II, Cambridge University Press, Cambridge.

Appendix: Continuity, Limits and Existence

Lemma 1. *Suppose we have a sequence of sets B_n^k, correlated profiles $\rho_n^{-k} \to \rho^{-k}$, scalars V_n^k and positive numbers $\gamma_n^k \to 0$ satisfying for any $\alpha_n^k \in B_n^k$*

1. $G^k(\alpha_n^k, \rho_n^{-k}) \leq \gamma_n^k$
2. $v^k(\alpha_n^k, \rho_n^{-k}) \geq V_n^k$

If B^k is the set of $\alpha^k \in B^k$ that satisfies

1. $G^k(\alpha^k, \rho^{-k}) = 0$
2. $v^k(\alpha^k, \rho^{-k}) \geq \liminf V_n^k$

then for any $\rho_n^k \in H(B_n^k)$ with $\rho_n^k \to \rho^k$ it is the case that $\rho^k \in H(B^k)$

Proof. Since G^k, v^k are continuous the closure of B_n^k satisfies the same inequalities so it suffices to prove the result for closed sets B_n^k.

We have $\rho_n^k \in H(B_n^k)$ if and only if there exists a probability measure μ_n^k over B_n^k with $\rho_n^k = \int \sigma \mu_n^k(d\sigma)$. Since B_n^k is closed $\mathcal{A}^k \backslash B_n^k$ is open and we can extend the measure to all of $\mathcal{A}^k$ by taking $\mu_n^k[\mathcal{A}^k \backslash B_n^k] = 0$. Since $\mathcal{A}^k$ is compact we may extract a weakly convergent subsequence that converges to μ^k and without loss of generality may assume the original sequence has this property. Because $\mu_n^k \to \mu^k$ it follows from weak convergence that $\rho^k = \int \sigma \mu^k(d\sigma)$. The result will follow if we can show that $\mu^k[B^k] = 1$.

Consider the sets B_v^k for which $v^k(\alpha^k, \rho^{-k}) \geq \liminf V_n^k$ and B_0^k for which $G^k(\alpha^k, \rho^{-k}) = 0$. We will show that $\mu^k[B_v^k] = 1$ and $\mu^k[B_0^k] = 1$ from which it follows that $\mu^k[B^k] = \mu^k[B_v^k \cap B_0^k] = 1$.

For B_v^k let $\epsilon > 0$ and let $D_{v\epsilon}^k$ be the set $v^k(\alpha^k, \rho^{-k}) < \liminf V_n^k - \epsilon$. For n sufficiently large $D_{v\epsilon}^k \cap B_n^k = \emptyset$, so $\mu_n^k[D_{v\epsilon}^k] = 0$. However since v^k is continuous $D_{v\epsilon}^k$ is an open set and if $\mu^k[D_{v\epsilon}^k] > 0$ then for all sufficiently large n we have $\mu_n^k[D_{v\epsilon}^k] > 0$, a contradiction. We conclude that for all $\epsilon > 0$ we have $\mu^k[D_{v\epsilon}^k] = 0$, so indeed $\mu^k[B_v^k] = 1$.

For B_0^k let $\epsilon > 0$ and let $D_{0\epsilon}^k$be the set $G^k(\alpha^k, \rho^{-k}) > \epsilon$. Because $\mathcal{A}^k \times R^{-k}$ is compact $G^k(\alpha^k, \rho^{-k})$ is uniformly continuous so $G^k(\cdot, \rho_n^{-k})$ converges uniformly to $G^k(\cdot, \rho^{-k})$. Hence for n sufficiently large $\alpha^k \in B_{0\epsilon}^k$ implies $G^k(\alpha^k, \rho_n^{-k}) > \epsilon/2$ and since $\gamma_n^k \to 0$ also for sufficiently large n this implies $\mu_n^k[D_{0\epsilon}^k] = 0$. However, since G^k is continuous $D_{0\epsilon}^k$ is an open set, and if $\mu^k[D_{0\epsilon}^k] > 0$ then for all sufficiently large n we have $\mu_n^k[D_{0\epsilon}^k] > 0$ a contradiction. We conclude that for all $\epsilon > 0$ we have $\mu^k[D_{0\epsilon}^k] = 0$, so indeed $\mu^k[B_0^k] = 1$. □

Corollary 1. *Let the sets B_n^k be defined by $G^k(\alpha_n^k, \rho_n^{-k}) \leq \gamma_n^k$ and $v^k(\alpha_n^k, \rho_n^{-k}) \geq V_{\epsilon_n}^k(\rho_n^{-k})$. If $\gamma_n^k, \epsilon_n \to 0$ and $\rho_n^k \in H(B_n^k) \to \rho^k$ for all k then ρ is a collusion constrained equilibrium.*

Proof. If $\epsilon_n \leq \epsilon/2$ and $|\rho_n^{-k} - \rho^{-k}| \leq \epsilon/2$ then $|\sigma_n^{-k} - \rho_n^{-k}| \leq \epsilon_n$ implies $|\sigma_n^{-k} - \rho^{-k}| \leq \epsilon$ whence $V_{\epsilon_n}^k(\rho_n^{-k}) \geq V_\epsilon^k(\rho^{-k})$. This gives $\liminf V_{\epsilon_n}^k(\rho_n^{-k}) \geq V_S^k(\rho^{-k})$. Therefore taking $V_n^k = V_{\epsilon_n}^k(\rho_n^{-k})$, Lemma 1 shows that ρ^k is contained in the convex hull of a set contained in $B_S^k(\rho^{-k})$ for all k, whence the conclusion. □

Collusion Constrained Equilibrium

Theorem 6 (Theorem 1 in text)**.** *Collusion constrained equilibria exist and are a subset of the set of group correlated equilibria of the game.*

Proof. For any sequence of correlated profiles $\rho_n^{-k} \to \rho^{-k}$, let $\gamma_n^k = 0$ and let $V_n^k = V_S^k(\rho_n^k)$. Notice that $\liminf V_n^k \geq V_S^k(\rho^k)$. Then by Lemma 1 we know that the convex hull of the shadow best response set, $H(B_S^k(\rho^{-k}))$ is UHC. Existence of collusion constrained equilibria then follows from Kakutani. The fact that collusion constrained equilibria are group correlated equilibria follows from the fact that the incentive constraints are satisfied for each individual given signals generated by the private and group randomizing devices. □

Random Belief Equilibria

Theorem 7 (Theorem 2 in text)**.** *If for each k we have group decision rules b^k and for each k and n we have random group belief models $f_{\epsilon_n}^k$ that are only ϵ_n-wrong then there exist random belief equilibria ρ_n with respect to* b^k and $f_{\epsilon_n}^k$*. Moreover if $\epsilon_n \to 0$ and $\rho_n \to \rho$ then ρ is a collusion constrained equilibrium.*

Proof. Remember that $\rho_n^k(a^k) = F^k(\rho^{-k})[a^k] = \int b^k(\sigma^{-k})[a^k] f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k}) d\sigma^{-k}$ where $f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k})$ is continuous as a function of ρ^{-k}. So $\rho_n^k(a^k)$ is a continuous function of ρ^{-k} by the Dominated Convergence Theorem, for every a^k . Existence then follows from the Brouwer fixed point theorem.

Turning to convergence, by definition

$$\begin{aligned}\rho_n^k &= \int b^k(\sigma^{-k}) f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k}) d\sigma^{-k} \\ &= \int_{|\sigma^{-k}-\rho^{-k}|\leq\epsilon_n} b^k(\sigma^{-k}) f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k}) d\sigma^{-k} + \int_{|\sigma^{-k}-\rho^{-k}|>\epsilon_n} b^k(\sigma^{-k}) f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k}) d\sigma^{-k}\end{aligned}$$

Let $e_n^k(\rho^{-k}) \equiv \int_{|\sigma^{-k}-\rho^{-k}|\leq\epsilon_n} f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k}) d\sigma^{-k}$ and

$$\overline{\rho}_n^k \equiv \int_{|\sigma^{-k}-\rho^{-k}|\leq\epsilon_n} b^k(\sigma^{-k}) \frac{f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k})}{e_n^k(\rho^{-k})} d\sigma^{-k}$$

then we may write

$$\rho_n^k = e_n^k(\rho^{-k})\overline{\rho}_n^k + (1 - e_n^k(\rho^{-k})) \int_{|\sigma^{-k}-\rho^{-k}|>\epsilon_n} b^k(\sigma^{-k}) \frac{f_{\epsilon_n}^k(\sigma^{-k}|\rho^{-k})}{1 - e_n^k(\rho^{-k})} d\sigma^{-k}.$$

Now assume $\epsilon_n \to 0$. By assumption $e_n^k(\rho^{-k}) \to 1$ and $\rho_n^k \to \rho^k$ it follows that $\overline{\rho}_n^k \to \rho^k$. Take then $B_n^k \equiv \{\alpha^k \in B^k(\sigma^{-k}) | |\sigma^{-k} - \rho_n^{-k}| \leq \epsilon_n\}$. Clearly $\overline{\rho}_n^k \in H(B_n^k)$. We now show that the sequence $(\overline{\rho}_n^k, \rho_n^{-k})$ satisfies the hypotheses of Corollary 1. For any $\alpha_n^k \in B_n^k$ there is σ_n^{-k} with $|\sigma_n^{-k} - \rho_n^{-k}| \leq \epsilon_n$ such that $G^k(\alpha_n^k, \sigma_n^{-k}) = 0$ and $v^k(\alpha_n^k, \sigma_n^{-k}) = V^k(\sigma_n^{-k})$. Taking

$$\gamma_n^k = \max_{\alpha^k\in\mathcal{A}^k} \max_{|\sigma^{-k}-\rho^{-k}|\leq\epsilon_n} |G^k(\alpha^k, \sigma^{-k}) - G^k(\alpha^k, \rho^{-k})|$$

we see that $G^k(\alpha_n^k, \rho_n^{-k}) \leq \gamma_n^k$. Since G^k is continuous on a compact set it is uniformly continuous so $\gamma_n^k \to 0$. Moreover if $\alpha_n^k \in B_n^k$ then clearly $v^k(\alpha_n^k, \rho_n^{-k}) \geq V_{\epsilon_n}^k(\rho_n^{-k})$. The result now follows from Corollary 1. □

Leadership Equilibrium

For $\nu > 0$ define $V_\nu^k(\rho^{-k}) = \sup_{\alpha^k\in\mathcal{A}^k | G^k(\alpha^k,\rho^{-k})<\nu} v^k(\alpha^k, \rho^{-k})$ and $B_\nu^k(\rho^{-k})$ to be the set of plans α^k satisfying $G^k(\alpha^k, \rho^{-k}) \leq \nu$ *and* $v^k(\alpha^k, \rho^{-k}) \geq V_\nu^k(\rho^{-k})$.

Definition 15. We say that ρ is a *strict ν- equilibrium* if $\rho^k \in H[B_\nu^k(\rho^{-k})]$ for all k.

Theorem 8. *Strict ν- equilibria* exist.

Proof. It is sufficient to show that B_ν^k is UHC. By Theorem 17.35 in Aliprantis and Border (2007) we then know that $H[B_\nu^k(\rho^{-k})]$ is also UHC. Existence of strict ν- equilibrium then follows by Kakutani's fixed point theorem.

