Collusion, Randomization and Leadership in Groups☆

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

Abstract

We propose a theory of collusive groups in the context of finite non-cooperative games. We first consider a simple setting in which players are exogenously partitioned into groups within which players are symmetric. Given the play of the other groups there may be several symmetric equilibria for a particular group. We develop the idea that if group can collude they will agree to choose the equilibrium most favorable for its members. We then consider an alternative model of a strictly non-cooperative meta-game played between "leaders" of groups. We establish equivalence between equilibria of the collusive group game and the leadership game - this also establishes existence for the collusive group game. Finally, we extend the leadership model to games where players within groups are not necessarily symmetric and groups are endogenously formed. In this model leaders bid for followers by making credible offers of what the followers will receive if they follow the instructions of the leader. We illustrate the model with a simple application in which there are three potential groups. There are two symmetric groups that play a prisoners' dilemma game with each other and there can also be a "group of the whole." We show that the equilibrium necessarily involves randomization and that both cooperation and defection take place in equilibrium. We also show that the frequency of cooperation increases as the benefit to cooperation increases and the benefit of deviating decreases.

JEL Classification Numbers: later

Keywords: Folk Theorem, Anti-Folk Theorem, Organization, Group

☆First Version: June 2, 2013. We are especially grateful to Michele Boldrin, Michael Chwe, Drew Fudenberg, and Zacharias Maniadis. We are grateful to the EIEF, to NSF Grant SES-08-51315 and to the MIUR for financial support.

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

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

[1]Department of Economics, McGill University

[2]Department of Economics, WUSTL

[3]Università di Palermo

1. Introduction

We propose a theory of collusive groups in the context of finite non-cooperative games. We first consider a simple setting in which players are exogenously partitioned into groups within which players are symmetric. Given the play of the other groups there may be several symmetric equilibria for a particular group. We develop the idea that if group can collude they will agree to choose the equilibrium most favorable for its members. This leads to an existence problem which we illustrate with an example and propose to overcome through a type of randomization, leading to what we call *collusion constrained equilibrium*. We then consider an alternative model of a strictly non-cooperative meta-game played between "leaders" of groups. We establish equivalence between equilibria of the collusive group game and the leadership game - this also establishes existence for the collusive group game. Finally, we extend the leadership model to games where players within groups are not necessarily symmetric and groups are endogenously formed. In this model leaders bid for followers by making credible offers of what the followers will receive if they follow the instructions of the leader. We illustrate the model with a simple application in which there are three potential groups. There are two symmetric groups that play a prisoners' dilemma game with each other and there can also be a "group of the whole." Here the dilemma is that if the group of the whole forms then the leaders of the individual groups can credibly offer more while if the two groups form independently the leader of the group of the whole can credibly offer more. We show that the equilibrium necessarily involves randomization and that both cooperation and defection take place in equilibrium. We also show that the frequency of cooperation increases as the benefit to cooperation increases and the benefit of deviating decreases.

The branch of the cooperative game theory literature that is most closely connected to the ideas we propose to develop here is the literature that uses non-cooperative methods to analyze cooperative games and in particular the endogenous formation of coalitions. One example is Ray and Vohra [74] who introduce a game in which players bargain over the formation of coalitions by making proposals to coalitions and accepting or rejecting those proposals within coalitions. This literature generally describes the game by means of a characteristic function and involves proposals and bargaining. Although our model of endogenous group formation also involves an element of bidding, we work in a framework of implicit or explicit coordination among group members in a non-cooperative game. This is similar in spirit to Bernheim, Peleg and Whinston [14]'s variation on strong Nash equilibrium, that they call coalition-proof Nash equilibrium, although the details of our model are rather different.

There is a long literature on collusion in mechanism design, and our model builds on those ideas. With a few exceptions the general idea is that within a mechanism a particular group - the bidders in an auction, the supervisor and agent in the Principal/Supervisor/Agent model, for example - must not wish to recontract in an incentive compatible way. In the case of the hierarchical models, the Principal/Agent/Supervisor model of collusion originates with Tirole [80] and the more general literature on hierarchical models is discussed in his survey Tirole [81]. For a recent contribution and an indication of the current state of the literature, see Celik [31]. In the auction literature, we have the papers of McAfee and McMillan [67] and Caillaud and Jéhiel [30] among many others. The theory has been pursued for other types of mechanisms, as in Laffont and Martimort [59]. In most of this work there is only one group recontracting, so the issue of a "game" among the groups does not arise. Our setting involves multiple groups on an equal footing. The closest model we know of is that of Che and Kim [33] in the auction setting - they allow multiple groups they refer to as cartels to recontract in an incentive compatible way among themselves. However, it does not appear that strictly speaking these cartels play a game.

Two other literatures are relevant as well. In applied work - for example by economic historians - the issue of how groups behave is usually dodged by examining a game in which an entire group is treated as a single individual. This is the case in the current literature on the role of taxation by the monarchy in bringing about more democratic institutions. Hoffman and Rosenthal [56] explicitly assume that the monarch and the elite act as single agents, and this assumption seems to be accepted by later writers such as Dincecco, Federico and Vindigni [39]. As the literature on collusion in mechanism design makes clear, by treating a group as an individual we ignore the fact that the group itself is subject to incentive constraints. Individuals wish other individuals to act in the group interest, but may not wish to do so themselves. That issue has been discussed as well in the literature on collective action (for example Olson [71]), but that literature has not provided a general framework for analysis, proposing instead particular solutions such as tying arrangements or other commitments to overcome incentive constraints.

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. If - given the play of other groups - there is more than one in-group equilibrium then a group should be able to agree or coordinate on their "most desired" equilibrium.

Example 1. We start with an example with three players. The first two players form a collusive group and the third acts independently. The simple theory is that given the play of player 3, players 1 and 2 should agree on the incentive compatible pair of (mixed) actions that give them the most utility. However, in the following game there is no equilibrium that satisfies this prescription. Specifically, 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 cooperate (C)

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

If player 3 plays D the payoff matrix for the actions of players 1 and 2 is a symmetric coordination game in which player 3 prefers that 1 and 2 defect (D)

	C	D
C	$6,6,0$	$4,4,0$
D	$4,4,0$	$5,5,5$

Let α^i denote the probability with which player i plays C. We examine the set of equilibria for players 1 and 2 given the strategy α^3 of player 3. If $\alpha^3 > 1/2$ then D is strictly dominant for both player 1 and 2 so there is a unique in-group equilibrium in which they play D, D. If $\alpha^3 = 1/2$ then there are two equilibria, both symmetric, one at C, C and one at D, D. If $\alpha^3 < 1/2$ then there are three equilibria, all symmetric, one at C, C, one at D, D and a strictly mixed equilibrium in which $\alpha^1 = \alpha^2 = (1/3)(1 + \alpha^3)/(1 - \alpha^3)$.[4]

How should the group of player 1 and player 2 collude given the play of player 3? If $\alpha^3 > 1/2$ they have no choice: there is only one in-group equilibrium at D, D. For $\alpha^3 \leq 1/2$ they each get 6 at the C, C equilibrium, no more than 5 at the D, D equilibrium, and strictly less than 6 at the strictly mixed equilibrium. So if $\alpha^3 \leq 1/2$ they should choose C, C. Notice that in this example there is no ambiguity about the preferences of the group: they unanimously agree in each case as to which is the best equilibrium.

We may summarize the play of the group by a kind of "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 $\alpha^3 = 0$ and in particular is not larger than $1/2$; when the group plays C, C player 3 should play C and so $\alpha^3 = 1$ and in particular is not less than or equal to $1/2$. In other words, there is no equilibrium of the game in which the group of player 1 and player 2 chooses the best in-group equilibrium given the play of player 3.

