Theoretical Economics 13 (2018), 307–340 1555-7561/20180307

Collusion constrained equilibrium

Rohan Dutta
Department of Economics, McGill University

David K. Levine
Department of Economics, European University Institute and Department of Economics, Washington University in Saint Louis

Salvatore Modica
Department of Economics, Università di Palermo

We study collusion within groups in noncooperative games. The primitives are the preferences of the players, their assignment to nonoverlapping 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 alternatives that are suboptimal for the group 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 that shows how collusion constraints may be socially costly.

Keywords. Collusion, organization, group.

JEL classification. C72, D70.

1. Introduction

As the literature on collective action (for example, Olson 1965) emphasizes, groups often behave collusively while the preferences of individual group members limit the possi-

Rohan Dutta: rohan.dutta@mcgill.ca
David K. Levine: david@dklein.com
Salvatore Modica: salvatore.modica@unipa.it
We are especially grateful to Marco Mariotti, Andrea Mattozzi, Kirill Pogorelskiy, Joel Sobel, Chris Tyson, seminar participants at Bocconi, Glasgow, Naples, Padua, QMUL, the Games Meetings in Maastricht, and a number of anonymous referees. We are grateful to the EIEF and to the MIUR PRIN 20103S5RN3 for financial support.

Copyright © 2018 The Authors. Theoretical Economics. The Econometric Society. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at http://econtheory.org.
https://doi.org/10.3982/TE2762

ble 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 noncooperative 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 (within-group equilibria). We model collusion within that group by supposing that the group will agree to choose the within-group equilibrium that best satisfies its objectives.

The idea of choosing a best outcome for a group subject to incentive constraints has not received a great deal of theoretical attention, but is important in applications. It has been used in the study of trading economics, for example, by Hu et al. (2009). In industrial organization, Fershtman and Judd (1986) study a duopoly where owners employ managers. Kopel and Löffler (2012) use a similar setting to explore asymmetries. Balasubramanian and Bhardwaj (2004) study a duopoly where manufacturing and marketing managers bargain with each other. In other settings, 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).[1] In political economy, Levine and Modica's (2016) 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. In mechanism design, a related idea is that within a mechanism, a particular group must not wish to recontract in an incentive compatible way. A theoretical study along these lines is Myerson (1982).[2]

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. That is, if groups take the actions of other groups as given and choose the best Nash equilibrium within the group, no equilibrium may exist. Example 1 illustrates. For this reason, applied theorists have generally either avoided imposing individual incentive constraints on group actions or else invented ad hoc solutions to the existence problem.[3] We show that the existence problem is due to the discontinuity of the within-group equilibrium correspondence and show how it can be overcome

[1]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) and Martimort and Moreira (2010). 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), collusion takes place implicitly within (many) clubs, but the clubs interact in a market rather than a game environment.

[2]Myerson also observes that there is an existence problem and introduces the notion of quasi-equilibrium to which our collusion constrained equilibrium is closely connected. This link is explored in greater detail below. We emphasize that while our notion of equilibrium and existence result are similar to Myerson's, unlike Myerson, our primary focus is on examining what is captured by the notion of equilibrium and, consequently, on whether it makes sense.

[3]For example, in the collusion in auction literature as stated in Harrington (2008), it is assumed that noncolluding firms will act the same in an industry with a cartel as they would without a cartel.

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*. This is a special type of correlated equilibrium of the underlying noncooperative game.

Our key goal is to motivate our definition of collusion constrained equilibrium. We argue that it is useful because it correctly captures several different types of small influences that might not be convenient to model explicitly. Specifically we consider three perturbations of the underlying model. We first consider models in which there is slight randomness in group beliefs about the play of other groups. This provides a formal version of the informal arguments we use to motivate the definition. We then consider models in which groups may overcome incentive constraints at a substantial enforcement cost; that is, group members are allowed take suboptimal nonequilibrium actions, but the group must bear enforcement costs to induce them to do so.[4] For both of these perturbations, strict collusion constrained equilibria exist[5]—in particular randomization occurs only when there is indifference—and as the perturbation vanishes, the equilibria of the perturbed games converge to collusion constrained equilibria of the underlying game. Finally, we explore the Nash program of motivating a cooperative concept as a limit of noncooperative games. Specifically, we consider a model in which there is a noncooperative metagame played between "leaders" and "evaluators" of groups and in which leaders have a slight valence.[6] If we call the leaders *principals*, this formulation is the closest to the models used in mechanism design. In the leader/evaluator game, perfect Bayesian equilibria exist and, as the valence approaches zero, once again the equilibrium play path converges to a collusion constrained equilibria of the underlying game.

These upper hemicontinuity results with respect to the three perturbations show that the set of collusion constrained equilibria is "big enough" in the sense of containing the limits of equilibria of several interesting perturbed models. The second key question we address is whether the set of collusion constrained equilibria is "too big" in the sense that perhaps not all collusion constrained equilibria arise as such limits: indeed, we could capture all relevant limits trivially by defining everything to be an equilibrium. Could there be a stronger notion of equilibrium that still captures the relevant limits? For any particular perturbation, the answer is "yes": we show in a simple example that

[4]Collusion constrained equilibrium allows groups to put positive probability on alternatives that are suboptimal for the group in certain cases. The model of costly enforcement does not allow this, but instead allows individual group members to play suboptimal but nearly optimal strategies.

[5]There is a certain irony here: using enforcement to overcome incentive constraints is quite natural in a principal–agent setting. This result shows that even if enforcement is quite costly, the existence problem noted by Myerson (1982) in the principal–agent setting goes away. Alternatively, if enforcement is quite costly, it is natural to work with the limiting case where enforcement is not possible, and our results show that collusion constrained—or quasi—equilibrium correctly captures what happens in that case.

[6]A related class of models, for example, Hermalin (1998), Dewan and Myatt (2008), and Bolton et al. (2013), examines leadership in which a group benefits from its members coordinating their actions in the presence of imperfect information about the environment. 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. For example, Bolton et al. (2013) find that the leader should not put too much weight on the information coming from followers (what they call *resoluteness* of the leader).

limits from perturbed games lead to strict refinements, that is, subsets, of collusion constrained equilibria, albeit different refinements depending on which perturbation we consider. Is it also the case that the set of collusion constrained equilibria is too big because some collusion constrained equilibria do not arise as any limit from interesting perturbed games? In our final theoretical result, we show that this is not the case. We consider a combination of two perturbations—a belief and an enforcement cost perturbation—and, to eliminate nongeneric preferences, we also allow a perturbation 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 our key conclusion: 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 broad yet relevant 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's (2000) 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 that leads to more democratic institutions, the game typically involves a monarch and a group (the elite).[7] In our leader/evaluator 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.

We should emphasize that there is an important territory between ignoring incentive constraints entirely and requiring as we do that they always be satisfied. An important example that we study explicitly is the possibility that incentive constraints can be overcome—for example, through an enforcement mechanism—albeit at some cost. Here we can view "no incentive constraints" as "no cost of enforcement" on the one extreme and "incentive constraints must always be satisfied" as "very high cost of enforcement" on the other. One result that we establish is to give conditions on costly enforcement such that strict collusion constrained equilibria do exist. More broadly our contribution is oriented toward applications where incentive constraints cannot easily be overcome.