Consider a sequence $(\alpha_n^k, \rho_n^{-k})$ such that $\alpha_n^k \in B_\nu^k(\rho_n^{-k})$. Suppose that $\lim_{n\to\infty} \alpha_n^k = \alpha^k$ and $\lim_{n\to\infty} \rho_n^{-k} = \rho^k$. By continuity, $G^k(\alpha_n^k, \rho_n^{-k}) \leq \nu$ for all n implies that $G^k(\alpha^k, \rho^{-k}) \leq$

ν. Suppose by contradiction, $v^k(\alpha^k, \rho^{-k}) < V^k_\nu(\rho^{-k})$. By the continuity of v^k it follows that for sufficiently large n we have $v^k(\alpha^k_n, \rho^{-k}_n) < V^k_\nu(\rho^{-k})$. Since $v^k(\alpha^k_n, \rho^{-k}_n) \geq V^k_\nu(\rho^{-k}_n)$ this implies $V^k_\nu(\rho^{-k}_n) < V^k_\nu(\rho^{-k})$. Hence there is some $\hat{\alpha}^k$ such that $G^k(\hat{\alpha}^k, \rho^{-k}) < \nu$ and $V^k_\nu(\rho^{-k}_n) < v^k(\hat{\alpha}^k, \rho^{-k})$. By continuity of G^k and v^k this in turn implies that for sufficiently large n we have $G^k(\hat{\alpha}^k, \rho^{-k}_n) < \nu$ and $V^k_\nu(\rho^{-k}_n) < v^k(\hat{\alpha}^k, \rho^{-k}_n)$ contradicting the definition of $V^k_\nu(\rho^{-k}_n)$. □

Theorem 9. *ρ is a Bayesian perfect equilibrium of the leader evaluator game if and only if it is a strict ν-equilibrium.*

Proof. Suppose ρ is Bayesian perfect. Let μ^k and η^i be the corresponding leader and evaluator strategies. It suffices to show that $\mu^k[B^k_\nu(\rho^{-k})] = 1$. Denote the equilibrium utility of leader k by U^k.

Let D^k_ν be the subset of $\mathcal{A}^k$ for which $G^k(\alpha^k, \rho^{-k}) > \nu$. For $\alpha^k \in D^k_\nu$ there is an i with $k(i) = k$ for whom it is optimal to choose $\eta^i(\alpha^k)[\alpha^k] = 0$, hence utility for the leader is at most $\max v^k - P$ for those choices of α^k. Suppose $d = \mu^k[D^k_\nu] > 0$. Let $\hat{\alpha}^k \in \mathcal{A}^k$ satisfy $G^k(\hat{\alpha}^k, \rho^{-k}) = 0$ which we know exists. Consider $\hat{\mu}^k$ that takes the weight from D^k_v and puts it on $\hat{\alpha}^k$. The utility from $\hat{\mu}^k$ is at least $(1-d)U^k + d(U^k + \min v^k - \max v^k + P)$ which is bigger than U^k since $P > \max v^k - \min v^k$. Hence $d = 0$.

Let now D^k_ν be the subset of $\mathcal{A}^k$ for which $v^k(\alpha^k, \rho^{-k}) < V^k_\nu(\rho^{-k}) - \epsilon$. Let $\hat{\alpha}^k \in \mathcal{A}^k$ satisfy $G^k(\hat{\alpha}^k, \rho^{-k}) < \nu$ and $v^k(\hat{\alpha}^k, \rho^{-k}) > V^k_\nu(\rho^{-k}) - \epsilon/2$ which we know exists. By evaluator optimality we have $\eta^i(\hat{\alpha}^k)[\hat{\alpha}^k] = 1$ for all $k(i) = k$. Consider $\hat{\mu}^k$ that takes the weight from D^k_v and puts it on $\hat{\alpha}^k$. The utility from $\hat{\mu}^k$ is at least $U^k + d\epsilon/2$ so again $d = 0$. This shows that indeed $\mu^k[B^k_\nu(\rho^{-k})] = 1$.

Now suppose that ρ is a strict ν-equilibrium. Since $\rho^k \in H[B^k_\nu(\rho^{-k})]$ there exist measures μ^k with $\mu^k[B^k_\nu(\rho^{-k})] = 1$ and $\rho^k = \int \sigma\mu^k(d\sigma)$ so it suffices to find η^i that together with μ^k form a Bayesian perfect equilibrium. Let $\hat{\alpha}^i(\alpha^{-k}) \in \arg\max_{\alpha^i} u^i(\alpha^i, \alpha^k, \rho^{-k})$ be measurable. Observe that it cannot be that $G^k(\alpha^k, \rho^{-k}) < \nu$ and $v^k(\alpha^k, \rho^{-k}) > V^k_\nu(\rho^{-k})$, so consider the following evaluator optimal choice of η^i

(i) if $G^k(\alpha^k, \rho^{-k}) > \nu$ then $\eta^i[\hat{\alpha}^i(\alpha^{-k})] = 1$ and note that in this case $\hat{\alpha}^i(\alpha^{-k}) \neq \alpha^i$ for at least one i

(ii) if $G^k(\alpha^k, \rho^{-k}) \leq \nu$ and $v^k(\alpha^k, \rho^{-k}) \leq V^k_\nu(\rho^{-k})$ then $\eta^i[\alpha^i] = 1$

(iii) if $G^k(\alpha^k, \rho^{-k}) = \nu$ and $v^k(\alpha^k, \rho^{-k}) > V^k_\nu(\rho^{-k})$ some evaluator j is indifferent between α^j and some $\tilde{\alpha}^j \neq \alpha^j$ (and this evaluator can be chosen in a measurable way). For $i \neq j$ take $\eta^i[\alpha^i] = 1$. For j choose $\eta^j[\tilde{\alpha}^j] = \left(v^k(\alpha^k, \rho^{-k}) - V^k_\nu(\rho^{-k})\right)/P$ and $\eta^j[\alpha^j] = 1 - \eta^j[\tilde{\alpha}^j]$.

Then if $\alpha^k \in B^k_\nu(\rho^{-k})$ the leader utility is exactly $V^k_\nu(\rho^{-k})$, while if $G^k(\alpha^k, \rho^{-k}) > \nu$ then leader utility is at most $\max v^k - P$. Hence α^k is at least as good as any other choice, and indifferent to any other choice in $B^k_\nu(\rho^{-k})$. It follows that μ^k is optimal for leader k. □

Lemma 2. *$V^k_\nu(\rho^{-k}) \geq V^k_\epsilon(\rho^{-k})$ for any $\epsilon > 0$.*

Proof. From

$$\begin{aligned} V^k_\nu(\rho^{-k}) &= \sup_{\alpha^k \in \mathcal{A}^k \,|\, G^k(\alpha^k,\rho^{-k})<\nu} v^k(\alpha^k,\rho^{-k}) \\ &\geq \sup_{\alpha^k \in \mathcal{A}^k \,|\, G^k(\alpha^k,\rho^{-k})=0} v^k(\alpha^k,\rho^{-k}) = V^k(\rho^{-k}) \geq V^k_\epsilon(\rho^{-k}) \end{aligned}$$

the stated inequality follows. □

Theorem 10. *If ρ_n is a sequence of strict ν_n-equilibria, $\nu_n \to 0$ and $\rho_n \to \rho$ then ρ is a collusion constrained equilibrium.*

Proof. Let $\gamma_n = \nu_n$ and notice that for any $\alpha^k_n \in B^k_{\nu_n}(\rho^{-k}_n)$ we have $v^k(\alpha^k_n, \rho^{-k}_n) \geq V^k_{\nu_n}(\rho^{-k}_n) \geq V^k_{\epsilon_n}(\rho^{-k}_n)$ by Lemma 2 for some sequence $\epsilon_n \to 0$. Result now follows from Corollary 1. □

Perturbed Equilibrium: Existence and Upper HemiContinuity

Theorem 11. *A perturbed equilibrium exists for any perturbation.*

Proof. Notice that for any perturbation $w^k(\alpha^k, r^{-k}_k(\rho^{-k})) - C^k(\alpha^k, r^{-k}_k(\rho^{-k}))$ is continuous in its arguments. By the Maximum Theorem we then get the correspondence $\arg\max_{\alpha^k} w^k(\alpha^k, r^{-k}_k(\rho^{-k})) - C^k(\alpha^k, r^{-k}_k(\rho^{-k}))$ to be UHC. In turn by Theorem 17.35 in Aliprantis and Border (2007), $H[\arg\max_{\alpha^k} w^k(\alpha^k, r^{-k}_k(\rho^{-k})) - C^k(\alpha^k, r^{-k}_k(\rho^{-k}))]$ is UHC. Existence of perturbed equilibria then follows from the Kakutani fixed point theorem. □

Theorem 12. *If ρ is justifiable then it is a collusion constrained equilibrium.*

Proof. Suppose ρ is justifiable. Then there exists a sequence of perturbations $r^{-k}_{kn}, C^k_n, w^k_n$ such that $\max_{\rho^{-k}} |r^{-k}_{kn}(\rho^{-k}) - \rho^{-k}| \to 0$, C^k_n is a high cost sequence, and $\max_{\alpha^k,\rho^{-k}} |w^k_n(\alpha^k,\rho^{-k}) - v^k(\alpha^k,\rho^{-k})| \to 0$, each with a perturbed equilibrium ρ_n that converges to ρ.

Let $B^k_{wcn} = \arg\max_{\alpha^k} w^k_n(\alpha^k, r^{-k}_{kn}(\rho^{-k})) - C^k_n(\alpha^k, r^{-k}_{kn}(\rho^{-k}))$. Let $\tilde{v} = \max v^k - \min v^k$. Let $\delta_{n1} = \max_{\alpha^k,\rho^{-k}} |w^k_n(\alpha^k, r^{-k}_{kn}(\rho^{-k})) - w^k_n(\alpha^k,\rho^{-k})|$ and $\delta_{n2} = \max_{\alpha^k,\rho^{-k}} |w^k_n(\alpha^k,\rho^{-k}) - v^k(\alpha^k,\rho^{-k})|$. Since C^k_n is a high cost sequence, for all large enough n, $G^k(\alpha^k,\rho^{-k}) > \gamma^k_n$ would imply $C^k_n(\alpha^k,\rho^{-k}) > 2(\tilde{v} + \delta_{n1} + \delta_{n2})$ and since $\max_{\rho^{-k}} |r^{-k}_{kn}(\rho^{-k}) - \rho^{-k}| \to 0$, also $C^k_n(\alpha^k, r^{-k}_{kn}(\rho^{-k})) > \tilde{v} + \delta_{n1} + \delta_{n2}$. So for all such n, $\alpha^k \in B^k_{wcn}$ would mean $G^k(\alpha^k,\rho^{-k}) \leq \gamma^k_n$.

Let $W^k_n(\rho^{-k}) = \max_{\alpha^k \in \mathcal{A}^k \,|\, G^k(\alpha^k, r^{-k}_{kn}(\rho^{-k}))=0} w^k_n(\alpha^k, r^{-k}_{kn}(\rho^{-k}))$. Suppose $\alpha^k_n \in B^k_{wcn}$; then for large enough n it must be that

$$w^k_n(\alpha^k_n, r^{-k}_{kn}(\rho^{-k})) \geq W^k_n(\rho^{-k}) \geq V^k_S(\rho^{-k}) - \delta_{n1} - \delta_{n2}$$

This in turn means

$$v^k(\alpha_n^k,\rho^{-k}) \geq W_n^k(\rho^{-k}) - \delta_{n1} - \delta_{n2} \geq V_S^k(\rho^{-k}) - 2\delta_{n1} - 2\delta_{n2}$$

Notice that the sets B_{wcn}^k therefore satisfy the premise of Lemma 1 if we set the scalars V_n^k equal to $W_n^k(\rho^{-k})-\delta_{n1}-\delta_{n2}$. So we know that ρ must be such that for all k, $\rho^k \in H(B^k)$ where B^k is the set of α^k that satisfies $G^k(\alpha^k,\rho^{-k}) = 0$ and $v^k(\alpha^k,\rho^{-k}) \geq \liminf V_n^k$. Finally note that

$$\liminf W_n^k(\rho^{-k}) - \delta_{n1} - \delta_{n2} \geq \liminf V_S^k(\rho^{-k}) - 2\delta_{n1} - 2\delta_{n2} \Rightarrow \liminf W_n^k(\rho^{-k}) \geq V_S^k(\rho^{-k}).$$

ρ is therefore a collusion constrained equilibrium. □

Perturbed Equilibrium: Lower HemiContinuity

Theorem 13. *If ρ is a collusion constrained equilibrium then it is justifiable.*

Proof. We are given a collusion constrained equilibrium ρ and want to find a sequence of perturbations with perturbed equilibria $\rho_n \to \rho$. In fact the construction we are going to suggest will do something stronger, the idea is to construct a series of perturbations with perturbed equilibria ρ which obviously converges to itself. Recall that $\rho^k \in H[B_S^k(\rho^{-k})]$. The idea is to find a perturbed equilibrium so that $\arg\max_{\alpha^k} w_n^k(\alpha^k, r_k^{-k}(\rho^{-k})) - C_n^k(\alpha^k, r_k^{-k}(\rho^{-k})) = B_S^k(\rho^{-k})$; then clearly ρ^k itself is in $H[\arg\max_{\alpha^k} w_n^k(\alpha^k, r_k^{-k}(\rho^{-k})) - C_n^k(\alpha^k, r_k^{-k}(\rho^{-k}))]$.