In this example, the non-existence of an equilibrium in which player 1 and player 2 collude is driven by the discontinuity in the group best response: a small change in the probability of α^3 leads to an abrupt change in the behavior of the group. The key idea of this paper is that this discontinuity is an artifact of the model and does not make sense from an economic point of view. In particular, it does not make much sense that as α^3 is increased slightly above .5 the C, C equilibrium for the group abruptly vanishes. To understand 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 [51] purification, or more concretely the limit of McKelvey and Palfrey [68]'s Quantal Response Equilibria. Here the underlying model is perturbed in such a way that as indifference is approached players begin to randomize and the probability with which each action is taken is a smooth function of beliefs. In the limit as the perturbation becomes small, like the Cheshire cat, only the randomization remains. Similarly, in the context of group behavior, it makes sense that as the beliefs of a group change the probability with which they play different equilibria varies continuously. Consider for example $\alpha^3 = 0.499$ versus $\alpha^3 = 0.501$. In a practical setting where nobody actually knows α^3 does it make sense to assert that in the former case player

[4] Here is the computation of the mixtures from the condition that player 1 must be indifferent between C and D:
$6\alpha^2 + (1-\alpha^2)4(1-\alpha^3) = \alpha^2(8\alpha^3 + 4(1-\alpha^3)) + (1-\alpha^2)(2\alpha^3 + 5(1-\alpha^3))$
$\left(6 - 4(1-\alpha^3)\right)\alpha^2 + 4(1-\alpha^3) = \left(8\alpha^3 + 4(1-\alpha^3) - 2\alpha^3 - 5(1-\alpha^3)\right)\alpha^2 + (2\alpha^3 + 5(1-\alpha^3))$
$\left(6 - 4(1-\alpha^3)\right)\alpha^2 + 4 - 4\alpha^3 = \alpha^2(6\alpha^3 - (1-\alpha^3)) + 5 - 3\alpha^3$
$\left(6 - 3(1-\alpha^3) - 6\alpha^3\right)\alpha^2 = 1 + \alpha^3$
$\left(3 - 3\alpha^3\right)\alpha^2 = 1 + \alpha^3$

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 agree that $\alpha^3 \leq 0.5$ with 90% probability and mistakenly agree that $\alpha^3 > 0.5$ with 10% probability in the first case and conversely in the second case. Consequently when $\alpha^3 = 0.499$ there would never-the-less be a 10% chance that the group would choose to play D, D not realizing that C, C is incentive compatible, while when $\alpha^3 = 0.501$ there would be a 10% chance that they would choose to play C, C incorrectly thinking that it is incentive compatible. We will develop below a formal model in which groups have beliefs that are a random function of the true play of the other groups and are only approximately correct. For the moment we expect, as in Harsanyi [51], that in that limit only the randomization will remain. Our first step is to introduce a model that captures the grin of the Cheshire cat - we will simply assume that randomization is possible at the critical point. In the example we assert that when $\alpha^3 = 0.5$ and the incentive constraint exactly binds, the equilibrium "assigns" a probability to C, C being the equilibrium that is chosen by the group.[5] That is, when the incentive constraint holds exactly we do not assume that the group can choose their most preferred equilibrium, but instead we assume that there is an endogenously determined probability that they will be able to choose that equilibrium.

Remark. Discontinuity and non-existence is not an artifact of restricting attention to Nash equilibrium. The same issue arises if we assume that players 1 and 2 can use correlated strategies. When the game is a PD, that is, $\alpha^3 > 1/2$ then strict dominance implies that the unique Nash equilibrium is also the unique correlated equilibrium. When $\alpha^3 \leq 1/2$ the Nash equilibrium at C, C Pareto dominates every other correlated strategy, hence remains the unique best choice for players 1 and 2. When $\alpha^3 \leq 1/2$ the correlated equilibrium set is indeed larger than the Nash equilibrium set (containing at the very least the public randomizations over the Nash equilibria), but these correlated equilibria are all inferior for players 1 and 2 to C, C 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-hemi-continuous - this example shows that the selection from that correspondence that chooses the best equilibrium for the group is never-the-less badly behaved - it is discontinuous. It is well known from the earliest work on competitive equilibrium Arrow and Debreu [9] that for the best choice from a constraint set to be well-behaved the constraint set needs to be lower-hemi-continuous and neither the Nash nor correlated equilibrium correspondence satisfies that property.

3. The Exogenous Group Model

We now introduce our model of exogenously specified homogeneous groups in which the groups pursue their own interest subject to incentive compatibility constraints.

There are players $i = 1, 2, \ldots I$ and groups $k = \mathbf{1}, \mathbf{2}, \ldots K$. The actions available to a player depend entirely on which group he is in; actions available for members of group k are A^k, assumed to be a finite set. We assume that there is a fixed assignment of players to groups $k(i)$. Notice that each individual is assigned to exactly one group and that the assignment is exogenous. All players within a group are symmetric - that is the groups are homogeneous - so the relevant utility

[5] This arbitrary assignment is similar to Simon and Zame [79] endogenous choice of sharing rules.

of player i is $u^{k(i)}(a^i, a^{-i})$ and is invariant with respect to within group permutations of the labels of other players within their respective groups. If we let $\mathcal{A}^k$ denote the mixed actions for a member of group k, profiles of play chosen from this set represent the universe in which in-group equilibria reside.[6] As should be clear from the example, we will need to consider randomizations over in-group equilibria: each group is assumed to possess a private randomizing device observed only by members of that group that can be used to coordinate group play.

Because $\mathcal{A}^k$ is infinite, randomization over this set by the group leads to technical and conceptual complications that we prefer to avoid, so we will restrict the set of possible choices for the group. Specifically, we fix a finite subset $A^{kR} \subseteq \mathcal{A}^k$ containing all pure strategies, and consider only in-group equilibria for group k in which all players choose the same action $a^k \in A^{kR}$. For example, with $A^k = \{H, T\}$ the actions in A^{kR} can be of the form: choose H, choose T, or randomize 50-50 between H and T. In other words, the model is consistent with individual randomization provided that individuals are limited to a finite grid of probabilities. Since in-group mixed equilibria may not be present in A^{kR} we will allow the group to choose in-group ϵ-equilibria in which small violations of the incentive constraints are allowed.

Given the symmetry restriction we can simplify notation and write $u^k(a^i, a^k, \alpha^{-k})$ for the expected utility of player i in group $k(i) = k$ when a^i is his choice, the other group members play the common group action $a^k \in A^{kR}$, and the other groups $\kappa \neq k$ assigns probability $\alpha^\kappa(a^\kappa)$ to all members of the group playing $a^\kappa \in A^{\kappa R}$.

Further, since only deviations from the common strategy matter, for player i in group $k(i) = k$ we need not allow a^i to take values in all of A^{kR} - it is sufficient to consider $a^i \in A^k \cup \{a_0^k\}$ where a_0^k means: "play the common mixed action $a^k \in A^{kR}$". That is, it is enough to consider deviations by player i to pure strategies A^k, letting $u^k(a_0^k, a^k, \alpha^{-k}) = u^k(a^k, a^k, \alpha^{-k})$ to be the utility when no deviation has taken place. Not only does this potentially greatly reduce the set of a^i that need be considered, but extends in a straightforward way when we come to consider correlated group strategies below. Notice that this formulation incorporates the use of randomizing devices that are private to the group: member i knows the result of the own group randomization $a^{k(i)}$ when choosing a^i, but does not know results of the randomization by other groups.

