Discussion Paper No. 1174R

PIVOTAL PLAYERS AND THE CHARACTERIZATION OF INFLUENCE

by

Nabil I. Al-Najjar*
and
Rann Smorodinski*

November 1996
Revised: April 1997

(*) Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL, 60208.

[e-mail: al-najjar@nwu.edu and rann@nwu.edu; Fax: 847-467-1220.]

We are grateful for the comments of Peter Klibanoff, Ehud Lehrer, Wolfgang Pesendorfer, and seminar participants at Tel-Aviv, Technion, Jerusalem, and Northwestern. We are especially grateful for Eilon Solan who pointed out an error in an earlier proof.

Abstract:

A player *influences* a collective outcome if his actions can change the probability of that outcome. He is *α-pivotal* if this change exceeds some threshold α. We study influence in general environments with N players and arbitrary sets of signals. It is shown that influence is maximized when players' signals are identically distributed and the outcome is determined according to simple majority rule. This leads to the surprising conclusion that majority rules already contain the maximal number of pivotal players. From this we derive a tight bound on average influence, as well as a tight bound on the number of α-pivotal players, which is independent of N.

This analysis is relevant to problems where players' influence is a key consideration in determining their strategic behavior. The applications we consider include the problem of designing a mechanism for the provision of public goods in the spirit of Mailath and Postlewaite (1990), partnership games, games with production complementarities, and cooperation in a noisy prisoner's dilemma.

1. INTRODUCTION

Strategic behavior often hinges on players' beliefs about the impact of their actions on a collective outcome. This dependence surfaces in many contexts ranging from the provision of public goods and optimal allocations in the presence of externalities, to voting, implementation and reputation building. In these settings, only players who believe they will be pivotal take into account the impact of their actions on a collective outcome; non-pivotal players ignore such strategic considerations and behave myopically. For example, in the classic public good problem, the only force countervailing the incentive to free-ride is the extent to which players believe their contributions will be pivotal in determining whether the public good will be provided.

How much influence can an individual player have? How many players can be pivotal in a given setting? And, what impact does the allocation of influence among players have on equilibrium outcomes in economic applications? To address these questions, we consider a general environment with N players, each with a random signal $\tilde{t}_n$. Signals can be interpreted as privately known types (as in implementation and mechanism design problems) or as noisy outcomes of unobserved actions (as in problems with moral hazard). An outcome function F maps the vector of observed or reported signals $\mathbf{t} = (t_1, \ldots, t_N)$ into a collective outcome $F(\mathbf{t}) \in [0, 1]$. The outcome may be interpreted as either the probability of a binary collective decision, *e.g.*, the probability that a public good is provided, or as an outcome in $[0,1]$, *e.g.*, the level of pollution. The outcome function F and the profile of signals $\tilde{\mathbf{t}} = (\tilde{t}_1, \ldots, \tilde{t}_N)$ can be arbitrary. In particular, signals may be correlated and asymmetrically distributed, and F can range from being anonymous, depending only on some aggregate statistic, to the other extreme of treating a few players as pivotal and ignoring all others. We define a player's influence to be the change in the probability of the outcome F caused by a change in his signal. A player is α-pivotal if his influence exceeds some threshold $\alpha \in (0, 1)$.

Our main result derives a tight bound on average influence which is uniform over all mechanisms and profiles. This bound is achieved in an environment where players' signals are identically distributed and the outcome function is simple majority rule. From this it immediately follows that average influence converges to zero at the rate $N^{-1/2}$. We then show that the number of α-pivotal players is bounded uniformly over N. This bound is tight and can be easily computed. Finally, we extend the results to environments with correlated signals and where players might have a continuum of possible signals.

Our work builds on an observation by Mailath and Postlewaite (1990) in their paper on the problem of provision of public goods. In their model, individuals have privately known valuations represented by non-degenerate independent random variables $\tilde{t}_1, \ldots, \tilde{t}_N$. Solving this mechanism design problem involves finding a schedule of individual contributions and an outcome function that maps reported valuations into a probability of provision. Mailath and Postlewaite derive an asymptotic result in which they show that the probability of provision goes to zero as the number of individuals increases to infinity. Their result can be intuitively understood in terms of an individual's reasoning in the interim stage after learning his type, but before learning the other individuals' types. This individual compares the private cost of his contribution with the *expected* impact of his report on the probability of provision. In the language of our paper, only players who believe they are sufficiently pivotal will make large contributions. Mailath and Postlewaite's insight was to recognize that in a large economy, "If valuations are independent ... [then] ... few agents can be pivotal" (p. 363).

The observation that there cannot be many pivotal players is central to arguments about strategic behavior in a wide range of problems. Our contribution is to provide, for a broad class of environments, characterizations showing that average influence and the number of pivotal players are maximized under simple majority rules, rather than more complex non-anonymous mechanisms.[1] We implement this approach in applications which highlight the central role of influence and 'pivotalness' as a common thread throughout the literature.

Overview of the Applications

In Section 3 we apply our results on influence to the problem of designing voluntary contribution mechanisms for the provision of public goods. We provide a simple proof establishing an upper bound on the probability of provision. A consequence of this bound is Mailath and Postlewaite's asymptotic result that the probability of providing a public good of fixed per capita cost converges to zero as the number of agents goes to infinity. We also identify conditions under which there is an amount C such that no community of any size can build a project costing more than C under any voluntary contribution scheme. This points out that the failure of such schemes may be more dramatic because it is not necessary to assume that the cost of the project is unbounded.

The connection between the mechanism design problem and our analysis of influence can be intuitively understood as follows: incentive compatibility constraints require that the private gain from declaring a lower valuation is smaller than the private loss from reducing the probability of provision. In the language of our paper, the monetary contribution of an individual of type t under an incentive compatible mechanism cannot exceed t times his influence. We use this observation to derive an upper bound on the probability of provision which is uniform over type distributions and all budget balanced, individually rational and incentive compatible mechanisms. The bound, which can be easily computed from the primitives of the model, clarifies the severity of the public good problem in settings with moderately small numbers of individuals, such as small organizations, neighborhoods, and so on. Many such settings involve too few players to make asymptotic arguments useful or informative. We provide numerical examples in Section 3.3.

The second class of applications is that of games with moral hazard. Here we study games with N players each generating a signal correlated with his unobserved actions.[2] The first two examples apply our results to a partnership game and an agency problem with production complementarities. The third example studies cooperation in a noisy prisoner's dilemma with random matching followed by a coordination game.

Related Papers

Our results may be compared to an alternative proof of Mailath and Postlewaite's (1990) result which they provide in the Appendix to their paper. That proof exploits the fact that a random variable f, viewed as a point in a linear space, cannot have high covariance with many members of an orthonormal basis for that space. Interpreting members of the basis as the random types of the agents, one can conclude that not many agents can accurately predict f based on their signals. While the notions of pivotal players and influence are not explicitly defined, the predictability of the public outcome is clearly in the same spirit. Another related approach is that of Fudenberg, Levine and Pesendorfer (1995). They study repeated games with a large player and many small players, provide an upper bound on the average predictability of an outcome and derive a result on the negligibility of small players.