One branch of the game theory literature that is closely connected to the ideas we develop here is the literature that uses noncooperative methods to analyze cooperative games. There, however, the emphasis has been on the endogenous formation of coalitions, generally in the absence of incentive constraints. The Ray and Vohra (1997)

[7]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 et al. (2011).

model of coalition formation contains in it a theory of how exogenously given groups play a game among themselves. The present paper may be viewed as a simple building block for their theory of coalition formation. With exogenous groups, an equilibrium in their model requires groups to play strategy profiles that, given the behavior of the other groups, cannot be Pareto improved. Notice that there are no incentive constraints.[8] Our analysis is at a lower level: we look for a reasonable solution when a partition of players is exogenously given and groups interact with each other strategically, keeping within-group incentive compatibility constraints. There is also an extensive literature that 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 noncooperative game among groups. This is similar in spirit to Bernheim et al.'s (1987) variation on strong Nash equilibrium, which they call coalition-proof Nash equilibrium. However, the details, goals, and analysis are quite different: while we analyze deviations by a fixed group for a given partition, Bernheim et al. (1987) consider any coalitional deviation and analyze which deviations are credible.[9]

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

[8]Haeringer (2004) points out that the assumption of quasi-concave utility is insufficient in guaranteeing equilibrium existence in the Ray and Vohra (1997) setup with exogenous groups unless the groups can play within-group correlated strategies. The nonexistence problem in our setup is of an entirely different nature and, in particular, is independent of whether groups can play correlated strategies.

[9]Because coalitions are not predefined, the existence problem for coalition-proof Nash equilibrium is more akin to the problem of an empty core than to the continuity problems discussed here.

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	$8, 0, 0$	$2, 2, 0$

If player 3 plays D, 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	$8, 0, 5$	$2, 2, 5$

Let α^i denote the probability with which player i plays C. We examine the set of within-group 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 within-group 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 within-group equilibria changes discontinuously with a second pure strategy within-group equilibrium at C, C; for $\alpha^3 < 1/2$, there is an additional symmetric strictly mixed within-group 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? Let us suppose that the group objective satisfies the Pareto criterion. If $\alpha^3 > 1/2$, they have no choice: there is only one within-group equilibrium at D, D. For $\alpha^3 \leq 1/2$, they each get $6 + 4(1-\alpha^3)$ at the C, C within-group equilibrium, 2 at the D, D within-group equilibrium, and strictly less than $6 + 4(1-\alpha^3)$ at the strictly mixed within-group 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 within-group 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 within-group equilibrium given the play of player 3. $\Diamond$

In this example, the nonexistence 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 0.5, the C, C within-group equilibrium 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's (1995) 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, 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 within-group 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, players 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 nevertheless 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 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 this limit: we 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 within-group equilibrium.[10] That is, when the incentive constraint holds exactly, we do not assume that the group can choose their most preferred within-group equilibrium, but instead we assume that there is an endogenously determined probability that they will choose that within-group equilibrium. In this case,

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

optimality for player 3 requires her to be indifferent between C and D, so in the "collusion constrained" equilibrium, we propose the group will mix 50–50 between C, C and D, D; player 3 mixes 50–50 between C and D.

The import of collusion constraints can be seen by comparing what happens in this game without collusion. This game has three Nash equilibria: one at D, D, D, one in which player 3 plays D and players 1 and 2 mix 50–50 between C and D, and a fully mixed one.[11] In the first, the group members each get 2, in the second get 5, and in the third get 6.25. By contrast, in the unique collusion constrained equilibrium, the group members each get 5. Moreover, in the completely mixed Nash equilibrium player 3 gets 2.5 exactly as in the unique collusion constrained equilibrium. Why do not the group members get together and promise player 3 not to collude and instead coordinate on the completely mixed Nash equilibrium? They will be better off and player 3 is indifferent. The problem is that by saying that player 1 and 2 form a group, we mean that they cannot credibly commit not to collude. If such an agreement was reached with player 3, as soon as the meeting was over players 1 and 2 would convene a second meeting among themselves and agree that rather than mixing they will play C, C. Anticipating this, player 3 would never make the original agreement. It would be convenient for public policy if lobbying groups, such as bankers and farmers, could credibly commit not to collude among themselves. Unfortunately this is not the world we live in; hence the need to consider collusion constraints.

Remark. Discontinuity and nonexistence are not an artifact of restricting attention to within-group 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 within-group Nash equilibrium is also the unique within-group correlated equilibrium. When $\alpha^3 \leq 1/2$, the within-group correlated equilibrium set is indeed larger than the within-group Nash equilibrium set (containing at the very least the public randomizations over the within-group Nash equilibria), but these within-group 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 nevertheless badly behaved: it is discontinuous.

The 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

[11] The game is analyzed in Appendix S2 of the Supplemental Material, available in a supplementary file on the journal website, http://econtheory.org/supp/2762/supplement.pdf.

a continuous correspondence.[12] However, neither the Nash nor the 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.[13]

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 a within-group 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 $i = 1, 2, \dots, I$; 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 $k = \mathbf{1}, \mathbf{2}, \dots, \mathbf{K}$. 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 use a to denote the profile of actions over all players. Like individuals, groups have well defined objectives given by a payoff function $v^k(a^k, a^{-k})$.[14]

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 a 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. Formally if α^i denotes probability distributions over A^i, then $\alpha^k[a^k] \equiv \prod_{k(i)=k} \alpha^i[a^i]$. 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 $H(B^k)$.

[12]This is the approach used by Myerson (1982) to prove the existence of quasi-equilibrium.

[13]A specific scenario in which the discontinuity and the nonexistence problem goes away requires all players belonging to the same group to have identical ordinal preferences over all outcomes.

[14]Notice that we are not restricting the group objective function. Depending on the application, some group objectives might be more natural than others. For example, we might have $v^k(a^k, a^{-k}) = \sum_{i|k(i)=k} \beta^i u^i(a^k, a^{-k})$ 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. In the special case of a group with just one individual, such a group objective function might be especially compelling. On the other hand, considerations of fairness might cause a group of more than one individual to prefer a Pareto inferior plan.

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.[15] 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

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

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

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

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 only properties that are used in the remainder of the paper.[16] For example, we might wish to allow groups to choose among within-group correlated equilibria rather than within-group Nash equilibria. In this case, we can take $\mathcal{A}^k$ to be all correlated strategies by group k.[17] Alternatively, if we thought that a homogeneous group might be restricted to anonymous play, we can take $\mathcal{A}^k$ to be the mixed strategy of a representative individual.

We are now in a position to give a comparison of our setup with that of Myerson (1982). Myerson adds to the model a finite set of types for each player. This in itself

[15]Note that since we are dealing here with ordinary mixed strategies, there is no need to consider deviations conditional on the outcome of the individual randomizing device. See also footnote 17

[16]It would not in general 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 that 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 that has no correlating devices available except public randomization cannot achieve the usual $(1/3, 1/3, 1/3)$ correlated equilibrium 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 $(1/3, 1/3, 1/3)$. We must thus dispense with a convexity assumption on $\mathcal{A}^k$ to properly account for incentive compatibility within groups.