Groups are assumed to be collusive - but they may collude only to choose plans that respect individual incentive constraints. The key reason that we start by considering homogeneous groups is that since group members are *ex ante* identical there is an "obvious" group objective, which is to assume that all members are treated equally and that the objective of the group is to maximize the common utility that they receive when all are treated equally. [7]

As indicated we allow a small amount of slack in the individual incentive constraints. Specifi-

[6] We will discuss also the possibility that this universe might encompass correlated strategies for the group but for expositional reason we defer that discussion.

[7] Incidentally: it may be that the group does best when rather than playing individual mixed strategies they agree on a common correlated strategy. This can be dealt with in a straightforward manner by allowing the group to use correlated strategies, but we defer discussion of this issue for the moment.

cally, we introduce strictly positive numbers $v^k > 0$ measuring in utility units the violation of incentive constraints that are allowed. For a mixed profile α^{-k} by other groups and an action a^k by group k we may define the gain function $G^k(a^k, \alpha^{-k}) = \max_{a^i \in A^k \cup \{a_0^k\}} [u^k(a^i, a^k, \alpha^{-k}) - u^k(a^k, a^k, \alpha^{-k})]$ as the degree to which the incentive constraint is violated at a^k (the smaller the gain the more stable the action). When the gain is strictly less than v^k then a^k must be chosen by the group if it is to the benefit of the group to do so. When the gain is greater than v^k then a^k the group cannot choose a^k. When the gain is exactly v^k then the group may mix with any probability onto a^k. This is the same Cheshire grin logic as in the example, except that in the example we took $v^k = 0$.

Define $U^k(\alpha^{-k}) = \max_{\{a^k | G^k(a^k, \alpha^{-k}) < v^k\}} u^k(a^k, a^k, \alpha^{-k})$ to be the most utility attainable against α^{-k} when the incentive constraints are violated by strictly less than v^k (it is equal to $-\infty$ if the constraint set is empty). Then we take the finite set $B^k(\alpha^{-k}) = \{a^k | G^k(a^k, \alpha^{-k}) \leq v^k, u^k(a^k, a^k, \alpha^{-k}) \geq U^k(\alpha^{-k})\}$ to represent actions that are feasible for the group given α^{-k}. We refer to this as the *shadow response set.* They are actions which violate the incentive constraints by strictly less than v^k and yield $U^k(\alpha^{-k})$, the most possible among such actions, plus those actions with $G^k(a^k, \alpha^{-k}) = v^k$ that yield at least $U^k(\alpha^{-k})$ - but possibly more. Observe that not all actions in $B^k(\alpha^{-k})$ need be indifferent, but that on the other hand all incentive compatible actions outside of $B^k(\alpha^{-k})$ are strictly worse for the group than any of those inside $B^k(\alpha^{-k})$.

Definition 1. A *collusion constrained equilibrium* is an α^k for each group that places weight only on $B^k(\alpha^{-k})$.

Define $\overline{B}^k(\alpha^{-k}) = \arg\max_{\{a^k | G^k(a^k, \alpha^{-k}) \leq v^k\}} u^k(a^k, a^k, \alpha^{-k}) \subseteq B^k(\alpha^{-k})$ to be the set of actions that maximize utility subject to the incentive constraints. Again, the key to collusion constrained equilibrium is that we allow a positive probability of actions in $B^k(\alpha^{-k})$ not merely in $\overline{B}^k(\alpha^{-k})$. If in a collusion constrained equilibrium α^k places positive weight on $B^k(\alpha^{-k}) \backslash \overline{B}^k(\alpha^{-k})$ we say that group k engages in *shadow mixing,* meaning that it is putting positive probability on alternatives it is not indifferent to. This may occur when best alternatives are not *strictly* incentive compatible, hence - this is our rationale for this equilibrium - they are not available to play with certainty within the group. This is to be contrasted with putting weight on $\overline{B}^k(\alpha^{-k})$ which are mixtures in the normal sense of indifference. Our example above shows that shadow mixing may be necessary in equilibrium.

Example 2. To illustrate the definition we apply it to the game of Example 1. If player 3 plays C with probability α^3 and the group plays D, D a player in the group who deviates to C gets $\alpha^3(-2) + (1-\alpha^3)(-1)$ so this deviation is never profitable, D, D being strictly incentive compatible. If the group plays C, C the player who deviated to D gets $\alpha^3 \cdot 2 + (1 - \alpha^3) \cdot (-2) = 2(2\alpha^3 - 1)$: the best in-group equilibrium is thus incentive compatible for $2(2\alpha^3 - 1) \leq v^1$, at equality incentive compatibility is just satisfied and the equilibrium vanishes for larger values. So the condition for shadow mixing between C, C and D, D is $2\alpha^3 - 1 = v^1/2$ or $\alpha^3 = (1 + (v^1/2))/2$. Formally, for this value of α^3 the shadow response set $B^1(\alpha^3) = \{C, D\}$ for D is the only, hence best, action satisfying incentive compatibility strictly. For player 3 to be indifferent between C and D, letting p the probability with which the group plays C, C we get the condition $5p = 5(1 - p)$ so $p = 1/2$. So equilibrium is that the group mixes 50-50 between C, C and D, D and player 3 plays C with probability $\alpha^3 = (1 + v^1/2)/2$. As $v^1 \to 0$ this converges to $1/2$.

The assumption that $v^k > 0$ plays a dual role in the model. First as indicated, we need to allow positive v^k if we wish to insure that in-group mixed equilibria are not excluded.[8] However, $v^k > 0$ plays a second role: it enables us to properly allow mixing only at "critical" points where small changes in beliefs lead to a discontinuous change in behavior.

Example 3. Group **1** has three actions H, M, L while group **2** has two actions H, L. For player i in group $k(i) = \mathbf{2}$ payoffs are $u^{\mathbf{2}}(a^i, a^{\mathbf{2}}, a^{\mathbf{1}}) = 0$, so group **2** has no active role and we concentrate on group **1**. For player i in group $k(i) = \mathbf{1}$ payoffs $u^{\mathbf{1}}(a^i, a^{\mathbf{1}}, a^{\mathbf{2}})$ are in the following matrix:

	$a^i = H, a^{\mathbf{2}} = H, L$	$a^i = M, a^{\mathbf{2}} = H, L$	$a^i = L, a^{\mathbf{2}} = H$	$a^i = L, a^{\mathbf{2}} = L$
$a^{\mathbf{1}} = H$	2	2	3	1
$a^{\mathbf{1}} = M$	1	0	1	1
$a^{\mathbf{1}} = L$	1	1	1	1

Action M is never part of an equilibrium: whatever the other group are doing, if the other members of your group play M you want to deviate. On the other hand no one ever wants to deviate from L - but incentive constraints are satisfied with exact equality there. Behavior against H is richer: you may want to deviate if the other group are playing H with high enough probability (to visualize as in the first example: if they work hard they cannot watch you). Specifically, equilibria are computed to be as follows. Let $\alpha^{\mathbf{2}}$ be the probability with which members of group **2** play H, and observe first that any $\alpha^{\mathbf{2}}$ is an equilibrium for group **2**.

If $\alpha^{\mathbf{2}} \leq 1/2$ there are two equilibria for group **1**: H and L; if $\alpha^{\mathbf{2}} > 1/2$ the only equilibrium is L. In all equilibria the incentive constraints are exactly satisfied (when $\alpha^{\mathbf{2}} \leq 1/2$ and group **1** action is H action M gives you the same utility as H; this is the role of M in the example).

