Peer Discipline and the Strength of Organizations☆

David K. Levine[1], Salvatore Modica[2]

Abstract

Groups do not act as individuals - as Olson and other have emphasized: incentives within groups matter. Here we study internal group discipline through a model of costly peer punishment. We investigate schemes that minimize the cost to a collusive group of enforcing particular actions, and use the model to determine how the strength of the group - as measured, for example, by ability to raise funds to provide a public good - depends on the size of the group, the size of the prize, and the heterogeneity of the group. We find that voluntary provisions models do not scale properly, while the peer discipline model does. The peer discipline model predicts that for a fixed size of an auctioned (rival) prize the strength of the group is single-peaked - increasing when the group is small then declining. The strength of the group also declines with heterogeneity. When groups compete for transfers which one group may have to make in favor of another we find that small groups have an advantage over larger ones.

JEL Classification Numbers:

C72 - Noncooperative Games

D7 - Analysis of Collective Decision-Making

D72 - Political Processes: Rent-Seeking, Lobbying, Elections, Legislatures, and Voting Behavior

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

☆First Version: November 26, 2012. We are especially grateful to Michele Boldrin, Michael Chwe, Drew Fudenberg, Zacharias Maniadis, Eric Maskin, Andy Postelwaite and seminar participants at the St. Louis Federal Reserve Bank and Venice Conference on Economic Theory. 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: `david@dklevine.com` (David K. Levine), `salvatore.modica@unipa.it` (Salvatore Modica)

[1]Department of Economics, WUSTL

[2]Università di Palermo

1. Introduction

Groups do not act as individuals - as Olson [46] and others have emphasized incentives within groups matter. Formal research is however rather scant on the topic, and as we report below most papers dealing with groups do not explicitly deal with internal working of group discipline. Group strength depends on a number of characteristics, including the size and cohesion of the group. Here we study self-sustaining discipline through a model of costly peer punishment. We investigate schemes that might be adopted by a collusive group to minimize the cost of enforcing actions - effectively social norms - which are possibly not Nash equilibria, and use the model to determine how the strength of the group - as measured, for example, by ability to raise funds to provide a public good - depends on the size of the group, the size of the prize, and the heterogeneity of the group.

Our model is one of an initial choice of action by group members in a base game followed by an open-ended game of peer punishment. In the open-ended peer punishment games group members repeatedly audit each other and determine punishments for violating the social norm and the prescribed behavior in audits. While this does not describe the method used to provide incentives in all groups at all times - for example, we do not consider dictatorship by a residual claimant or organizations that monitor one another - the setting is an important one for many of the groups studied by political economists. The essential feature is that groups members agree that not punishing deviators can in turn be punished, a condition which has been found crucial also in field work such as that conducted by Elinor Ostrom and reported for instance in Ostrom [47] and Ostrom, Walker and Gardner [48]. From a conceptual point of view punishment cannot have a definite end even if there are third party enforcement agencies. As Juvenal was aware in the 2nd Century CE "Quis custodiet ipsos custodes?" - who will guard the guardians? Repeated game theorists know the answer to that question: they must guard each other and for that to be possible the game should not have a definite ending.

Several considerations point to a model of peer discipline. As argued by Dixit [17], using the formal legal system is costly (in terms of money, time and uncertainty of outcomes) and requires verifiable information which is not always available. Moreover:

- In many settings third party enforcement is not even feasible - for example, in the case of criminal gangs, or collusive oligopolists concerned about anti-trust law. In other contexts the underlying action that may be enforced through peer punishment may be membership in an organization - for example, members of an interest group may through peer discipline encourage each other to pay dues and join a lobbying group. Contractual relations are irrelevant where the action undertaken is signature of the contract itself.

- In some settings the activities encouraged through peer discipline may be inimical to the ostensible interests of the organization. For example, a corrupt police force may collude to protect corrupt law officers by a "code of silence" that for obvious reasons cannot easily be enforced by third parties.

- Even if contracting enforced by third parties is possible, the writing and enforcing of contracts is expensive - this is especially likely to be the case in a group with many people.[3] Moreover, contracts are not a magic wand. Somebody must determine whether or not the contract has been violated and if legal sanctions are required, and somebody else must check that this determination was made properly, especially if it is decided that no punishment is called for. Hence, while contracts with third parties may reduce the cost to the group of enforcement, it does not alleviate the need for peer discipline.

The type of scheme we study - group members who seamlessly monitor each other - is frequently seen in practice. Besides the procedures observed by Ostrom in her work on common property management, it is the type of enforcement mechanism used by modern internet groups such as Wikipedia and Slashdot. It is also typical of how criminal organizations operate - failure to carry out a punishment is itself a punishable offense in the typical codes of criminal conduct. By contrast enforcement mechanisms that involve only a single round of punishment tend to be less effective: in cases such as the Better Business Bureau - a single agency charged with monitoring business behavior - we find numerous complaints about the organization.[4]

The calculation of equilibria of our setting is essentially based on the strongly symmetric computation of Abreu, Pearce and Stachetti [1].[5] When punishment is costly to the punisher the Nash equilibria of the "base game" will also be equilibria of the punishment game in which the signals are ignored and nobody ever punishes. But depending on the parameters of the cost of carrying out and receiving punishments and the quality of information there can be other "more cooperative" equilibria. We characterize equilibria of the peer-punishment game, and derive the implementation that maximizes group utility. This gives rise to a simple expression for the "cost of peer punishment" that can be used to analyze the effectiveness of groups.

The use of punishments and rewards to induce desired behavior over basic actions is not a new idea. It is the basis of the efficiency wage model, for example as described in Shapiro and Stiglitz [52], and also the models of collusion proofness, for example, as described by Laffont [32]. Generally speaking, these models have not had costs associated with enforcement - in the efficiency wage model there is generally no punishment on the equilibrium path, while in the Laffont model punishments and rewards take the form of transfer payments so that there is no net cost. In practice however, there is punishment on the equilibrium path - as in Laffont - but practical forms of punishment, such as exclusion, generally have a net cost associated with them. In our setup we allow the possibility of costs on the equilibrium path, and in addition model the enforcement of first

[3]Li [38] appears to argue the opposite. However we rarely see large groups in lobbying organization bound together by contractual relationships, while for individual employment and business arrangements involving small numbers of individuals are typically contractual when contractual arrangements are feasible.

[4]The 17th century pirates constitute an interesting exception. Their decisions had to be taken quite swiftly, and to avoid abuses they had a balance of power whereby abuses of the captain were judged by the quartermaster. Of course the matter ended up there, in at most two rounds of judgment.

[5]The highly structured nature of the game avoids the types of complications found in more general repeated games as described in Fudenberg Levine and Maskin [22] and Sugaya [54].

stage punishments through subsequent actual rounds of auditing rather than through commitment to carry out punishments (or transfers).

To determine group effectiveness we analyze willingness to pay, first in a context where a single group bids for a given prize. Empirical results on the relation between group size and strength in this case are mixed, see the survey by Potters and Sloof [51]. The theory indicates that the results differ depending on the rivalry of the prize. Our main findings are the following:

- In the case where prizes are non-rival, so that all group members receive the full benefit of the prize, a group is stronger the larger it is.
- In the case of rival prizes, where each group member receives a proportional share of a prize of fixed value, the insight of Olson [45] is true only up to a point. When the group is small, willingness to pay increases with group size up to a threshold. Above this threshold size willingness to pay decreases with group size.[6]
- Homogeneous groups are stronger than heterogeneous ones.

We then turn to the case of two groups of different size that compete for transfers, as is typically the case in the political arena. Modeling bids as second price auctions we find that:

- For a prize of fixed size and of equal value to both groups there is an "optimal" group size, and the stronger group is the one "closer" to that size. Moreover, various inefficiencies are likely to emerge.
- When one of the groups - or the seller (a politician for example) - can take the initiative in setting the size of the prize, small groups are stronger in that they will obtain transfers from larger groups at low "political" cost, while large groups will opt out of the competition. When the seller sets the agenda the large group is favored and the transfers to the seller are more substantial. One interpretation of results such as Olson [45] who give evidence that small groups are often successful in practice is that this is because they take the initiative.

The model also makes specific quantitative predictions - that can be examined, for example, using the laboratory methods of Dal Bo [13].

There is of course a large literature on lobbying and other interest groups. Generally these models have fallen into four categories. Some treat the strength of the group as a black box and proceed with a working assumption, generally one in which strength decreases with size (Olson [45] Becker [6], Becker [9]), or in the case of Acemoglu [2] that strength increases with group size for a relatively small and a relatively large group.[7] A second class of models treats collusive groups as

[6]Of course different models may deliver different predictions. Dixit [17], Chapter 3 for example considers a two-period model of bilateral trading where misbehavior by a given individual in the first random match can be punished if her second partner knows that when the second match occurs. Assuming that this information is harder to come the larger the group yields the result that larger groups are less capable of enforcing 'fair' trade.

[7]This is consistent with our results, since we show that strength increases with size for a small group, and for a relatively large group, the opposition is small, and therefore weak. Acemoglu is also more focused on non-rival case, such as the National Rifle Association - where our results indicate that strength increases with size.

individuals - effectively ignoring internal incentive constraints - and focuses instead on information differences between the groups: examples are Nti [44], Persson and Tabellini [50], Kroszner and Stratman [31] Laffont and Tirole [34], Austen Smith and Wright [4], Banks and Weingast [5], Damania, Frederiksson and Mani [14], Green and Laffont [26], Laffont [33] and Di Porto, Persico and Sahuguent [16]. Dixit, Grossman and Helpman [18] is similar, but allows the endogenous possibility that groups either act non-collusively, or collusively as a single individual. A few papers assume that leaders of the group can distribute benefits differentially (this may or may not be what Olson [45] has in mind by "selective incentives"[8]) so that there is no public goods problem: see for example Nitzan and Ueda [43] and Uhlaner [55]. Finally Pecorino [49] and Lohmann [39] treat the problem of individual contribution within a group as a voluntary public goods contribution problem - a problem which other authors often indicate informally they view as key. As we shall point out, such a model does not behave very well in understanding the reason why groups of similar relative size, but very different absolute size (farmers in Belgium versus farmers in the United States) seem to be similarly effective.

Finally, we should mention a paper that goes in the direction opposite of ours: in Mitra [42] there is a fixed cost of forming a group - in contrast to our conclusion that there is a fixed cost per person in the group - so the more people there are the easier it is to overcome the fixed cost.

In the next two sections we lay out the model and characterize feasible and optimal discipline implementations. In section 4 we then apply the results to analyze the relation between group size, heterogeneity and strength. In section 5 we turn to competing groups and agenda setting. Section 6 concludes.

2. The Discipline Model

There are $N > 2$ identical players $i = 1, \ldots, N$ in the group. We first describe the *peer discipline environment,* in which group members monitor each other. Along the lines of Kandori [29]'s information systems we allow for the self-referential nature of punishment equilibria by supposing that the group plays a potentially unlimited number of rounds.