Step 1: Choose, for each k, a sequence σ_{kn}^{-k} with $\sigma_{kn}^{-k} \to \rho^{-k}$ and $V^k(\sigma_{kn}^{-k}) \to V_S^k(\rho^{-k})$. We know that we can find such a sequence by the definition of $V_S^k(\rho^{-k})$: it is the limit of the worst of the local best, so there must be some sequence of local best that converges to it.

Constants: Define $\overline{G}^k(\sigma^{-k}) = \max_{\alpha^k} |G^k(\alpha^k,\sigma^{-k}) - G^k(\alpha^k,\rho^{-k})|$, $\overline{G}_n^k = \overline{G}^k(\sigma_{kn}^{-k})$, and similarly $\overline{V}(\sigma^{-k}) = \max\{0, V^k(\sigma^{-k}) - V_S^k(\rho^{-k})\}$, $\overline{V}_n^k = \overline{V}(\sigma_{kn}^{-k})$ and note that both $\overline{G}_n^k$ and $\overline{V}_n^k$ go to zero as $n \to \infty$. Also let $\overline{v}^k(\sigma^{-k}) = \max_{\alpha^k} |v^k(\alpha^k,\sigma^{-k}) - v^k(\alpha^k,\rho^{-k})|$ and $\overline{v}_n^k = \overline{v}^k(\sigma_{kn}^{-k})$; observe that $\overline{v}_n^k \to 0$. Take $\lambda_n^k = 1/\sqrt{\overline{G}_n^k}$ which goes to infinity, $\overline{\kappa}_n^k = 3(\overline{v}_n^k + \overline{V}_n^k + \lambda_n^k \overline{G}_n^k)$ which goes to zero and $\overline{\gamma}_n^k = 1/\sqrt{\lambda_n^k}$ which goes to zero.

The functions $\overline{w}_n^k(\alpha^k,\sigma^{-k})$ and $\overline{C}_n^k(\alpha^k,\sigma^{-k})$: Define first $D_n^k(\alpha^k) = \max\{0, v^k(\alpha^k,\rho^{-k}) - V_S^k(\rho^{-k})\} + \lambda_n^k G(\alpha^k,\rho^{-k})$ and $d_n^k(\alpha^k) = \min\{D_n^k(\alpha^k), \overline{\kappa}_n^k\}$. This converges uniformly to zero. We then take $\overline{C}_n^k(\alpha^k,\sigma^{-k}) = D_n^k(\alpha^k) - d_n^k(\alpha^k)$ and $\overline{w}_n^k(\alpha^k,\sigma^{-k}) = v^k(\alpha^k,\rho^{-k}) - d_n^k(\alpha^k)$. Observe that

$$\begin{aligned}
\overline{w}_n^k(\alpha^k,\sigma^{-k}) - \overline{C}_n^k(\alpha^k,\sigma^{-k}) =& v^k(\alpha^k,\rho^{-k}) - D_n^k(\alpha^k) \\
=& v^k(\alpha^k,\rho^{-k}) - \max\{0, v^k(\alpha^k,\rho^{-k}) - V_S^k(\rho^{-k})\} - \lambda_n^k G(\alpha^k,\rho^{-k}) \\
=& \min\{v^k(\alpha^k,\rho^{-k}), V_S^k(\rho^{-k})\} - \lambda_n^k G(\alpha^k,\rho^{-k})
\end{aligned}$$

Key fact: $\arg\max_{\alpha^k} \overline{w}_n^k(\alpha^k, \sigma^{-k}) - \overline{C}_n^k(\alpha^k, \sigma^{-k}) = B_S^k(\rho^{-k})$. To see this consider the maximizers of $\min\{v^k(\alpha^k, \rho^{-k}), V_S^k(\rho^{-k})\} - \lambda_n^k G(\alpha^k, \rho^{-k})$. For the elements of $B_S^k(\rho^{-k})$, that is the α^k for which $G(\alpha^k, \rho^{-k}) = 0$ and $v^k(\alpha^k, \rho^{-k}) \geq V_S^k(\rho^{-k})$, the expression equals $V_S^k(\rho^{-k})$. Outside $B_S^k(\rho^{-k})$, that is for α^k such that $G^k(\alpha^k, \rho^{-k}) > 0$ or $v^k(\alpha^k, \rho^{-k}) < V_S(\rho^{-k})$, the expression is lower than that value. This proves the assertion.

Properties: There exists $\epsilon_n^k > 0$ *such that* $|\sigma^{-k} - \sigma_{kn}^{-k}| \leq \epsilon_n^k$ *implies*

(i) if $G^k(\alpha^k, \sigma^{-k}) > \overline{\gamma}_n^k$ then $\overline{C}_n^k(\alpha^k, \sigma^{-k}) \geq \lambda_n^k \overline{\gamma}_n^k - \kappa_n^k - 2\lambda_n^k \overline{G}_n^k \to \infty$

(ii) if $G^k(\alpha^k, \sigma^{-k}) = 0$ then $\overline{C}_n^k(\alpha^k, \sigma^{-k}) = 0$

(iii) $|\overline{w}_n^k(\alpha^k, \sigma^{-k}) - v^k(\alpha^k, \sigma^{-k})| \leq 2\overline{v}_n^k + \overline{\kappa}_n^k \to 0$

Proof of these:

(i) $\overline{C}_n^k(\alpha^k, \sigma^{-k}) \geq \lambda_n^k G(\alpha^k, \rho^{-k}) - \overline{\kappa}_n^k \geq \lambda_n^k G(\alpha^k, \sigma^{-k}) - \overline{\kappa}_n^k - \lambda_n^k \overline{G}^k(\sigma^{-k})$, so choose ϵ_n^k small enough that $\overline{G}^k(\sigma^{-k}) \leq 2\overline{G}_n^k$.

(ii) Choose $\epsilon_n^k > 0$ such that for all $|\sigma^{-k} - \sigma_{kn}^{-k}| < \epsilon_n^k$ we have $\max_{\alpha^k} |G^k(\alpha^k, \sigma^{-k}) - G^k(\alpha^k, \sigma_{kn}^{-k})| \leq \overline{G}_n^k$. Note that $\max_{\alpha^k} |G^k(\alpha^k, \sigma_{kn}^{-k}) - G^k(\alpha^k, \rho^{-k})| = \overline{G}_n^k$. Hence by the triangle inequality $G^k(\alpha^k, \sigma^{-k}) = 0$ implies $G^k(\alpha^k, \rho^{-k}) \leq 2\overline{G}_n^k$.

Also choose $\epsilon_n^k > 0$ such that for all $|\sigma^{-k} - \sigma_{kn}^{-k}| < \epsilon_n^k$ we have $V_S^k(\sigma^{-k}) \leq V_S^k(\sigma_{kn}^{-k}) + \overline{v}_n^k$. Note that $V_S^k(\sigma_{kn}^{-k}) \leq V_S^k(\rho^{-k}) + \overline{V}_n^k$. Hence $V_S^k(\sigma^{-k}) \leq V_S^k(\rho^{-k}) + \overline{v}_n^k + \overline{V}_n^k$. Therefore $G^k(\alpha^k, \sigma^{-k}) = 0$ implies $v^k(\alpha^k, \sigma^{-k}) \leq V_S^k(\rho^{-k}) + \overline{v}_n^k + \overline{V}_n^k$.

Finally choose $\epsilon_n^k > 0$ such that for all $|\sigma^{-k} - \sigma_{kn}^{-k}| < \epsilon_n^k$ we have $\max_{\alpha^k} |v^k(\alpha^k, \sigma^{-k}) - v^k(\alpha^k, \sigma_{kn}^{-k})| \leq \overline{v}_n^k$. Hence by the triangle inequality $\max_{\alpha^k} |v^k(\alpha^k, \sigma^{-k}) - v^k(\alpha^k, \rho^{-k})| \leq 2\overline{v}_n^k$.

Putting these inequalities together we see that $G^k(\alpha^k, \sigma^{-k}) = 0$ implies that $D_n^k(\alpha^k) = \max\{0, v^k(\alpha^k, \rho^{-k}) - V_S^k(\rho^{-k})\} + \lambda_n^k G(\alpha^k, \rho^{-k}) \leq 3\overline{v}_n^k + \overline{V}_n^k + 2\lambda_n^k \overline{G}_n^k \leq \overline{\kappa}_n^k$, which in turn implies $\overline{C}_n^k(\alpha^k, \sigma^{-k}) = 0$.

(iii) Recalling that $\epsilon_n^k > 0$ is such that for all $|\sigma^{-k} - \sigma_{kn}^{-k}| < \epsilon_n^k$ we have $\max_{\alpha^k} |v^k(\alpha^k, \sigma^{-k}) - v^k(\alpha^k, \sigma_{kn}^{-k})| \leq \overline{v}_n^k$, property (iii) follows from

$$
\begin{aligned}
&|\overline{w}_n^k(\alpha^k, \sigma^{-k}) - v^k(\alpha^k, \sigma^{-k})| \\
&\leq |v^k(\alpha^k, \sigma^{-k}) - v^k(\alpha^k, \sigma_{kn}^{-k})| + |v^k(\alpha^k, \sigma_{kn}^{-k}) - v^k(\alpha^k, \rho^{-k})| + d_n^k(\alpha^k) \leq 2\overline{v}_n^k + \overline{\kappa}_n^k
\end{aligned}
$$

Step 2: We now have $\overline{w}_n^k(\alpha^k, \sigma^{-k})$ and $\overline{C}_n^k(\alpha^k, \sigma^{-k})$ which are defined in a ϵ_n^k-neighborhood of σ_{kn}^{-k} and have the right properties there. For $|\sigma^{-k} - \rho^{-k}| < \epsilon_n^k$ we define $\overline{r}_{kn}^{-k}(\sigma^{-k}) = \sigma_{kn}^{-k}$ (taking advantage of the fact that these need not be the same for all k). We must now extend these to functions $w_n^k(\alpha^k, \sigma^{-k}), C_n^k(\alpha^k, \sigma^{-k}), r_{kn}^{-k}(\sigma^{-k})$ on all of R^{-k} while preserving the right properties and the values of $\overline{w}_n^k(\alpha^k, \sigma_{kn}^{-k})$, $\overline{C}_n^k(\alpha^k, \sigma_{kn}^{-k})$ and $\overline{r}_{kn}^{-k}(\rho^{-k})$. We can do this with a simple pasting. Let $\beta_n^k(x)$ be a non-negative continuous real valued function taking the value of 1 at $x = 0$ and the value of 0 for $x \geq \epsilon_n^k$. Then we define

$$w_n^k(\alpha^k, \sigma^{-k}) = \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|)\overline{w}_n^k(\alpha^k, \sigma^{-k}) + (1 - \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|))v^k(\alpha^k, \sigma^{-k})$$

$$C_n^k(\alpha^k, \sigma^{-k}) = \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|)\overline{C}_n^k(\alpha^k, \sigma^{-k}) + (1 - \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|))\lambda_n^k G^k(\alpha^k, \sigma^{-k})$$

$$r_n^k(\sigma^{-k}) = \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|)\overline{r}_n^k(\sigma^{-k}) + (1 - \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|))\sigma^{-k}.$$