Our approach (and proof) differs from these references in a number of respects. First, we provide a tight bound on average influence and explain what sort of mechanism achieves this bound. Second, the bound covers the cases of correlated signals and when players' have a continuum of possible signals. Both cases are important in many applications. Third, we show that the *number* of α-pivotal players is bounded for any $\alpha \in (0, 1)$. This is a stronger conclusion than asserting that the *ratio* of such players goes to zero as N goes to infinity. Finally, our proof provides a better perspective on how the non-anonymity of an outcome function affects average influence and the number of pivotal players. An implication is that the problem of characterizing influence for potentially complex, non-anonymous mechanisms can be reduced to looking at simple, anonymous mechanisms which take the form of majority rule. This leads to the surprising conclusion that majority rules already contain the maximal number of pivotal players; so, no further gain can be made by considering more complex mechanisms.

2. THE MODEL AND MAIN RESULTS

2.1. The Model

We consider an environment with N players, each with a random signal $\tilde{t}_n$, taking values in a set T_n of cardinality M_n (we use t_n to denote the realized signals). We focus here on the case where M_n is finite. In section 2.8 we consider the case where one or more players has infinitely many signals. The profile of random signals is denoted $\tilde{\mathbf{t}} = (\tilde{t}_1, \ldots, \tilde{t}_N)$ and takes values in $T = T_1 \times \cdots \times T_N$. We use $\tilde{\mathbf{t}}_{-n}$ and $\mathbf{t}_{-n}$ to denote, respectively, profiles of random signals and vectors of signal realizations for all players other than player n, and write $T_{-n} = T_1 \times \cdots \times T_{n-1} \times T_{n+1} \times \cdots \times T_N$.

Player n observes his private signal t_n, but does not observe the signals of the others. Signals are generated according to a joint probability distribution P on T, with $P(\mathbf{t})$ denoting the probability of the vector $\mathbf{t} \in T$. We use E to denote expectations with respect to P and $E(F \mid t_n)$ to denote the expectation of F conditional on individual n's signal being t_n. To simplify the exposition, we assume for the moment that the signals are independent. Independence allows us to express P as the product of its marginals on T_{-n} and T_n. Section 2.7 extends the results to environments with correlated signals.

Let $\Delta(T_n)$ denote the simplex representing all probability distributions on the set of player n's actions, and $\bar{t}_n \in \Delta(T_n)$ denote the distribution of $\tilde{t}_n$. Define

$$\Delta_\epsilon(T_n) = \{(\bar{t}_n^{\,1}, \ldots, \bar{t}_n^{\,M_n}) \in \Delta : \bar{t}_n^{\,m} > \epsilon, \text{ for all } m = 1, \ldots, M_n\}$$

to be the set of totally mixed distributions in which each signal has probability at least ϵ. Denote profiles of such signals by

$$\Delta_\epsilon^N = \Delta_\epsilon(T_1) \times \cdots \times \Delta_\epsilon(T_N).$$

The observed vector of signals $\mathbf{t}$ is mapped into a collective outcome or decision represented by an *outcome function* $F : T \to [0, 1]$.[3] Here, $F(\mathbf{t})$ can denote either the value of an aggregate outcome (*e.g.*, the level of pollution, output of team production, a principal's reward, ... etc.), or the probability of a binary outcome (*e.g.*, the probability that a public project is undertaken).

For Theorems 1-3 we need the assumption (which we drop in Theorem 4) that all signals have positive probability:

Assumption A1: There is $\epsilon > 0$ such that $\bar{t}_n \in \Delta_\epsilon(T_n)$ for all n.

We call a signal distribution $\bar{t}_n$ *ϵ-extremal* if it coincides with a vertex of the trimmed simplex $\Delta_\epsilon(T_n)$. That is, $\bar{t}_n$ is ϵ-extremal if it puts probability ϵ on all but possibly one signal. Denote the set of extremal distributions for player n by $ext\,\Delta_\epsilon(T_n)$, and profiles of extremal distributions by $ext\,\Delta_\epsilon^N$.

2.2. *Symmetric Environments*

An environment is *symmetric* if all players have the same signal sets: *i.e.*, if $T_n = T_{n'}$ for any pair of players n and n'. In this case we denote the number of signals for each player by M. Symmetric environments are of interest because they arise naturally in many applications (*e.g.*, voting), and because their simplicity often provides intuition for the results in more general settings. More significantly, we will show that in studying bounds on influence and the number of pivotal players, one can restrict attention to symmetric environments without loss of generality.

For symmetric environments we also define a profile $\bar{\mathbf{t}}$ to be *symmetric* if the agents' signals are identically distributed: $\bar{t}_n = \bar{t}_m$ for all m, n. This can be done with no ambiguity since agents' signal distributions in a symmetric environment are defined on the same signal space.

2.3. *Anonymous Mechanisms and Majority Rules*

In symmetric environments it makes sense to talk about anonymous mechanisms that ignore the agents' names. Specifically, a mechanism F is *anonymous relative to a subset of players K* if for any permutation σ of the names of players that maintains the names of players outside K, $F(\mathbf{t}) = F(\sigma(\mathbf{t}))$.[4]

We call F *anonymous* if $K = N$. Anonymity can be equivalently defined in terms of the vector $d(\mathbf{t}) = (d_1(\mathbf{t}), \ldots, d_M(\mathbf{t}))$ of empirical frequencies of the M signals in $\mathbf{t}$ (that is, $d_m(\mathbf{t})$ is the number of times the mth signal is observed, divided by N). It is easy to verify that F is anonymous if and only if $F(\mathbf{t})$ depends on $\mathbf{t}$ only through the vector of frequencies $d(\mathbf{t})$.

A special class of anonymous mechanisms is that of majority rules. Formally, F is a *majority rule* if there are two signals m and m' such that for any $\mathbf{t}$,

$$d_m(\mathbf{t}) > d_{m'}(\mathbf{t}) \implies F(\mathbf{t}) = 1, \quad \text{and}$$
$$d_m(\mathbf{t}) \leq d_{m'}(\mathbf{t}) \implies F(\mathbf{t}) = 0.$$

It will be useful to interpret signal m as 'Yes', signal m' as 'No', and pool all remaining signals into a single signal which we refer to as 'All-other-signals', or simply 'Abstain'. Thus, a majority rule ignores abstaining players and determines a binary YES/NO outcome based on a simple count of Yes's and No's.