In the initial round - we call it round zero - group members choose primitive actions $a^i \in A$ representing production decisions and the like. We let $a^R \in A$ denote the action of a representative member of the group. By playing a^i against a^R player i gets payoffs $u(a^i, a^R)$, and also generates a binary good/bad signal $z_0^i \in \{0, 1\}$ where the probability of a bad signal 0 is $\pi_0(a^i, a^R)$.[9]

[8] Olson's concept is a bit slippery. He may have in mind people who are not in a group benefiting from the activity of the group - although this view of voluntary group participation runs somewhat counter to his notion of what constitutes a group. He argues that the group should devise auxiliary services (free lawyers, insurance) which "selectively" benefit only group members. It is not entirely clear why it would not be better to free ride on the group and pay directly for the auxiliary services, unless the group has some cost advantage in providing those services. In our setting members to not have the option of leaving the group - which is to say that they can not avoid being punished by group members. For example, farmers cannot avoid being shunned by neighboring farmers by refusing to join a farm association.

[9] Note that we do not specify what happens if more than one player deviates from a common action chosen by group members either for signal probabilities or for utility. As we will restrict attention to symmetric equilibria in

Following the primitive actions and the corresponding signals and utilities, a sequence of audit rounds $t = 1, 2, \ldots$ commences. During these rounds players are matched in pairs as *auditor* i and *auditee* j.[10] These matches may be *active* or *inactive*. A player cannot audit himself, so if $i = j$ a match is inactive. Second, auditing is possible only if it is the first audit round or if in the previous round the auditee was herself an auditor in an active match. Hence if the current auditee was an auditor in an inactive match in the previous round, the current match is also inactive. The remaining matches are active.

In round $t \geq 1$ in an active match an auditor i assigned to audit member j observes the signal $z^j_{t-1} \in \{0, 1\}$ of the behavior of the auditee in the previous round and has two choices r^i: to recommend punishment ($r^i = 1$) or not to recommend punishment ($r^i = 0$), while the auditee j does not get a move. Based on a member i's behavior as an auditor, another signal $z^i_t \in \{0, 1\}$ is generated. If the auditor recommends punishment on a bad signal or does not recommend punishment on a good signal, then the bad signal is generated with probability π; otherwise with probability $\pi^p \geq \pi$.

Note that the consequences of auditing decisions are assumed to be symmetric in the sense that $\pi = \Pr(z^i_t = 0 \mid z^j_{t-1} = 1, r^i = 0) = \Pr(z^i_t = 0 \mid z^j_{t-1} = 0, r^i = 1)$, similarly with π^p. We refer to this as *signal symmetry*. In other words the distribution of z^i_t depends only on whether the player "follows the social norm" (punish on bad signal or not punish on good) but not on which right thing she does. We should also mention that neither $\pi_0(a^i, a^R)$ nor π and π^p depend on the size of the group - in other words we assume that auditors are close to the auditees regardless of size.[11]

Remark. What is important is not what is true but what players think. Even if signal symmetry fails, it will be difficult for players to learn the consequences of the asymmetry. If their learning process merges information about the particular way in which the social norm was followed, they will act as if the distribution is symmetric. This can be formalized as the model of analogy based expectations developed in Jehiel [28]. Hence a broader interpretation of this assumption is that we are studying an analogy based equilibrium in which players do not distinguish different deviations from the social norm.

2.1. Costs and Punishments

Payoffs are additively separable between the initial primitive utilities and costs incurred or imposed during auditing - that is, we assume quasi-linearity. There is no discounting, as we imagine these rounds take place relatively quickly.

Following a recommendation of punishment a punishment is imposed, once or several times with some probability. Suppose that the auditor is i. The corresponding auditee then suffers an expected utility penalty of $P^i_t \geq 0$, the auditor incurs an expected utility cost of $C^i_t = \theta P^i_t$, and

which all group members play the same way, we only need to specify the pair (a^i, a^R) for determining equilibrium, so for notational simplicity and because it does not matter, we do not give a complete specification of the game.

[10] We do not have in mind that only a single person assesses whether the auditee has adhered to the social norm, rather that there is a single auditor who evaluates the evidence, possibly provided by many other people, and makes a determination of "guilt" or "innocence."

[11] This picture of localized monitoring is different in spirit from Dixit [17], which in terms of our model may be interpreted as arguing that the differences $\pi_0(a^i, a^R) - \pi_0(a^R)$ and $\pi^p - \pi$ decrease with group size.

in addition the other $N-2$ members of the group share an expected utility cost of $C_t^{iG} = \psi P_t^i$. [12] We assume that it is feasible to choose any value of P_t^i, but that the corresponding costs to the auditor and other group members are scaled accordingly, that is that θ and ψ are constants. We refer to this as a *linear punishment scheme*. It includes the case in which there is one particular punishment that may be repeated several times,[13] possibly randomly: for example a coin can be flipped so that with some probability the player is punished twice. The scheme is also consistent with the possibility of actual punishment being possible only once: for example losing a job or the death penalty. In this case P_t^i can be a probability (or "demerits") that is cumulated over time, with the actual punishment being issued at the time the game ends. Notice that it makes sense for probabilities of indivisible punishments that utility is additively separable.

With the exception of the auditee whose loss must be non-negative, the other costs may be either positive or negative, allowing the possibility that particular individuals may benefit from the punishment - for example if the auditee is demoted, some other group member may be promoted to fill the vacated position. However, we assume that the group as a whole cannot benefit from punishment, so that $1 + \theta + \psi \geq 0$.

Discussion. Punishment may have many possible forms. For example, if the auditee is fired from his job, removed from the organization or demoted this may have an adverse effect on the organization and hence lower the utility of those group members who are not directly involved in the punishment. On the other hand, some punishments may involve the collaboration of the entire group - for example shunning or refusing to speak to a group member. These costs are in a sense "avoidable" by an individual group member who may refuse to go along with the "social norm" of carrying out the punishment, and so avoid the cost of doing so. Rather than giving each player several decisions: whether to punish in a particular audit and also whether to carrying out their own "share" of a punishment, we compress the decision into a single decision "whether to follow the social norm." Roughly speaking then we regard the individual cost C_t^i as including the expected costs imposed by carrying out the player's individual share of punishments mandated in other matches as well as their own.

There are two specific limitations with linear punishment schemes. First, there may be more than one different type of punishment: for example you may either be fired or moved to a less nice office. These different punishments may have different values of θ, ψ. However, if multiple values of θ, ψ are available, after applying our results on group cost minimization for each value of θ, ψ we may then pick the combination of θ, ψ that yields the least cost. Second, the linear punishment schemes does not recognize that there may be an upper bound on the worst possible punishment. Earlier literature, such as work on efficiency wage, has focused on a higher level of informational perfection and the size of the upper bound. As a practical matter, the fact that the worst possible

[12] There may also be a utility cost to individuals outside the group, however as this plays no role in analyzing equilibrium we ignore it for the time being.

[13] Since we assume that a player can potentially be punished in each round, it makes sense that he can be punished several times in one round.

punishments that are used have diminished over time - at one time private organizations could use torture as a punishment; now torture is usually banned, and in many cases even governments are prohibited from using the death penalty, while private organizations are severely limited in the punishments that they can carry out - suggests that extreme punishments are generally not required to provide adequate incentives. Moreover, the set of applications we have in mind, such as membership or contributions to lobbying organizations, or the decision whether to accept a single bribe, generally do not bring great benefits or cost to the individual, so that large punishments are not generally relevant. Hence our focus not on worst possible punishments, but rather the costs to punisher and the group of carrying out the punishments - that is on the ratios θ, ψ rather than the maximum levels.

There are two other limitations of the model. First, it allows only punishments and not rewards, and second it does not allow for a fixed cost of carrying out an audit. This is primarily for simplicity. The use of rewards and not punishments is essentially a normalization. If it is possible to reward group members for good behavior, we can attribute the expected value of receiving a reward in every audit round to the initial utility $u(a^i, a^R)$, and consider not receiving the reward as a "punishment" in subsequent rounds. Take the example of an end of year bonus that might be withheld. We can on the one hand consider this as a reward. But we can equally well consider it part of income, and treat the withholding of the bonus as a punishment.

Fixed costs of carrying on audits is more complex because these costs can potentially be avoided by not carrying out the audit, and we need to consider how easy this is to observe. Our view is that it is relatively easy to see whether an audit has been conducted, and relatively more difficult to see if it has been done with due diligence, so we focus on strategic misrepresentation of recommendations rather than failing to carry out audits. If auditing is perfectly observable, then there is no cost in equilibrium of carrying out punishments since they are off the equilibrium path only, and we may ignore the incentive effects of avoiding the fixed cost of auditing. Although our computations of audit costs ignore fixed costs we may easily incorporate them into the model: if C^F is the fixed cost of carrying out an audit, since on the equilibrium path punishments are carried out with probability π we may simply add C^F/π to the group cost C^G and this gives the correct expected cost of the audit.[14]

2.2. Implementations and Equilibrium

An implementation specifies the procedure for matching and a profile of punishment costs. At the beginning of each audit round either the game ends or a *matching* and *punishment profile* are established. A matching is simply a 1-1 map from the set of players to the set of players with the argument being the auditor and the image being the auditee. A *punishment profile* is the assignment of P^i (and implicitly $C^i = \theta P^i, C^{iG} = \psi P^i$) to each auditor i's match.

[14]Notice that in this case if punishments are to be randomized it should be done by randomizing the audit itself rather than randomizing following the recommendation - the former leads to a proportionate reduction of the fixed cost, the latter does not.

An *implementation* is a determination at the beginning of each round whether the game ends, and if not of a matching and punishment profile. That determination can depend (randomly) on the history of previous matchings and punishment profiles. However it may not depend on the private signals or past recommendations of punishments.

A pure strategy for a player is an initial action and subsequently choices of signal dependent punishment recommendations. In general these choices depend both on the private history of the player (the signals he has seen and the recommendations he has made) as well as the public history, that is previous matchings and punishment profiles. A profile of strategies are a *Nash Equilibrium* if given the strategies of the others no player can improve his payoff. A *public strategy* depends only on the public history and a *perfect public equilibrium* is a Nash equilibrium in public strategies with the additional property that following every public history the strategies are a Nash equilibrium in the subsequent game. A *peer discipline equilibrium* is a perfect public equilibrium in which all players follow the strategy of punishing on the bad signal and not punishing on the good signal, and we say that we say that a^R is *incentive compatible in the implementation* if there is a peer discipline equilibrium with a^R as common initial action.

Note that the rules of the implementation are not strategically manipulable in the sense that future matching and punishment profiles do not depend on the recommendations made by auditors - the chances of future participation are exogenous, and an auditor need only care about the chances of being audited and punished in the next round.

An obvious and important fact is that a possible implementation is not to have any auditing rounds. In this case a^R is incentive compatible in the implementation if and only if it is a symmetric Nash equilibrium of the primitive game with payoffs $u(a^i, a^R)$. Our interest will be in establishing which non-Nash outcomes in the primitive game can be sustained by non-trivial peer punishment - and at what cost.