So given the mixing rule we have specified above, with $v^{\mathbf{1}} = 0$ the collusion constrained equilibria consist of $\alpha^{\mathbf{2}} \leq 1/2$ and any vector $\alpha^{\mathbf{1}} = (a, 0, b)$, and $\alpha^{\mathbf{2}} > 1/2$ together with $\alpha^{\mathbf{1}} = (0, 0, 1)$. The group cannot guarantee that it will collude on the preferred action H.

With $v^{\mathbf{1}} > 0$ observe that $2 + v^{\mathbf{1}} = (1/2 + v^{\mathbf{1}}/2) \cdot 3 + (1/2 - v^{\mathbf{1}}/2) \cdot 1$ so that members of group **1** are indifferent between the payoff $2 + v^{\mathbf{1}}$ they get from agreeing with the group at H and deviating to L against group **2** playing $\alpha^{\mathbf{2}} = 1/2 + v^{\mathbf{1}}/2$.[9] Hence the collusion constrained equilibria consist of: (1) $\alpha^{\mathbf{2}} < 1/2 + v^{\mathbf{1}}/2$ and $\alpha^{\mathbf{1}} = (1, 0, 0)$, where H is strictly incentive compatible and best group alternative; (2) $\alpha^{\mathbf{2}} = 1/2 + v^{\mathbf{1}}/2$ and any vector $(a, 0, b)$, where the only strictly incentive compatible action is L hence $B^{\mathbf{1}}(\alpha^{\mathbf{2}}) = \{H, L\}$; and (3) $\alpha^{\mathbf{2}} > 1/2 + v^{\mathbf{1}}/2, \alpha^{\mathbf{1}} = (0, 0, 1)$. As we see, for $\alpha^{\mathbf{2}}$ slightly larger than $1/2$ incentive constraints are violated but the violation is small enough to make collusion on H viable. Using $v^{\mathbf{1}} > 0$ captures the difference between $\alpha^{\mathbf{2}} < 1/2$ and the critical economy where a small change in $\alpha^{\mathbf{2}}$ makes H no longer viable. In a sense it captures the fact that indifference for $\alpha^{\mathbf{2}} < 1/2$ is not fundamental - it occurs just because there is an action M to which individuals are indifferent - but small perturbations in $\alpha^{\mathbf{2}}$ leave that indifference unchanged. Put differently, if we think that the inability of the group to coordinate perfectly is due to the fact that a small randomization in beliefs about the other group may cause indifference to be violated, then the "razor edge" equilibria for $\alpha^{\mathbf{2}} < 1/2$ are not vulnerable while the critical economy at $1/2 + v^{\mathbf{1}}/2$ is and this is correctly picked up when we make $v^{\mathbf{1}}$ strictly positive.

[8] The importance of this issue is underscored by the possibility of a unique in-group equilibrium, which is mixed.

[9] We are abusing terminology a bit: they do not "get $2 + v^{\mathbf{1}}$", but as long as the left member is larger than the right one the gain to deviating to L is less than $v^{\mathbf{1}}$.

Incentive Compatible Games

There are two kinds of mixing: the group can mix between different actions chosen by the group using the group randomization device, but also individuals can mix. As we noted above individual mixing is included in the finite set A^{kR}, so the group mixes over a finite rather than continuous set. From an economic and empirical point of view dealing with approximate equilibria within the group does not pose a problem - in the field, laboratory or computationally we cannot expect individuals to achieve more than an approximate equilibrium.

If A^{kR} contains a relatively fine grid of mixtures there will be an ϵ-Nash equilibrium with a small value of ϵ. As long as v^k is strictly bigger than ϵ the group can find an action that is guaranteed to satisfy the incentive constraints to the required degree. Specifically, define $g^k = \max_{\alpha^{-k}} \min_{a^k \in A^{kR}} G^k(a^k, \alpha^{-k})$ so that regardless of the behavior of the other groups there is always a g^k approximate equilibrium within the group.

Definition 2. A game is *incentive compatible* if $v^k > g^k$ for all k.

Hereafter we will restrict attention to incentive compatible games: roughly this means that we chose a "fine enough" grid for each group.

4. Analysis of the Model

Having defined collusion constrained equilibrium we now want to show that they exist and make sense. In this section we consider how collusion constrained equilibria arise as the limits of fully collusive equilibria with random group beliefs and analyze more closely the role of shadow mixing. In the next section we will consider a concrete non-cooperative game involving representative or virtual players from each group and show that it gives rise exactly to collusion constrained equilibria.

4.1. The Existence of Collusion Constrained Equilibria

In this subsection we show that the basic problem of non-existence that arises when group try to choose actions in $\overline{B}(\alpha^{-k})$ is resolved by collusion constrained equilibrium by establishing a basic existence result.

Theorem 1. *An incentive compatible game has a collusion constrained equilibrium.*

This result follows from the basic properties of the shadow response set

Lemma 1. *(i) In an incentive compatible game $\overline{B}^k(\alpha^{-k})$ is non-empty for all α^{-k}; (ii) every α^{-k} has an open neighborhood $\mathcal{A}$ such that $\tilde{\alpha}^{-k} \in \mathcal{A}$ implies that $B^k(\tilde{\alpha}^{-k}) \subseteq B^k(\alpha^{-k})$.*

Proof. Assertion (i) is obvious from the definition. (ii) If not there must be a sequence $\alpha_n^{-k} \to \alpha^{-k}$ and points $a_n^k \in B(\alpha_n^{-k}), a_n^k \notin B(\alpha^{-k})$. Since A^{kR} is a finite set, we may assume that we have chosen a subsequence along which $a_n^k = a^k$ is constant. Since G^k is continuous in α_n^{-k} any a^j such that $G^k(a^j, \alpha^{-k}) < v^k$ satisfies $G^k(a^j, \alpha_n^{-k}) < v^k$ for n large enough, so since A^{kR} is finite all those which satisfy the constraint strictly in the limit do so for n large enough, which implies that for such n it is $U(\alpha_n^{-k}) \geq U(\alpha^{-k})$. Let $\tilde{a}^k \in \arg\max_{\{a^k | G^k(a^k, \alpha^{-k}) < v^k\}} u^k(a^k, a^k, \alpha^{-k})$. Then $U^k(\alpha^{-k}) = u^k(\tilde{a}^k, \tilde{a}^k, \alpha^{-k})$ and since $a^k \in B(\alpha_n^{-k})$ for all n we then have

$$u^k(a^k, a^k, \alpha^{-k}) \geq u^k(\tilde{a}^k, \tilde{a}^k, \alpha^{-k}) = U^k(\alpha^{-k}).$$

By continuity of G^k it is also the case that $G^k(a^k, \alpha^{-k}) \leq v^k$ so we obtain $a^k \in B(\alpha^{-k})$, a contradiction. □

We can now prove the existence theorem

Proof of Theorem 1. Call $C(\alpha^{-k})$ the set of distributions over $B(\alpha^{-k})$. A profile α is a collusion constrained equilibrium if $\alpha^k \in C(\alpha^{-k})$ for all k, that is if $\alpha \in C(\alpha) \equiv \times_k C(\alpha^{-k})$, in other words if α is a fixed point of the correspondence $\alpha \twoheadrightarrow C(\alpha)$. Since the game is incentive compatible $C(\alpha^{-k})$ is non empty for any α^{-k}. Further, by construction, it is a convex valued correspondence. As a result, the correspondence $C(\alpha)$ is non empty and convex valued. By Lemma 1 we know that that $B(\alpha^{-k})$ is upper hemicontinuous. In turn this implies that both $C(\alpha^{-k})$ and $C(\alpha)$ are upper hemicontinuous. Hence the fixed point sought for exists by the Kakutani fixed point theorem. □

4.2. Random Beliefs

In this section we show that collusion constrained equilibria are limit points of standard equilibria when beliefs of each group about behavior of the other groups are random and randomness tends to vanish. We start by describing a random belief model. The idea is that given the true play α^{-k} of the other groups, there is a common belief $\tilde{\alpha}^{-k}$ 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 give rise to common knowledge - looking each other in the eye, a handshake or whatever - then it makes sense that during these discussions a consensus emerges not just on what action to take, but underlying that choice, a consensus on what the other groups are thought to be doing. We must emphasize: our model is a model of the consequences of groups successfully colluding - we do not attempt to model the underlying processes of communication, negotiation and consensus that leads to their successful collusion.