[17]In this case we must broaden the definition of a deviation to be a contingent deviation, $d^i : A^i \to A^i$, reflecting a choice of how to play contingent on a particular recommendation.

does not change anything: our actions can easily be the finite set of maps from types to individual decisions. However, Myerson's types are reported to a group coordinator (the principal) who can then make recommendations to individual group members. We do not allow this so that our model corresponds to Myerson's model where there is a single type of each group member. In this sense, our model is a specialization of Myerson. However, because the principal can make private recommendations, in Myerson the space $\mathcal{A}^k = R^k$ is the space of all correlated strategies for the group. As indicated, our model is consistent with this possibility but it is not our base model and we do not require this. In the setting of groups rather than agents organized by principals, it is not particularly natural to allow such a broad set of group strategies; neither is it particularly natural in the applied game settings described in the Introduction. For example, if the groups are groups of voters, it makes sense that they might coordinate their play by communicating a degree of enthusiasm for voting that day (randomizing as a group), but less sense that they would communicate their individual difficulty of voting that day to a central coordinator who would then make individualized voting recommendations.

The bottom line of this comparison is that our model in which $\mathcal{A}^k = R^k$ (and the appropriate set of deviations are used; see footnote 17) coincides with the Myerson model in which there is a single type. As we indicate below, in this case, his notion of quasi-equilibrium is exactly our definition of collusion constrained equilibrium.

3.2 *Equilibrium*

We first give a formal definition of the notion of strict collusion constrained equilibrium. As we have already shown that these may not exist, we then go on to consider collusion constrained equilibrium.

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

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

For the solutions to the maximization problem, we state the following two definitions.

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 $v^k(\alpha^k, \rho^{-k}) = V^k(\rho^{-k})$.

Note that $B^k(\rho^{-k})$ is closed.

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

As these may not exist, we now give our definition of *collusion constrained equilibrium*. We adopt the motivation given in Myerson (1982) for his notion of quasi-equilibrium.[18] Recall that in the proposed collusion constrained equilibrium of our

[18] As indicated above in the case where the two models coincide, the definition of quasi-equilibrium and collusion constrained equilibrium coincide as well.

example, the third player was randomizing 50–50 and that as a consequence it was a within-group equilibrium for the group to either both cooperate or both defect. However, cooperation is not a safe option in the sense that a small perturbation in beliefs can cause it to fail to be incentive compatible. Hence group members might be concerned that after an agreement is reached, some small change in beliefs would lead members to violate the agreement. Alternatively, defection is safe in the sense that if such an agreement is reached, no small change in beliefs would lead any group member to wish to violate the agreement. Let us first define the ϵ-worst best utility for group k for beliefs near ρ^{-k} as

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

Observe that this is nonincreasing in ϵ, so we may take the limit and define $V_S^k(\rho^{-k}) = \lim_{\epsilon\to 0} V_\epsilon^k(\rho^{-k})$ as the *group safety utility*. Our basic premise is that there would be no reason for the group to choose a plan that gives less group utility than the group safety utility. For incentive compatible plans yielding higher utility, we are agnostic: perhaps the group can reach agreement on such plans, perhaps not. This leads us to the next definition.

Definition 3. The *shadow*[19] *response set* $B_S^k(\rho^{-k})$ is the set of plans α^k that satisfy $G^k(\alpha^k, \rho^{-k}) = 0$ and $v^k(\alpha^k, \rho^{-k}) \geq V_S^k(\rho^{-k})$.

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 hemicontinuous since a sequence of incentive compatible best plans may converge to a plan that is not best. We show in the Appendix that by contrast the correspondence $B_S^k(\rho^{-k})$ must be upper hemicontinuous. The key intuition is that the group safety level $V_S^k(\rho^{-k})$ can jump down but not up so that a sequence of safe plans converges to a safe plan.[20] Because $B_S^k(\rho^{-k})$ is upper hemicontinuous, $B_S^k(\rho^{-k}) = B^k(\rho^{-k})$ 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.

Our premise is that the group will place weight only on incentive compatible plans that provide at least the group safety utility, that is, on $B_S^k(\rho^{-k})$, so we adopt the following definition.

Definition 4. The correlated profile $\rho \in R$ 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 $B^k(\rho^{-k})$. In a collusion constrained equilibrium, if $\rho^k \notin H[B^k(\rho^{-k})]$, we say

[19]The set $B_S^k(\rho^{-k})$ is a kind of shadow of nearby best within-group equilibria.

[20]Basically this solves the problem of discontinuity by forcing Reny's (1999) better reply security condition. Bich and Laraki (2017) demonstration that Reny solutions are Nash is similar to the fact here that when the shadow best response set is the same as the group best response set, then collusion constrained equilibria are strict collusion constrained equilibria.

that group k engages in *shadow mixing*. This means that the group puts positive probability on within-group equilibria in $B_S^k(\rho^{-k}) \setminus B^k(\rho^{-k})$ 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 $k(1) = k(2) = \mathbf{1}$ and $k(3) = \mathbf{2}$. In this and all subsequent use of this example, we take group utility to be defined by equal welfare weights on individual utility functions $v^{\mathbf{1}}(a^{\mathbf{1}}, a^{\mathbf{2}}) = u^1(a^{\mathbf{1}}, a^{\mathbf{2}}) + u^2(a^{\mathbf{1}}, a^{\mathbf{2}})$ and $v^{\mathbf{2}}(a^{\mathbf{1}}, a^{\mathbf{2}}) = u^3(a^{\mathbf{1}}, a^{\mathbf{2}})$.

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 within-group 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 within-group 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 is that 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 within-group equilibria C, C and D, D; that is, $B_S^{\mathbf{1}}(\rho^{\mathbf{2}}) = \{(C, C), (D, D)\}$. For the group $k = \mathbf{2}$ consisting solely of individual 3, the shadow response set is just the usual best response set.

Clearly there is no equilibrium with $\alpha^3 \neq 1/2$. Alternatively, 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 equilibrium with $\rho^{\mathbf{1}}$ a 50–50 mixture over $\{(C, C), (D, D)\}$ and $\rho^{\mathbf{2}}$ a 50–50 mixture over $\{C, D\}$. $\Diamond$

As the example shows, collusion constrained equilibrium may require that the group sometimes agree to plans that are "unsafe." Whether this makes sense is not clear: one of our main tasks in the remainder of the paper is to establish whether it does indeed make sense.

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*.[21] Formally, let $\mathscr{B}^k(\rho^{-k})$ be the set of plans α^k satisfying $G^k(\alpha^k, \rho^{-k}) = 0$. Then a group correlated equilibrium is a $\rho \in R$ such that $\rho^k \in H[\mathscr{B}^k(\rho^{-k})]$ for all k.

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

[21] These types of equilibria as well as others where groups have more sophisticated correlating devices for internal use have been investigated in the context of voting models by Pogorelskiy (2014).

The theorem is proved in the Appendix.[22] It makes clear the sense in which collusion constrained equilibria are constrained: there are many group correlated equilibria, but those that are interesting from the point of view of collusion are those in which groups are constrained to play in their shadow response sets.

4. Three model perturbations

We now study how collusion constrained equilibrium arises as a limit of equilibria in perturbed models. The key point is that equilibria in the perturbed models will be strict: groups make best choices and there is no shadow mixing. There is no issue of the group sometimes sacrificing utility for safety and sometimes not. Neither is there an issue of existence: in each case, strict equilibria are shown to exist.

We consider three different types of perturbations. First, based loosely on the earlier discussions of perturbations of beliefs and safety, we consider the possibility that group beliefs are random. Second, we 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 tat 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 lead to their successful collusion.

[22]Despite the close relationship, the existence of collusion constrained equilibrium does not follow from the existence of quasi-equilibrium in Myerson (1982); neither can we use his argument since he assumes that principals have finitely many choices, while our groups choose from a continuum.