We should note that in the general case there is no assumption of anonymity and players may be treated differently based on their name. Indeed, it may be that some people are audited less frequently than others, so must be punished more when "caught." Or it may be that only a subset of the population carry out audits - or that there is a hierarchy, with only "managers" conducting audits, and only a subset of "top managers" conducting audits in the next round, and so forth.[15]

2.3. Enforceability

Recall that in the initial primitive round the probability of a "bad" signal 0 is $\pi_0(a^i, a^R)$ and utility is $u(a^i, a^R)$. Following the repeated game literature such as Fudenberg Levine and Maskin [22] we say that a^R is *enforceable* if there is some punishment scheme based on the signal such

[15] While we consider groups that collude to minimize costs to the group of punishment, we do not impose renegotiation proofness. In the case of groups that collude by communication in setting up rules, it is typically expensive and impractical to convene a meeting every time a decision must be made. We imagine that such meetings are infrequent relative to the frequency with which decisions are taken. Moreover, in some cases the equilibrium and mechanism that minimizes cost may evolve rather than be explicitly agreed to, or agreed to at a far distant time, possibly by people quite different from those currently in the group. These tendencies are strengthened in light of evidence showing that there is strong inertia in Nash equilibrium when many people are involved.

that a^R is incentive compatible. In the case of a binary signal, this means there must be some punishment P_1 such that for all a^i we have $u(a^R, a^R) - \pi_0(a^R, a^R)P_1 \geq u(a^i, a^R) - \pi_0(a^i, a^R)P_1$. If for all a^i we have $u(a^i, a^R) - u(a^R, a^R) \leq 0$ we say that a^R is *static Nash.* This case is not terribly interesting since no peer discipline is required to implement it as an outcome.

We now characterize enforceability. Let $\sigma(a^i, a^R) \equiv \operatorname{sgn}(\pi_0(a^i, a^R) - \pi_0(a^R, a^R))$. For $\sigma(a^i, a^R) = 0$, that is for actions indistinguishable from a^R, if $u(a^i, a^R) = u(a^R, a^R)$ define the *gain function* $\tilde{G}(a^i, a^R) = 0$, otherwise $\tilde{G}(a^i, a^R) = \operatorname{sgn}(u(a^i, a^R) - u(a^R, a^R)) \cdot \infty$. For actions that are distinguishable from a^R define the gain function to be

$$\tilde{G}(a^i, a^R) = \frac{u(a^i, a^R) - u(a^R, a^R)}{\pi_0(a^i, a^R) - \pi_0(a^R, a^R)}.$$

Let now $G(a^R) \equiv \max_{\sigma(a^i, a^R) \geq 0} \tilde{G}(a^i, a^R)$ and $G^-(a^R) \equiv \min_{\sigma(a^i, a^R) < 0} \tilde{G}(a^i, a^R)$. Note that $G^-(a^R) \geq 0$ if and only if $\sigma(a^i, a^R) \geq 0$ for all a^i with $u(a^i, a^R) - u(a^R, a^R) > 0$.

Lemma 1. *The group action a^R is enforceable with the punishment $P_1 \geq 0$ if and only if* $\max\{0, G(a^R)\} \leq P_1 \leq G^-(a^R)$, *hence it is enforceable if and only if* $\max\{0, G(a^R)\} \leq G^-(a^R)$.

Proof. Spelling out the inequality $u(a^R, a^R) - \pi_0(a^R, a^R)P_1 \geq u(a^i, a^R) - \pi_0(a^i, a^R)P_1$ and applying the definitions just given yields that a^R is enforceable with punishment $P_1 \geq 0$ if and only if $\max\{0, G(a^R)\} \leq P_1 \leq G^-(a^R)$. The conclusion follows. □

Observe that enforceability only concerns the first audit round hence does not imply peer discipline in the full sequence of audit rounds: an agent will deviate from an enforceable a^R if she thinks that punishment will not be inflicted in the future. For peer discipline equilibrium we need more.

3. Feasible and Optimal Implementations

We first examine the simple case of a two-stage implementation. This both serves to illustrate how the model works, and, as we shall see, as far as optimality goes, is perfectly general. In the *two stage implementation* a single punishment level P_1 is chosen along with a probability of continuing the game after each audit round $0 < \delta < 1$. When audit rounds take place, matchings are symmetric: for example we may place players randomly on the circle and have each player audit the adjacent opponent in the clockwise direction.

3.1. Characterization of Equilibrium in the Two-Stage Implementation

For notational simplicity, we set $\pi_0^R = \pi_0(a^R, a^R), G^R = G(a^R), u^R = u(a^R, a^R)$.

Theorem 1. *If the action a^R is not static Nash it can be incentive compatible in the two-stage implementation only if a^R is enforceable and $P_1 \geq \max\{G^R, |\theta P_1|/[\pi^p - \pi]\}$. In this case, a^R is incentive compatible if and only if*

$$P_1 \geq G^R\,,\ \delta \geq |\theta|/[\pi^p - \pi].$$

The resulting equilibrium utility level is

$$u^R - \left[\pi_0^R + (\delta/(1-\delta))\pi\right](1+\theta+\psi)P_1.$$

Proof. Recall that in the present case punishment is inflicted only if the game continues, which occurs with probability one in the initial round and δ in the other ones. The condition $P_1 \geq G^R$ thus follows from Lemma 1.

In an audit round, if you are an auditor and see a bad signal you are willing to punish if and only if the cost you would save by deviating is no greater than the extra expected punishment you would get next round, that is $C_1 \leq \delta[\pi^p - \pi]P_1$, or $\theta \leq \delta[\pi^p - \pi]$. Similarly, if you are an auditor and see a good signal you are willing not to punish if and only if $\theta \geq -\delta[\pi^p - \pi]$. Combining these two gives the second incentive constraint.

The necessary conditions follow from $\delta \leq 1$. The expected present values on the equilibrium path are then computed by summing the geometric series given the parameters. Indeed the individual player gets θP_1 as an auditor and P_1 as an auditee, each in the event of a bad signal; the group cost term is ψP_1, equal to $\psi P_1/(N-2)$ times the number of times it is incurred which is $N-2$ (the audits where the player does not participate). □

Notice that equilibrium utility for group members goes to minus infinity as $\delta \to 1$: this is because punishment occur essentially forever - it can be viewed as a kind of Hatfield and McCoy equilibrium.[16]

3.2. Optimal Punishment Plans in the Two-stage Case

We now consider choosing P_1, δ to maximize the utility of a representative group member for a given initial action a^R. The condition $P_1 \geq \max\{G^R, |C_1|/[\pi^p - \pi]\}$ is maintained.

Theorem 2. *The non static-Nash enforceable initial action a^R is incentive compatible for some two-stage implementation if and only if $|\theta| < \pi^p - \pi$. In this case to maximize the average expected utility of the group over two-stage implementations it is necessary and sufficient that the incentive constraints hold with equality, that is*

$$P_1 = G^R\,,\ \delta = \frac{|\theta|}{\pi^p - \pi}$$

The equilibrium utility level is

$$u^R - \left[\pi_0^R + \frac{|\theta|}{[\pi^p - \pi] - |\theta|}\pi\right](1+\theta+\psi)\,G^R.$$

Proof. From Theorem 1 the objective is decreasing in both P_1, δ hence it is maximized when these parameters are minimized - that is, when the constraints bind. □

Remark. Observe that we can always increase the δ's slightly and obtain an equilibrium that is strict in all the audit rounds at the price of a small reduction in group welfare. Such an equilibrium

[16] Notice also that if $1+\theta+\psi > 0$ and $\theta, \psi > 0$ these equilibria are not renegotiation proof - in this case all group members will prefer that the punishments not be carried out. If either θ or ψ are negative then after the recommendations are made at least some group members will wish to carry out the punishments, while before the signals are seen all will agree to dispense with auditing rounds. See also footnote 15.

can be more robust as it does not require individuals to "make the right choice" when indifferent. Notice however that, as we shall see below, it is not always possible to avoid indifference in the first period.

3.3. The General Linear Case

Theorem 2 in fact holds not only for all two-stage implementation, but as shown in the Appendix, for all implementations.

Theorem 3. *The non static-Nash enforceable initial action a^R is incentive compatible for some implementation if and only if $|\theta| < \pi^p - \pi$. In this case to maximize the per capita expected utility of the group it is necessary and sufficient that the incentive constraints hold with equality for each positive probability public history. The per capita expected equilibrium utility level per person is*

$$u^R - \left[\pi_0^R + \frac{|\theta|}{[\pi^p - \pi] - |\theta|}\pi\right](\theta + \psi + 1)G^R.$$

The expression

$$\Pi \equiv \left[\frac{|\theta|}{[\pi^p - \pi] - |\theta|}\pi\right]$$

can be viewed as a measure of the frequency of peer punishment on the equilibrium path, and we can rewrite per capita group utility as

$$u^R - (\pi_0^R + \Pi)(\theta + \psi + 1)G^R.$$

Notice that even if θ is zero, there is still a cost $\left[\pi_0^R(\psi + 1)\right] G^R$ from the need to punish in the first round.

4. Group Size and the Strength of Groups

What determines strength of a group? A simple measure of group effectiveness is its ability to mobilize resources. In this section we analyze a simple model of willingness to pay and focus particularly on the role of group size in determining strength. This group effectiveness can have a variety of interpretations - for example the group might be attempting to corrupt a politician as in Ades and DiTella [3] or Slinko and Yakovlev [53], or it could be a consortium bidding on a contract.

Throughout this analysis we will maintain the assumption that $|\theta| < \pi^p - \pi$ so that peer punishment is feasible. We study simple model of linear cost of effort and a prize worth S that will be divided equally among the group, each group member getting a benefit of $s = S/N$. The question we ask is: how much effort is the group willing to provide to get the prize? To determine willingness to provide effort, we use a Becker-DeGroot-Marschak (BDM) elicitation procedure.[17] The bid itself is a commitment to an implementation and basic actions that are incentive compatible with respect to that implementation; this includes also the possibility of not using peer discipline

[17] Basically this is a second price auction. In the political economy setting the all-pay auction studied, for example, by Krishna and Morgan [30] may be more relevant, but here we are trying to abstract from strategic considerations.

at all, but having instead (incentive compatible) voluntary contributions. Effort is provided only after the bid is accepted; otherwise the situation would be strategically one of an all-pay auction. Hence we have in mind a lobbyist who goes to a politician and says "my group will provide so many campaign contributions and provide so many volunteers in your next election if you provide us with S." We will consider also the case in which the prize can be rescinded if a threshold effort level is not provided.

Effort may be divisible or indivisible. Strategically this makes a difference: when effort is divisible, everyone can contribute equally a small amount, and it is relatively easy to monitor whether individuals made the agreed upon contribution. As a practical matter effort is not indivisible: lobbying, protesting, bribing and so forth require an overhead cost of thinking about and organizing oneself to participate in the activities. In general it is not feasible to spend two minutes a year contributing to a group effort in an effective way. Hence we will focus on the case where each group member j can provide either 0 or 1 unit of effort: $e^j \in \{0, 1\}$.

With indivisible effort in order to bid B the group should appoint a subset of B members each to provide an effort level of 1. To maintain symmetry of effort provision across group members we assume this is done through a messaging technology by which the group sends messages to individuals indicating whether they are expected to contribute: each individual j receives an independent signal $\sigma^j = \{0, 1\}$ of whether or not to provide effort, where $\mu \equiv \mathrm{pr}(\sigma^j = 1) = B/N$. Notice that the actual effort level that will be provided is random - however the bid is evaluated according to the expected value. The BDM mechanism chooses a random number $\tilde{B}$ and if $\tilde{B} \leq B$ the bid is accepted. For reasons that will emerge below - groups may prefer a higher effort level to a lower one - when the bid is accepted we take $\tilde{B}$ to be a floor on effort - the group is free to deliver a higher level if it chooses to do so. To model imperfect monitoring we assume that auditors can tell whether or not the auditee has contributed effort, but observe whether or not they received a signal with noise. That is, it is observable whether or not the auditee turned up at the rally, but if he did not, he may say "I never got the phone call" and the auditor cannot perfectly determine the truth of this. Specifically, in the first audit round auditor i observes auditee j's action and a signal $s^j \in \{0, 1\}$ which is equal to σ^j with probability $1 - \epsilon$ and to the opposite $1 - \sigma^j$ with probability $\epsilon \leq 1/2$.