We will be interested later in describing the players' influence when the outcome is determined according to a majority rule. Define the maximal binomial probability of K independent Bernoulli trials with probability of success $q \in (0, 1)$ by:

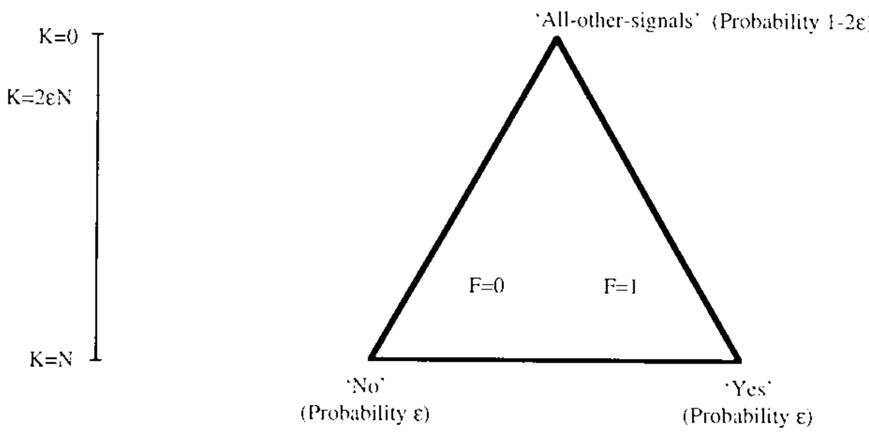
$$\rho_{q,K} = \max_{k=0,\ldots,K} \binom{K}{k} q^k \, (1-q)^{K-k}.$$

Under this rule each player assesses probability $\rho_{q,K}$ of being in a position where his vote will be pivotal in determining the outcome.

Consider now a three-signal environment with N players, each having signals (Yes, No, Abstain) with probabilities $(p, p, 1-2p)$, with $0 < p < \frac{1}{3}$. The outcome is now determined by applying a simple majority rule for the voting (*i.e.*, non-abstaining) players. What is the probability a player assesses to his vote being pivotal? Let $K(\mathbf{t})$ be the random variable denoting the number of non-abstaining players. Then $P(K(\mathbf{t}) = K)$ is the probability that precisely K players out of the remaining $N-1$ players make a Yes/No vote. Then, the probability that the vote of a particular player is pivotal is the probability of being pivotal with K voting players averaged over the range of the number of voting players:

$$R_{p,N} = \sum_{K=0}^{N-1} P(K)\ \rho_{n,\tilde{t},K}$$

The figure below gives a simple geometric representation of R and ρ. Points in the simplex represent frequencies of Yes's, No's and Abstentions. Horizontal lines represent outcomes in which the number of voting players is held constant. A players faces two types of uncertainty. First he is unsure about the number of voting players K (*i.e.*, uncertainty about the horizontal line he is on). Second, conditional on knowing K, there remains the uncertainty about the number of Yes's and No's. In a majority rule, the outcome is determined by a simple 50/50 count at each K, so $F = 1$ to the right of the vertical line and $F = 0$ to its left. Roughly, a player is pivotal if the actions of all other players correspond to a point on the vertical line where a single vote can change the outcome. The probability of the vertical line is $R_{p,N}$.



2.4. Influence

Define the *influence* of player n relative to an outcome function F, a profile $\tilde{\mathbf{t}}$, and a pair of signals t, t' by

$$V_n(F,\tilde{\mathbf{t}};t,t') = E(F \mid t_n = t) - E(F \mid t_n = t')$$

That is, the influence of player n is the expected impact on the outcome caused by changing his signal from t to t'. The expectation is calculated from the perspective of this player after uncertainty about his own signal has been resolved, but before uncertainty about the others is resolved.[5]

To give an economic intuition, think of the problem of designing a mechanism F to determine the value of a public outcome (*e.g.*, whether a public good is provided). Requiring F to be incentive compatible means that an individual of type t must evaluate the implications on the outcome of the mechanism if he reports some other type t'. The mechanism can, in principle, depend on the vector of reports of all individuals in a very complex way, so the effect of individual n's reporting t' instead of t may be quite sensitive to the signals $\mathbf{t}_{-n}$ of the remaining individuals. For example, while reporting t' might increase F at a particular $\mathbf{t}_{-n}$, it may well decrease it at some other vectors $\mathbf{t}'_{-n}$. The key observation is that since individual n is unsure about the realization of $\mathbf{t}_{-n}$, the relevant object in his decision problem is the change in the *expected impact* on the collective outcome his report will have. This is the notion of influence captured in the definition above.

We define a player's influence as the maximal influence he can have over all his possible signals:[6]

$$\begin{aligned} V_n(F, \tilde{\mathbf{t}}) &= \max_{t,t'} V_n(F, \tilde{\mathbf{t}}; t, t') \\ &= \max_{t \in T_n} E(F \mid t) - \min_{t \in T_n} E(F \mid t). \end{aligned}$$

2.5. *The Bound on Average Influence*

One measure of aggregate influence is given by *average influence* relative to $\tilde{\mathbf{t}}$ and F:

$$V(F; \tilde{\mathbf{t}}) = \frac{1}{N} \sum_{n=1}^{N} V_n(F, \tilde{\mathbf{t}}).$$

This measure is useful in dealing with problems like the provision of a public good. Our first theorem says that average influence is maximized using a simple majority rule applied to a symmetric profile. It then follows that average influence is bounded by R_N:

THEOREM 1. *For any $\epsilon > 0$, $\tilde{\mathbf{t}} \in \Delta_\epsilon^N$ and F,*

$$V(F, \tilde{\mathbf{t}}) \leq R_{\epsilon, N}.$$

This bound is tight: It is achieved in a symmetric environment by a symmetric ϵ-extremal profile and an outcome function F that takes the form of a majority rule.[7]

The exact value of the bound $R_{\epsilon, N}$ for low values of N can be derived from the definition of the binomial distribution. For large N, estimates of $R_{\epsilon, N}$ can be found as follows: Given a symmetric extremal profile and a majority rule relative to two ϵ-probability signals, Chebyshev's inequality implies that with high probability the number of voting players is approximately $K = 2\epsilon N$. Conditional on this value of K, the Yes's and no's are i.i.d. with mean 0.5, so the application of a majority rule means that a player's influence depends on the odds that there are exactly $k = \epsilon N$ Yes's. Using Stirling's formula in estimating binomial probabilities, we get the following approximation:[8]

$$R_{\epsilon, N} \simeq \frac{1}{\sqrt{\epsilon\pi}} \frac{1}{\sqrt{N}}.$$

2.6. The Number of Pivotal Players

In many applications there is a critical threshold of influence below which a player ignores the impact of his actions on the mechanism (for such applications, see Section 4). With this motivation, for $\alpha \in (0, 1)$, we define a player n to be *α-pivotal* if his influence is at least α:

$$V_n(F, \tilde{\mathbf{t}}) \geq \alpha.$$

For a fixed ϵ, define K_α^* to be the smallest integer K satisfying

$$R_{\epsilon, K} \geq \alpha.$$

Note that K_α^* is completely determined by two parameters ϵ and α and is otherwise independent of N, $\tilde{\mathbf{t}}$, and F. Our next result provides a tight bound on the number of α-pivotal players:

THEOREM 2. *For any $0 < \alpha < 1$, $N, \bar{\mathbf{t}} \in \Delta_\epsilon^N$, and F the number α-pivotal players is at most K_α^*.*

The bound K_α^* is achieved in a symmetric environment by a symmetric extremal profile and an outcome function that applies a simple majority rule relative to K_α^* players and ignores the signals of the remaining players. A useful implication of Theorem 2 is that the ratio of α- pivotal players to the remaining players is bounded by $\frac{K_\alpha^*}{N}$, and so it converges to zero at the rate of $\frac{1}{N}$.