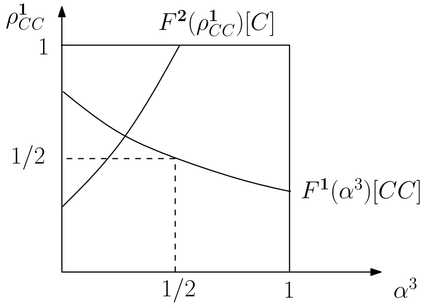

FIGURE 1. Beliefs equilibrium.

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 Appendix S1, we give, for every positive ϵ, an example based on the Dirichlet distribution of a random group belief model that is only ϵ-wrong.

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 $B^k_S(\rho^{-k}) \setminus B^k(\rho^{-k})$ 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 theorem is proved.

THEOREM 2. *For each k, n, and ϵ_n, given group decision rules b^k and random group belief models $f^k_{\epsilon_n}$ that are only ϵ_n-wrong, there are 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 Appendix S2, we analyze the random belief model corresponding to the Dirichlet belief model defined in Appendix S1. Figure 1 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. $\Diamond$

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 costs 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 nonnegative and continuous in α^k, ρ^{-k} and adopt the following definition.

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. Moreover, since nearby plans have similar gains to deviating, we assume also that the cost of deterring those deviations is similar, that is, we assume that enforcement costs are continuous. A particular example of such a cost function is $C^k(\alpha^k, \rho^{-k}) = G^k(\alpha^k, \rho^{-k})$, that is, the cost of deterring a deviation is proportional to the biggest benefit any player receives by deviating. We give below an alternative example based on a technology for monitoring deviations. Notice that we allow the possibility that incentive incompatible plans have zero cost.

With this technology, we state the following definition.

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 following usual definition of equilibrium.

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

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.

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

In the Appendix, we prove[23] the following theorem.

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

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 that generates a noisy signal of whether 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.[24]

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

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

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, $C_n^k(\alpha^k, \rho^{-k}) = (\pi_n/(1 - \pi_n)) \sum_{k(i)=k} G^i(\alpha^k, \rho^{-k})$.

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 Appendix S2, we show that the costly enforcement equilibrium of our three-player game consists of the group randomizing 50–50 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 tackle collusion constrained equilibrium from the perspective of the Nash program, showing how this partially cooperative notion arises from a limit of stan-

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

dard noncooperative games. We do so by introducing leaders. 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 do 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 and P. The parameter ν 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.[25] 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.[26] Provided P is large enough, we show that when valence tends to zero, the limits of perfect Bayesian equilibria of the leader/evaluator game are collusion constrained equilibria of the original game.

Our noncooperative game goes as follows.

Stage 1. Each leader chooses a plan $\alpha^k \in \mathcal{A}^k$ that is communicated only to members of group k: conceptually these are orders given to the members who must obey them.

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

Payoffs. 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. 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 that the other players in the group have followed orders and gets a bonus of ν if he agrees with the leader's decision.

Note that the leader and evaluator do not learn what the other groups did until the game is over. In interpreting this game it is important to realize that the actions taken by group members are those ordered by the leader: the choices they make as evaluators are simply statements of regret. So, for example, if the leader recommended a mixed strategy, mix 50–50 between C and D, the choice $d^i = 0$ is a statement by the evaluator of satisfaction with that recommendation and the choice $d^i = C$ is a statement of dissatisfaction, i.e., the evaluator regrets not having chosen C. If the plan of the leader is regretted, the evaluator then imposes a punishment on the leader.

Definition 12. We say that ρ is a *perfect Bayesian equilibrium* of the leader/evaluator game if for each leader k, there is a mixed plan μ^k over $\mathcal{A}^k$, and for each evaluator i in

[25] 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.

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

each group k and each plan α^k, there is a mixed action $\eta^i(\alpha^k)$ over $A^k \cup \{0\}$, measurable as a function of α^k, such that the following statements hold:

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

(ii) The plan μ^k (that is to say ρ^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 "not signaling what you do not know"[27] that beliefs about the play of leaders of other groups is independent of the plan chosen by the leader of the own group.[28] Note that we have not explicitly defined a system of beliefs, since the not signaling what you do not know condition makes the beliefs of evaluators over α^{-k} constant across all their information sets.

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 $\nu > 0$ 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 perfect Bayesian 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 within-group equilibrium DD when $\alpha^3 > 1/2$ given that they are willing to forgo gains no larger than ν. Appendix S2 shows that the equilibrium is $\hat{\alpha}^3 = (2+\nu)/4 >$

[27]It is known for finite games that this is an implication of sequentiality, and Fudenberg and Tirole (1991) use this condition to define perfect Bayesian 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 perfect Bayesian directly with the not signaling what you do not know condition.

[28]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.

1/2 and that the group 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. Limits of perturbations

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 and in which different perturbations pick different points out of this set.

The example is a variation of Example 1, where player 3 gets zero for sure if he plays C, and the good within-group equilibrium in the coordination game for the group that results if player 3 plays D is only weakly incentive compatible. We continue to set $v^{\mathbf{1}}(a^{\mathbf{1}}, a^{\mathbf{2}}) = u^1(a^{\mathbf{1}}, a^{\mathbf{2}}) + u^2(a^{\mathbf{1}}, a^{\mathbf{2}})$ and $v^{\mathbf{2}}(a^{\mathbf{1}}, a^{\mathbf{2}}) = u^3(a^{\mathbf{1}}, a^{\mathbf{2}})$.

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

	C	D			C	D
C	6, 6, 0	0, 8, 0		C	8, 8, 0	0, 8, 5
D	8, 0, 0	2, 2, 0		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 within-group equilibria, CC and DD, and any mixture between them is a collusion constrained equilibrium. To see this, observe that for any belief perturbation around $\alpha^3 = 0$, the worst within-group equilibrium for the group is always DD, so $V_S^{\mathbf{1}}(\alpha^3 = 0)$ is the utility the group obtains in that within-group 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 enforcement equilibrium, alternatively, the better within-group 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 (CCE) set, but the collusion constrained equilibria involving strict mixtures do not arise as a limit from any of the perturbed models. $\Diamond$

This example is nongeneric 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 is a continuum of shadow mixtures by 1 and 2 that are consistent with player 3 playing D.[29]

6. Lower hemicontinuity and a characterization of CCE

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's (1973) notion of purification of a mixed equilibrium. It is the question addressed by Fudenberg et al. (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.

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 $C^k(\alpha^k, \rho^{-k})$, and a continuous objective function $w^k(\alpha^k, \rho^{-k})$. A *perturbed equilibrium* ρ is defined by the condition $\rho^k \in H[\arg\max_{\alpha^k} w^k(\alpha^k, r_k^{-k}(\rho^{-k})) - C^k(\alpha^k, r_k^{-k}(\rho^{-k}))]$.

[29]We do not know if generic examples exist: genericity is quite difficult to analyze in this model. That our results on lower hemicontinuity in the next section make use of perturbations of the group utility function suggests that examples in which the limits fail to coincide may well be nongeneric.

The belief perturbation is a simplification of the random belief model that 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 nongeneric examples. As we are only interested in small perturbations, we state the following definition.

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, follows.

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, to coordinate their play.[30] 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. So, in addition to static Nash equilibrium, a second useful benchmark is to analyze the case in which there is free (costless) enforcement equilibrium (FEE), 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) framework (see also Levine and Mattozzi 2016). There are two parties: the "large" party has two voters, players 1 and 2; the "small" party has 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 to vote. The party that wins receives a transfer payment of $2\tau > 0$ from the losing party: if the large party wins, player 3 loses 2τ, which the large party members split; if player 3 wins, she gets τ from each member of the large party. Usually it is assumed that a tie means that each party has a 50–50 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

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

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 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 group objective of either party is to maximize the sum of payoffs of its members.