4.1. Enforceability and Signal Compression

So far we have allowed only binary signals of behavior in the initial round. Here the auditor observes a pair (e^j, s^j) where e^j is the effort provided by the auditee and s^j is the auditors garbled version of the signal received by the auditee. Hence there are four - rather than two - possible values of the signal. However: in general if enforceability is satisfied it is always possible using randomization to reduce a multi-value signal to a binary signal without consequence for the cost of punishment or incentives. Rather than proving this in general the idea becomes clear by working it through in this specific instance.

In the primitive action game each group member has the following four possible strategies corresponding to chosen effort level depending on signal value. [01]: do not contribute on $\sigma = 0$, contribute on $\sigma = 1$; [00]: never contribute; [11]: always contribute; [10]: contribute on $\sigma = 0$, do not contribute on 1. We are interested in the enforceability of [01]. Since we are assuming linear cost of effort, individual expected utilities of the alternative strategies are easily computed. The group member gets the expected value of s given the bid, call it $\bar{s}$, minus probability of effort. So for example utility of conforming to [01] is $u([01],[01]) = \bar{s} - \mu$; while $u([10],[01]) = \bar{s} - (1-\mu)$; and so forth.

Since there are four possible signal combinations which the auditor may observe, $(e^j, s^j) = (1,1),(0,1),(1,0),(0,0)$, there are four possible corresponding punishments $P_{11}, P_{01}, P_{10}, P_{00}$. Let $P_1 = \max\{P_{00}, P_{01}, P_{10}, P_{11}\}$ and let $\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}$ be defined by $P_{00} = \beta_{00}P_1, P_{01} = \beta_{01}P_1, P_{10} = \beta_{10}P_1, P_{11} = \beta_{11}P_1$. Note that the β's may be interpreted as probabilities, and the one corresponding to the highest P is equal to 1. The expected value of punishment contingent on strategy choice is then

$$\begin{aligned}
&[01]\text{: } [\mu(1-\epsilon)\beta_{11} + (1-\mu)\epsilon\beta_{01} + \mu\epsilon\beta_{10} + (1-\mu)(1-\epsilon)\beta_{00}]P_1 \equiv \pi_0([01],[01])P_1 \\
&[00]\text{: } [(\mu(1-\epsilon) + (1-\mu)\epsilon)\,\beta_{01} + (\mu\epsilon + (1-\mu)(1-\epsilon))\,\beta_{00}]P_1 \quad \equiv \pi_0([00],[01])P_1 \\
&[11]\text{: } [(\mu(1-\epsilon) + (1-\mu)\epsilon)\,\beta_{11} + (\mu\epsilon + (1-\mu)(1-\epsilon))\,\beta_{10}]P_1 \quad \equiv \pi_0([11],[01])P_1 \\
&[10]\text{: } [(1-\mu)\epsilon\beta_{11}, +\mu(1-\epsilon)\beta_{01} + (1-\mu)(1-\epsilon)\beta_{10} + \mu\epsilon\beta_{00}]P_1 \equiv \pi_0([10],[01])P_1
\end{aligned}$$

where the π_0's are the effective probabilities of bad signal. More precisely, the underlying signal with four outcomes is compressed to a single binary signal where if the underlying signal has the value (e^j, s^j) the bad signal has probability $\beta_{e^j s^j}$; and punishment occurs if a bad signal is generated.

It remains to say how the probabilities β are chosen. Concordant with our earlier assumption that the group chooses an implementation to maximize per capita group utility, we imagine that the β's are chosen to maximize

$$s - \mu - [\pi_0([01],[01]) + \Pi](\theta + \psi + 1)G^R$$

where G^R is determined by all the π_0's. The following is proved in the Appendix.

Theorem 4. *If*

$$\Pi \leq \frac{(1-\mu)(1-2\epsilon)}{\epsilon}$$

group utility maximization implies $\beta_{00} = \beta_{10} = \beta_{11} = 0, \beta_{01} = 1$. *Otherwise* $\beta_{10} = \beta_{11} = 0$ *and* $\beta_{00} = \beta_{01} = 1$.

Per capita group utility is equal to

$$s - \mu - (\theta + \psi + 1) \cdot \min\{\frac{(1-\mu)\epsilon + \Pi}{1-\epsilon}, (1-\mu) + \Pi\}$$

Thus there are two cases depending on the condition $\epsilon\Pi \leq (1-\mu)(1-2\epsilon)$. If it holds then punishment should occur only when the auditors signal shows a contribution was supposed to be made and it was not. If it fails then punishment should occur if a contribution was not made regardless of whether the auditor thinks it should have been. It may be slightly puzzling in the latter case that the optimal strategy is in fact to contribute only when asked to - would it not be better to contribute regardless of whether or not asked to? But the punishment is calibrated so that the group member is exactly indifferent between contributing and not contributing, so is content to contribute exactly when asked to. Notice that here, unlike in the audit rounds, indifference plays a key role.

This theorem has a surprising implication: group utility may be increasing in μ - that is, it may prefer to provide a higher level of effort. To understand why, examine the expression for per capita group utility. Effort level μ has two effects. First, as μ goes up everyone has to contribute a greater amount of expected effort. This lowers utility accordingly. On the other hand, as μ goes up the cost of punishing the basic actions is proportional to $1-\mu$, and this goes down. To understand why, consider the case $\mu = 1$ in which case there is no cost of punishing the basic action. This is transparent: everyone is asked to contribute and punishment only occurs when there is a failure to contribute - which never happens on the equilibrium path. [18] By contrast when μ is smaller sometimes people are erroneously punished, with a corresponding social cost. The next result shows that decreased cost due to lower monitoring cost as μ rises may exceed the increased cost of providing the effort.

Theorem 5. *Group utility is increasing in μ if and only if $(\theta + \psi + 2)\epsilon \geq 1$ or*

$$\mu \geq 1 - \frac{\Pi\epsilon}{1-2\epsilon}.$$

Proof. See the proof of Theorem 6 in the Appendix. □

Notice that there are two cases. In the former case the cost of punishment is high - so dominates the cost of effort. In the second case the effort level is high and the noise ϵ is large, so that the frequency of false signals is high.

Remark. The fact that utility may be increasing in μ has a subtle implication for policy. Consider a policy to discourage or weaken an organization such as a rent-seeking lobbyist or criminal gang. It is natural to engage in policies that raise the cost of punishment - for example, if a criminal gang punishes members by murdering them, then by cracking down with sweeps when the murder rate is high (large ψ) or vigorously prosecuting murderers (large θ). However if the increased cost to the gang of peer discipline causes a shift from the regime in which utility is decreasing in μ to one in which utility is increasing in μ it could actually trigger an increase in gang activity.

4.2. Willingness to Pay and Group Size

We now turn to the issue of willingness to pay in a BDM mechanism. If the group bids an effort level $B = N\mu$ group utility is

$$S - N\mu - N(\theta + \psi + 1)\min\left\{\frac{(1-\mu)\epsilon + \Pi}{1-\epsilon}, (1-\mu) + \Pi\right\}.$$

[18] Note however, that there is still the overhead of operating the peer punishment system $\Pi(\theta+\psi+1)/(1-\epsilon)$. This is required as a deterrent so that people will actually choose to contribute.

The alternative is not to bid. Hence the group should bid the highest value of $N\mu$ that gives positive utility, or not bid at all if there is no such value. The following is our main first result.

Theorem 6. *The bid function is single peaked as a function of N. Define*

$$\hat{N} = \frac{S}{1 + (\theta + \psi + 1)\Pi} \qquad \hat{\hat{N}} = \frac{S}{(\theta + \psi + 1)\frac{\epsilon + \Pi}{1 - \epsilon}}$$

For $N \leq \hat{N}$ the bid is N. If $\epsilon(\theta + \psi + 2) \leq 1$ and

$$(\theta + \psi + 1)\frac{\epsilon(1 + \Pi)}{1 - \epsilon} < 1$$

then $\hat{\hat{N}} > \hat{N}$, and for $\hat{N} \leq N \leq \hat{\hat{N}}$ the bid is given by the following function, decreasing in N:

$$\tilde{B} \equiv \frac{(1 - \epsilon)S - N\big[(\epsilon + \Pi)(\theta + \psi + 1)\big]}{1 - \epsilon(\theta + \psi + 2)}$$

In all other cases the group does not bid.

Proof. See the Appendix. □

Observe that the group always bids less than the full value S, and that $\hat{N}$ is "optimal" in the sense that for the specific prize S a group of size $\hat{N}$ will make the highest possible bid. However the word "optimal" should be interpreted cautiously here: the size of a group may be determined from many considerations - it may have to participate in many different auctions and they may not all be second price auctions, for example. Moreover, so far we have assumed homogeneous groups, which puts an additional constraint on the size of a group.

With that cautionary note, we observe that "optimal" group size $\hat{N}$ increases linearly with the size of the prize S but at a rate that is less than 1. Moreover $\hat{N}$ falls as the cost of peer punishment $\Pi(\psi + \theta + 1)$ rises.

Remark. The general result that there is an "optimal" group size neither too big or too small holds for general increasing marginal cost of effort provision.[19] Here we have greatly simplified computations and found a simple closed form solution by a sharp form of increasing marginal cost of effort provision: it is constant up to the capacity constraint, then is infinite.

4.3. The Case of a Non-rival Prize

Lobbying often concerns non-rival goods, as in the case of laws affecting all group members equally. Here a prize is worth S - not S/N - to each member of the group. In this case, as the reader can guess, the result is a group is stronger the larger it is. For simplicity we shall specialize to the *low punishment cost case* in which $\epsilon(\theta + \psi + 2) \leq 1$ and

$$(\theta + \psi + 1)\frac{\epsilon(1 + \Pi)}{1 - \epsilon} < 1.$$

[19] With the indivisibility of effort provision there cannot be increasing marginal cost for individuals, but the group will face increasing marginal cost if individual costs of effort provision are independently drawn from a distribution, and the "message" is a threshold of cost below which members are expected to contribute full effort.

The result is the following:

Corollary 1. *Assume the prize is non-rival. Then the group bid increases linearly with N. Precisely, let*

$$S^l = \frac{(\epsilon + \Pi)(\theta + \psi + 1)}{1 - \epsilon} \quad S^h = \frac{1 - \epsilon + \Pi(\theta + \psi + 1)}{1 - \epsilon}$$

then

$$\tilde{B} = \begin{cases} 0 & if\ S \leq S^l \\ \frac{(1-\epsilon)S - [(\epsilon+\Pi)(\theta+\psi+1)]}{1-\epsilon(\theta+\psi+2)} N & if\ S^l < S < S^h \\ N & if\ S \geq S^h \end{cases}$$

Proof. First observe that $\epsilon(\theta + \psi + 2) < 1$ is the same as $\epsilon(\theta + \psi + 1) < 1 - \epsilon$ so that $S^l < S^h$. Following the same argument as in Theorem 6 the bid becomes

$$\tilde{B} = \frac{(1 - \epsilon)S - [(\epsilon + \Pi)(\theta + \psi + 1)]}{1 - \epsilon(\theta + \psi + 2)} N$$

Taking into account the constraint $0 \leq \mu \leq 1$ simple algebra gives the result. □

This is not particularly deep. When additional members do not reduce the per-capital value of the prize the per-capita bid is constant, and all those interested in the prize should be part of the lobby, and the more they are the stronger the group will be. The thresholds S^l and S^h will go up with discipline cost $\Pi(\theta + \psi + 1)$, but this is all there is to it.