It is easy to check that these pasted functions have the correct properties. Note that requiring $\overline{w}_n^k(\alpha^k, \sigma^{-k})$ and $\overline{C}_n^k(\alpha^k, \sigma^{-k})$ to have the right properties in the ϵ_n^k-neighborhood of σ_{kn}^{-k} ensures that the above convex combinations inherit those properties. $\square$

Web Appendix 1: Perturbations in the Leading Example

Recall the payoff matrices if player 3 plays C (left) or D (right)

	C	D			C	D
C	$6, 6, 5$	$0, 8, 0$		C	$10, 10, 0$	$0, 8, 5$
D	$8, 0, 0$	$2, 2, 0$		D	$8, 0, 5$	$2, 2, 5$

We ease notation a bit. Group $\mathbf{2}$ is just player 3 who has to choose between C and D; we let $\alpha^3 = \rho^{\mathbf{2}}[C] = \rho^{-\mathbf{1}}[C]$. Then recall that given α^3 the payoff matrix for players $1, 2$ is

	C	D
C	$6 + 4(1 - \alpha^3), 6 + 4(1 - \alpha^3)$	$0, 8$
D	$8, 0$	$2, 2$

so that if $\alpha^3 < 1/2$ they play CC, if $\alpha^3 > 1/2$ they play DD. We will drop the superscript from $\rho^{\mathbf{1}} = \rho^{-\mathbf{2}}$ so this is going to be ρ, with ρ_{CC}, ρ_{DD} the probabilities that group $\mathbf{1}$ plays CC or DD. For individual play we will also use α^i for the probability that $i = 1, 2$ plays C.

Player 3 payoff from C is $5\rho_{CC}$, from D it is $5(1-\rho_{CC})$ so indifference imposes $\rho_{CC} = 1/2$: if $\rho_{CC} > 1/2$ he plays C, if $\rho_{CC} < 1/2$ he plays D.

Belief Equilibrium

Assume Dirichlet belief model (defined in Web Appendix 2). What do the group response functions look like? Recall that σ indicates the beliefs variable. For group $\mathbf{1}$ they play only CC and DD, and the probability $F^{\mathbf{1}}(\alpha^3)[CC]$ of playing CC is the probability that the belief $\sigma^{-\mathbf{1}}[C] < 1/2$; this is strictly between 0 and 1, symmetric around $\alpha^3 = 1/2$ where it is equal to $1/2$ and strictly decreasing in α^3.

For player 3 the probability $F^{\mathbf{2}}(\rho)[C]$ of playing C is the probability that the belief $\sigma^{-\mathbf{2}}[CC] > 1/2$; this is strictly between 0 and 1 and strictly increasing in ρ_{CC}.

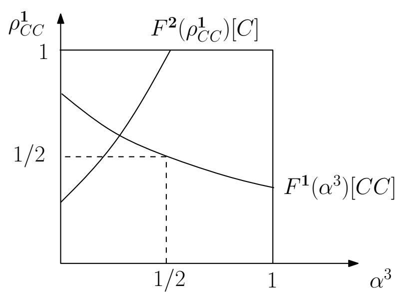
Consider what happens at $\rho_{CC} = \rho_{DD} = 1/2$ and write $f^{\mathbf{2}}_{1/2}(\sigma^{-\mathbf{2}})$ for the density of $\mathbf{2}$'s beliefs. Then by symmetry

$$f^{\mathbf{2}}_{1/2}(\sigma^{-\mathbf{2}}[CC] = s|\sigma^{-\mathbf{2}}[CC]+\sigma^{-\mathbf{2}}[DD] = S) = f^{\mathbf{2}}_{1/2}(\sigma^{-\mathbf{2}}[DD] = s|\sigma^{-\mathbf{2}}[CC]+\sigma^{-\mathbf{2}}[DD] = S)$$

so that

$$f^{\mathbf{2}}_{1/2}(\sigma^{-\mathbf{2}}[CC] = s|\sigma^{-\mathbf{2}}[CC]+\sigma^{-\mathbf{2}}[DD] = S) = f^{\mathbf{2}}_{1/2}(\sigma^{-\mathbf{2}}[CC] = S-s|\sigma^{-\mathbf{2}}[CC]+\sigma^{-\mathbf{2}}[DD] = S)$$

In other words given $\sigma^{-\mathbf{2}}[CC] + \sigma^{-\mathbf{2}}[DD] = S$ then $\sigma^{-\mathbf{2}}[CC]$ is symmetric around $S/2$, hence $\sigma^{-\mathbf{2}}[CC] > 1/2$ occurs less than $1/2$ the time so $F^{\mathbf{2}}(\rho_{CC})[C] < 1/2$. Hence the intersection of $F^{\mathbf{1}}, F^{\mathbf{2}}$ occurs for $\alpha^3 < 1/2$ and and $\rho_{CC} > 1/2$, with $\rho_{CD} = \rho_{DC} = 0$, as illustrated in the picture below:



As beliefs converge to true values the $F^{\mathbf{2}}$ function shifts to the right and the intersection occurs at $(1/2, 1/2)$.

Player 3 in Leadership and Costly Collusion Equilibrium

Player 3's incentive constraint is the same as his objective function: he has the standard best response function, if $\rho^{\mathbf{1}}_{CC} > 1/2$ he plays C, if $\rho^{\mathbf{1}}_{CC} < 1/2$ he plays D and if $\rho^{\mathbf{1}}_{CC} = 1/2$ he is indifferent. Because player 3 is the only one in his group he faces no incentive constraint and hence ν does not matter.

Costly Collusion Equilibrium

We use the high cost sequence defined in example 2 which is

$$C^k_n(\alpha^k, \rho^{-k}) = \frac{\pi_n}{1-\pi_n} \sum_{k(i)=k} G^i(\alpha^k, \rho^{-k})$$

with $\pi_n \to 1$. To pin down the group's best response correspondence note that for $\alpha^3 \leq 1/2$, it is simply CC. If the group chooses CC, the objective function takes a value of $2[6 + 4(1-\alpha^3)] - 2\frac{\pi_n}{1-\pi_n}[2 - 4(1-\alpha^3)]$. This turns out to be higher than the value of 4 achieved by playing DD if and only if $\alpha^3 < \frac{4-3\pi_n}{2}$. It turns out that no other mixed strategy profile is ever an element of the best response set. Consider any mixed strategy profile for the group. The group payoff would then be

$$\alpha^1\alpha^2 2[6+4(1-\alpha^3)] + [\alpha^1(1-\alpha^2)+\alpha^2(1-\alpha^1)]8 \\ + (1-\alpha^1)(1-\alpha^2)4 - \frac{\pi_n}{1-\pi_n}[2\alpha^1\alpha^2[2-4(1-\alpha^3)] + [\alpha^1(1-\alpha^2)+\alpha^2(1-\alpha^1)]2]$$

which can be rewritten as

$$(\alpha^1\alpha^2)\left\{2[6+4(1-\alpha^3)] - \frac{\pi_n}{1-\pi_n}2[2-4(1-\alpha^3)]\right\} \\ + \left[\alpha^1(1-\alpha^2)+\alpha^2(1-\alpha^1)\right]\left\{8-2\frac{\pi_n}{1-\pi_n}\right\} + \left[(1-\alpha^1)(1-\alpha^2)\right]4$$

For $\pi_n > 4/5$ the term $8-2\frac{\pi_n}{1-\pi_n}$must be negative. So the value to the group from such a mixed strategy profile is the convex combination of the group's value from playing CC, the negative quantity $8-2\frac{\pi_n}{1-\pi_n}$ and 4. When $\alpha^3 > \frac{4-3\pi_n}{2}$ then the group's value from playing CC is strictly less than 4. Consequently every mixed strategy profile other than DD must give a value strictly less than 4. Hence the unique group best reply is DD. When $\alpha^3 < \frac{4-3\pi_n}{2}$ then the group's value from playing CC is strictly greater than 4. So every mixed strategy profile other than CC must have a value strictly less than that from playing CC. The unique group best reply is therefore CC. Similarly when $\alpha^3 = \frac{4-3\pi_n}{2}$ CC and DD are the only elements of the group best reply correspondence.

It follows immediately that the costly collusion equilibrium consists of the group randomizing half half between CC and DD while player 3 plays $\alpha^3 = \frac{4-3\pi_n}{2}$, for all $\pi_n > 4/5$. It is easy to see how this equilibrium converges to the CCE as $\pi_n \to 1$.

Leadership Equilibrium

For $\alpha^3 < 1/2$ playing CC is incentive compatible for the group, the question is how much can they mix out of the unique bad equilibrium DD when $\alpha^3 > 1/2$ given that they are willing to forgo gains not larger than ν.

From the payoff matrix of group **1** we see that utility for player 1 is given by $u^1(\alpha^1,\alpha^2,\alpha^3) = 4\alpha^1\alpha^2(1-\alpha^3)-2\alpha^1+6\alpha^2+2$. The group utility (with weights $\beta^1=\beta^2=1$) is $v^{\mathbf{1}}(\alpha^1,\alpha^2,\alpha^3) = u^1+u^2 = 8\alpha^1\alpha^2(1-\alpha^3)+4\alpha^1+4\alpha^2+4$; notice that it is increasing in α^1 and α^2 for any α^3.

Consider the utility gained by player 1 upon deviating from $(\alpha^1,\alpha^2,\alpha^3)$ to $(0,\alpha^2,\alpha^3)$, namely $2\alpha^1[1-2\alpha^2(1-\alpha^3)]$. This is strictly positive when $\alpha^3 > 1/2$ for any positive value of α^1 and so the optimal deviation from such profiles is precisely to play D with utility

$6\alpha^2 + 2$ and utility gain $2\alpha^1[1 - 2\alpha^2(1 - \alpha^3)]$. Group **1** must play ν-incentive compatible profiles, that is profiles with gain not larger than ν.

When $\alpha^3 > 1/2$ increasing α^2 reduces the utility gain from player 1's optimal deviation and hence relaxes the incentive constraint for any ν. So in a strict ν-equilibrium we should choose $\alpha^1 = \alpha^2$ and either the constraint binds in that $2\alpha^1[1 - 2\alpha^1(1 - \alpha^3)] = \nu$ or $\alpha^1 = \alpha^2 = 1$ since group utility is increasing in both α^1 and α^2 for any α^3.

Notice that the utility gain $G(\alpha^1) = -4(\alpha^1)^2(1 - \alpha^3) + 2\alpha^1$ is quadratic concave with $G(0) = 0$, $G' = 2[1 - 4\alpha^1(1 - \alpha^3)]$ so that $G'(0) > 0$ and $G'(1) = 2[1 - 4(1 - \alpha^3)]$ meaning $G'(1) < 0$ for $\alpha^3 < 3/4$.

Since group utility increases in α^1 and α^2, if the utility gain at $\alpha^1 = \alpha^2 = 1$ that is $G(1) = 2[1 - 2(1 - \hat{\alpha}^3)]$ turns out to be less than ν group **1** plays CC and player 3 plays C - not an equilibrium. If this is greater than ν then regardless of the sign of $G'(1)$, $G(\alpha^1)$ reaches ν while increasing, and group **1** plays $\hat{\alpha}^1 = \hat{\alpha}^2$ such that $G(\hat{\alpha}^1) = \nu$ - that is, both players mix a little just until the incentive constraint is satisfied with equality. For small enough ν the solution to $G(\hat{\alpha}^1) = \nu$ must be an $\hat{\alpha}^1$ so small that $\rho^{\mathbf{1}}_{CC} < 1/2$. This in turn would make player 3 play D - again not an equilibrium.