Definition 3. An *ϵ-random group belief model* is a density function $f^k(\tilde{\alpha}^{-k}|\alpha^{-k})$ that is continuous as a function of $\tilde{\alpha}^{-k}, \alpha^{-k}$ and satisfies $\int_{|\tilde{\alpha}^{-k}-\alpha^{-k}|\leq\epsilon} f_\epsilon^k(\tilde{\alpha}^{-k}|\alpha^{-k}) d\tilde{\alpha}^{-k} \geq 1-\epsilon$.

It is important to know that there are ϵ-random belief models for every positive value of ϵ. An obvious idea is to take a smooth family of probability distributions with mean equal to the truth and small variance. A good candidate for a smooth family is the Dirichlet since we can easily control the precision by increasing the "number of observations." However using an unbiased probability distribution will not work - it is ill-behaved on the boundary: if we try to keep the mean equal to the truth, then as we approach the boundary the variance has to go to zero, and on the boundary there will be a spike. A simple alternative is is bias to the mean slightly towards a fixed strictly positive probability vector alpha with a small weight on that vector, and then let that weight go to zero as we take the overall variance to zero. The next example shows that this works.

Example 4. Let M^{-k} be the number of actions in A^{-k} and set $h(\epsilon) = (\epsilon/2)^2 M^{-k}/(M^{-k} - (\epsilon/2)^2)$. Fix a strictly positive probability vector over A^{-k} denoted by β^{-k} and call the *ϵ-Dirichlet belief model* the Dirichlet distribution with parameters

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

Theorem 2. *The ϵ-Dirichlet belief model is an ϵ-random belief model.*

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

$$Pr[|\hat{\alpha}^{-k} - \overline{\alpha}^{-k}| > \epsilon/2] \leq E|\tilde{\alpha}^{-k} - \beta^{-k}|^2/(\epsilon/2)^2 \leq M^{-k}h(\epsilon)/[\epsilon(M^{-k} + h(\epsilon))] \leq \epsilon/2.$$

Observe that $|\hat{\alpha}^{-k} - \overline{\alpha}^{-k}| = (1-\frac{\epsilon}{2\sqrt{2}})|\tilde{\alpha}^{-k} - \beta^{-k}| \geq |\tilde{\alpha}^{-k} - \beta^{-k}| - \frac{\epsilon}{2}$. Hence $Pr(|\tilde{\alpha}^{-k} - \beta^{-k}| > \epsilon) \leq \epsilon/2 \leq \epsilon$ which shows that this is indeed an ϵ-random belief model. □

Fix some probability distribution $F^k(\alpha^{-k})$ over $\overline{B}^k(\alpha^{-k})$ measurable as a function of α^{-k}. Define $R^k(a^k|\alpha^{-k}) = \int F^k(\tilde{\alpha}^{-k})[a^k] f^k(\tilde{\alpha}^{-k}|\alpha^{-k}) d\tilde{\alpha}^{-k}$. Notice that for given beliefs $\tilde{\alpha}^k$ we are assuming that the group colludes on a response in $\overline{B}^k(\tilde{\alpha}^{-k})$ which are the best choices for the group that weakly satisfy the incentive constraints, and not on points in $B^k(\tilde{\alpha}^{-k})\backslash\overline{B}^k(\tilde{\alpha}^{-k})$ as would be permitted by shadow mixing. We define an *ϵ-random belief equilibrium* as an α_ϵ such that $\alpha^k_\epsilon = R^k(\alpha^{-k}_\epsilon)$. The key result is

Theorem 3. *Fix a family of ϵ-random group belief models, an $F^k(\alpha^{-k})$ and an incentive compatible game. Then for all $\epsilon > 0$ there exist ϵ-random belief equilibria. Further, if α_ϵ are ϵ-random group equilibria and $\lim_{\epsilon\to 0}\alpha_\epsilon = \alpha$ then α is a collusion constrained equilibrium.*

Proof. By the Lebesgue dominated convergence Theorem R^k is continuous, so we may apply the Brouwer fixed point to get existence of ϵ-random group equilibria. Now consider a sequence of ϵ-random group equilibria with $\lim_{\epsilon\to 0}\alpha_\epsilon = \alpha$. By Lemma 1 we know that for sufficiently small ϵ, $|\alpha^{-k}_\epsilon - \alpha^{-k}| \leq \epsilon$ implies $B^k(\alpha^{-k}_\epsilon) \subseteq B^k(\alpha^{-k})$. Hence for such α^k_ϵ and ϵ it must be that $\alpha^k_\epsilon(B^k(\alpha^{-k})) = 1$ with $\alpha^k(B^k(\alpha^{-k})) = 1$ at the limit - which is the condition for a collusion constrained equilibrium. □

We should emphasize that this result is not an equivalence result: random belief equilibria converge as $\epsilon \to 0$ to collusion constrained equilibria. However, there is no assertion that all collusion constrained equilibria arise this way. This is similar to the result for Harsanyi [51] where convergence of random utility equilibria to Nash equilibria is assured, but only under additional conditions do we know that Nash equilibria arise as limits of random utility equilibria. In cases such as quantal response indeed, limits of quantal response equilibria are a refinement of Nash equilibrium.

4.3. When Does Shadow Mixing Matter?

For applications it is useful to know when groups do not engage in shadow mixing. There are two important cases where groups will engage only in ordinary mixing.

1. The action that maximizes group utility without constraint is always an in-group equilibrium. Since the action is an equilibrium, it strictly satisfies the relaxed constraint with $v^k > 0$. Since it maximizes group utility without any constraint, it certainly maximizes group utility with the constraint, so $\overline{B}^k(\alpha^{-k}) = B^k(\alpha^k)$. Notice that in case the group has a single player, or more generally the game is a game of common interest so that group members always get the same payoffs as each other regardless of the actions chosen this assumption is satisfied.[10]

2. Separable games in which $u(a^i, a^k, a^{-k}) = w(a^{-k}) - c(a^i, a^k)$ so that the incentive constraints do not depend on what the other groups do. Here $G(a^k, \alpha^{-k}) = \max_{a^i \in A^k} c(a^k, a^k) - c(a^i, a^k)$ independent of α^{-k}. Hence for generic v^k there will be no a^k for which $G(a^k, \alpha^{-k}) = v^k$. These models can be important for applications because they can be thought of as approximation in political economy games such as voting or lobbying games where the group size is large so individuals perceive that their own action has no impact on the common public good w - for example, the outcome of a vote.

4.4. What Difference Do Collusion Constraints Make?

We return to example 1 to illustrate how accounting for incentive and collusion constraints may impact on the strategic analysis of a game.