4.4. Heterogeneity

To model a heterogeneous group, we assume that there are N different prizes. If the group wins one of the prizes is chosen at random. Player i is primarily interested in prize i which she values at $s = S/N$ and less and less in prizes preferred by more "distant" individuals. Specifically, for each player i and distance $d = 0, \ldots, N - 1$ assume that there is exactly one player i' at distance d, and the type at distance 0 is type i.[20] These distances may be interpreted as an index of "affinity," not necessarily symmetric. Let α_d for $d = 0, \ldots, N - 1$ be non-increasing weights, with $\alpha_0 = 1$. These weights may or may not be negative for large d. The value for player i of the prize targeted at the player at distance d from her is $\alpha_d s$, so the expected value of the prize to individual i is

$$\frac{\sum_{d=0}^{N-1} \alpha_d s}{N}$$

An obvious definition of greater heterogeneity in this context, means that players care less for prizes at great distances - that we define weights α' to exhibit *greater heterogeneity* than α if for $d > 0$ weights $\alpha'_d < \alpha_d$. Since this simply reduces the value of the prize to every individual, we see immediately from Theorem 6 that the group bids less. Heterogeneity weakens the group.

[20] For example players are arranged on the circle, and distance is measured in the clockwise direction.

4.5. Voluntary Contributions With Prizes That Can Be Withdrawn

We now compare the predictions of the previous section with those of the alternative "usual" voluntary public goods contribution model. Realized group effort level $\tilde{E}$ is a noisy signal of intended group effort μN. When a peer mechanism is implemented there is no useful information in this signal since individual contributions are directly observed. However, if it is possible to withdraw the prize based on $\tilde{E}$ a voluntary contribution mechanism may be used to provide an incentive for a positive level of contributions.

Suppose the group commits to a voluntary level of per capita contribution μ. It does so by sending independent messages with probability μ indicating that individual effort should be provided. Such a bid is feasible if it is in fact incentive compatible to follow those recommendations. If these recommendations are followed the realized effort level $\tilde{E}$ will follow a binomial with parameters μ as success probability and N as number of trials. If the prize cannot be withdrawn then there is no incentive to provide effort. Suppose instead that the bid includes a threshold $\hat{\mu}$ with the agreement that if effort level $\tilde{E}$ falls below $N\hat{\mu}$ the prize will be withdrawn.

Observe first that a necessary condition for voluntary contribution is $s \geq 1$. Otherwise even an individual who will lose the prize for sure if he does not contribute will not do so. Consider first the case $\mu = 1$, where everyone is called to contribute. If the threshold $\hat{\mu} = 1$ then the prize will be withdrawn if any single person fails to contribute regardless of N. Hence if $s \geq 1$ voluntary contribution by everyone is incentive compatible. However, this result is heavily dependent on the fact that the aggregate statistic $\tilde{E}$ is a perfect signal of individual behavior. In practice some people will be unable to contribute because of other obligations, health problems, and so forth and so on, and indeed the size of the group may not be known with perfect accuracy. A simple way to model this in the current framework is to assume that there is an upper bound $\overline{\mu} < 1$ and that it is infeasible to send signals with probability $\mu > \overline{\mu}$. The remaining probability $1 - \overline{\mu}$ may be thought of as "the chance that the person is unable to contribute." This leads to a radically different result:

Theorem 7. *For all s and $1 > \overline{\mu} > \underline{\mu} > 0$ there exists an $\overline{N}$ such that $N > \overline{N}$ implies that any incentive compatible $\mu < \underline{\mu}$.*

The key feature of this result is that the number $\overline{N}$ depends on s rather than S. In particular, if the per capita size of the reward s is fixed, then if the group size is large enough there will be no voluntary contribution. By contrast in the peer discipline case the threshold $\hat{N}$ is linear in S.[21] If the per capita size of reward s is fixed and the group size is expanded, then the threshold $\hat{N}$ increases in proportion. Think of this in terms of, for example, the farm lobby seeking a per farmer subsidy of s. In the voluntary contributions model, a large country with many farmers would have an ineffective farm lobby, since N is very large. In the peer discipline case the size of the country does not matter for the effectiveness of the farm lobby. As we observe that the farm lobby has

[21] With the upper bound on μ the maximum bid $\hat{N}$ changes because instead of $\tilde{B} > N$ we must consider $\tilde{B} > \bar{\mu}N$ and the optimal group size becomes $\hat{N} = (1-\epsilon)S/[\bar{\mu} - \epsilon(2\bar{\mu} - 1 - (1-\mu)(\theta + \psi)) + \kappa]$ which is increasing in $\overline{\mu}$ but more important remains linear in S.

similar effectiveness in large countries such as the United States and smaller ones such as Japan, there is some reason to believe that the absolute size of the group is not important. Moreover, in the U.S. there are about 3 million farmers and 2 million farms. In particular, the public goods problem that must be overcome by the group if there are to be voluntary contributions seems nearly insurmountable.

Remark. One alternative explanation as to why the effectiveness of a group does not depend on the absolute size for a fixed per capita prize is that contributions are voluntary but players are altruistic. A standard model of altruism distinguishes between personal utility u_i for player i and and $u_G = \sum_{j \neq i} u_j$ for other group members, and makes the hypothesis that the individual maximizes a weighted sum $u_i + \alpha_N u_G$. In the voluntary contribution model this amounts to having the individual prize valued at $(1 + \alpha_N N)s$. There are two cases. If $\alpha_N N$ remains bounded for large N then the basic result goes through unchanged. However, if $\alpha_N N$ grows without bound then the result can be overturned. However it is implausible that $\alpha_N N$ grows without bound. Consider the problem of voluntary provision of taxes τ_i to provide a public good. Let $\tau_G = \sum_{j \neq i} \tau_j$ and assume the value of the public good is $v(\tau_i + \tau_G)$. The individual objective function is $(1+\alpha_N N)v(\tau_i + \tau_G) - \tau_i$. The corresponding first order condition for voluntary contribution is $v'(\tau_i + \tau_G) = 1/(1 + \alpha_N N)$, so in a large group voluntary contributions are so large that v' is nearly zero, that is, the saturation point for the public good is reached. However, the condition for efficiency is that $v'(\tau_i + \tau_G) = 1$ so there are far too many contributions. Hence there should be no need in a large country like the United States to tax in order to provide (say) for defense spending - indeed, the problem should be to discourage the inefficiently high level of private contributions. This is the case even if we assume that individuals care only for people of a similar type - even if farmers care only for other farmers, the fact that there 3 million of them would probably be adequate to raise enough money to pay for the defense budget through voluntary contributions. The point is that if farmers are contributing voluntarily to farm lobbying because they are altruistic towards other farmers, they should be very happy indeed to contribute voluntarily to the national defense. Of course public goods contributions are not carried out through voluntary contributions, but using punishment for non-compliance.

Remark. Theorem 7 is a variation on the literature on the negligibility of agents, especially Mailath and Postlewaite [40], Levine and Pesendorfer [36] and Fudenberg, Levine and Pesendorer [23]. This literature asks when a Stackelberg leader can punish a group of N followers based on an aggregate signal to induce a desired behavior. The result - which is quite robust - is that this is not possible if there is noise in the individual signal of behavior that is aggregated.

5. Competing Groups

Because the BDM mechanism is compatible with the second price auction, we can extend the previous analysis to consider two groups that compete in a second price auction. We suppose that the groups are identical except for their size. We are interested in which group will win the bidding, and in the resulting efficiency implications. We are also interested in endogenizing the size of the prize - we consider how the different groups and the seller might choose the size of the prize.

We shall work again in the *low punishment cost case* in which $\epsilon(\theta + \psi + 2) \leq 1$ and

$$(\theta + \psi + 1)\frac{\epsilon(1 + \Pi)}{1 - \epsilon} < 1.$$

Let $N_1 < N_2$ be the group size, and consider first the case where the prize has a fixed value S, the same for both groups. From Theorem 6 we immediately find

Corollary 2. *When both groups value the prize equally*

1. *if* $N_1 \geq \hat{\hat{N}}$ *both groups bid zero; otherwise*
1. *if* $N_1 \geq \hat{N}$ *or* $N_2 \geq \hat{\hat{N}}$ *the small group wins*
2. *if* $N_2 \leq \hat{N}$ *the large group wins*
3. *if* $N_1 < \hat{N} < N_2 \leq \hat{\hat{N}}$ *there are cases where either group may win*

Basically this shows that it is advantageous to be small but not too small - that it is advantageous to be close to $\hat{N}$.

Suppose now that the two groups value the prize differently, bearing in mind that $\hat{N} = \hat{N}(S)$. If one group values the prize more, that increases the chances of winning, but a group closer to $\hat{N}$ may never-the-less win despite the fact it values the prize less. We can identify three different kinds of efficiency losses as a result of this bidding procedure.

1. If the prize is valued differentially by the two groups the prize can to the wrong group.

2. If the winning group pays a positive price, then there is a social cost to the punishment that takes place on the equilibrium path.

3. If the winning group pays a positive price the effort provided may be socially costly. It seems usual in the literature about lobbying - for example Hillman and Samet [27] or Ehrlick and Lui [20] - that this effort is socially costly. Whether this is true in general depends on the value of the effort to the group versus the recipient of the effort: if the effort is “getting out the vote” it is a social loss; if the effort is ”shining the shoes of the politician” and he likes his shoes shined it may be partially a transfer payment (hence no efficiency loss) depending on the value to him of having his shoes shined, versus the value of the effort to the provider.

Related to the final point - there is an important issue with laws designed to prevent bribery. They may simply result in inefficient bribes. That is, cash transfers are relatively efficient, while “shoe shining” isn’t. If the bribe is simply that I work at my highest valued occupation and give you some of the proceeds there isn’t an efficiency loss. If the bribe requires me to work at a lower valued occupation (“shoe shining” when I’m a heart surgeon) then there is an efficiency loss from my working in a less valuable occupation.

5.1. Agenda Setting: Endogenous Prizes

We continue to study two groups of size $N_1 < N_2$ competing for a prize in a second price auction. Now however, as in the political area, we will assume that the prize is a transfer payment between the groups and analyze agenda setting. Specifically we will consider respectively the agenda being set by the small group, by the large group, and by the seller - the politician. The agenda setter will determine two things: which group j will pay for the transfer, and how large the transfer - the prize S - is. Given the agenda the politician auctions the prize in a second price auction, collects the winning bid, and commands the designated group j to pay the prize (more precisely $\min\{S, N_j\}$) to the winner. Note that the winner may be group j itself; in this case they are paying the politician to avoid being taxed even more.

Obviously if it is one of the groups setting the agenda they will designate the other to pay - that is, they will propose the legislator a transfer from the other group to their advantage - and strategically choose the size of the prize so that they win the auction. The maximum group j can pay is N_j (when they provide full effort), so the largest prize that group 2 can choose is N_1 and the largest prize that group 1 can choose is N_2. The politician on the other hand can choose which group will make the transfer as well as the size of the prize.

We also assume that for a given utility all groups lexicographically prefer a smaller prize to a larger one. In case of a tie in the bidding, we impose the continuity requirement that the group that would win when the prize was slightly higher wins the tie. The following may be seen as the second main result of the paper.