Note that if there is no noise, *i.e.*, $\epsilon = 0$, then $\rho_{0,K} = 1$ for any K. Thus, every player is 1-pivotal and average influence equals 1 regardless of how large N is. In a public good setting with publicly known valuations, this can be exploited to induce individuals to contribute their true valuations via 'discontinuous' mechanisms that makes building the project contingent on every player reporting the truth. Such mechanisms are clearly fragile to the introduction of private information.

2.7. Correlated Signals

In this section we extend our results on influence to environments with correlated signals. We study correlation structures in which signals are generated as the outcome of a two-stage lottery in the following sense: Let Θ be a finite set of *aggregate parameters* and P a joint distribution on $T \times \Theta$. We assume, without loss of generality, that $P(\theta) > 0$ for all θ since we can simply remove θ's with zero probability. For $\theta \in \Theta$, let $P(\cdot \mid \theta)$ denote the conditional probability on T given θ. We make the following assumption:

Assumption A2: (Conditional Independence) For every $\theta \in \Theta$, the signals $(\tilde{t}_1, \ldots, \tilde{t}_N)$ are independent given θ, and their distribution $P(\cdot \mid \theta)$ belongs to Δ_ϵ^N.

Roughly, assumption A2 says that: (1) θ is a sufficient statistic summarizing all that a player hopes to infer about $\mathbf{t}_{-n}$ from his private signal t_n; and (2) even if player n knew θ, then there is still enough residual randomness about other players' signals.[9]

Let $P(\mathbf{t}_{-n} \mid t_n)$ and $P(\theta \mid t_n)$ denote the posterior distributions of player n on T_{-n} and Θ respectively. The conditional independence assumption implies that $P(\mathbf{t}_{-n} \mid t_n, \theta) = P(\mathbf{t}_{-n} \mid \theta)$. In particular, for any random variable f that is measurable with respect to $\mathbf{t}_{-n}$, we have $E(f \mid t_n, \theta) = E(f \mid \theta)$. We will apply this fact later to the random variables $F(\mathbf{t}_{-n}, t_n = t)$, for those values of t that achieve player n's maximum influence.

We introduce two notions of influence, each corresponding to a different assumption about the information available to a player. Under the first scenario, player n is first informed that his signal is t_n, then determines the pair of actions that generates the greatest expected impact on F. Note that knowing t_n may be useful in increasing influence because the player can use that information to better forecast other players' signals. In this setting, a natural analogue of our definition for the case of independent signals is to require that player n computes his influence as before, using the posterior belief $P(t_{-n} \mid t_n)$. That is:

$$\text{Player } n\text{'s influence at } t_n: \qquad V_n(F; t_n) = \max_{t \in T_n} E(F(\mathbf{t}_{-n}, t) \mid t_n) - \min_{t \in T_n} E(F(\mathbf{t}_{-n}, t) \mid t_n).$$

$$\text{Average influence at } \mathbf{t}: \qquad V(F; \mathbf{t}) = \frac{1}{N} \sum_n V_n(F; t_n)$$

$$\text{Expected average influence:} \qquad V(F) = \sum_{\mathbf{t}} P(\mathbf{t}) V(F; \mathbf{t}).$$

Note that under this definition, different players will typically have different posteriors about distribution of signals of other players.

In our second definition, player n is informed of both his signal t_n and the aggregate parameter θ. He then determines the pair of actions that yields the greatest expected influence. Our assumption of conditional independence implies that his private signal t_n is superfluous because all information relevant to forecasting the signals of others is contained in θ. Thus, we can define player n's influence at (t_n, θ) in terms of the posterior $P(\cdot \mid \theta)$:

$$V_n(F; \theta) = \max_{t_n \in T_n} E(F(\mathbf{t}_{-n}, t_n) \mid \theta) - \min_{t_n \in T_n} E(F(\mathbf{t}_{-n}, t_n) \mid \theta)$$

and define

$$\text{Average influence at } \theta: \qquad V(F;\theta) = \frac{1}{N}\sum_n V_n(F;\theta)$$

$$\text{Expected average influence:} \qquad V^{\theta}(F) = \sum_{\theta} P(\theta)V(F;\theta).$$

Note that players agree on the probability distribution used in computing expectations. Thus, conditional on θ, this setting is covered by our result for the independent signal case applied to the distribution $P(\cdot\,|\,\theta)$. Our bound for the case of independent signals applies to $V(F;\theta)$ and $V^{\theta}(F)$.

THEOREM 3. *Under assumption A2.*

i) $\sum_{\mathbf{t}} P(\mathbf{t})V(F;\mathbf{t}) \leq \sum_{\theta} P(\theta)V(F;\theta)$;

ii) *For any* $\epsilon > 0$,

$$\sum_{\theta} P(\theta)V(F;\theta) \leq R_{\epsilon,N}.$$

2.8. Continuum of Signals

We now extend the framework to the case in which agents have continuous signal spaces, a case that is relevant in many models in information economics (*e.g.*, a continuum of possible valuations for a public good as in Section 3). We also want to understand what happens to the bound on influence as each player has a finite but increasingly large number of possible signals.

A key parameter in our bound is the probability ϵ of the least likely signal. With M signals per player, this probability cannot exceed $\frac{1}{M}$. As M increases, the minimum probability decreases so our bound on average influence becomes weaker. In the limit when all players have a continuum of possible signals it is easy to design a mechanism F such that each player has maximum influence of 1. For example, suppose that $\tilde{t}_n$ is uniformly distributed on $T_n = [0,1]$ for every n. $F(\mathbf{t}) = 0$ if at least one signal is 0, $F(\mathbf{t}) = 1$ if at least one signal is 1, and $F = 0.5$ otherwise. Using our earlier definition, every player has influence equal to 1 as he moves from signal 0 to signal 1. Average influence is then equal to 1 regardless of how large the number of players may be. This conclusion overlooks the fact that $F(\mathbf{t}) = 0.5$ with probability 1, so F is in fact independent of the signal of any individual player, suggesting that no player has any influence in this case. The formal definition we provide below captures the observation that a player's influence should reflect the change he can cause in the outcome over a wide range (but not necessarily all) signals.