First, the only Nash equilibrium of the game consists of all players to play D. To see this observe that as shown in Footnote 4 players 1 and 2 can mix only if $\alpha^3 \leq 1/2$ and then $\alpha^1 = \alpha^2$ are increasing in α^3; so the smallest value of α^1 occurs when $\alpha^3 = 0$ and it is $\alpha^1 = 1/3$. But for $\alpha^1 = \alpha^2 \geq 1/3$ player 3's best response is to play C for sure; hence there is no equilibrium in which player 1 and 2 mix. The two of them playing C, C is not an equilibrium because 3's best response to it is C for sure, but in that case they will play D, D. Profile D, D, D on the other hand is Nash. In this equilibrium payoffs are $(5, 5, 5)$.

On the other hand, ignoring individual incentive constraints, that is assuming that the group will collude on best group action, leads to predict that players 1 and 2 will play C, C in which case 3 also chooses C. Predicted payoffs would be $(6, 6, 5)$.

Consider now collusion constrained equilibrium. We have seen in Example 2 that in this equilibrium the group mixes 50-50 between C, C and D, D and player 3 plays C with probability $\alpha^3 = (1+\epsilon/2)/2$. In equilibrium player 3 gets 2.5. Players 1 and 2 get $4(\frac{1}{2}+\frac{\epsilon}{4})+\frac{11}{2}(\frac{1}{2}-\frac{\epsilon}{4}) = 4\frac{3}{4}-\frac{3}{8}\epsilon$. As $\epsilon \to 0$ the limit payoff vector is a much lower$(4.75, 4.75, 2.5)$.

As can be expected, ignoring individual constraints lead to an unrealistically optimistic conclusion. But the remarkable point is that in the example the same is true for Nash equilibrium: ignoring collusion constraints also leads to predicting higher utilities for the players. Incidentally, this is why we call our equilibrium collusion *constrained*: in general collusion makes the group of the whole worse off.

[10] In these games an action profile maximizing the utility of some group member does the same for each group member and must therefore be an in-group equilibrium too.

Notice that a benevolent mechanism designer who could choose between having players play the game and a safe alternative that gave payoffs of $(4.9, 4.9, 4.9)$ who either analyzed the game ignoring collusion or who analyzed the game assuming that players could collude would choose the game over the safe alternative, while a designer who recognized that collusion is subject to incentive constraints would reach the opposite conclusion.

5. Leadership Equilibrium

To give a concrete way in which collusion constrained equilibria can arise, we give a non-cooperative model of leadership which gives rise to collusion constrained equilibria. Leaders lead their group to act when several groups interact - they tell their followers things such as "let's go on strike" or "let's vote against that law." The idea is that group leaders serve as explicit coordinating devices for groups - and we will model them in a way that gives rise exactly to collusion constrained equilibrium. Each group will have a leader who tells group members what to do, and since he is to serve as an effective coordination device for group members these instructions cannot be optional for group members. However, we do not want leaders to issue instructions that members would not wish to follow - that is, that are not incentive compatible. Hence we give them incentives to issue instructions that are incentive compatible by allowing group members to "punish" their leader. As in the previous section incentive compatibility will mean that constraints can be violated by no more than v^k, and here this value has a concrete interpretation as the leader's "valence": the higher v^k the more members are ready to give up to follow the leader. While this is intended as an abstract model of how groups can reach decisions, we observe that in fact it is often the case that groups follow orders given by a leader but engage in *ex post* evaluation of the leader's performance.

Specifically, we will consider the following non-cooperative game. Each group is represented by two virtual players: a *leader* and an *evaluator,* each of whom has the same underlying preferences as the group members. Each leader has a punishment utility $\underline{u}^k < \min_{a^j, a^k, a^{-k}} u^k(a^j, a^k, a^{-k})$. The game goes as follows:

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

Stage 2: In each group, the evaluator observes the action plan of the leader and chooses a response a^i [11]

Payoffs: The evaluator receives utility $u^k(a^i, a^k, a^{-k}) + v^k \cdot I(a^i = a^k)$ where I is the indicator function, that is he gets the v^k bonus only if he chooses a^k. As to the leader, if the evaluator chooses a^k he gets $u^k(a^k, a^k, a^{-k})$, otherwise heis deposed and gets $\underline{u}^k$. Note that the leader and evaluator do not learn what the other groups did until the game is over.

Theorem 4. *In an incentive compatible game α are sequential equilibrium choices by the leaders if and only if $\alpha^k(a^k) > 0$ implies $a^k \in B^k(\alpha^{-k})$.*

[11] The evaluation need not be done by a single evaluator, but by consensus or some other aggregation method by all or a subset of group members. It makes no difference to the results.

Proof. The key implication of sequentiality is that the beliefs of the evaluator about the mixtures of other leaders must be independent of the signal received from his own leader - since his leader has no information about the signals of the other leaders. Suppose first that α is sequential. Then the beliefs the evaluator for group k about other groups is α^{-k} independent of the signal that they receive from his own leader - so in effect from the perspective of the evaluator this is treated as a constant.

Because the game is incentive compatible, the leader can insure himself a utility of $U^k(\alpha^{-k})$ by choosing the best a^k that strictly satisfies the incentive constraints since he will not be deposed in that case. If he makes an announcement that violates the incentive constraints he is deposed with probability one and gets $\underline{u}^k < U^k(\alpha^{-k})$, so it must be that any announcement with $\alpha^k(a^k) > 0$ has $a^k \in B^k(\alpha^{-k})$.

Suppose conversely that any announcement with $\alpha^k(a^k) > 0$ has $a^k \in B^k(\alpha^{-k})$. There are two kinds of $a^k \in B^k(\alpha^{-k})$: those for which the incentive constraints hold exactly and those for which they hold strictly. If they hold strictly, then the benevolent leader gets $U^k(\alpha^{-k})$ by the definition of U^k. If they hold weakly, then the evaluator is indifferent between choosing a^k and keeping the leader and picking an alternate best response and deposing him. Hence the probability that the leader is deposed $p^k(a^k, \alpha^{-k})$ may be any number between zero and one, and in particular may be chosen so that $(1 - p^k(a^k, \alpha^{-k}))u^k(a^k, a^k, \alpha^{-k}) + p^k(a^k, \alpha^{-k})\underline{u}^k = U^k(\alpha^{-k})$ since by definition of B^k we have $u^k(a^k, a^k, \alpha^{-k}) \geq U^k(\alpha^{-k})$. This means the leader is indifferent between all actions in $B^k(\alpha^{-k})$ and in particular it is optimal for him to choose α^k since that places weight only on $B^k(\alpha^{-k})$. □

6. Correlation and Symmetry

We have so far supposed that the groups are homogeneous and that they choose only symmetric mixed strategies. We now wish to relax both of those assumptions. We first continue to assume that the group is homogeneous but allow a broader set of strategies. Then we show how the resulting model can be extended to heterogeneous groups in a way that is consistent with the homogeneous group model.

We have assumed that the strategies available to group k are a finite subset A^{kR} of symmetric mixed strategies, while the deviations available to individual members are the pure strategies A^k or the special strategy a_0^k meaning play the group mixed strategy a^k. Notice, however, that the assumption of symmetric mixed strategies is limiting. For example, if a group of two members is playing a hunter-gatherer game in which members choose between hunter and gatherer, and get 0 for agreeing, and the hunter gets 2 and the gatherer gets 1 if they specialize, the unique symmetric mixed equilibrium gives an expected utility to each member of 2/3 while a public randomization over the two asymmetric pure Nash equilibria gives an expected utility to each member of 3/2. In the game of chicken, for another example, there is a correlated equilibrium that gives both players more than any public randomization over Nash equilibria. It seems plausible that groups would choose to use correlating devices to achieve these superior results. This leads us to extend the model to include correlated strategies by each group.