Theorem 8. *We consider again the low punishment cost case. A transfer takes place only if the small group sets the agenda, in which case it sets the prize S to*

$$S_2^L \equiv (\theta + \psi + 1)\frac{\epsilon + \Pi}{1 - \epsilon}N_2,$$

bids a positive amount and pays zero; the large group pays $\min\{S_2^L, N_2\}$ to the small group. If the large group sets the agenda, it sets the prize equal to zero.

The winning group pays a positive amount only if the politician sets the agenda, in which case she chooses the large group to pay for the prize, which she sets equal to

$$\hat{S} \equiv \frac{N_1[1 - \epsilon(\theta + \psi + 2)] + N_2\left[(\epsilon + \Pi)(\theta + \psi + 1)\right]}{1 - \epsilon}$$

Both groups bid N_1 and the large group wins the bidding, so there is no transfer between the groups but simply a payment by the large group of N_1 to the politician.

Proof. If the small group sets the agenda it will call the highest prize for which the large group bids zero, that is the S for which $N_2 = \hat{\hat{N}}$, that is S_2^L. It will not set a higher prize because the transfer to their advantage goes up more slowly (slope 1) than the other group's bid (slope $(1-\epsilon)/[1-\epsilon(\theta+\psi+2)]$). The large group has nothing to win in this game because to win they would have to bid at least N_1 to the politician (from Corollary 2) to get at most N_1 from the small group. The politician on the other hand maximizes the lower bid (which is what she gets), whose upper bound is N_1. This is achieved by having both groups bid N_1, which occurs when the larger one bids that much. From the equation $\tilde{B}(N_2) = N_1$ we then get $\hat{S}$. □

It is useful to have an idea of how much bribes are actually paid. Bribes include campaign expenditure and also things like employment of officials after they leave office. Data on presidential campaigns from Floating Path [21] shows that this is quite small, less than 0.1% of GDP. This suggests that the agenda setting ability lies with the outside groups and not with the politicians: that politicians cannot simply threaten transfer payments in order to collect bribes. It also suggests that, in accordance with our result, in general it is small groups who obtain transfers from larger groups, at low "political" cost.

5.2. Differential Productivity

A group of size N may be more productive, so each unit of effort translates into A units of bid. If the prize S is in dollar units, then the prize in efficiency units is S/A and the dollar bid is derived by multiplying the effort bid by A so that in the low punishment cost case

$$A\tilde{B} \equiv A\frac{(1-\epsilon)(S/A) - N\big[(\epsilon+\Pi)(\theta+\psi+1)\big]}{1-\epsilon(\theta+\psi+2)} = \frac{(1-\epsilon)S - AN\big[(\epsilon+\Pi)(\theta+\psi+1)\big]}{1-\epsilon(\theta+\psi+2)}$$

which is identical to the dollar bid for a group of size AN and unitary productivity. Note that this determines also $\hat{N}$ and $\hat{\hat{N}}$, and that when $\epsilon(\theta+\psi+2) \geq 1$ the groups bid AN. Hence a group of sizes N and productivity A chooses the same bid as a group of size AN.

This says that a group that is twice as productive, but has half as many members is exactly as effective as a group twice as big but half as productive. It is the total resources controlled by the group that matters (in effect). You might think that a group twice as productive but half the size would be more effective. But it is not: the problem is that while it has half as many members, so that halves the fixed cost of peer discipline, the per-capita discipline costs it twice as much since the cost in effort is the same but each unit of effort has a double opportunity cost in dollars. Or to say it differently: the productive guys do not have to work as hard to produce a given bid, but it is pretty expensive for them to waste their time monitoring each other.

As an example, think about influence within an academic department. There are those who are very productive and those who are not. The productive ones can be more influential per unit time, but don't waste their time organizing their productive peers to lobby with the administration. So the department is run by the unproductive ones.

6. Conclusion

The broad issue we have addressed is the relation between size, cohesion and the strength of groups. Since the pioneering work of Olson [45] the presumption has been that smaller, more homogeneous groups may perform better owing to the increased easiness of enforcing group discipline. This idea is of course challenged by the observation that large groups may be harder to coordinate but have nonetheless more weight in the political arena.

To address this we have developed a model of group behavior with explicit account of individual incentives inside the group. The model seems more descriptively realistic than the voluntary contribution model, and makes more appropriate predictions with respect to large groups. To analyze the strength of the group we look at the amount of effort the group can enforce on its members in a second price auction. We have considered two different contexts. The first is one where the prize comes from outside a single group or a pair of competing groups, typically a politician; the second is a situation where two groups compete for a prize which is a transfer from a group to the other - at a cost to be paid to the politician.

In the first case we have found that Olson's original idea that small groups are stronger needs to be refined. It is true that larger groups bid more the larger the prize, but for a given prize there is an “optimal”, highest-bidding group size, which outperforms both smaller and larger groups. Moreover, the analysis reveals that when the same prize is valued differently by different competing groups inefficiencies in allocation of the prize are likely to emerge, and this has subtle implications for anti-corruption policy.

In the case of two groups contending a transfer granted by a politician from one group to another - where the politician awards the group which grants her a higher payoff - we find instead that indeed smaller groups have an advantage, and usually secure themselves the transfer from the larger group, at low “political” cost, by taking the initiative and strategically choosing the size of the transfer.

We have not pursued all aspects of the model, focusing largely on the issue of group size. The model has implications as well for the role of information in internal discipline, issues which we hope to examine in greater detail in future work. Also, we have treated the costs of punishment as exogenous. To the extent that they can take of form of transfer payments as in Laffont [33] the cost on the equilibrium path can be reduced. If the cost to the auditor can be made equal to zero ($\theta = 0$) and the cost to the auditee is a transfer payment to the rest of the group ($\psi = C^G/P = -1$) then the cost to the group of peer discipline is zero. This is analogous to the usual folk theorem in repeated games with private information, as in Fudenberg Levine and Maskin [22] where the discount factor approaching one makes it possible for future rewards and punishments to effectively act as transfer payments.

The peer discipline setting allows the consideration of different types of monitoring schemes. Trial by press or peer pressure to conform are simple informal schemes. Often more formal judicial schemes are used - for example firing an employee may require formal hearings. Formal schemes systematically consider evidence. On the one hand, this means that the signals of malfeasance are more accurate. On the other hand, a higher level of evidence is generally required, so while punishment of the innocent is less likely, punishment of the guilty is less likely as well. Using an underlying signal structure with thresholds such as that in Fudenberg and Levine [24] we can see how different formal and informal schemes will impact on collusion constraints. Another common method of reducing incentives for collusion is through rotation. Military and police organizations often rotate officials from one position and location to another and politicians have term limits. This can have the effect of degrading the quality of service provided while increasing the quality of audits.

Notice that the parameters of signal quality and the chances of being caught depend also on social characteristics. Groups that have been in existence for a long time and are suspicious of outsiders will be better able to provide incentives - and also harder to prevent from engaging in bad behavior - than shorter-lived more open groups. Through this channel, culture may have an impact on group incentives; and conversely as people specialize in ways most suited to the group they are in. Policy can also play a role - for example by rules against nepotism. There are many cross-cultural experiments on trust and other simple games - we would like also to see what impact culture has on our games played between groups.

Ultimately we may anticipate a better theory of mechanism design - a theory of mechanisms that are robust to group collusion. Colluding groups and their internal incentives are important for two reasons. For some groups - productive groups such as business firms, or regulatory agencies - efficiency will demand strong incentives. For other groups - rent seekers and lobbying groups - efficiency will demand weak incentives. While a formal theory of peer discipline is a start on such a theory, we need also to understand how groups coordinate or may fail to coordinate, and when and how groups are formed and membership determined.

Appendix: Proofs not included in text

Theorem. [3 in text] *The non static-Nash enforceable initial action a^R is incentive compatible for some implementation if and only if $|\theta| < [\pi^p - \pi]$. In this case to maximize the average expected utility of the group it is necessary and sufficient that the incentive constraints hold with equality for each positive probability public history. The average expected equilibrium utility level per person is*

$$u^R - \left[\pi_0^R + \frac{|\theta|}{[\pi^p - \pi] - |\theta|}\pi\right](\theta + \psi + 1)G^R.$$

Proof. By Theorem 2 the stated utility level is attained by a two-stage implementation. Hence it suffices to show that no incentive compatible implementation can attain a higher utility level.

The the objective function is

$$u(a^i, a^R) - (1/N)E_0 \sum_{t=1}^{\infty}\sum_{i=1}^{N}(\theta + \psi + 1)P_t^i$$

where E_0 is expectation conditional on information at beginning of round 0. Defining $\overline{P}_t = (1/N)E_0 \sum_{i=1}^{N} P_t^i$ we can further write this as

$$u(a^i, a^R) - E_0 \sum_{t=1}^{\infty}(\theta + \psi + 1)\overline{P}_t.$$

Fix any equilibrium and consider now the decision by auditor i to punish j on B and not punish on G in round t. Under our assumption the only consequence of this decision is the expected cost of punishment in this round $E_t C_t^i$ and the expected punishment conditional on getting a bad signal in the subsequent round $E_t P_{t+1}^{j(i)}$ where $j(i)$ is the unique person who audits i in period $t+1$ - note that this is a random variable. The decision has no consequence for future decisions as an auditor and no effect on the distribution of the signals accruing to the current auditor. On the equilibrium path utility associated with being audited in period t conditional on the signal z_t^j is $E_t[C_t^i|z_t^j] + E_t[P_{t+1}^{j(i)}|z_t^j]$. Consider now those audit rounds for which $t > 1$. Conditional on the bad signal, deviating to not punishing results in punishment increment $([\pi^p - \pi]/\pi)E_t[P_{t+1}^{j(i)}|B] = ([\pi^p - \pi]/\pi)E_t P_{t+1}^{j(i)}$, equality following from the fact that on the equilibrium path your chance of being punished next period is by assumption independent of the signal your auditee has this period. Cost saving on the other hand is $E_t[\theta P_t^i|B]$; so that the incentive constraint is $([\pi^p - \pi]/\pi)E_t[P_{t+1}^{j(i)}] \geq E_t[\theta P_t^i|B]$. Since $E_t(P_t^i|B) = E_t(P_t^i)/\pi$, we may write this as $[\pi^p - \pi]E_t P_{t+1}^{j(i)} \geq \theta E_t P_t^i$, . A similar calculation conditioning on the good signal yields $[\pi^p - \pi]E_t P_{t+1}^{j(i)} \geq -\theta E_t P_t^i$. Combining the two gives $[\pi^p - \pi]E_t P_{t+1}^{j(i)} \geq |\theta| E_t P_t^i$.