Finally consider the case of $G(1) = \nu$ so that group **1** shadow mixes between CC and the smaller solution of $-4(\hat{\alpha}^1)^2(1 - \hat{\alpha}^3) + 2\hat{\alpha}^1 = \nu$. For this to be an equilibrium, since player 3 is mixing, player 1 must mix so that $\rho^{\mathbf{1}}_{CC} = 1/2$. Letting p be the probability of shadow mixing on CC we may compute $p + (1 - p)(\hat{\alpha}^1)^2 = \rho^{\mathbf{1}}_{CC} = 0.5$ from which we get

$$p = \frac{0.5 - (\hat{\alpha}^1)^2}{(1 - (\hat{\alpha}^1)^2)}.$$

So in this equilibrium player 3 has a greater than 50% chance of playing C and the group has a less than 50% chance of playing DD, a 50% chance of playing CC and some small chance of playing CD, DC. Here the solution for player 3 is on the opposite side of 1/2 from the belief equilibrium case.

Thus equilibrium has $G(1) = G(\hat{\alpha}^1) = \nu$ that is $2\hat{\alpha}^1[1 - 2\hat{\alpha}^1(1 - \hat{\alpha}^3)] = 2[1 - 2(1 - \hat{\alpha}^3)] = \nu$. As $\nu \to 0$ we get $\hat{\alpha}^3 \to 1/2$ and the smaller solution $\hat{\alpha}^1 \to 0$ so that in the limit the group shadow mixes half half between CC and DD.

Web Appendix 2: A Dirichlet Based Family of Random Belief Models

We show here that there are ϵ-random belief models for every positive value of ϵ. An obvious idea is to take a smooth family of probability distributions with mean equal to the truth and small variance. A good candidate for a smooth family is the Dirichlet since we can easily control the precision by increasing the "number of observations." However using an unbiased probability distribution will not work - it is ill-behaved on the boundary: if we try to keep the mean equal to the truth, then as we approach the boundary the variance has to go to zero, and on the boundary there will be a spike. A simple alternative is to bias the mean slightly towards a fixed strictly positive probability vector alpha with a small weight on that vector, and then let that weight go to zero as we take the overall variance to zero.Set $h(\epsilon) = (\epsilon/2)^3$. Fix a strictly positive probability vector over A^{-k} denoted by β^{-k} and call the *ϵ-Dirichlet belief model* the Dirichlet distribution with parameter vector (dimension cardinality of A^{-k})

$$\frac{1}{h(\epsilon)}\left[(1-\frac{\epsilon}{2\sqrt{2}})\alpha^{-k}(a^{-k})+\frac{\epsilon}{2\sqrt{2}}\beta^{-k}(a^{-k})\right]$$

Theorem 14. *The ϵ-Dirichlet belief model is an ϵ-random belief model.*

Proof. Since the parameters are away from the boundary by at least $\epsilon/2$ this has the requisite continuity property. The random variable $\tilde{\alpha}$ has mean $\overline{\alpha}^{-k} = (1-\frac{\epsilon}{2\sqrt{2}})\alpha^{-k}+\frac{\epsilon}{2\sqrt{2}}\beta^{-k}$. Since the covariances of the Dirichlet are negative, $E|\tilde{\alpha}^{-k}-\overline{\alpha}^{-k}|^2$ is bounded by the sum of the variances and we may apply Chebyshev's inequality to find

$$Pr[|\tilde{\alpha}^{-k}-\overline{\alpha}^{-k}|>\epsilon/2]\leq E|\tilde{\alpha}^{-k}-\bar{\alpha}^{-k}|^2/(\epsilon/2)^2$$

To evaluate the last expression let $\delta_\epsilon(a^{-k}) \equiv \frac{1}{h(\epsilon)}\left[(1-\frac{\epsilon}{2\sqrt{2}})\alpha^{-k}(a^{-k})+\frac{\epsilon}{2\sqrt{2}}\beta^{-k}(a^{-k})\right]$ and observe that $\sum_{a^{-k}}\delta_\epsilon(a^{-k}) = 1/h(\epsilon)$. Then by the standard Dirichlet variance formula we have

$$\begin{aligned}\frac{E|\tilde{\alpha}^{-k}-\bar{\alpha}^{-k}|^2}{(\epsilon/2)^2} &= \frac{1}{(\epsilon/2)^2}\frac{\left(\sum_{a^{-k}}\delta_\epsilon(a^{-k})\right)^2-\sum_{a^{-k}}\left(\delta_\epsilon(a^{-k})\right)^2}{\left(\sum_{a^{-k}}\delta_\epsilon(a^{-k})\right)^2\left(\sum_{a^{-k}}\delta_\epsilon(a^{-k})+1\right)}\\ &\leq \frac{1}{(\epsilon/2)^2}\frac{(1/h(\epsilon))^2}{(1/h(\epsilon))^2(1/h(\epsilon)+1)}\leq\frac{h(\epsilon)}{(\epsilon/2)^2}=\frac{\epsilon}{2}\end{aligned}$$

We also have $|\bar{\alpha}^{-k}-\alpha^{-k}| = \frac{\epsilon}{2\sqrt{2}}|\alpha^{-k}-\beta^{-k}|\leq\frac{\epsilon}{2}$; then $|\tilde{\alpha}^{-k}-\alpha^{-k}|>\epsilon$ implies $|\tilde{\alpha}^{-k}-\overline{\alpha}^{-k}|>\epsilon/2$; hence $Pr(|\tilde{\alpha}^{-k}-\alpha^{-k}|>\epsilon)\leq Pr[|\tilde{\alpha}^{-k}-\overline{\alpha}^{-k}|>\epsilon/2]\leq\epsilon/2\leq\epsilon$, which shows that this is indeed an ϵ-random belief model. □

Web Appendix 3: Analysis of the Voting Game in Section 7

We start by summarizing the results concerning the different types of equilibria and payoffs. The first table summarizes the different types of equilibria using the notation of the text. The first column is the designation of the equilibrium. The second column gives the equilibrium strategies. The final three columns give the total payoff of the group, player 3 and the sum of all the payoffs respectively. The probability of voting in the group's mixed strategy is denoted by p.

Table 1: Equilibrium Table

	Equilibrium Strategies	*Group Payoff*	*Pl. 3 Payoff*	*Total Payoff* (W)
N	$\alpha^3 = \rho_{00} = 1$	0	0	0
L	$\alpha^3 = 1, \rho_{10} + \rho_{01} = 1$	$2\tau - 1$	-2τ	-1
S	$\alpha^3 = 0, \rho_{00} = 1$	-2τ	$2\tau - 1$	-1
m_1	$\alpha^3 = \frac{1}{\tau}, \rho_{00} = \frac{1}{2\tau}, \rho_{10} + \rho_{01} = 1 - \frac{1}{2\tau}$	$3 - 2\tau - \frac{1}{2\tau}$	$1 - 2\tau$	$4 - 4\tau - \frac{1}{2\tau}$
M_1	$\alpha^3 = \frac{1}{2\tau}, \rho_{00} = \frac{1}{2\tau}, \rho_{10} + \rho_{01} = 1 - \frac{1}{2\tau}$	$1 - 2\tau$	$1 - 2\tau$	$2 - 4\tau$
m_2	$\alpha^3 = 1 - \frac{1}{2\tau}, \rho_{00} = \frac{1}{2\tau}, \rho_{11} = 1 - \frac{1}{2\tau}$	$-3 + 2\tau + \frac{1}{2\tau}$	$1 - 2\tau$	$-2 + \frac{1}{2\tau}$
M_2	$\alpha^3 = 2(1 - \frac{1}{2\tau}), \rho_{00} = \frac{1}{2\tau}, \rho_{11} = 1 - \frac{1}{2\tau}$	$2\tau - 2$	$1 - 2\tau$	-1
L_2	$\alpha^3 = 1, p = 1 - \frac{1}{\tau}$	$2\tau - 2$	$-2\tau + \frac{2}{\tau}$	$-2 + \frac{2}{\tau}$
L_3	$\alpha^3 = 0, p = \frac{1}{2\tau}$	-2τ	$2\tau - 5 + \frac{1}{\tau}$	$-5 + \frac{1}{\tau}$
M	$\alpha^3 = \frac{1}{\tau}\frac{2p\tau-1}{3p-1}, p = 1 - \frac{1}{\sqrt{2\tau}}$	$2\frac{\sqrt{2\tau}-\tau\sqrt{2\tau}-1+3\tau}{3-2\sqrt{2\tau}}$	$1 - 2\tau$	$1 - 2\tau + 2\frac{\sqrt{2\tau}-\sqrt{2}\tau^{\frac{3}{2}}-1+3\tau}{3-2\sqrt{2\tau}}$
A	$\alpha^3 = \frac{1}{\tau}, p_i = 1 - \frac{1}{2\tau}, p_j = 0, i \neq j = 1, 2$	$3 - 2\tau - \frac{1}{2\tau}$	$1 - 2\tau$	$4 - 4\tau - \frac{1}{2\tau}$

Next we give the ranges of τ for which these equilibria exist, where as in the text $\tilde{\tau} \approx 1.31$.

Table 2: Existence Table

lower τ	upper τ	CCE	Nash	FEE
0	1/2	N	N	N
1/2	3/4	m_2	S	L, M_1, M_2
3/4	1	m_2	S	L
1	$\tilde{\tau}$	m_2, m_1, L	S, L, A	L
$\tilde{\tau}$	3/2	m_2, m_1, L	S, L, M, A, L_3	L
3/2	2	L	S, L, L_3	L
2	∞	L	S, L, L_2, L_3	L

The next table contains payoffs comparisons: we compare payoffs from the point of view of the whole set of players, represented by the total payoff, and from the point of view of the large group. We use $\succ_W$ and $\succ_{\mathbf{1}}$ to denote respectively welfare and large group preference. We neglect M, L_2 and L_3 (notice that A is a special case of m_1). Then we have:

The last table contains information about the electoral outcome. We let $H = \rho_{11}(1-\alpha^3)$ denote the probability of all voting (High turnout); $D = (1-\alpha^3)(1-\rho_{00}-\rho_{11})$ the probability of deadlock; and $\Lambda = \alpha^3(1 - \rho_{00}) + (1 - \alpha^3)\rho_{11}$ the probability that large group wins. In the table the rows denote different types of equilibria and the columns provide the relevant values of H, D, Λ.