In Section 3 we took the space of deviations to be $A^k \cup \{a_0^k\}$. By redefining A^{kR} and, the space of deviations we can extend the model to incorporate correlated strategies in a straightforward way. First we take A^{kR} to be an arbitrary finite subset of symmetric correlated strategies for the group:

that is, a probability distribution over profiles of individual actions. Then we define the space of deviations D^k to be maps $d^i : A^k \to A^k$ from pure actions to pure actions with the interpretation that $d^i(a^k)$ is the action chosen by member i when he is told to play a^k. Here the identity map plays exactly the role that a_0^k played in the original model. With this change all the existing results and definitions remain unchanged.

Extending the model to correlated strategies also enables us to incorporate asymmetries in a straightforward way. First, take A^{kR} to be an arbitrary finite subset of the correlated equilibria - not necessarily symmetric. We assume utility has the form $u^i(a^i, a^{k(i)}, a^{-k(i)})$ where $d^i \in D^{k(i)}$and $a^{k(i)} \in A^{k(i)R}$, $a^{-k(i)} \in A^{-k(i)R}$ are no longer required to be symmetric, and individuals may no longer be homogeneous. The group is now assumed to have an exogenously specified objective of weighted sum of individual utility: $U^k(a^k, a^{-k}) = \sum_{i|k(i)=k} \omega_i u^i(a_0^k, a^k, a^{-k})$, and if we wish we may index the valences $v^i > 0$ by individual rather than by group. From a mathematical point of view, the only change needed to the existing model is that in the leadership version the evaluator must choose a vector of deviations $a^i|_{k(i)=k}$ and should equally weight the utility of each member of the group[12], while the leader should be punished if the evaluator chooses any deviation other than a_0^i on behalf of any group member. We refer to this notion as *asymmetric collusion constrained equilibrium.* Notice, however, that there is no longer a compelling commonality of interest to explain why group members should obey orders from a leader who does not share their own preferences. Indeed this raises the issue that there might be competition within the group over who should be leader - this is the topic of the second half of the paper.

Given the asymmetric model, suppose the game is in fact symmetric - we would like to know that the new notion of equilibrium is consistent with the old notion. Suppose that the weights $\omega_i = 1$ and that the valences $v^i = v^{k(i)}$. Suppose also that for every correlated strategy $a^k \in A^{kR}$ the set A^{kR} also includes the uniform public randomization over all correlated strategies which permute the identities of the group members in a^k. In this case we say that A^{kR} *contains a symmetric model.* Then we can show that the new notion of asymmetric collusion constrained equilibrium is consistent with the old notion of symmetric collusion constrained equilibrium in the following sense:

Theorem 5. *Suppose that A^{kR}contains a symmetric model. Then there exists an asymmetric collusion constrained equilibrium $\tilde{\alpha}$ that is symmetric and is a collusion constrained equilibrium with respect to the subset of A^{kR} that is symmetric. Conversely if $\tilde{\alpha}$ is a collusion constrained equilibrium with respect to the subsets of A^{kR} that are symmetric then it is an asymmetric collusion constrained equilibrium.*

Proof. To show asymmetry implies symmetry, we construct the symmetric equilibrium from an arbitrary asymmetric equilibrium. Given a collusion constrained (or leadership) equilibrium - not necessarily symmetric - for each positive probability realization of the group public randomization device (or equivalently recommendation of the leader) we may replace the recommended profile a^k with the uniform public randomization over all permutations of the names of the group members,

[12] Any strictly positive vector of weights is fine: we specify equal weights for definiteness. The point is that for the evaluator the optimal choice of each d^i is independent of the other choices.

$\tilde{a}^k$. By assumption no other group cares about this, and since the incentive constraints are violated by no more than v^k at a^k for any group member $k(i) = k$ the same remains true for $\tilde{a}^k$. Moreover, $U^k(\tilde{a}^k, \alpha^{-k}) = U^k(a^k, \alpha^{-k})$ since each permutation of group member utilities yields exactly the same value. Hence $\tilde{a}^k$ is also an asymmetric collusion constrained equilibrium. Moreover, if $\tilde{a}^k$ gave less utility than some symmetric $\hat{a}^k$ that violates the incentive constraints by strictly less than v^k then so would a^k. Hence it is a symmetric collusion constrained equilibrium.

Now suppose that $\tilde{\alpha}$ is a collusion constrained equilibrium with respect to the subsets of A^{kR} that are symmetric and let $\tilde{a}^k$ be a positive probability realization of the group public randomization device. We have to show that there is no $\hat{a}^k \in A^{kR}$ that violates the incentive constraints by strictly less than v^k and has $U^k(\hat{a}^k, \tilde{\alpha}^{-k}) > U^k(\tilde{a}^k, \tilde{\alpha}^{-k})$. Suppose instead that there is such a $\hat{a}^k \in A^{kR}$. Consider the uniform randomization over permutations of group members of $\hat{a}^k$ and denote it by a^k. Then this also violates the incentive constraints by strictly less than v^k and has $U^k(a^k, \tilde{\alpha}^{-k}) = U^k(\hat{a}^k, \tilde{\alpha}^{-k}) > U^k(\tilde{a}^k, \tilde{\alpha}^{-k})$. But by construction a^k is symmetric and this then contradicts the fact that $\tilde{a}^k$ had positive probability in equilibrium. □

References

[1] Abreu, Dilip and Ariel Rubinstein (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata" *Econometrica* 56: 1259-1281.

[2] Acemoglu, Daron (2002): "Directed Technical Change", *The Review of Economic Studies* 69: 781-809.

[3] Acemoglu, Daron and Alexander Wolitzky (2011): "The Economics of Labor Coercion", *Econometrica* 79: 555–600.

[4] 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.

[5] Alesina, Alberto and Enrico Spolaore (2003): *The Size of Nations*, MIT University Press.

[6] Alger, I. and J. Weibull (2010): "Kinship, Incentives and Evolution," American Economic Review 100(4): 1725-1758.

[7] Aoki, K. (1982) "A condition for group selection to prevail over counteracting individual selection," *Evolution* **36**: 832-42.

[8] Aoki, K, L. Lehmann and M. W. Feldman (2011): "Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission," *Theoretical Population Biology* **79**: 192-202.

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

[10] Ashraf , Quamrul and Oded Galor (2011): "Dynamics and Stagnation in the Malthusian Epoch", *American Economic Review* 101: 2003–41.

[11] Bacci, Massimo (2006) *A Concise History of World Population.*

[12] Benoit, Jean-Pierre and Krishna, Vijay (1987): "Nash equilibria of finitely repeated games," *International Journal of Game Theory* 16: 197-204.

[13] Bergstrom, Ted (2003): "The Algebra of Assortative Encounters and the Evolution of Cooperation," *International Game Theory Review* 5: 1-18.

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

[15] Binmore, Kenneth (1992): "Evolutionary stability in repeated games played by finite automata" *Journal of Economic Theory* 57: 278–305.

[16] Bisin, A. (2001): "The economics of cultural transmission and the evolution of preferences," Journal of Economic Theory 97, 298-319.

[17] Bisin, A. and G. Topa (2004): "Cooperation as a transmitted cultural trait," Rationality and Society 16: 477-507.

[18] Blume, L. and D. Easley (1992): "Evolution and Market Behavior," Journal of Economic Theory 58: 9-40.