We will find it convenient to be explicit about the underlying probability space on which the random signals are defined. Specifically, we will assume that each agent's signal is defined on a non-atomic probability space (Ω, Σ, P). Player n's signal is a random variable $\tilde{t}_n : \Omega \to T_n$, where T_n is an arbitrary set. It is clear that what is relevant for this player's influence is the σ-algebra $\mathcal{G}_n$ generated by $\tilde{t}_n$ rather than the signal space T_n itself. A mechanism F is a random variable $F : \Omega \to [0, 1]$ and the ability of player n to influence F based on his signal t is represented by the conditional expectation $E(F \mid \mathcal{G}_n)(t)$.

Our treatment of the finite-signal case can be viewed as the special case where each $\mathcal{G}_n$ is a finite partition in which each atom has probability at least ϵ. In this case, the conditional expectation $E(F \mid \mathcal{G}_n)(t)$ takes only finitely many possible values (in which conditional expectation is uniquely defined). Our earlier definition of influence of player n, $V_n(F)$, is just the difference between the highest and the lowest values of the conditional expectation of F viewed as a function of t_n. With a continuum of signals and using this definition, player n can have maximum influence of 1 even though F may be independent of his signal with probability 1.

One way to eliminate this problem can be roughly explained as follows: Fix $0 < \epsilon < 1$, and remove a set of signals $A^+{}_n$ of measure ϵ on which $E(F \mid \mathcal{G}_n)(t)$ assumes its highest values. In the example discussed earlier, this would be any subset containing the point $t_n = 1$ and excluding the point $t_n = 0$. Similarly, we remove a set of signals $A^-{}_n$ of measure ϵ on which $E(F \mid \mathcal{G}_n)(t)$ assumes its lowest values. Let $A_n = T_n - (A^+{}_n \cup A^-{}_n)$. That is, A_n represents the signal space after removing the two extreme sets $A^+{}_n$ and $A^-{}_n$. We then define influence as the difference between the highest and lowest value of $E(F \mid \mathcal{G}_n)(t)$ as t ranges over the set A_n. With this definition, a player has small influence if he cannot change by much the conditional expectation of the outcome by moving his signal over a set of large measure.

To make this definition precise, fix versions of the conditional probabilities $E(F \mid \mathcal{G}_n)(t)$ and a parameter $0 < \epsilon < 1$. We define the influence of player n relative to F and ϵ:

$$V_n(F, \epsilon) = \inf_{\{A \in \Sigma : P(A) < \epsilon\}} \sup_{t \notin A} E(F \mid \mathcal{G}_n)(t) - \sup_{\{A \in \Sigma : P(A) < \epsilon\}} \inf_{t \notin A} E(F \mid \mathcal{G}_n)(t).$$

The complicated appearance of this definition stems from the need to satisfy two desired properties. First, for a player with a finite number of signals and distribution $\bar{t}_n \in \Delta_\epsilon(T_n)$, this definition coincides with the one we provided in Section 2.4. Second, the definition does not depend on the particular version of conditional expectations selected, so in particular it is independent of the values of $E(F \mid \mathcal{G}_n)(t)$ over sets of signals of measure zero.[10] Average influence is defined in the usual way:

$$V(F, \epsilon) = \frac{1}{N} \sum_{n=1}^{N} V_n(F, \epsilon).$$

We call an agent (ϵ, α)-*pivotal* if $V(F, \epsilon) \geq \alpha$.

THEOREM 4. *Fix $\epsilon > 0$ and N. Then for any set of signals $\{\mathcal{G}_1, \ldots, \mathcal{G}_N\}$ and any mechanism F,*

$$V(F, \epsilon) \leq R_{\epsilon, N}.$$

An alternative approach is to define the influence of a player to be the change in the average value of the random variable$E(F \mid \mathcal{G}_n)(t)$ as this player moves between sets of measure at least ϵ. Formally,

$$\hat{V}_n(F) = \max_{\{A \in \mathcal{G}_n : P(A) \geq \epsilon\}} E(F|A) - \min_{\{A \in \mathcal{G}_n : P(A) \geq \epsilon\}} E(F|A)$$

$$\text{and} \qquad \hat{V}(F) = \frac{1}{N} \sum_{n=1}^{N} \hat{V}_n(F).$$

With this definition, following the proof of Theorem 1 we can show that: $\hat{V}(F) \leq R_{\epsilon, N}$ For any $N, \mathcal{G}_1, \ldots \mathcal{G}_n$ and F (the simple proof is omitted).

3. THE PROVISION OF A PUBLIC GOOD

In this section we use our analysis of influence to study the problem of designing a voluntary contribution mechanism to provide a public good. We first derive an upper bound on the probability of provision that is uniform over all feasible, individually rational and incentive compatible mechanisms. From this bound the asymptotic result of Mailath and Postlewaite (1990) follows: namely, that as the number of agents increases the probability of sustaining a positive per capita contribution goes to zero. We then discuss variations on the basic model and provide numerical examples illustrating the results.

3.1. The Basic Model

Consider a public good economy $\mathcal{E}_N$ with N individuals. In economy $\mathcal{E}_N$ there is a public project costing C_N. In this subsection, we consider (as do Mailath and Postlewaite) the case where the per capita cost of the project is bounded away from zero. Specifically, we assume there is $\beta > 0$ such that $C_N \geq \beta N$ uniformly in N. Under this assumption, the total project cost is unbounded across economies.

To model uncertainty about valuations and the mechanism, it is useful to be explicit about the underlying probability space (Ω, Σ, P) on which all uncertainty is defined. Individual n has a privately known valuation for the public project, which we model as a random variable $t_n(\omega)$ taking values in a compact interval of possible valuations $[t_n^-, t^+]$. We assume that agents' valuations are independent and that there is a non-vanishing uncertainty that an agent has the minimum valuation t_n^-. More precisely, we assume that there is $\epsilon > 0$ such that $P(t_n = t_n^-) \geq \epsilon$ uniformly across all agents n and economies $\mathcal{E}_N$.

For simplicity, we normalize $t_n^- = 0$ for all N: If the minimum positive-probability valuation of individual n differs from zero, then a mechanism can always extract $\sum_n t_n^-$ by assigning to each individual his minimal valuation. The problem then becomes whether a mechanism can extract contributions exceeding the minimum valuations. Thus, one can view the assumption that $t_n^- = 0$ as a normalization of the original problem in which it is assumed that the minimal contributions have already been extracted.[11]

There are two possible collective outcomes corresponding to whether the project is built or not. A voluntary contribution mechanism in $\mathcal{E}_N$ is a pair (δ, c), where $\delta : \Omega \rightarrow \{0, 1\}$ is a random variable indicating whether the public good is provided, and c is a vector of contributions $(c_1, \ldots, c_N)$, where $c_n : \Omega \rightarrow R$ denotes the amount contributed by individual n. This is a setting in which the Revelation Principle (see, for example, Myerson and Satterthwaite (1983)) applies so we can, without loss of generality, restrict attention to direct revelation mechanisms in which each agent truthfully reports his type.