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

$$\begin{array}{ccc} & 1 & 0 \\ 1 & \tau-1, \tau-1, -2\tau & \tau-1, \tau, -2\tau \\ 0 & \tau, \tau-1, -2\tau & 0, 0, 0 \end{array}$$

This game between players 1 and 2 has a unique dominant strategy equilibrium at which neither votes if $\tau < 1$. If player 3 does vote, the payoff matrix for the actions of players 1 and 2 becomes

$$\begin{array}{ccc} & 1 & 0 \\ 1 & \tau-1, \tau-1, -2\tau-1 & -\tau-1, -\tau, -2\tau-1 \\ 0 & -\tau, -\tau-1, -2\tau-1 & -\tau, -\tau, 2\tau-1 \end{array}$$

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

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 player 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.

The full analysis of the structure of collusion constrained, Nash and free enforcement equilibria in this model can be found in Appendix S3. To appreciate the usefulness of CCE, focus on the range $3/4 < \tau < 1$. Here there is a unique Nash equilibrium S in which only the small group votes and a unique FEE L in which the small group abstains and the large group wins by casting a single vote. In this range there is also a unique CCE in which the small group mixes on voting and not voting with positive probability and the large group shadow mixes between staying out with probability $1/2\tau$ and casting two votes. In the CCE, the small group does better than FEE and worse than Nash while the large group does worse than FEE and better than Nash. The CCE more accurately captures the behavior of a collusive group as one that is in between the Nash prediction of extreme free riding and the FEE prediction of complete disregard of individual incentives. A more subtle implication relates to the equilibrium behavior of the small party. Despite consisting of a single player, the CCE aptly captures how equilibrium behavior depends significantly on whether the player faces an individual or a group, and in the latter case then whether it is collusive.

Varying τ provides a richer but similar picture. First note that among all equilibria of all types, when they are equilibria, S is always best for the small group and L is always best for the large group. Start with $\tau < 1/2$ in which case nobody votes. As we increase τ, Nash always allows S, although for $\tau > 1$ there are additional equilibria less favorable to the small player, including L. CCE and FEE both shift gradually in favor of the large group, but CCE changes more slowly than does FEE: for FEE, once $\tau > 3/4$ the unique equilibrium is L, while for CCE, this is true only for $\tau > 3/2$.

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 the existence of equilibrium in simple games. We show that the existence problem is due to a discontinuity of the equilibrium set and we propose a "fix" that 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 within-group equilibrium, the incentive constraints are satisfied with equality. This "tremble" implies that the group may put positive probability on actions that are worse for the group 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's 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 a step in that direction.

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 $\alpha_n^k \in B_n^k$,*

(i) $G^k(\alpha_n^k, \rho_n^{-k}) \le \gamma_n^k$,

(ii) $v^k(\alpha_n^k, \rho_n^{-k}) \ge V_n^k$.

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

(a) $G^k(\alpha^k, \rho^{-k}) = 0$,

(b) $v^k(\alpha^k, \rho^{-k}) \ge \liminf V_n^k$,

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 and 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, $\mathcal{A}^k \setminus B_n^k$ is open and we can extend the measure to all of $\mathcal{A}^k$ by taking $\mu_n^k[\mathcal{A}^k \setminus B_n^k] = 0$. 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 follows 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 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, $D_{v\epsilon}^k \cap B_n^k = \varnothing$, 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, $\alpha^k \in D_{0\epsilon}^k$ 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 satisfy $G^k(\alpha_n^k, \rho_n^{-k}) \leq \gamma_n^k$ 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 1 (restatement). *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 response set $H(B_S^k(\rho^{-k}))$ is upper hemicontinuous (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 2 (restatement). *For each k, n, and ϵ_n, given group decision rules b^k and random group belief models $f^k_{\epsilon_n}$ that are only ϵ_n-wrong, there are 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.*

PROOF. Remember that $\rho^k_n(a^k) = F^k(\rho^{-k})[a^k] = \int b^k(\sigma^{-k})[a^k] f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})\, d\sigma^{-k}$, where $f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})$ is continuous as a function of ρ^{-k}. So $\rho^k_n(a^k)$ 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

$$\begin{aligned}\rho^k_n &= \int b^k(\sigma^{-k}) f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})\, d\sigma^{-k} \\ &= \int_{|\sigma^{-k}-\rho^{-k}|\le\epsilon_n} b^k(\sigma^{-k}) f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})\, d\sigma^{-k} \\ &\quad + \int_{|\sigma^{-k}-\rho^{-k}|>\epsilon_n} b^k(\sigma^{-k}) f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})\, d\sigma^{-k}.\end{aligned}$$

Let $e^k_n(\rho^{-k}) \equiv \int_{|\sigma^{-k}-\rho^{-k}|\le\epsilon_n} f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})\, d\sigma^{-k}$ and

$$\overline{\rho}^k_n \equiv \int_{|\sigma^{-k}-\rho^{-k}|\le\epsilon_n} b^k(\sigma^{-k}) \frac{f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})}{e^k_n(\rho^{-k})}\, d\sigma^{-k}.$$

Then we may write

$$\rho^k_n = e^k_n(\rho^{-k})\overline{\rho}^k_n + \bigl(1-e^k_n(\rho^{-k})\bigr)\int_{|\sigma^{-k}-\rho^{-k}|>\epsilon_n} b^k(\sigma^{-k}) \frac{f^k_{\epsilon_n}(\sigma^{-k}|\rho^{-k})}{1-e^k_n(\rho^{-k})}\, d\sigma^{-k}.$$

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

$$\gamma^k_n = \max_{\alpha^k\in\mathcal{A}^k} \max_{|\sigma^{-k}-\rho^{-k}|\le\epsilon_n} \bigl|G^k(\alpha^k, \sigma^{-k}) - G^k(\alpha^k, \rho^{-k})\bigr|,$$

we see that $G^k(\alpha^k_n, \rho^{-k}_n) \le \gamma^k_n$. Since G^k is continuous on a compact set, it is uniformly continuous and so $\gamma^k_n \to 0$. Moreover, if $\alpha^k_n \in B^k_n$, then clearly $v^k(\alpha^k_n, \rho^{-k}_n) \ge V^k_{\epsilon_n}(\rho^{-k}_n)$. The result now follows from Corollary 1. □

Leadership equilibrium

For $\nu > 0$, define $V^k_\nu(\rho^{-k}) = \sup_{\alpha^k\in\mathcal{A}^k | G^k(\alpha^k,\rho^{-k})<\nu} v^k(\alpha^k, \rho^{-k})$ and $B^k_\nu(\rho^{-k})$ to be the set of plans α^k satisfying $G^k(\alpha^k, \rho^{-k}) \le \nu$ and $v^k(\alpha^k, \rho^{-k}) \ge V^k_\nu(\rho^{-k})$.

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 6. *Strict ν equilibria* exist.