[19] Bocquet-Appel, Jean-Pierrre (2002): "The paleoanthropological traces of the Neolithic demographic transition," *Current Anthropology* **43**: 638-50.

[20] Bocquet-Appel, Jean-Pierrre (2006): "Testing the Hypothesis of a Worldwide Neolithic Demographic Transition: Corroboration from American Cemeteries," *Current Anthropology* **47**: 341-65.

[21] Bocquet-Appel, Jean-Pierrre (2011): "When the world's population took off: The springboard of the Neolithic demographic transition," *Science* **333**: 560-61.

[22] Boorman, S. and P.R. Levitt (1973): "Group Selection on the Boundary of a Stable Population," *Theoretical Population Biology* **4**: 85-128.

[23] Bottazzi G. and P. Dindo (2011): "Evolution and Market Behavior with Endogenous Investment Rules," LEM Papers Series 2010/20, Sant'Anna School of Advanced Studies, Pisa, Italy.

[24] Bowles, Samuel (2004) *Microeconomics: Behavior, Institutions, and Evolution*, Princeton: Princeton University Press.

[25] Bowles, Samuel (2006): "Group competition, reproductive leveling and the evolution of human altruism," *Science* **314**: 1569-72.

[26] Bowles, Samuel (2009): "Did Warfare among Ancestral Hunter-Gatherer Groups Affect the Evolution of Human Social Behaviors", *Science* 324: 1293-98.

[27] Bowles, Samuel (2011a): "Better mousetraps? Beware simple history," *New Scientist*, 2823.

[28] Bowles, Samuel (2011b): "The cultivation of cereals by the first farmers was not more productive than foraging," *Proceedings of the National Academy of Science*, **108**: 4760-4765.

[29] Bowles, Samuel, Jung-Kyoo Choi and Astrid Hopfensitz (2003): "The coevolution of individual behaviors and group level institutions," *Journal of Theoretical Biology* **223**: 135-47.

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

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

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

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

[34] Choi, Jung-Kyoo and Samuel Bowles (2007): "The coevolution of parochial altruism and war," *Science* **318**: 636-40.

[35] Cipolla, Carlo (1965): *Guns, Sails, and Empires: Technological Innovation and the Early Phases of European Expansion, 1400- 1700.*

[36] Clark, Gregory (2007): *A Farewell to Alms,* Princeton University Press.

[37] Dekel, E. J. C. Ely and O.Yilankaya (2007): "Evolution of Preferences," The Review of Economic Studies 74(3):685-704.

[38] Diamond, Jared (1998): *Guns, Germs, and Steel: The Fates of Human Societies.*

[39] 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

[40] Ellison, Glenn (2000): "Basins of Attraction, Long Run Stochastic Stability and the Speed of Step-by-step Evolution," *Review of Economic Studies* 67: 17-45.

[41] Ely, Jeffrey (2002): "Local conventions," *The BE Journal of Theoretical Economics.*

[42] Finley, Moses I. (1954): The World of Odysseus. Reprinted in the New York Review of Books Classics, 2002

[43] Foster, Dean and Peyton Young (2003): "Learning, hypothesis testing, and Nash equilibrium" Games and Economic Behavior.

[44] Friedlin and Wentzell (1984): *Random Perturbations of Dynamical Systems*, Springer-Verlag.

[45] Fudenberg, Drew and Eric Maskin (1986): "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica.*

[46] Fudenberg, Drew, David Levine and Eric Maskin (1994): "The folk theorem with imperfect public information," *Econometrica.*

[47] Fudenberg, Drew and David Levine (2009): "Repeated Games with Frequent Signals," *Quarterly Journal of Economics*, 124 (2009), 149-195.

[48] Fukuyama, Francis (2011), *The Origins of Political Order*, Farrar, Straus and Giroux, New York.

[49] Garfinkel, Michelle and Stergios Skaperdas (2007): "Economics of Conflict: An Overview," *Handbook of Defense Economics*, Elsevier.

[50] Haeringer, G. (2004): "Equilibrium binding agreements: a comment," *Journal of Economic Theory.*

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

[52] Hansen, Gary and Edward Prescott (2002): "Malthus to Solow," *The American Economic Review.*

[53] Hausken, Kjell (2005): "Production and Conflict Models Versus Rent Seeking Models," *Public Choice* 123: 59-93.

[54] Hirshleifer, Jack (2001): *The dark side of the force: economic foundations of conflict theory*, Cambridge University Press.

[55] Hirshleifer, Jack (1992): "Anarchy and its Breakdown", Working Paper No. 674, Dept. of Economics, UCLA, September.

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

[57] Kandori, Michihiro (1992): "Social norms and community enforcement," *The Review of Economic Studies.*

[58] Kandori, Michihiro, George Mailath, and Rafael Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica* **61**: 29-56.

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

[60] Ladurie, Le Roy (1974): *The Peasants of Languedoc.*

[61] Levine, David K. and Salvatore Modica (2011) "Anti-Malthus: Conflict and the Evolution of Societies", mimeo

[62] Levine, David, Salvatore Modica, Federico Weinschelbaum and Felipe Zurita (2011): "Evolving to the Impatience Trap: The Example of the Farmer-Sheriff Game," mimeo Washington University of St. Louis.

[63] Levine, David and Wolfgang Pesendorfer (2007): "The Evolution of Cooperation Through Imitation," *Games and Economic Behavior.*

[64] Maddison, Angus (2007): *Contours of the World Economy 1-2030 AD.*

[65] Malthus, Thomas (1798): *An Essay on the Principle of Population*, J. Johnson: London.

[66] Maynard Smith, John (1982): *Evolution and the Theory of Games*, Cambridge University Press.

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

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

[69] McNeill, William (1963): *The Rise of the West*, University of Chicago Press.

[70] Nozick, Robert (1974): *Anarchy, State, and Utopia*

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

[72] Price, George R. (1970): "Selection and Covariance," *Nature* **227**: 520-21.

[73] Price, George R. (1972): "Extension of Covariance Selection Mathematics," *Annals of Human Genetics* **35**: 485-90.

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

[75] Richerson, Peter J. and Robert Boyd (2001) "Commitment to Groups: A Tribal Instincts Hypothesis" in *Evolution and the Capacity for Commitment*, ed. Randolph M. Nesse, Russell Sage Foundations.

[76] Rogers D. S., O. Deshpande and M. W. Feldman (2011): "The Spread of Inequality," PLoS ONE 6(9): e24683. doi:10.1371/journal.pone.0024683

[77] Rowthorn, Robert and Paul Seabright (2010): "Property rights, warfare and the neolithic transition," Toulouse School of Economics Working Paper 10-207.

[78] Rubinstein, Ariel (1986): "Finite automata play the repeated prisoner's dilemma" *Journal of Economic Theory* 39: 83–96.

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

[80] Tirole, J. (1986): "Hierarchies and breaucracies: on the role of collusion in organizations" *Journal of Law, Economics and Organization* 2: 181-214.

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

[82] Thompson, Earl and Charles Hickson (2000): *Ideology and the evolution of vital economic institutions: groups, the gold standard, and modern international cooperation.*

[83] Tilly, Charles (1985): "War Making and State Making as Organized Crime", in *Bringing the State Back In,* Evans, Rueschemeyer and Skocpol, Cambridge University Press

[84] Weightman, Gavin (2009): *Industrial Revolutionaries: The Making of the Modern World 1776-1914,* Grove Press

[85] Wilson, Edward O. (1975): Sociobiology. The New Synthesis. Harvard University Press

[86] Young, Peyton (1993): "The Evolution of Conventions," *Econometrica* **61**: 57-84.