Adding up the constraints across individuals and taking time zero expectations, we discover that $[\pi^p - \pi]\overline{P}_{t+1} \geq |\theta|\overline{P}_t$. Examining the objective function, we see that if this holds with strict inequality the objective function is increased by reducing $\overline{P}_{t+1}$ until it holds with equality. This may not be feasible, depending on the matching and so forth, but the constraints holding with exact equality provides an upper bound on the objective function given by solving the difference equation $[\pi^p - \pi]\overline{P}_{t+1} = |\theta|\overline{P}_t$ for $t > 1$. In the first audit round, punishment is $\overline{P}_1 = \pi_0^R G^R$. Also, for $t = 1$ it is $E_1(P_1^i|B) = E_1(P_1^i)/\pi_0^R$ since the probability of the bad signal in the first audit round is π_0^R, so the incentive constraint $([\pi^p - \pi]/\pi)E_1[P_2^{j(i)}|B] \geq E_1[\theta P_1^i|B]$ reads $([\pi^p - \pi]/\pi)E_1 P_2^{j(i)} \geq E_1\theta P_1^i/\pi_0^R$; combining this with the constraints for $t > 1$ and taking expectations yields $\overline{P}_2 = (\pi/\pi_0^R)(|\theta|/[\pi^p - \pi])\overline{P}_1$, equality as before being a lower bound. Substituting these plus the solution to the difference equation into the objective function then gives the desired bound. Note that the condition $|\theta| < [\pi^p - \pi]$ is necessary and sufficient for the sum in the objective function to converge. □

Theorem. [4 in text] *If*

$$\Pi \leq \frac{(1-\mu)(1-2\epsilon)}{\epsilon}$$

group utility maximization implies $\beta_{00} = \beta_{10} = \beta_{11} = 0, \beta_{01} = 1$. *Otherwise* $\beta_{10} = \beta_{11} = 0$ *and* $\beta_{00} = \beta_{01} = 1$.

Per capita group utility being equal to

$$s - \mu - (\theta + \psi + 1)\min\{\frac{(1-\mu)\epsilon + \Pi}{1-\epsilon}, (1-\mu) + \Pi\}$$

Proof. The conditions for enforceability of [01] are

$$u(a^i, [01]) - u([01], [01]) \leq [\pi_0(a^i, [01]) - \pi_0([01], [01])]P_1, \quad a^i = [00], [11], [10]$$

First $u([00], [01]) - u([01], [01]) = \mu > 0$ implies $P_1 > 0$. The conditions for [00] and [11] can be written respectively as

$$1/P_1 \leq (1-\epsilon)(\beta_{01} - \beta_{11}) + \epsilon(\beta_{00} - \beta_{10})$$
$$1/P_1 \geq \epsilon(\beta_{01} - \beta_{11}) + (1-\epsilon)(\beta_{00} - \beta_{10})$$

and they are the only relevant ones because as is easily checked by summing them one gets the condition for [10]. Since $u([11], [01]) - u([01], [01]) = -(1-\mu) < 0$ we then have $G(a^R) = G([00], [01])$. The objective is to minimize

$$[\pi_0([01], [01]) + \Pi](\theta + \psi + 1)G([00], [01])$$

subject to enforceability.

The two above enforceability inequalities imply - together with $P_1 > 0$ - that [01] is enforceable iff

$$\beta_{01} - \beta_{11} \geq \beta_{00} - \beta_{10}$$
$$(1-\epsilon)(\beta_{01} - \beta_{11}) + \epsilon(\beta_{00} - \beta_{10}) > 0$$

Observe that they imply $\beta_{01} - \beta_{11} > 0$. Incidentally: the set of vectors defined by the conditions

$\beta_{10} = \beta_{11} = 0, 1 = \beta_{01} \geq \beta_{00}$ satisfy the conditions. Now

$$G([00],[01]) = \frac{u([00],[01]) - u([01],[01])}{\pi_0([00],[01]) - \pi_0([01],[01])} = \frac{1}{(1-\epsilon)(\beta_{01} - \beta_{11}) + \epsilon(\beta_{00} - \beta_{10})}$$
$$\pi_0([01],[01]) = \mu(1-\epsilon)\beta_{11} + (1-\mu)\epsilon\beta_{01} + \mu\epsilon\beta_{10} + (1-\mu)(1-\epsilon)\beta_{00}$$

Clearly both $\pi_0([01],[01])$ and $G([00],[01])$ go down with β_{11} and β_{10} hence $\beta_{11} = \beta_{10} = 0$. First observe that if $\mu = 1$ then $\pi_0 = 0$ so the minimum of the objective function is achieved by setting $\beta_{00} = \beta_{01} = 1$. Suppose now that $\mu < 1$.

Dividing by the constant $(1-\mu)(\theta + \psi + 1)$ the objective function is maximized by minimizing the function

$$\frac{\epsilon\beta_{01} + (1-\epsilon)\beta_{00} + \frac{\Pi}{(1-\mu)}}{(1-\epsilon)\beta_{01} + \epsilon\beta_{00}}$$

The derivative with respect to β_{01} is

$$\frac{-(1-2\epsilon)\beta_{00} - (1-\epsilon)\frac{\Pi}{(1-\mu)}}{[(1-\epsilon)\beta_{01} + \epsilon\beta_{00}]^2},$$

hence the optimal $\beta_{01} = 1$. Then the derivative with respect to β_{00} is

$$\frac{(1-2\epsilon) - \frac{\epsilon\Pi}{(1-\mu)}}{[(1-\epsilon) + \epsilon\beta_{00}]^2}.$$

From this it follows that if $1 - 2\epsilon \geq \epsilon\Pi/(1-\mu)$ then $\beta_{00} = 0$, and $\beta_{00} = 1$ otherwise. Plugging this into the objective function we obtain that: if

$$\Pi \leq \frac{(1-\mu)(1-2\epsilon)}{\epsilon}$$

we have $\pi_0([01],[01]) = (1-\mu)\epsilon$ and $G^R = 1/(1-\epsilon)$, hence per-capita utility is given by

$$s - \mu - [(1-\mu)\epsilon + \Pi](\theta + \psi + 1)\frac{1}{1-\epsilon};$$

otherwise we have $\pi_0([01],[01]) = 1 - \mu$ and $G([00],[01]) = 1$, thus per-capita utility is

$$s - \mu - [(1-\mu) + \Pi](\theta + \psi + 1)$$

Now it is easily checked that the condition $\Pi \leq (1-\mu)(1-2\epsilon)/\epsilon$ is equivalent to

$$\frac{(1-\mu)\epsilon + \Pi}{1-\epsilon} \leq (1-\mu) + \Pi$$

and this yields the desired result. □

Theorem. *[6 in text] The bid function is single peaked as a function of N. Define*

$$\hat{N} = \frac{S}{1 + (\theta + \psi + 1)\Pi} \qquad \hat{\hat{N}} = \frac{S}{(\theta + \psi + 1)\frac{\epsilon + \Pi}{1-\epsilon}}$$

For $N \leq \hat{N}$ the bid is N. If $\epsilon(\theta+\psi+2) \leq 1$ and

$$(\theta+\psi+1)\frac{\epsilon(1+\Pi)}{1-\epsilon} < 1$$

then $\hat{\hat{N}} > \hat{N}$ and for $\hat{N} \leq N \leq \hat{\hat{N}}$ the bid is given by the following function, decreasing in N:

$$\tilde{B} \equiv \frac{(1-\epsilon)S - N\big[(\epsilon+\Pi)(\theta+\psi+1)\big]}{1-\epsilon(\theta+\psi+2)}$$

In all other cases the group does not bid.

Proof. It is easily seen that the group payoff from a bid of μ can be written as

$$\text{Payoff}(\mu) = S - N\mu - N(\theta+\psi+1) \times \begin{cases} \frac{(1-\mu)\epsilon+\Pi}{1-\epsilon} & \mu \leq \frac{1-(2+\Pi)\epsilon}{1-2\epsilon} \\ 1-\mu+\Pi & \text{otherwise} \end{cases}$$

whose derivative with respect to μ is

$$\begin{cases} N\big(\frac{(\theta+\psi+1)\epsilon}{1-\epsilon} - 1\big) & \mu \leq \frac{1-(2+\Pi)\epsilon}{1-2\epsilon} \\ N(\theta+\psi) & \text{otherwise} \end{cases}$$

Case 1 (high punishment cost): $(\theta+\psi+2)\epsilon > 1$. In this case the payoff is increasing in μ throughout its interval, so the group bids $\mu = 1$ if Payoff$(1) \geq 0$ and nothing otherwise. And the non-negativity condition amounts to $N \leq \hat{N}$.

Case 2 (low punishment cost): $(\theta+\psi+2)\epsilon \leq 1$. In this case the payoff is U-shaped. If Payoff$(1) \geq$ Payoff(0) again the group will bid 1 if Payoff$(1) \geq 0$ and zero otherwise; the non-negativity condition is the same as before, $N \leq \hat{N}$, while the inequality is satisfied iff

$$(\theta+\psi+1)\frac{\epsilon(1+\Pi)}{1-\epsilon} \geq 1$$

If this last inequality is reversed, then for $N \leq \hat{N}$ we still have Payoff$(1) \geq 0$ so the group still bids 1. For $N > \hat{N}$ on the other hand Payoff$(1) < 0$ so the group's bid is given by the μ at which the decreasing part of the payoff crosses zero, which is easily computed as the function $\tilde{B} = \tilde{B}(N)$ given in the statement, which decreases and reaches zero at $N = \hat{\hat{N}} > \hat{N}$, the latter inequality following from $(\theta+\psi+1)\epsilon(1+\Pi) < (1-\epsilon)$. Theorem now follows. □

Theorem. [7 in text] *For all s and $1 > \overline{\mu} > \underline{\mu} > 0$ there exists an $\overline{N}$ such that $N > \overline{N}$ implies that any incentive compatible $\mu < \underline{\mu}$.*

Proof. Let $\Phi^N(x)$ be the cdf of the standardized binomial, that is, the probability that a binomial $\tilde{x}$ with success probability μ and N trials has $(\tilde{x} - \mathbb{E}\tilde{x})/\sqrt{\text{var}(\tilde{x})} \leq x$. With E representing total effort, the prize is awarded if $E \geq N\hat{\mu}$. Passing to standardized variables one finds that if you get a message and you contribute you get s with probability $1 - \Phi^N((\hat{\mu}-\mu)\sqrt{N}/\sqrt{\mu(1-\mu)})$, and pay an effort cost of 1. On the other hand if you get a message and you do not contribute you get s with probability

$$1 - \Phi^{N-1}([\frac{N}{N-1}\hat{\mu} - \mu]\frac{\sqrt{N-1}}{\sqrt{\mu(1-\mu)}})$$

Hence the gain to not contributing is

$$G^N = s\left[\Phi^N((\hat{\mu}-\mu)\frac{\sqrt{N}}{\sqrt{\mu(1-\mu)}}) - \Phi^{N-1}([\frac{N}{N-1}\hat{\mu}-\mu]\frac{\sqrt{N-1}}{\sqrt{\mu(1-\mu)}})\right] + 1.$$

Suppose $\underline{\mu} \leq \mu \leq \overline{\mu}$. It is easy to show that when $|\hat{\mu}-\mu|\sqrt{N}$ is large enough - say $N > \bar{N}_1$ - the two probabilities are either both close to one (if $\hat{\mu} < \mu$) or both close to zero (if $\hat{\mu} > \mu$), using the Chebyshev inequality. Thus in that case it will be optimal not to contribute. If $\hat{\mu} \notin [\underline{\mu}, \overline{\mu}]$ this ends the proof. Suppose not, and that for some $\bar{\chi}$ it is $|\hat{\mu}-\mu|\sqrt{N} \leq \bar{\chi}$ (so that $\mu \to \hat{\mu}$). Now we will use the Central Limit Theorem to argue that since both Φ^{N-1} and Φ^N are close to a standard normal Φ they are close to each other. Because $\underline{\mu} \leq \mu \leq \bar{\mu}$ and the bound $\overline{\chi}$ the Central Limit Theorem applies uniformly. The difference

$$[\frac{N}{N-1}\hat{\mu}-\mu]\sqrt{N-1} - (\hat{\mu}-\mu)\sqrt{N} = \frac{1}{\sqrt{N-1}}\hat{\mu} - (\hat{\mu}-\mu)\frac{1}{\sqrt{N}+\sqrt{N-1}}$$

is uniformly close to zero. Since Φ is continuous it then follows that there exists an $\bar{N}_2$ such that for $N > \bar{N}_2$ we again have $G^N > 0$. □

Theorem. [8 in text] *A transfer takes place only if the small group sets the agenda, in which case it sets the prize S to*

$$S_2^L \equiv \frac{\epsilon(\theta+\psi+1)+\kappa}{1-\epsilon}N_2,$$

bids a positive amount and pays zero; the large group pays $\min\{S_2^L, N_2\}$ to the small group. If the large group sets the agenda, it sets the prize equal to zero.