Individual n's payoff under the mechanism (δ, c) in state ω when he truthfully reports his type is:

$$u_n(\omega) = t_n(\omega)\,\delta(\omega) - c_n(\omega).$$

This formulation covers the case of mechanisms which may reimburse parts of the contributions if the project is not built. For example, the case in which there is a full reimbursement of contributions if the project is not built can be expressed by requiring that $c_n(\omega) = 0$ whenever $\delta(\omega) = 0$.

We will call a mechanism (δ, c) (ex-ante) *budget balanced* if

$$C_N \; E\delta \leq \sum_n Ec_n. \tag{BB}$$

This requires budget balancing on average rather than for each state. The stronger ex post budget balancing condition may be more appropriate if the mechanism designer does not have access to risk-neutral credit markets. We use the weaker requirement (BB) in Proposition 1 so the bound we obtain on the probability of provision is valid for a larger class of mechanisms.

Next, we impose ex ante *individual rationality* on the mechanism (δ, c):

$$E(u_n \,|\, t_n)(\omega) \geq 0 \qquad \text{for all } n, \; P-a.e. \tag{IR}$$

This version of the assumption allows the mechanism designer to offer fair lotteries to the agent if doing so improves his incentives to contribute.

The final requirement is that the mechanism be *incentive compatible*. Specifically, there is a subset $\Omega' \subset \Omega$, with $P(\Omega') = 1$ such that

$$E(u_n \mid t_n)(\omega) \geq t_n(\omega)\, E(\delta \mid t_n)(\omega') - E(c_n \mid t_n)(\omega'), \qquad \text{for all } \omega,\ \omega' \in \Omega'. \tag{IC}$$

That is, the payoff when individual n truthfully reports his valuation is at least as large as his payoff if he mis-reports his type to be $t_n(\omega')$.

PROPOSITION 1:

i) *For every $0 < \eta \leq \epsilon$, the probability of provision $E\delta$ satisfies:*

$$\sup_{(\delta,c)} E\delta \leq \frac{t^*}{\beta}\,[R_{\eta,N} + \eta]$$

where the sup is taken over all mechanisms (δ, c) satisfying IR, IC, and BB;

ii)

$$\lim_{N\to\infty} \sup_{(\delta,c)} E\delta = 0.$$

Proof: To prove part (i), let $A_n = \{\omega : E(\delta \mid t_n)(\omega) \geq E(\delta \mid t_n = 0)\}$ denote the set of states at which individual n's announcement of a type higher than 0 increases the conditional probability of provision. Note that (IR) implies that $E(c_n \mid t_n = 0) \leq 0$, so (IC) in the special case of reporting $t = 0$ can be rewritten as:

$$E(c_n \mid t_n)(\omega) \;\leq\; t_n(\omega)\,[E(\delta \mid t_n)(\omega) - E(\delta \mid t_n = 0)].$$

This implies $E(c_n \mid t_n)(\omega) \leq 0$, for $\omega \in A_n^c$, so

$$E(c_n) = \int_\Omega E(c_n \mid t_n)(\omega)\, dP \leq \int_{A_n} E(c_n \mid t_n)(\omega)\, dP \leq \int_{A_n} t_n[E(\delta \mid t_n)(\omega) - E(\delta \mid t_n = 0)]\, dP.$$

The assumption that $P(t_n = 0) \geq \epsilon$, implies that $\sup_{\{A:\lambda(A)<\eta\}} \inf_{t\notin A} E(\delta \mid t_n) \leq E(\delta \mid t_n = 0)$. From the definition of influence, there is a sequence of sets $B_k \subset \Omega$, $k = 1, 2 \ldots$ such that $P(B_k) < \eta$ and $\sup_{t\notin B_k} E(\delta \mid t_n)(\omega) - \frac{1}{k} \leq \inf_{\{B:\lambda(B)<\eta\}} \sup_{t\notin B} E(\delta \mid t_n)$. Thus, for every k, we have

$$V_n(\delta, \eta) \geq E(\delta \mid t_n)(\omega) - E(\delta \mid t_n = 0) - \frac{1}{k}, \qquad \text{for all } \omega \in B_k^c.$$

These inequalities imply that for every k:

$$E(c_n) \leq \int_{A_n \cap B_k^c} t_n \, [E(\delta \,|\, t_n)(\omega) - E(\delta \,|\, t_n = 0)] \, dP$$
$$+ \int_{A_n \cap B_k} t_n \, [E(\delta \,|\, t_n)(\omega) - E(\delta \,|\, t_n = 0)] \, dP$$

$$\leq t^* \left[V_n(\delta, \eta) + \frac{1}{k} \right] + t^* \eta.$$

Since this is true for any k, we have $E(c_n) \leq t^* [V_n(\delta, \eta) + \eta]$. From (BB), we have

$$E(\delta) \leq \frac{\sum E c_n}{C_N} \leq \frac{\sum_n t^* [V_n(\delta, \eta) + \eta]}{\beta N} \leq \frac{t^*}{\beta} [V(\delta, \eta) + \eta] \leq \frac{t^*}{\beta} [R_{\eta, N} + \eta].$$

To prove part (ii), take a sequence of $\eta \to 0$ and $N = N(\eta)$ such that $R_{\eta, N} \to 0$.

Q.E.D.

Part (ii) is the counterpart of Mailath and Postlewaite's conclusion of asymptotic inefficiency (the main difference is that the rate of convergence here is $N^{-1/2}$ instead of $N^{-1/4}$). Part (i) provides bounds on the probability of providing the public good that can give some insight into the severity of the free-riding problem in settings where asymptotic arguments may be inappropriate. See Section 3.3 for numerical examples.[12]

One draw back of the bound in Proposition 1 is that it depends on the uniform upper bound on valuations t^*. This undermines the bound in settings where players' valuations have unbounded support, or if some players have a vanishingly small probability of a very high valuation. To see the nature of the problem, let $t_n^N(\omega)$ denote the valuation of individual n in economy N, and let B be a subset of individual n's types of measure η. Then the second component of the bound on $E(c_n)$ is

$$\int_{A_n \cap B} t_n^N \, [E(\delta \,|\, t_n^N)(\omega) - E(\delta \,|\, t_n^N = 0)] dP \leq \int_{A_n \cap B} t_n^N \, dP.$$

If the expected valuations are uniformly bounded (*i.e.*, $\sup_{n,N} E t_n^N < \infty$) then Chebyshev's inequality implies that the most the probability mass is concentrated in some compact interval $[0, \bar{t}]$, for large enough $\bar{t}$, uniformly across agents and economies. This, however, is not enough to guarantee that the integrals $\int_B t_n^N \, dP$ are unbounded across economies. The reason is that some individuals may have increasingly larger valuations with vanishingly small probabilities in such a way that this integral is unbounded. This situation does not occur in many settings (for example, if there is a uniform upper bound t^* or if valuations are identically distributed). For every $x > 0$ define

$$b(x) = \sup_{n,N} \int_{\{t_n^N \geq x\}} t_n^N \, dP.$$

A weaker condition than the existence of a uniform bound is to require that $b(x) \to 0$ as $x \to \infty$. This condition, known as *uniform integrability* ensures that the integrals $\int_{A_n \cap B} t_n^N \, dP$ over sets of types of small probability become uniformly small.