In the following: we first relate the tables to the assertions made in the text. Analysis of collusion constrained, Nash and free enforcement equilibria in the game follows. Then

Table 3: Payoffs comparisons

τ	CCE	Nash	FEE	$\succ_W$	$\succ_{\mathbf{1}}$
$1/2 < \tau < 3/4$	m_2	S	L, M_1, M_2	$M_1 \succ_W L \sim_W S \sim_W M_2 \succ_W m_2$	$L \succ_{\mathbf{1}} M_1 \succ_{\mathbf{1}} M_2 \succ_{\mathbf{1}} m_2 \succ_{\mathbf{1}} S$
$3/4 < \tau < 1$	m_2	S	L	$L \sim_W S \succ_W m_2$	$L \succ_{\mathbf{1}} m_2 \succ_{\mathbf{1}} S$
$1 < \tau \leq 9/8$	m_1, m_2, L	S, L	L	$m_1 \succ_W L \sim_W S$	$L \succ_{\mathbf{1}} m_1 \succ_{\mathbf{1}} m_2 \succ_{\mathbf{1}} S$
$9/8 \leq \tau < \tilde{\tau}$	m_1, m_2, L	S, L	L	$L \sim_W S \succ_W m_1 \succ_W m_2$	$L \succ_{\mathbf{1}} m_1 \succ_{\mathbf{1}} m_2 \succ_{\mathbf{1}} S$
$\tilde{\tau} < \tau < 3/2$	m_1, m_2, L	S, L	L	$L \sim_W S \succ_W m_2 \succ_W m_1$	$L \succ_{\mathbf{1}} m_2 \succ_{\mathbf{1}} m_1 \succ_{\mathbf{1}} S$
$3/2 < \tau < 2$	L	S, L	L	$S \sim_W L$	$L \succ_{\mathbf{1}} S$
$\tau > 2$	L	S, L	L	$S \sim_W L$	$L \succ_{\mathbf{1}} S$

Table 4: Electoral outcome probabilities

	ρ_{11}	H	D	Λ
S	0	0	0	0
L	0	0	0	1
m_1	0	0	$(1-\frac{1}{2\tau})(1-\frac{1}{\tau})$	$\frac{1}{\tau}(1-\frac{1}{2\tau})$
m_2	$1-\frac{1}{2\tau}$	$(1-\frac{1}{2\tau})\frac{1}{2\tau}$	0	$1-\frac{1}{2\tau}$
M_1	0	0	$(1-\frac{1}{2\tau})^2$	$\frac{1}{2\tau}(1-\frac{1}{2\tau})$
M_2	$1-\frac{1}{2\tau}$	$(1-\frac{1}{2\tau})(\frac{1}{\tau}-1)$	0	$1-\frac{1}{2\tau}$

we provide payoff comparisons, and lastly electoral outcome probabilities. Throughout this appendix we write ρ_{ab} for $\rho^{\mathbf{1}}[a,b]$.

Assertions in the Text

From Tables 1 and 2 the total payoff W is negative except for the non-voting equilibrium N.

From Table 3 M_1 gives welfare greater than -1 and m_1 gives welfare greater than -1 in the range $1 \leq \tau \leq 9/8$.

From Tables 1 and 2 all equilibria other than M_1, m_1 and N give welfare no more than -1.

From Tables 1 and 2 in CCE and FEE the small player always gets a negative utility.

In the range $3/4 < \tau < 1$ from Table 3 m_2 is less efficient than S or L but the large group does better than S and does worse than L.

In the range $3/4 < \tau < 1$ from Table 4 as the stakes τ increase at m_2 the probability of both members of the large group voting, the probability of everyone voting and the probability of the large group winning all increase, while from Table 1 total welfare decreases.

In the range $1/2 < \tau < 3/2$ in m_1 and m_2 from Table 1 the small group does better than at L with utility of $1-2\tau$ versus -2τ while from Table 3 the large group does worse than L.

In the range $1/2 < \tau < 3/4$ from Table 1 in m_2 and M_2 the group mixes the same way, but the third player must vote more frequently in m_2 and M_2 .

In the range $1/2 < \tau < 3/4$ for FEE we have from Table 3 that $M_1 \succ_{\mathbf{1}} M_2$ but from Table 4 gives them a smaller probability of winning Λ.

In the range $1 < \tau < \tilde{\tau}$ for CCE we have from Table 3 that $m_1 \succ_{\mathbf{1}} m_2$ but from Table 4 gives them a smaller probability of winning Λ.

Equilibria

It is convenient in the analysis of equilibria to create a single group **1** payoff matrix as a function of α^3 by averaging together the two matrices corresponding to 3 not voting and voting.

	1	0
1	$\tau - 1, \tau - 1$	$(2\alpha^3 - 1)\tau - 1, (2\alpha^3 - 1)\tau$
0	$(2\alpha^3 - 1)\tau, (2\alpha^3 - 1)\tau - 1$	$(\alpha^3 - 1)\tau, (\alpha^3 - 1)\tau$

To the matrix above we add the constant $1 + \tau(1 - 2\alpha^3)$ since this is independent of group **1** play; this gives the following payoff matrix for group **1**:

	1	0
1	$2\tau(1 - \alpha^3), 2\tau(1 - \alpha^3)$	$0, 1$
0	$1, 0$	$1 - \alpha^3\tau, 1 - \alpha^3\tau$

We also make the observation that optimality of the small group (player 3) depends only on ρ_{00} and that if $\rho_{00} < 1/(2\tau) \equiv \Upsilon$ then $\alpha^3 = 1$, if $\rho_{00} > \Upsilon$ then $\alpha^3 = 0$ and if $\rho_{00} = \Upsilon$ then player 3 is indifferent. Notice also that $\Upsilon \le 1$ if and only if $\tau \ge 1/2$. Hence if $\tau < 1/2$ then $\alpha^3 = 1$ in any equilibrium.

Collusion Constrained Equilibria

Case 1: $\tau < 1/2$. Nobody votes, equilibrium N. It is easy to check that this is the only group correlated equilibrium.

Case 2: $1/2 < \tau < 1$. There is a unique CCE where $\alpha^3 = 1 - \Upsilon$, $\rho_{00} = 1/(2\tau) = \Upsilon$ and $\rho_{11} = 1 - \Upsilon$. This is m_2. This CCE has shadow mixing. The remaining group correlated equilibria are: $\rho_{00} = \Upsilon, \rho_{11} = 1 - \rho_{00}$ and $0 < \alpha^3 < 1 - 1/(2\tau)$; and $\alpha^3 = 0, \rho_{00} \ge \Upsilon, \rho_{11} = 1 - \rho_{00}$.

Proof. If $2\tau(1 - \alpha^3) < 1$ that is $\alpha^3 > 1 - 1/2\tau$ the only equilibrium for **1** is 00 and then 3 would prefer to vote whence $\alpha^3 = 0$. It must then be $2\tau(1 - \alpha^3) \ge 1$ that is $\alpha^3 \le 1 - 1/2\tau$

in any group correlated equilibrium. In this case the group faces a coordination game with three Nash equilibria: both vote, neither votes and the symmetric mixed equilibrium.

Let p be the probability of voting in the symmetric mixed equilibrium. The indifference is $2\tau(1-\alpha^3)p = p + (1-p)(1-\alpha^3\tau)$ whence $p = (1-\alpha^3\tau)/[\tau(2-3\alpha^3)]$. This increases in α^3 from $p(0) = 1/2\tau > 1/2$ to $p(1-1/2\tau) = 1$. Since $\alpha^3 < 1$ for this to be part of an equilibrium 3 should weakly prefer voting (otherwise $\alpha^3 = 1$) and this means $-[1-(1-p)^2] + (4\tau-1)(1-p)^2 \geq 2\tau(1-p)^2$ which is equivalent to $p \leq 1 - 1/\sqrt{2\tau} < 1 - 1/2\tau < 1/2$; this is not in the range of equilibrium p's for group **1**. Hence **1** playing their mixed Nash in any group correlated equilibrium is ruled out.

Next: in any group correlated equilibrium the probability that **1** plays $(0,0)$ must be positive, otherwise 3 prefers not voting ($\alpha^3 = 1$) and **1** would play $(0,0)$ for sure. And also the probability that **1** plays $(1,1)$ must be positive, otherwise when 1 is told to vote he knows 2 is not voting and would deviate. So $\rho_{00}, \rho_{11} > 0$. For the possible values of ρ_{10} and ρ_{01} we are left to consider there are the two cases where correlated equilibrium probability is concentrated on $(1,1),(1,0),(0,0)$ or on $(1,1),(0,1),(0,0)$. They are essentially the same, we consider the first. Player 1 indifference gives $\rho_{11} \cdot 2\tau(1-\alpha^3) = \rho_{11} + \rho_{10}(1-\alpha^3\tau)$ that is $\rho_{11}[2\tau(1-\alpha^3)-1] = \rho_{10}(1-\alpha^3\tau)$ and analogously from player 2 we get $\rho_{10}[2\tau(1-\alpha^3)-1] = \rho_{00}(1-\alpha^3\tau)$; from $\rho_{11}+\rho_{10}+\rho_{00} = 1$, letting $A = [2\tau(1-\alpha^3)-1]/(1-\alpha^3\tau)$ we get in particular $\rho_{00} = A^2/(1+A+A^2)$. Again player 3 should weakly prefer voting, which in this case gives $-(\rho_{11}+\rho_{10}) + (4\tau-1)\rho_{00} \geq 2\tau\rho_{00}$ that is $\rho_{00} \geq 1/2\tau$. Thus for **1**'s CE to be part of an equilibrium it must be $2\tau \geq (1+A+A^2)/A^2$. Now the RHS decreases in A and A reaches its maximum for $\alpha^3 = 0$ where its value is $A_0 = 2\tau-1$. So $(1+A+A^2)/A^2 \geq 1 + 2\tau/(2\tau-1)^2$. But since $0 < 2\tau-1 < 1$ we have $(2\tau-1)^3 < 2\tau$ which is equivalent to $2\tau < 1+2\tau/(2\tau-1)^2$, whence $2\tau \geq (1+A+A^2)/A^2$ is false for all admissible values of A. This shows that $\rho_{01} = \rho_{10} = 0$ in any group correlated equilibrium.

Summing up, group correlated equilibria have $\alpha^3 \leq 1 - 1/2\tau$ and $\rho_{00} + \rho_{11} = 1$ with $\rho_{00}, \rho_{11} > 0$. That player 3 should weakly prefer voting gives $\rho_{00} \geq \Upsilon$, with equality if $\alpha^3 > 0$. This yields the equilibrium set in the statement.

For CCE: The threshold between dominant strategy and coordination game occurs when given that one party member votes the other is indifferent to voting: the condition is $2\tau(1-\alpha^3) = 1$ so that $\alpha^3 = 1 - 1/(2\tau)$. This is strictly positive, so $\rho_{00} = \Upsilon$. The equilibria with smaller α^3 are not CCE because collusion would lead the group to play the voting equilibrium for sure. □

Case 3: $1 < \tau < 3/2$. There are three sets of CCEs: (a) a continuum of CCEs where player 3 does not vote and the group mixes with any probability over $(1,0)$ and $(0,1)$, which is L; *(b) a CCE where* $\alpha^3 = 1 - 1/2\tau$ and the group plays $(1,1)$ with probability $1 - 1/2\tau$ and $(0,0)$ with probability $1/2\tau$, which is m_2 and (c) a CCE with $\alpha^3 = 1/\tau$ where with probability $1-1/2\tau$ the group mixes over $(1,0)$ and $(0,1)$ while with probability $1/2\tau$ they play $(0,0)$, which is m_1.

Proof. For $\alpha^3 \le 1-1/2\tau$, $(1,1)$ and $(0,0)$ are Nash equilibria along with a mixed strategy equilibrium. The highest payoff for the group comes from $(1,1)$. For $1-1/2\tau < \alpha^3 < 1/\tau$ the game becomes dominance solvable with the unique equilibrium $(0,0)$. For all higher values of α^3, the Nash equilibria are $(1,0)$ and $(0,1)$ along with the mixed equilibrium. The highest payoff for the group in this case turns out to be any of the group correlated equilibria with mixing over $(1,0)$ and $(0,1)$. For these higher values of α^3 where $1/\tau < \alpha^3$ and $1 < \tau < 3/2$ the expected payoff to each player from the mixed Nash is always strictly less than that from the group correlated equilibrium average payoff. Indeed, the inequality is $2p\tau(1-\alpha^3) < 1/2$, which since $\alpha^3 > 1/\tau > 2/3$ reads $4(\alpha^3\tau - 1)(1-\alpha^3) < 3\alpha^3 - 2$ that is $4\alpha^3\tau(1-\alpha^3) < 2-\alpha^3$; the left member is decreasing in α^3, and using this and $\tau < 3/2$ we get $4\alpha^3\tau(1-\alpha^3) < \frac{4}{3} < 2-\alpha^3$, last inequality from $\alpha^3 > 2/3$.