Proof. It is sufficient to show that B_ν^k is UHC. By Theorem 17.35 in Aliprantis and Border (2006), we then know that $H[B_\nu^k(\rho^{-k})]$ is also UHC. The 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, $G^k(\alpha_n^k, \rho_n^{-k}) \le \nu$ for all n implies that $G^k(\alpha^k, \rho^{-k}) \le \nu$. Suppose by contradiction, $v^k(\alpha^k, \rho^{-k}) < V_\nu^k(\rho^{-k})$. By the continuity of v^k, it follows that for sufficiently large n, we have $v^k(\alpha_n^k, \rho_n^{-k}) < V_\nu^k(\rho^{-k})$. Since $v^k(\alpha_n^k, \rho_n^{-k}) \ge V_\nu^k(\rho_n^{-k})$, this implies $V_\nu^k(\rho_n^{-k}) < V_\nu^k(\rho^{-k})$. Hence there is some $\hat{\alpha}^k$ such that $G^k(\hat{\alpha}^k, \rho^{-k}) < \nu$ and $V_\nu^k(\rho_n^{-k}) < v^k(\hat{\alpha}^k, \rho^{-k})$. By continuity of G^k and v^k, this in turn implies that for sufficiently large n, we have $G^k(\hat{\alpha}^k, \rho_n^{-k}) < \nu$ and $V_\nu^k(\rho_n^{-k}) < v^k(\hat{\alpha}^k, \rho_n^{-k})$, contradicting the definition of $V_\nu^k(\rho_n^{-k})$. □

Theorem 7. *The variable ρ is a perfect Bayesian equilibrium of the leader/evaluator game if and only if it is a strict ν equilibrium.*

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

Let D_ν^k be the subset of $\mathcal{A}^k$ for which $G^k(\alpha^k, \rho^{-k}) > \nu$. For $\alpha^k \in D_\nu^k$ 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 $d = \mu^k[D_\nu^k] > 0$. 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_v^k 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 $\tilde{D}_\nu^k$ be the subset of $\mathcal{A}^k$ for which $v^k(\alpha^k, \rho^{-k}) < V_\nu^k(\rho^{-k}) - \epsilon$. Suppose $\tilde{d} = \mu^k[\tilde{D}_\nu^k] > 0$. Let $\tilde{\alpha}^k \in \mathcal{A}^k$ satisfy $G^k(\tilde{\alpha}^k, \rho^{-k}) < \nu$ and $v^k(\tilde{\alpha}^k, \rho^{-k}) > V_\nu^k(\rho^{-k}) - \epsilon/2$, which we know exists. By evaluator optimality, we have $\eta^i(\tilde{\alpha}^k)[\hat{\alpha}^k] = 1$ for all $k(i) = k$. Consider $\tilde{\mu}^k$ that takes the weight from $\tilde{D}_v^k$ and puts it on $\tilde{\alpha}^k$. The utility from $\tilde{\mu}^k$ is at least $U^k + d\epsilon/2$, so $\tilde{d} = 0$. Since $B_\nu^k(\rho^{-k}) \subseteq D_\nu^k \cup \tilde{D}_\nu^k$, we see that that indeed $\mu^k[B_\nu^k(\rho^{-k})] = 1$.

Now suppose that ρ is a strict ν equilibrium. Since $\rho^k \in H[B_\nu^k(\rho^{-k})]$, there exist measures μ^k with $\mu^k[B_\nu^k(\rho^{-k})] = 1$ and $\rho^k = \int \sigma\mu^k(d\sigma)$ so it suffices to find η^i that together with μ^k forms a perfect Bayesian equilibrium. Specifically for the given ρ^k, we show how to choose evaluator optimal responses η^i to each α^k such that the given μ^k are optimal for the leader with respect to those evaluations. Start by choosing an optimal default for the evaluators, i.e., a function $\bar{\alpha}^i(\alpha^k) \in \arg\max_{\alpha^i} u^i(\alpha^i, \alpha^k, \rho^{-k})$ that is measurable. We define evaluator optimal responses η^i to α^k by the leader in three cases depending on the size $G^k(\alpha^k, \rho^k)$ and $v^k(\alpha^k, \rho^{-k})$. Observe that it cannot be that $G^k(\alpha^k, \rho^{-k}) < \nu$ and $v^k(\alpha^k, \rho^{-k}) > V_\nu^k(\rho^{-k})$, so we may omit consideration of that case.

(i) If $G^k(\alpha^k, \rho^{-k}) > \nu$, then $\eta^i[\bar{\alpha}^i(\alpha^k)] = 1$. Note that in this case $\bar{\alpha}^i(\alpha^k) \neq \alpha^i$ 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 $\bar{\bar{\alpha}}^j \neq \alpha^j$ (and this evaluator can be chosen in a measurable way). For $i \neq j$, take $\eta^i[\alpha^i] = 1$. For j, choose $\eta^j[\bar{\bar{\alpha}}^j] = (v^k(\alpha^k, \rho^{-k}) - V^k_\nu(\rho^{-k}))/P$ and $\eta^j[\alpha^j] = 1 - \eta^j[\bar{\bar{\alpha}}^j]$.

We now establish that for this evaluator, optimal response μ^k is indeed optimal for the leader. 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 $\alpha^k \in B^k_\nu(\rho^{-k})$ is at least as good as any other choice and is indifferent to any other choice in $B^k_\nu(\rho^{-k})$. Since μ^k is a randomization over $B^k_\nu(\rho^{-k})$ for the leader, it follows that it is optimal. □

Lemma 2. *We have $V^k_\nu(\rho^{-k}) \geq V^k_\epsilon(\rho^{-k})$ 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 8. *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 $\epsilon_n \to 0$. The result now follows from Corollary 1. □

Perturbed equilibrium: Existence and upper hemicontinuity

Theorem 9. *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 that the correspondence $\arg\max_{\alpha^k} w^k(\alpha^k, r^{-k}_k(\rho^{-k})) - C^k(\alpha^k, r^{-k}_k(\rho^{-k}))$ is UHC. In turn, by Theorem 17.35 in Aliprantis and Border (2006), $H[\arg\max_{\alpha^k} w^k(\alpha^k, r^{-k}_k(\rho^{-k})) - C^k(\alpha^k, r^{-k}_k(\rho^{-k}))]$ is UHC. The existence of perturbed equilibria then follows from the Kakutani fixed point theorem. □

Theorem 10. *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

$\max_{\alpha^k,\rho^{-k}} |w_n^k(\alpha^k,\rho^{-k}) - v^k(\alpha^k,\rho^{-k})| \to 0$, each with a perturbed equilibrium ρ_n that converges to ρ.

Let $B_{wcn}^k = \arg\max_{\alpha^k} w_n^k(\alpha^k, r_{kn}^{-k}(\rho^{-k})) - C_n^k(\alpha^k, r_{kn}^{-k}(\rho^{-k}))$. Let $\tilde{v} = \max v^k - \min v^k$. Let $\delta_{n1} = \max_{\alpha^k,\rho^{-k}} |w_n^k(\alpha^k, r_{kn}^{-k}(\rho^{-k})) - w_n^k(\alpha^k,\rho^{-k})|$ and $\delta_{n2} = \max_{\alpha^k,\rho^{-k}} |w_n^k(\alpha^k,\rho^{-k}) - v^k(\alpha^k,\rho^{-k})|$. Since C_n^k is a high cost sequence, for all large enough n, $G^k(\alpha^k,\rho^{-k}) > \gamma_n^k$ implies $C_n^k(\alpha^k,\rho^{-k}) > 2(\tilde{v} + \delta_{n1} + \delta_{n2})$ and, since $\max_{\rho^{-k}} |r_{kn}^{-k}(\rho^{-k}) - \rho^{-k}| \to 0$, also $C_n^k(\alpha^k, r_{kn}^{-k}(\rho^{-k})) > \tilde{v} + \delta_{n1} + \delta_{n2}$. So for all sufficiently large n, $\alpha^k \in B_{wcn}^k$ means $G^k(\alpha^k,\rho^{-k}) \le \gamma_n^k$.