3.2. Bounded Project Cost

Theorem 4 shows that the expected per capita contributions converge to zero uniformly over all mechanisms as N increases to infinity. While this implies that it is increasingly difficult to build a project whose per capita cost is bounded below, it leaves open the possibility that free-riding in financing public projects of a given fixed cost might not be severe in large communities, *i.e.*, when N is sufficiently large relative to the cost of the project. If correct, this suggests that larger communities might have an advantage over smaller ones in resolving public good problems.

To shed light on this issue we investigate conditions under which the conclusions of Proposition 1 can be strengthened so that the maximum aggregate contribution is bounded. Under the conditions we introduce, there is a bound C such that the probability of building a project whose cost exceeds C is zero regardless of the size of the community. These additional conditions, while restrictive, may be relevant in many public good environments. Analytically, our conditions help explain what might enable mechanisms in the general setting of Section 3.1 to extract unbounded contributions.

For simplicity, we restrict attention to the case of where each individual has a finite number of possible types, each with probability at least ϵ. Our first condition is that a mechanism (δ, c) satisfies Minimal Contributions (MC) relative to a parameter $c^* > 0$:

$$c_n(\omega) < c^* \implies c_n(\mathbf{t}) = 0. \qquad (MC)$$

This rules out mechanisms with vanishingly small contributions in the limit. One motivation for this assumption is the existence of a collection cost (*e.g.*, of administrative nature) that make very small contributions impractical.

Second, we assume that the contribution $c_n(\omega)$ of individual n depends only on $t_n(\omega)$. We call this assumption measurability as it amounts to requiring that the random variable c_n is measurable with respect to the signal t_n:

$$c_n(\omega) \text{ is measurable with respect to } t_n \text{ for each } n \tag{M}$$

This condition restrict the mechanism in two ways. First, it eliminates the possibility of making the contribution of one agent dependent on the contributions of others. Second, it eliminates the possibility of randomized contributions.

The next theorem shows that under these two conditions imply that it is impossible to extract an aggregate contribution that exceeds the number of $\frac{c}{t^*}$-pivotal players times the maximum valuation t^*.

PROPOSITION 2: *For any economy $\mathcal{E}_N$ such that $C_N > t^* K_\alpha^*$ and $\alpha = \frac{c^-}{t^*}$,*

$$\sup_{(\delta,c)} E\delta = 0$$

where the sup is taken over all mechanisms (δ, c) satisfying IR, IC, BB, MC and M.

Proof: Fix $\alpha = \frac{c}{t^*}$, an economy $\mathcal{E}_N$ and a mechanism (δ, c). By Theorem 2, there are at most K_α^* individuals $\mathcal{E}_N$ for whom $V_n(\delta) \geq \alpha$. We assume that $N - K_\alpha^* > 0$ (otherwise the claim of the theorem is vacuous) and let n be an individual who is not α-pivotal.

Consider an arbitrary mechanism (δ, c) satisfying the conditions of the Theorem. The incentive compatibility constraint for a type t_n misrepresenting his type to be $t_n = 0$ becomes

$$t_n E(\delta \mid t_n) - E(c_n \mid t_n) \;\geq\; t_n E(\delta \mid t_n = 0) - E(c_n \mid t_n = 0).$$

Rearranging terms, and using the definition of influence and the assumption that c_n is t_n-measurable, we have

$$E(c_n \mid t_n) = c_n(t_n) \leq t^+ V_n(\delta) \leq t^+ \alpha < c^+ .$$

The assumption of a lower bound on positive contributions implies

$$c_n(t_n) = 0$$

for all individuals who are not α-pivotal. The maximum any α-pivotal individual will contribute is t^+ so no project which costs more than $C_N > t^+ K_\alpha^*$ will be built.

Q.E.D.

3.3. Numerical Examples

Propositions 1 and 2 can be used to compute bounds on the probability of provision $E\delta$. In illustrating Proposition 1, it is easier to work with the case where each player has finitely many types each with probability ϵ. The following is a corollary to the proof of Proposition 1:

Corollary. *Suppose that there is $\epsilon > 0$ such that for each n, T_n is finite and $P(t_n) \geq \epsilon$ for every $t_n \in T_n$. Then*

$$\sup_{(\delta, c)} E\delta \leq \frac{\max_n Et_n}{\beta} R_{\epsilon, N}$$

where the sup is taken over all mechanisms (δ, c) satisfying IR, IC, and BB.

Proof: From the definition of influence, for every n, $E(\delta \mid t_n)(\omega) - E(\delta \mid t_n = 0) \leq V_n(\delta)$ for all n and t_n. This implies $E(c_n) \leq \sum_{t_n} t_n [E(\delta \mid t_n)(\omega) - E(\delta \mid t_n = 0)] P(t_n) \leq \max_n Et_n V_n(\delta)$. From (BB), we have $E(\delta) \leq \frac{\sum Ec_n}{C_N} \leq \frac{\max_n Et_n}{\beta} V(\delta) \leq \frac{\max_n Et_n}{\beta} R_{\epsilon, N}$.

Q.E.D.

To take an example, suppose that $C_N = N$ so that $\beta = 1$, that $\epsilon = .10$ and $E_n t_n = 2$ for all n, so that on average each individual values the public good twice as much as expected per capita cost. The following table provides the value of $R_{\epsilon,N}$ and the maximum probability of provision for various values of N (The estimated values are calculated using the approximation formula in Section 2.5, and the exact value for $N = 10,000$ was difficult to compute on a personal computer):

N	$R_{\epsilon,N}$ (exact)	$R_{\epsilon,N}$ (estimate)	$E\delta$ at most
30	0.320	0.325	64%
100	0.177	0.178	36%
200	0.125	0.126	25%
500	0.079	0.079	16%
1000	0.056	0.056	12%
10,000	-	0.017	3.5%

The bound on the probability of provision declines with N slowly at the rate $N^{-1/2}$. Note that the Weak Law of Large Numbers implies that for moderately large values of N, with high probability, the sum of individual valuations for the public good is approximately twice as large as its cost, in which case it is efficient to build it.

To illustrate Proposition 2, suppose that $\epsilon = 0.10$, $t^+ = \$100$ and $c^- = \$1$. Then for $\alpha = 0.01$, one can calculate that $K_\alpha^* < 32,000$. This means that no public project costing more than \$3.2 million can be financed through a voluntary contribution mechanism satisfying the assumptions of Proposition 2. Note that this bound is independent of both the size of the community and the distribution of valuations. For example, if valuations are identically distributed with $P(t^+) = 0.90$ and $P(t_0) = 0.10$, then in any community of size $N \geq 40,000$ there is probability almost 1 that total valuation will exceed \$3.2 million. Thus, in such communities there is probability almost 1 that the project should be built, but probability 0 that it will be built under any voluntary contribution mechanism.