Thus in this case the group best response correspondence is as follows:

$$\begin{array}{rl} (1,1) & \text{if } \alpha^3 \le 1-1/2\tau \\ (0,0) & \text{if } 1-1/2\tau \le \alpha^3 \le 1/\tau \\ \textit{correlated} & \text{if } 1/\tau \le \alpha^3 \end{array}$$

So for any $1 < \tau < 3/2$, we get three sets of CCEs. (a) $\alpha^3 = 1$ and the group mixes over $(1,0)$ and $(0,1)$, (b) $\alpha^3 = 1-1/2\tau$ and the group plays $(1,1)$ with probability $1-1/2\tau$ and $(0,0)$ with probability $1/2\tau$ and (c) $\alpha^3 = 1/\tau$ and with probability $1-1/2\tau$ the group mixes over $(1,0)$ and $(0,1)$ while with probability $1/2\tau$ they play $(0,0)$, as asserted. □

Case 4: $\tau > 3/2$. There is a continuum of CCEs, where player 3 does not vote and the group mixes with any probability over $(1,0)$ and $(0,1)$.

Proof. It is seen from group **1** payoff matrix that for $\alpha^3\tau \le 1$, $(1,1)$ and $(0,0)$ are Nash equilibria along with a mixed strategy symmetric equilibrium where the probability say p that a player votes is given by

$$p = \frac{1-\alpha^3\tau}{\tau(2-3\alpha^3)}$$

The highest payoff for the group comes from $(1,1)$. For $1/\tau < \alpha^3 < 1-1/2\tau$ the game becomes dominance solvable with the unique equilibrium $(1,1)$. For $\alpha^3 = 1-1/2\tau$ there are three equilibria: $(1,1),(1,0)$ and $(0,1)$ and again the best equilibrium for the group is $(1,1)$.

For $\alpha^3 > 1 - 1/2\tau$ the equilibria are $(1,0)$ and $(0,1)$ and the mixed equilibrium as above. Turning to the group payoff, the two pure NE give the same payoff hence so does any mixture of the two; the alternative to consider is the mixed equilibrium. In the latter the expected payoff to each player (say when player 1 plays 1) is $2p\tau(1-\alpha^3)$; in the former per-player payoff is $1/2$. Recalling that in the range under consideration $\alpha^3\tau > 1$, the condition for the mixed to be better than the correlated mixtures becomes

$$\frac{2-\alpha^3}{4\alpha^3(1-\alpha^3)} \leq \tau$$

In the relevant range - $\tau > 3/2$ and $\alpha^3 > 1 - 1/2\tau$ imply $\alpha^3 > 2/3$ - the left hand side is increasing, so letting $\hat{\alpha}^3(\tau)$ solve the above with equality we get that: the mixed Nash is better for $\alpha^3 \leq \hat{\alpha}^3(\tau)$, while the mixture over the two pure Nash is better for $\alpha^3 > \hat{\alpha}^3(\tau)$. So the group best response correspondence is as follows:

$$\begin{array}{rl} (1,1) & \textit{if } \alpha^3 \leq 1 - 1/2\tau \\ \textit{mixed} & \textit{if } 1 - 1/2\tau < \alpha^3 \leq \hat{\alpha}^3(\tau) \\ \textit{correlated} & \textit{if } \alpha^3 > \hat{\alpha}^3(\tau) \end{array}$$

Now we can search for collusion constrained equilibria. Player 3's best response to the group playing $(1,1)$ is to set $\alpha^3 = 1$. So there cannot be a CCE with $\alpha^3 \leq 1 - 1/2\tau$. Since Player 3's best response to the group mixing over $(1,0)$ and $(0,1)$ is to again play $\alpha^3 = 1$, we must also rule out CCE where $\hat{\alpha}^3(\tau) < \alpha^3 < 1$. The group mixing over $(1,0)$ and $(0,1)$ with some probability and player 3 choosing $\alpha^3 = 1$, is indeed a CCE. Consider the possibility of a CCE that involves the group playing the mixed Nash equilibrium and player 3 mixing too. For Player 3 to be indifferent (in order to mix) it must be that $p = 1 - 1/\sqrt{2\tau}$. Now the equilibrium p in the mixed Nash is decreasing in α^3 over the relevant region: it takes values from 1 when $\alpha^3 = 1 - 1/2\tau$ to $(\tau-1)/\tau$ when $\alpha^3 = 1$. Since $(\tau - 1)/\tau > 1 - 1/\sqrt{2\tau}$ for $\tau > 2$, for such values of τ we cannot have such a CCE. For $3/2 < \tau \leq 2$ there does exist an α^3 that solves

$$\frac{1-\alpha^3\tau}{\tau(2-3\alpha^3)} = 1 - \frac{1}{\sqrt{2\tau}}$$

but the solution has $\alpha^3 > \hat{\alpha}^3(\tau)$ whence there is no CCE in the range $1 - 1/2\tau < \alpha^3 \leq \hat{\alpha}^3(\tau)$ either.[18] □

[18] *Proof of this:* the displayed equality can be re-written as $3\sqrt{\frac{\tau}{2}}(\alpha^3 - \frac{2}{3}) = 1 - 2\tau(1-\alpha^3)$, while $\alpha^3 \leq \hat{\alpha}^3(\tau)$ reads $\alpha^3[1 + 4\tau(1-\alpha^3)] \geq 2$. Since $\tau > 3/2$ we have $3\sqrt{\frac{\tau}{2}}(\alpha^3 - \frac{2}{3}) > \frac{3}{2}\sqrt{3}(\alpha^3 - \frac{2}{3}) = \sqrt{3}(\frac{3}{2}\alpha^3 - 1)$ so the

Nash

Recall that $\tilde{\tau} \equiv 1/(3-\sqrt{5}) \approx 1.31$. Reiterating the payoff matrix for the group for visibility:

	1	0
1	$2\tau(1-\alpha^3), 2\tau(1-\alpha^3)$	$0, 1$
0	$1, 0$	$1-\alpha^3\tau, 1-\alpha^3\tau$

Case 1: $\rho_{00} < \Upsilon$ and $\alpha^3 = 1$. The payoff matrix for the group is

	1	0
1	$0, 0$	$0, 1$
0	$1, 0$	$1-\tau, 1-\tau$

If $\tau < 1$ then it is dominant to play 0 and this is not an equilibrium. If $\tau > 1$ then there are two pure equilibria where one voter in the group votes and these imply $\rho_{00} < \Upsilon$ so this corresponds to the equilibrium L. The other equilibrium is symmetric and mixed, continuing to use p for the probability of voting, the indifference condition is $p + (1-p)(1-\tau) = 0$ or

$$p = \frac{\tau - 1}{\tau} = 1 - 2\Upsilon.$$

Here $p > 0$ requires $\Upsilon \leq 1/2$. The probability that neither player votes is $4\Upsilon^2$ which must satisfy $4\Upsilon^2 < \Upsilon$ or $\Upsilon < 1/4$. Hence we have an equilibrium of this type (it is L_2) if $1/(2\tau) < 1/4$ or $\tau > 2$. Notice that in this equilibrium the probability that the group wins $1 - 4\Upsilon^2$ is larger than $3/4$.

Case 2: $\rho_{00} > \Upsilon$ and $\alpha^3 = 0$. Recall that this requires $\tau \geq 1/2$ (otherwise $\alpha^3 = 1$). The payoff matrix for the group is

	1	0
1	$2\tau, 2\tau$	$0, 1$
0	$1, 0$	$1, 1$

This coordination game has one pure strategy equilibrium where both vote, which contradicts $\rho_{00} > \Upsilon$ and one where neither vote, corresponding to the equilibrium S which

equality implies $1 - 2\tau(1-\alpha^3) > \sqrt{3}(\frac{3}{2}\alpha^3 - 1)$ that is $2\tau(1-\alpha^3) < 1 - \sqrt{3}(\frac{3}{2}\alpha^3 - 1)$, whence

$$\alpha^3[1 + 4\tau(1-\alpha^3)] < \alpha^3[3 - 2\sqrt{3}(\frac{3}{2}\alpha^3 - 1)] = \alpha^3\sqrt{3}[\sqrt{3} - (3\alpha^3 - 2)] < \frac{2}{3}\sqrt{3}[\sqrt{3} - (3\frac{2}{3} - 2)] = 2$$

where the last inequality follows from the fact that in the relevant range $\alpha^3 \geq 2/3$ the function $\alpha^3\sqrt{3}[\sqrt{3} - (3\alpha^3 - 2)]$ is decreasing.

therefore exists for all values of $\tau \geq 1/2$. It also has a unique symmetric mixed equilibrium where the indifference condition is $p2\tau = 1$ or $p = \Upsilon$. The probability that neither vote is then $(1-\Upsilon)^2$ and the condition is $(1-\Upsilon)^2 > \Upsilon$. This is $1 - 3\Upsilon + \Upsilon^2 > 0$ which has roots at $(3 \pm \sqrt{5})/2$ and is positive only for Υ smaller than the lesser root $(3-\sqrt{5})/2 \approx 0.38$. That is to say, we have an equilibrium of this type when $\tau > 1/(3-\sqrt{5}) = \tilde{\tau}$. This is L_3

Case 3: $\rho_{00} = \Upsilon$. Indifferences give the same values of p and α^3 as in the case of $1/2 < \tau < 1$ that is

$$p = 1 - 1/\sqrt{2\tau} = 1 - \sqrt{\Upsilon} \quad \alpha^3 = \frac{1}{\tau}\frac{2p\tau - 1}{3p - 1}$$

This equilibrium - labeled M - exists for

$$\tilde{\tau} < \tau < 3/2$$

In addition, for $1 < \tau < \frac{3}{2}$ there is an asymmetric partially mixed equilibrium where one of the players in the group does not vote and the other votes with probability $1-\Upsilon$ while $\alpha^3 = 2\Upsilon$. This is equilibrium A. Notice that this is a special case of m_1.

Proof. If both group members mix we must have symmetry and this gives $(1-p)^2 = \Upsilon$, or $p = 1 - \sqrt{\Upsilon} > 0$. From the group payoff matrix we see that if $\tau < 1/2$ then 0 is strictly dominant, so this is impossible. Assume $\tau > 1/2$. For $\tau > 1/2$ the indifference condition of player 1 between voting and not when 2 votes with probability p gives $p2\tau(1-\alpha^3) = p + (1-p)(1-\alpha^3\tau)$ which yields

$$\alpha^3 = \frac{1}{\tau}\frac{2p\tau - 1}{3p - 1}$$

We then plug $p = 1 - \sqrt{\Upsilon}$ and look at the sign of numerator and denominator of this expression. The numerator is $2\tau(1 - 1/\sqrt{2\tau}) - 1 = 2\tau - \sqrt{2\tau} - 1$. This is positive if and only if $2\tau - 1 > \sqrt{2\tau}$, which since $\tau > 1/2$ is equivalent to $(2\tau-1)^2 > 2\tau$ that is $4\tau^2 - 6\tau + 1 > 0$. This has roots $(3 \pm \sqrt{5})/4$ and is negative in between. Note that the lesser root is $< 1/2$. The denominator is positive for $3(1 - 1/\sqrt{2\tau}) - 1 > 0$ that is for $\tau > 9/8$. Note that $(3+\sqrt{5})/4 = 1/(3-\sqrt{5}) > 9/8$ hence for $\tau > 1/2$ the numerator and denominator have the same sign if and only if $1/2 < \tau < 9/8$ (both negative) or $\tau > 1/(3-\sqrt{5})$ (both positive). In the latter case $\alpha^3 < 1$ requires $2p\tau - 1 < 3p - 1$ which is to say $2\tau < 3$ or $\tau < 3/2$, and in this range this equilibrium exists. In the former case $\alpha^3 \leq 1$ would require $2p\tau - 1 \geq 3p - 1$ which is true only for $\tau \geq 2$ so this range is ruled out.