The winning group pays a positive amount only if the politician sets the agenda, in which case she chooses the large group to pay for the prize, which she sets equal to

$$\hat{S} \equiv \frac{N_1[1-\epsilon(\theta+\psi+2)] + N_2\left[\epsilon(\theta+\psi+1)+\kappa\right]}{1-\epsilon}$$

Both groups bid N_1 and the large group wins the bidding, so there is no transfer between the groups but simply a payment by the large group of N_1 to the politician.

Proof. Recall that

$$\tilde{B}_j \equiv \frac{(1-\epsilon)S - N_j\left[\epsilon(\theta+\psi+1)+\kappa\right]}{1-\epsilon(\theta+\psi+2)}.$$

Observe that since we always assume $\theta+\phi+1 > 0$ we have

$$\frac{1-\epsilon}{1-\epsilon(\theta+\psi+2)} > 1$$

so that the bid increases faster than the prize. The following is the prize size below which group j bids zero:

$$S_j^L \equiv \frac{\epsilon(\theta+\psi+1)+\kappa}{1-\epsilon}N_j$$

Note that $S_j^L > N_j$ for κ large enough. The following is the prize size for which group j posts the highest bid:

$$S_j^H \equiv \frac{1-\epsilon+\kappa}{1-\epsilon}N_j$$

Observe that $\epsilon(\theta+\psi+2)<1$ is the same as $\epsilon(\theta+\psi+1)<1-\epsilon$, so that $S_j^L<S_j^H$. A group of size N_j bids zero for $S\leq S_j^L$ then the bid increases linearly, it reaches N_j for $S=S_j^H$ and remains constant for larger S.

Below $S<S_1^L$ both groups bid zero. Between S_1^L and S_2^L group 1 bids $\min\{\tilde{B}_1,N_1\}$, wins the auction and pays zero, getting a surplus of S. If $S_1^H>S_2^L$ then between S_2^L and S_1^H group 1 bids $\tilde{B}_1$, wins the auction, pays the bid $\tilde{B}_2$, getting a surplus of $S-\tilde{B}_2$ which decreases in S. Above S_1^H group 1 bids N_1, wins the auction and pays $\tilde{B}_2$ until S equals

$$\hat{S}=\frac{N_1[1-\epsilon(\theta+\psi+2)]+N_2\left[\epsilon(\theta+\psi+1)+\kappa\right]}{1-\epsilon}.$$

Above $\hat{S}$ group 2 wins and pays N_1. If $S_1^H<S_2^L$ the only change is that the range over which the small group pays zero is larger. The winning group is unchanged.

Now consider who sets the agenda. Because the bid increases faster than the prize, when the small group has the agenda, they should choose a prize of S_2^L and pay zero (regardless of the relative sizes of S_1^H and S_2^L). The large group will have to pay a transfer of $\min\{S_2^L,N_2\}$ to the smaller one.

Suppose the large group chooses S. It can only win by choosing $S>\hat{S}$ in which case it pays N_1, however the size of the prize cannot be larger than N_1, so as it can never get any surplus it prefers the smallest prize $S=0$. No transfer will take place.

Suppose lastly that the politician chooses S. She will do so to maximize the lower bid, which is what she collects. The bids are increasing in the size of the prize, but limited by the ability of the small group to pay no more than N_1. The large group is willing to pay N_1 only above $\hat{S}$, while the small group is willing to pay N_1 above $S_1^H<\hat{S}$. To get N_1 in a second price auction both groups have to be willing to pay N_1, which means that we must have $S\geq\hat{S}$. Notice that $S_1^H>N_1$ so $\hat{S}$ is as well. Hence it is only feasible to charge $\hat{S}$ if the prize is paid for by the large group. Above $\hat{S}$ the lower bid remains constant at N_1 (and the large group wins by bidding strictly more), so since both groups prefer lower prize values the politician will set $S=\hat{S}$ and get N_1 from the large group - which by assumption wins the tie and pays N_1 to the politician to avoid paying $\hat{S}>N_1$ to the small group. □

References

[1] Abreu, D., D. Pearce, E. Stacchetti (1986): "Optimal cartel equilibria with imperfect monitoring," *Journal of Economic Theory*

[2] Acemoglu, Daron (2001): "Inefficient redistribution", American Political Science Review Vol. 95, pp. 649-661.

[3] Ades, A. and R. DiTella (2002): "The new economics of corruption: a survey and some new results," Political Studies.

[4] Austen-Smith, David and J. R. Wright (1992): "Competitive lobbying for a legislator's vote," Social Choice and Welfare.

[5] Banks, Jeffrey S. and Barry R. Weingast (1992): "The political control of bureaucracies under asymmetric information," American Journal of Political Science.

[6] Becker, G. S. (1983): "A theory of competition among pressure groups for political influence," *Quarterly Journal of Economics* 98: 371-400.

[7] Becker, Gary S. (1985): "Public policies, pressure groups, and dead weight costs", Journal of public economics, 1985 - Elsevier

[8] Becker, Gary S. and Casey B. Mulligan (1998): "Deadweight costs and the size of government," NBER.

[9] Becker, Gilbert (1986): "The Public Interest Hypothesis Revisited: A New Test of Peltzman's Theory of Regulation", Public Choice Vol. 49, pp. 223-234.

[10] Boadway, R. and M. Keen (2000): "Redistribution," Handbook of income Distribution.

[11] Border, Kim C. and Joel Sobel (1987): "Samurai Accountant: A Theory of Auditing and Plunder", The Review of Economic Studies Vol. 54, pp. 525-540

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

[13] Dal Bo, P. (2005): "Cooperation under the shadow of the future: experimental evidence from infinitely repeated games," *American Economic Review.*

[14] Damania, R., P. G. Frederiksson and M. Mani (2004): "The persistence of corruption and regulatory compliance failures: Theory and evidence," Public Choice.

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

[16] Di Porto, Edoardo, Nicola Persico and Nicolas Sahuguet (2013): "Decentralized Deterrence, with an Application to Labor Tax Auditing", American Economic Journal: Microeconomics Vol. 5, pp. 35-62

[17] Dixit, Avinash (2004): *Lawlessness and Economics*, Princeton University Press

[18] Dixit, Avinash, Gene M. Grossman and Elhanan Helpman (1997): "Common Agency and Coordination: General Theory and Application to Government Policy Making", Journal of Political Economy Vol. 105, pp. 752-769

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

[20] Ehrlick, I. and F. Lui (2000): "Bureaucratic corruption and endogenous economic growth," Journal of Political Economy.

[21] `www.floatingpath.com/2012/03/05/campaign-spending-percentage-gdp/`

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

[23] Fudenberg, D., D. Levine and W. Pesendorfer (1998): "When are nonanonymous players negligible?" *Journal of Economic Theory.*

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

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

[26] Green, Jerry and Jean-Jacques Laffont (1979): "On Coalition Incentive Compatibility", The Review of Economic Studies Vol. 46, pp. 243-254

[27] Hillman, A. L. and D. Samet (1987): "Dissipation of contestable rents by small numbers of contenders," Public Choice.

[28] Jehiel, Philippe [2005]: "Analogy-based Expectation Equilibrium", *Journal of Economic Theory* 123: 81-104.

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

[30] Krishna, V. and J. Morgan (1997): "An analysis of the war of attrition and the all-pay auction," Journal of Economic Theory.

[31] Kroszner, R. S. and T. Stratmann (1998): "Interest-group competition and the organization of congress: theory and evidence from financial services' political action committees,"American Economic Review.

[32] Laffont, Jean-Jacques (1999): "Political economy, information and incentives," European Economic Review.

[33] Laffont, Jean-Jacques (2000): Incentives and political economy, Oxford University Press.

[34] Laffont, Jean-Jacques and Jean Tirole (1991): "The politics of government decision-making: A theory of regulatory capture," The Quarterly Journal of Economics.

[35] Levine, David K. and Salvatore Modica (2012) "Conflict and the Evolution of Societies", mimeo

[36] Levine, D. K. and W. Pesendorfer [1995]: "When Are Agents Negligible?" *American Economic Review*, 85: 1160-1170

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

[38] Li, John Shuhe (2003): "The Benefits and Costs of Relation-based Governance: an Explanation of the East Asian Miracle and Crisis", *Review of International Economics* 11: 651-673

[39] Lohmann, Susanne (1998): "An Informational Rationale for the Power of Special Interests", American Political Science Review Vol. 92, pp. 809-28

[40] Mailath, George J. and Andrew Postlewaite (1990): "Asymmetric Information Bargaining Problems with Many Agents", *Review of Economic Studies* 57: 351-367

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

[42] Mitra, D. (1999): "Endogenous lobby formation and endogenous protection: a long-run model of trade policy determination," American Economic Review.

[43] Nitzan, Shmuel and Ueda, Kaoru (2011): "Prize sharing in collective contests," *European Economic Review* 55, pp. 678-687

[44] Nti, K. O. (1999): "Rent-seeking with asymmetric valuations," *Public Choice* 98, pp. 415-430.

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

[46] Olson, Mancur (1982): *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*, Yale University Press.

[47] Ostrom, Elinor (1990): *Governing the Commons: The Evolution of Institutions for Collective Action,* Cambridge University Press.

[48] Ostrom, Elinor, James Walker and Roy Gardner (1992): "Covenants with and Without a Sword: Self-Governance Is Possible", *The American Political Science Review* (86), pp. 404-417

[49] Pecorino, Paul (2009): "Public goods, group size, and the degree of rivalry", *Public Choice* 138, pp. 161-169.

[50] Persson, T. and G. Tabellini (2002): "Political economics and public finance," Handbook of public economics.

[51] Potters, Jan and Randolph Sloof (1996): "Interest groups: A survey of empirical models that try to assess their influence", European Journal of Political Economy Vol. 12, pp. 403-442

[52] Shapiro, Carl and Stiglitz, Joseph E. (1984): "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review* 74, pp. 433-444

[53] Slinko, Irina, Evgeny Yakovlev, and Ekaterina Zhuravskaya (2005), "Laws for sale: evidence from Russia," *American Law and Economics Review* 7.1: 284-318.

[54] Sugaya T. (2011): "Folk Theorem in Repeated Games with Private Monitoring," Princeton University.

[55] Uhlaner, C. J. (1989): "Rational turnout: The neglected role of groups," American Journal of Political Science.

[56] Wittman, Donald (1989): "Why Democracies Produce Efficient Results", Journal of Political Economy Vol. 97, pp. 1395-1424