Let $W_n^k(\rho^{-k}) = \max_{\alpha^k \in \mathcal{A}^k | G^k(\alpha^k, r_{kn}^{-k}(\rho^{-k}))=0} w_n^k(\alpha^k, r_{kn}^{-k}(\rho^{-k}))$. Suppose $\alpha_n^k \in B_{wcn}^k$. Then for large enough n, it must be that

$$w_n^k\left(\alpha_n^k, r_{kn}^{-k}\left(\rho^{-k}\right)\right) \ge W_n^k\left(\rho^{-k}\right) \ge V_S^k\left(\rho^{-k}\right) - \delta_{n1} - \delta_{n2}.$$

This in turn means

$$v^k\left(\alpha_n^k, \rho^{-k}\right) \ge W_n^k\left(\rho^{-k}\right) - \delta_{n1} - \delta_{n2} \ge V_S^k\left(\rho^{-k}\right) - 2\delta_{n1} - 2\delta_{n2}.$$

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 $v^k(\alpha^k,\rho^{-k}) \ge \liminf V_n^k$. Finally note that

$$\begin{aligned}
&\liminf W_n^k\left(\rho^{-k}\right) - \delta_{n1} - \delta_{n2} \ge \liminf V_S^k\left(\rho^{-k}\right) - 2\delta_{n1} - 2\delta_{n2} \\
&\quad\Rightarrow \quad \liminf W_n^k\left(\rho^{-k}\right) \ge V_S^k\left(\rho^{-k}\right).
\end{aligned}$$

Therefore, ρ is a collusion constrained equilibrium. □

Perturbed equilibrium: Lower hemicontinuity

Theorem 11. *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 does something stronger: the idea is to construct a series of perturbations with perturbed equilibria $\rho_n = \rho$ that 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 bests 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

$\overline{G}^k_n$ and $\overline{V}^k_n$ 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 $\overline{v}^k_n = \overline{v}^k(\sigma^{-k}_{kn})$; observe that $\overline{v}^k_n \to 0$. Take $\lambda^k_n = 1/\sqrt{\overline{G}^k_n}$, which goes to infinity, $\overline{\kappa}^k_n = 3(\overline{v}^k_n + \overline{V}^k_n + \lambda^k_n \overline{G}^k_n)$, which goes to zero, and $\overline{\gamma}^k_n = 1/\sqrt{\lambda^k_n}$, which goes to zero.

The functions $\overline{w}^k_n(\alpha^k, \sigma^{-k})$ *and* $\overline{C}^k_n(\alpha^k, \sigma^{-k})$. Define first $D^k_n(\alpha^k) = \max\{0, v^k(\alpha^k, \rho^{-k}) - V^k_S(\rho^{-k})\} + \lambda^k_n G(\alpha^k, \rho^{-k})$ and $d^k_n(\alpha^k) = \min\{D^k_n(\alpha^k), \overline{\kappa}^k_n\}$. This converges uniformly to zero. We then take $\overline{C}^k_n(\alpha^k, \sigma^{-k}) = D^k_n(\alpha^k) - d^k_n(\alpha^k)$ and $\overline{w}^k_n(\alpha^k, \sigma^{-k}) = v^k(\alpha^k, \rho^{-k}) - d^k_n(\alpha^k)$. Observe that

$$\begin{aligned}
&\overline{w}^k_n(\alpha^k, \sigma^{-k}) - \overline{C}^k_n(\alpha^k, \sigma^{-k}) \\
&\quad = v^k(\alpha^k, \rho^{-k}) - D^k_n(\alpha^k) \\
&\quad = v^k(\alpha^k, \rho^{-k}) - \max\{0, v^k(\alpha^k, \rho^{-k}) - V^k_S(\rho^{-k})\} - \lambda^k_n G(\alpha^k, \rho^{-k}) \\
&\quad = \min\{v^k(\alpha^k, \rho^{-k}), V^k_S(\rho^{-k})\} - \lambda^k_n G(\alpha^k, \rho^{-k}).
\end{aligned}$$

Key fact: $\arg\max_{\alpha^k} \overline{w}^k_n(\alpha^k, \sigma^{-k}) - \overline{C}^k_n(\alpha^k, \sigma^{-k}) = B^k_S(\rho^{-k})$. To see this consider the maximizers of $\min\{v^k(\alpha^k, \rho^{-k}), V^k_S(\rho^{-k})\} - \lambda^k_n G(\alpha^k, \rho^{-k})$. For the elements of $B^k_S(\rho^{-k})$, that is, the α^k for which $G(\alpha^k, \rho^{-k}) = 0$ and $v^k(\alpha^k, \rho^{-k}) \geq V^k_S(\rho^{-k})$, the expression equals $V^k_S(\rho^{-k})$. Outside $B^k_S(\rho^{-k})$, 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* $\epsilon^k_n > 0$ *such that* $|\sigma^{-k} - \sigma^{-k}_{kn}| \leq \epsilon^k_n$ *implies the following statements*:

(i) If $G^k(\alpha^k, \sigma^{-k}) > \overline{\gamma}^k_n$, then $\overline{C}^k_n(\alpha^k, \sigma^{-k}) \geq \lambda^k_n \overline{\gamma}^k_n - \kappa^k_n - 2\lambda^k_n \overline{G}^k_n \to \infty$.

(ii) If $G^k(\alpha^k, \sigma^{-k}) = 0$, then $\overline{C}^k_n(\alpha^k, \sigma^{-k}) = 0$.

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

Proofs of the implications.

(i) We have $\overline{C}^k_n(\alpha^k, \sigma^{-k}) \geq \lambda^k_n G(\alpha^k, \rho^{-k}) - \overline{\kappa}^k_n \geq \lambda^k_n G(\alpha^k, \sigma^{-k}) - \overline{\kappa}^k_n - \lambda^k_n \overline{G}^k(\sigma^{-k})$, so choose ϵ^k_n small enough that $\overline{G}^k(\sigma^{-k}) \leq 2\overline{G}^k_n$.

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

Since V^k_S cannot jump up, we may choose $\epsilon^k_n > 0$ such that for all $|\sigma^{-k} - \sigma^{-k}_{kn}| < \epsilon^k_n$, we have $V^k_S(\sigma^{-k}) \leq V^k_S(\sigma^{-k}_{kn}) + \overline{v}^k_n$. Note that $V^k_S(\sigma^{-k}_{kn}) \leq V^k_S(\rho^{-k}) + \overline{V}^k_n$. Hence $V^k_S(\sigma^{-k}) \leq V^k_S(\rho^{-k}) + \overline{v}^k_n + \overline{V}^k_n$. Therefore, $G^k(\alpha^k, \sigma^{-k}) = 0$ implies $v^k(\alpha^k, \sigma^{-k}) \leq V^k_S(\rho^{-k}) + \overline{v}^k_n + \overline{V}^k_n$.