Now consider the possibility of only one group member mixing. Say player 1 mixes while player 2 plays 0 with certainty. It must be that $1 - p_1 = \rho_{00} = \Upsilon$. For player 1 to be so indifferent we need $\alpha^3 = 2\Upsilon$. For player 2 to prefer not voting to voting, we need $(1-\frac{1}{2\tau})(3-2\tau) \geq 0$. Satisfying this inequality along with $\alpha^3 \leq 1$, gives the range $1 < \tau < \frac{3}{2}$. So for each $1 < \tau < \frac{3}{2}$, we get two more mixed equilibria, in each of which one group member plays 0 for sure while the other does so with probability Υ and $\alpha^3 = 2\Upsilon$. $\square$

Free Enforcement Equilibrium

Assuming uniform weights in the group utility, group **1** payoffs are $1-\alpha^3\tau$ if neither votes, $1/2$ if one votes and $2\tau(1-\alpha^3)$ if both vote. Recalling that if $\rho_{00} < 1/(2\tau) \equiv \Upsilon$ then $\alpha^3 = 1$, if $\rho_{00} > \Upsilon$ then $\alpha^3 = 0$ and if $\rho_{00} = \Upsilon$ then player 3 is indifferent, equilibrium analysis goes as follows.

Case 1: $\rho_{00} < \Upsilon$ and $\alpha^3 = 1$. Group payoffs are $1-\tau, 1/2, 0$. If $1-\tau > 1/2$ that is $\tau < 1/2$ the optimum is not to vote and this is an equilibrium, since $\Upsilon > 1$ for $\tau < 1/2$. If $\tau > 1/2$ the optimum is for exactly one to vote leading to the equilibrium L - hence this is the equilibrium for $\tau > 1/2$.

Case 2: $\rho_{00} > \Upsilon$ and $\alpha^3 = 0$. Group payoffs are $1, 1/2, 2\tau$. If $\tau > 1/2$ optimum is vote, not an equilibrium given $\rho_{00} > 0$. For $\tau < 1/2$ notice that $\alpha^3 = 0$ cannot be optimal. So, no equilibrium corresponds to this case.

Case 3: $\rho_{00} = \Upsilon$, this requires that $1-\alpha^3\tau \geq 1/2, 2\tau(1-\alpha^3)$ with at least one equality.

Case 3a: $1-\alpha^3\tau = 1/2$, $1/2 \geq 2\tau(1-\alpha^3)$. The first solves as $\alpha^3 = \Upsilon$ which we know requires $\tau \geq 1/2$. The inequality becomes $1/2 \geq 2\tau(2\tau-1)/(2\tau) = 2\tau - 1$ that is $\tau \leq 3/4$. Hence for $1/2 < \tau < 3/4$ there is an equilibrium with $\rho_{11} = 0$ and $\alpha^3 = \Upsilon$. This is M_1.

Case 3b: $2\tau(1-\alpha^3) = 1-\alpha^3\tau$, $1-\alpha^3\tau \geq 1/2$. The first one gives $\alpha^3 = 2-1/\tau = 2(1-\Upsilon)$. For Υ we need as usual $\tau \geq 1/2$. We also need $2-1/\tau \leq 1$ or $1 \leq 1/\tau$ or $\tau \leq 1$. Plugging into the inequality we get $1-(2-1/\tau)\tau \geq 1/2$ which gives $\tau \leq 3/4$. Hence if $1/2 < \tau < 3/4$ there is another equilibrium with $\rho_{11} = 1-\Upsilon$ and $\alpha^3 = 2(1-\Upsilon)$. This is M_2.

Payoff comparisons

For the welfare of all three players combined we have

$$L, S \succ_W m_2 \iff \tau > 1/2, \quad L, S \succ_W m_1 \iff \tau \gtrapprox 1.14$$
$$m_1 \succ_W m_2 \iff \tau < \tilde{\tau}, \quad M_1 \succ_W L, S \iff \tau < 3/4$$

For the large group the inequalities are as follows:

$$L \succ_{\mathbf{1}} S \iff \tau > 1/4, \quad L \succ_{\mathbf{1}} m_1 \Longleftarrow \tau > 1, \quad m_1 \succ_{\mathbf{1}} m_2 \iff 0.2 < \tau < \tilde{\tau}$$
$$M_1 \succ_{\mathbf{1}} M_2 \iff \tau < 3/4, \quad L \succ_{\mathbf{1}} M_1 \iff \tau > 1/2, \quad M_1 \succ_{\mathbf{1}} m_2 \Longrightarrow \tau \lessapprox 0.85$$
$$M_2 \succ_{\mathbf{1}} m_2 \iff \tau > 1/2, \quad m_2 \succ_{\mathbf{1}} S \Longleftarrow \tau > 1/2, \quad m_1 \succ_{\mathbf{1}} S \iff \tau > 1/6$$

Going in the order of the last display, for the three players we have:

$L, S \succ_W m_2 \iff -1 > -2 + \frac{1}{2\tau} \iff \frac{1}{2\tau} < 1 \iff \tau > 1/2$

$L, S \succ_W m_1 \iff -1 > 4 - 4\tau - \frac{1}{2\tau} \iff 8\tau^2 - 10\tau + 1 > 0 \iff .11 \precapprox \tau \precapprox 1.14$

$m_1 \succ_W m_2 \iff 4 - 4\tau - \frac{1}{2\tau} > -2 + \frac{1}{2\tau} \iff 6 - 4\tau - \frac{1}{\tau} > 0 \iff 4\tau^2 - 6\tau + 1 > 0 \iff .19 \precapprox \tau \leq \tilde{\tau}$

$M_1 \succ_W L, S \iff 2 - 4\tau > -1 \iff 3 > 4\tau \iff \tau < 3/4$

For the large group:

$L \succ_{\mathbf{1}} S \iff 2\tau - 1 > -2\tau \iff 4\tau > 1 \iff \tau > 1/4$

$L \succ_{\mathbf{1}} m_1 \iff 2\tau - 1 > 3 - 2\tau - \frac{1}{2\tau} \iff 4\tau - 4 + \frac{1}{2\tau} > 0 \iff 8\tau^2 - 8\tau + 1 > 0 \Longleftarrow \tau > 0.85$

$m_1 \succ_{\mathbf{1}} m_2 \iff 3 - 2\tau - \frac{1}{2\tau} > -3 + 2\tau + \frac{1}{2\tau} \iff 6 - 4\tau - \frac{1}{\tau} > 0 \iff 4\tau^2 - 6\tau + 1 < 0 \iff 0.2 < \tau < \tilde{\tau}$

$M_1 \succ_{\mathbf{1}} M_2 \iff 1 - 2\tau > 2\tau - 2 \iff 3 > 4\tau$

$L \succ_{\mathbf{1}} M_1 \iff 2\tau - 1 > 1 - 2\tau \iff 4\tau > 2$

$M_1 \succ_{\mathbf{1}} m_2 \iff 1 - 2\tau > -3 + 2\tau + \frac{1}{2\tau} \iff 8\tau^2 - 8\tau + 1 < 0 \iff 0.15 \precapprox \tau \precapprox 0.85$

$M_2 \succ_{\mathbf{1}} m_2 \iff 2\tau - 2 > -3 + 2\tau + \frac{1}{2\tau} \iff 1 > \frac{1}{2\tau} \iff \tau > 1/2$

$m_2 \succ_{\mathbf{1}} S \iff -3 + 2\tau + \frac{1}{2\tau} > -2\tau \iff 8\tau^2 - 6\tau + 1 > 0 \Longleftarrow \tau > 1/2$

$m_1 \succ_{\mathbf{1}} S \iff 3 - 2\tau - \frac{1}{2\tau} > -2\tau \iff 3 > \frac{1}{2\tau} \iff \tau > 1/6$

We check that it is always the case that $M \prec_W S, L$. Indeed this is equivalent to $1 - 2\tau + 2\frac{\sqrt{2\tau} - \sqrt{2}\tau^{\frac{3}{2}} - 1 + 3\tau}{3 - 2\sqrt{2\tau}} < -1$ that is $2 - 2\tau + 2\frac{\sqrt{2\tau} - \sqrt{2}\tau^{\frac{3}{2}} - 1 + 3\tau}{3 - 2\sqrt{2\tau}} < 0$. In the relevant range the denominator in the fraction is always negative so after multiplying we get $(2 - 2\tau)(3 - 2\sqrt{2\tau}) + 2(\sqrt{2\tau} - \tau\sqrt{2\tau} - 1 + 3\tau) > 0$ which simplifies to $2\sqrt{2}[\sqrt{2} - \sqrt{\tau} + \tau\sqrt{\tau}] > 0$ which is true for every $\tau > 0$.

Electoral outcome probabilities

Electoral outcome probabilities are also elementarily obtained. Recall that $H = \rho_{11}(1 - \alpha^3), D = (1 - \alpha^3)(1 - \rho_{00} - \rho_{11})$ and $\Lambda = \alpha^3(1 - \rho_{00}) + (1 - \alpha^3)\rho_{11}$; we just have to apply these formulas.

We follow the order of the table. In S we have $H = D = \Lambda = 0$. In L the only difference is $\Lambda = 1$.

In m_1 we have $\alpha^3 = \frac{1}{\tau}, \rho_{00} = \frac{1}{2\tau}, \rho_{10} + \rho_{01} = 1 - \frac{1}{2\tau}$. So $H = 0$, $D = (1 - \frac{1}{\tau})(1 - \frac{1}{2\tau})$ and $\Lambda = \frac{1}{\tau}(1 - \frac{1}{2\tau})$.

In m_2 it is $\alpha^3 = 1 - \frac{1}{2\tau}, \rho_{00} = \frac{1}{2\tau}, \rho_{11} = 1 - \frac{1}{2\tau}$ so $H = (1 - \frac{1}{2\tau})\frac{1}{2\tau}, D = 0$ and $\Lambda = (1 - \frac{1}{2\tau})^2 + (1 - \frac{1}{2\tau})\frac{1}{2\tau} = 1 - \frac{1}{2\tau}$

In M_1 we have $\alpha^3 = \frac{1}{2\tau}, \rho_{00} = \frac{1}{2\tau}, \rho_{10} + \rho_{01} = 1 - \frac{1}{2\tau}$ so $H = 0$, $D = (1 - \frac{1}{2\tau})^2$ and $\Lambda = \frac{1}{2\tau}(1 - \frac{1}{2\tau})$.

Finally, in M_2 we have $\alpha^3 = 2(1 - \frac{1}{2\tau}), \rho_{00} = \frac{1}{2\tau}, \rho_{11} = 1 - \frac{1}{2\tau}$ so $H = (1 - \frac{1}{2\tau})[1 - 2(1 - \frac{1}{2\tau})] = (1 - \frac{1}{2\tau})(\frac{1}{\tau} - 1), D = 0$, and $\Lambda = 2(1 - \frac{1}{2\tau})(1 - \frac{1}{2\tau}) + [1 - 2(1 - \frac{1}{2\tau})](1 - \frac{1}{2\tau}) = 1 - \frac{1}{2\tau}$.

For the ranges of H in m_2 and M_2 and of D in m_1 and M_1 we have:

H in m_2: up from 0 for $\tau = 1/2$ to $2/9$ for $\tau = 3/4$, still up to $1/4$ for $\tau = 1$ then down to $2/9$ again for $\tau = 3/2$.

H in M_2: up from 0 for $\tau = 1/2$ to $1/8$ for $\tau = 2/3$, then down to $1/9$ for $\tau = 3/4$

D in m_1: up from 0 for $\tau = 1$ to $2/9$ for $\tau = 3/2$

D in M_1: up from 0 for $\tau = 1/2$ to $1/9$ for $\tau = 3/4$