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

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

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

Step 2. We now have $\overline{w}_n^k(\alpha^k, \sigma^{-k})$ and $\overline{C}_n^k(\alpha^k, \sigma^{-k})$ that are defined in an 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 $w_n^k(\alpha^k, \sigma^{-k})$, $C_n^k(\alpha^k, \sigma^{-k})$, and $r_{kn}^{-k}(\sigma^{-k})$ 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 nonnegative 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

$$\begin{aligned}w_n^k(\alpha^k, \sigma^{-k}) &= \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|)\overline{w}_n^k(\alpha^k, \sigma^{-k}) + (1 - \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|))v^k(\alpha^k, \sigma^{-k}), \\ C_n^k(\alpha^k, \sigma^{-k}) &= \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|)\overline{C}_n^k(\alpha^k, \sigma^{-k}) + (1 - \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|))\lambda_n^k G^k(\alpha^k, \sigma^{-k}), \\ r_n^k(\sigma^{-k}) &= \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|)\overline{r}_n^k(\sigma^{-k}) + (1 - \beta_n^k(|\sigma^{-k} - \sigma_{kn}^{-k}|))\sigma^{-k}.\end{aligned}$$

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. □

References

Acemoglu, Daron (2001), "Inefficient redistribution." *American Political Science Review*, 95, 649–661. [330]

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. [310]

Aliprantis, Charalambos D. and Kim C. Border (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third edition. Springer-Verlag, Berlin, Heidelberg. [333, 334]

Arrow, Kenneth J. and Gerard Debreu (1954), "Existence of an equilibrium for a competitive economy." *Econometrica*, 22, 265–290. [314]

Balasubramanian, Sridhar and Pradeep Bhardwaj (2004), "When not all conflict is bad: Manufacturing-marketing conflict and strategic incentive design." *Management Science*, 50, 489–502. [308]

Bernheim, B. Douglas, Bezalel Peleg, and Michael D. Whinston (1987), "Coalition-proof Nash equilibria. I. Concepts." *Journal of Economic Theory*, 42, 1–12. [311]

Bich, Philippe and Rida Laraki (2017), "On the existence of approximate equilibria and sharing rule solutions in discontinuous games." *Theoretical Economics*, 12, 79–108. [318]

Bolton, Patrick, Markus K. Brunnermeier, and Laura Veldkamp (2013), "Leadership, coordination and corporate culture." *Review of Economic Studies*, 80, 512–537. [309]

Caillaud, Bernard and Philippe Jéhiel (1998), "Collusion in auctions with externalities." *RAND Journal of Economics*, 29, 680–702. [308]

Celik, Gorkem (2009), "Mechanism design with collusive supervision." *Journal of Economic Theory*, 144, 69–95. [308]

Che, Yeon-Koo and Jinwoo Kim (2009), "Optimal collusion-proof auctions." *Journal of Economic Theory*, 144, 565–603. [308]

Cole, Harold L. and Edward C. Prescott (1997), "Valuation equilibrium with clubs." *Journal of Economic Theory*, 74, 19–39. [308]

Dewan, Torun and David P. Myatt (2008), "The qualities of leadership: Direction, communication, and obfuscation." *American Political Science Review*, 102, 351–368. [309]

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. [310]

Ellickson, Bryan, Birgit Grodal, Suzanne Scotchmer, and William R. Zame (2001), "Clubs and the market: Large finite economies." *Journal of Economic Theory*, 101, 40–77. [308]

Fershtman, Chaim and Kenneth L. Judd (1986), *Strategic Incentive in Manipulation in Rivalrous Agency*, volume 496 of Technical Report Series: The Economics Series. Institute for Mathematical Studies the Social Sciences, Stanford University. [308]

Fudenberg, Drew, David M. Kreps, and David K. Levine (1988), "On the robustness of equilibrium refinements." *Journal of Economic Theory*, 44, 354–380. [327]

Fudenberg, Drew and Jean Tirole (1991), "Perfect Bayesian equilibrium and sequential equilibrium." *Journal of Economic Theory*, 53, 236–260. [325]

Haeringer, Guillaume (2004), "Equilibrium binding agreements: A comment." *Journal of Economic Theory*, 117, 140–143. [311]

Harrington, Joseph (2008) "Detecting cartels." In *Handbook of Antitrust Economics* (P. Buccirossi, ed.), 213–258. MIT Press. [308]

Harsanyi, John C. (1973), "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points." *International Journal of Game Theory*, 2, 1–23. [313, 327]

Hermalin, Benjamin E. (1998), “Toward an economic theory of leadership: Leading by example.” *American Economic Review*, 88, 1188–1206. [309]

Hoffman, Philip T. and Jean-Laurent Rosenthal (2000), “Divided we fall: The political economy of warfare and taxation.” Report, California Institute of Technology. [310]

Hu, Tai-wei, John Kennan, and Neil Wallace (2009), “Coalition-proof trade and the Friedman rule in the Lagos-Wright model.” *Journal of Political Economy*, 117, 116–137. [308]

Kopel, Michael and Clemens Löffler (2012), “Organizational governance, leadership, and the influence of competition.” *Journal of Institutional and Theoretical Economics*, 168, 362–392. [308]

Laffont, Jean-Jacques and David Martimort (1997), “Collusion under asymmetric information.” *Econometrica*, 65, 875–911. [308]

Levine, David and Andrea Mattozzi (2016), “Voter participation with collusive parties.” Working paper, Ademu Working Paper Series WP 2016/021. [308, 311, 322, 328]

Levine, David K. and Salvatore Modica (2016), “Peer discipline and incentives within groups.” *Journal of Economic Behavior and Organization*, 123, 19–30. [308, 322, 323]

Martimort, David and Humberto Moreira (2010), “Common agency and public good provision under asymmetric information.” *Theoretical Economics*, 5, 159–213. [308]

McAfee, R. Preston and John McMillan (1992), “Bidding rings.” *American Economic Review*, 82, 579–599. [308]

McKelvey, Richard D. and Thomas R. Palfrey (1995), “Quantal response equilibria for normal form games.” *Games and Economic Behavior*, 10, 6–38. [313]

Myerson, Roger B. (1982), “Optimal coordination mechanisms in generalized principal-agent problems.” *Journal of Mathematical Economics*, 10, 67–81. [308, 309, 315, 316, 317, 320]

Olson, Mancur (1965), *The Logic of Collective Action: Public Goods and the Theory of Groups*, volume 124 of Harvard Economic Studies. Harvard University Press. [307, 330]

Palfrey, Thomas R. and Howard Rosenthal (1985), “Voter participation and strategic uncertainty.” *American Political Science Review*, 79, 62–78. [311, 328]

Pogorelskiy, Kirill (2014), “Correlated equilibria in voter turnout games.” Unpublished paper, APSA 2014 Annual Meeting Paper SSRN 2451740. [319]

Ray, Debraj and Rajiv Vohra (1997), “Equilibrium binding agreements.” *Journal of Economic Theory*, 73, 30–78. [310, 311]

Reny, Philip J. (1999), “On the existence of pure and mixed strategy Nash equilibria in discontinuous games.” *Econometrica*, 67, 1029–1056. [318]

Simon, Leo K. and William R. Zame (1990), “Discontinuous games and endogenous sharing rules.” *Econometrica*, 58, 861–872. [313]

Tirole, Jean (1986), "Hierarchies and bureaucracies: On the role of collusion in organizations." *Journal of Law, Economics and Organization*, 2, 181–214. [308]

Tirole, Jean (1992), "Collusion and the theory of organizations." In *Advances in Economic Theory: Sixth World Congress, Vol. II* (J. J. Laffont, ed.). Cambridge University Press, Cambridge. [308]

Co-editor George J. Mailath handled this manuscript.

Manuscript received 12 January, 2017; final version accepted 11 May, 2017; available online 12 May, 2017.