3.4. Discussion

i) *Mechanism Design Problems with Separable Payoffs:* A key feature of the public good problem is that individual payoffs can be decomposed into a component which depends on the collective outcome only, and another component which depends on costs and transfers. Mechanism design problems that share this feature arise naturally in many contexts including, for example, the externality and spill-over problems studied in Rob (1989) and Klibanoff and Morduch (1995). For this type of problem the incentive compatibility constraint can be decomposed (as in the public good case) into a component that depends on the transfers and a component that reflects the expected effect of a misrepresentation of types on the probability of the collective outcome. Our results on influence simplify analysis of such problems when the second component can be bounded by a (linear) function of the individual's influence. If so, then most individuals will regard the influence of their reports on the collective outcome to be so small that their behavior is almost entirely driven by the transfers component of their payoffs.

ii) *Allocation of Property Rights:* In our analysis we focused on the case in which the minimal valuation of any individual is non-negative. An interesting case is that where for some individuals the project may be harmful (*e.g.*, a garbage incineration plant or a shopping mall which may cause local traffic congestion) so $t_n^- < 0$. This raises an important point concerning the specification of property rights in the economy. We have so far assumed that each individual has the right to opt out of the project. If $t_n^- < 0$, then the issue of whether such an individual is entitled to compensation must be addressed. Our analysis can be easily adapted to other property rights structures. Alternative property right structures are studied in Chari and Jones (1994) and by Neeman (1994). Neeman examines a setting similar to the one considered above; his main result is the identification of a property right structure that can achieve an ex post efficient outcome.

iii) *Correlated Valuations:* The case of correlated valuations can be handled using Theorem 3 under the additional restriction that the mechanisms are required to be ex post individually rational.[13] The restriction to ex post individually rational mechanisms is necessary to rule out the sort of bets, familiar from the work of Cremer and McLean (1985), that can be used to induce truthful reporting at minimal cost in a way consistent with interim individual rationality. This point clarifies the scope of our analysis which provides bounds on the extent of individuals' influence on the collective-outcome component of their payoffs. Other aspects of the mechanism design problem, such as lotteries with outcomes which depend on the reports of other individuals, are in principle unrestricted by our bounds on influence.

4. APPLICATIONS TO GAMES WITH MORAL HAZARD

We provide three examples of games with moral hazard in which our results on influence can be used to make predictions about the set of equilibria. Throughout this section we maintain the following notation: Each of the N players has a binary action set $B = \{b^+, b^-\}$, where each action $b_n \in B$ of player n generates a distribution $\pi(\cdot \mid b)$ on a finite set of signals $X = \{x_1, \dots, x_M\}$. We assume that there is $\epsilon > 0$ such that $\pi_n(x, b) > \epsilon$ for all $b \in B$, $x \in X$ and uniformly in N. Player n's mixed strategy is denoted $\tilde{b}_n \in [0, 1]$ which we will interpret as the probability of action b^+. The sets of profiles of actions and signals are denoted by B^N and X^N respectively. Mixed profiles will be denoted by $\tilde{\mathbf{b}}$ and a profile in which every player plays b^+ by $\tilde{\mathbf{b}}^+$. There are two collective outcomes in a set $A = \{a^+, a^-\}$ and a decision or outcome rule $\sigma : X^N \to \Delta(A)$ that maps vectors of signals into a distribution on the set of collective outcomes. The interpretation of the collective outcomes and the decision rule will depend on the specific applications considered.

4.1. A Simple Partnership Game

We examine a simple partnership problem in which free-riding in a moral hazard context forces any equilibrium outcome to be inefficient. For simplicity, set $a^- = 0$ and interpret a^+ as the per capita output, so total output is Na^+. Given a vector of signals $\mathbf{x}$, we interpret the outcome rule $\sigma(\mathbf{x})$ as a "production function" which determines the probability of high output a^+ as a function of the individuals' signals. We impose no restrictions on σ. In particular, σ can be non-monotonic in the signals and may exhibit complex patterns of substitutability and/or complemenatrities in efforts.

Output is divided among individuals according to the sharing rule $S = (s_1, \dots, s_N)$ which allocates to individual n the share $s_n N a^+$ when output is realized. We assume that S is balanced and feasible in the sense that $0 \leq s_n \leq 1$ for all n, and $\sum_n s_n = 1$. Given a sharing rule S and action b, the payoff of agent n conditional on high output being realized is $s_n N a^+ - c_n(b)$, where $c_n(b)$ is the cost of effort when action b is taken. We assume that higher effort is more costly uniformly across agents, so there is l such that $c_n(b^+) - c_n(b^-) \geq l > 0$ for all n. With this description of payoffs, any sharing rule S defines a non-cooperative game $\Gamma(S)$ between the N agents.

Given a sharing rule S and a profile of actions of other players $\tilde{\mathbf{b}}_{-n}$, player n takes the action b^+ only if the following incentive constraint is satisfied:

$$s_n N a^* \sum_n \pi(x_n \mid b^+) \underbrace{\sum P(\mathbf{x}_{-n} \mid \tilde{\mathbf{b}}_{-n}) \sigma(\mathbf{x}_{-n}, x_n)}_{g(x_n)} - c(b^+) \geq$$
$$s_n N a^* \sum_n \pi(x_n \mid b^-) \sum P(\mathbf{x}_{-n} \mid \tilde{\mathbf{b}}_{-n}) \sigma(\mathbf{x}_{-n}, x_n) - c(b^-). \qquad (**)$$

Note that for any two signals x and x', $g(x) - g(x') \leq V_n = V_n(\sigma, \tilde{\mathbf{b}})$. Therefore, any two averages over g cannot differ by more than V_n, so (**) implies

$$s_n a^* N V_n \geq c(b^+) - c(b^-) \geq l. \qquad (***)$$

PROPOSITION 3: *For any a^*, l, $\epsilon > 0$ and N satisfying $R_{\epsilon,N} < \dfrac{l}{a^*}$, there exists no sharing rule S for which the profile $\mathbf{b}^+$ is an equilibrium for $\Gamma(S)$, for any $\{\pi_n\}$ and σ.*

Proof: From Theorem 1, for any $\{\pi_n\}$ and σ, $V = V(\sigma, \tilde{\mathbf{b}}) \leq R_{\epsilon,N}$. By the choice of N, we have $a^* V(\sigma, \tilde{\mathbf{b}}) < l$. combining this with (***), we have $s_n a^* N V_n > a^* V = a^* \frac{\sum V_n}{N}$ for all n. Therefore, $s_n > \frac{\sum V_n}{N} \frac{1}{N} \frac{1}{V_n}$ for all n. Summing over all agents, we get that $1 = \sum_n s_n > \frac{\sum V_n}{N} \frac{1}{N} \sum_n \frac{1}{V_n}$. This implies that $\frac{N}{\sum_n \frac{1}{V_n}} > \frac{\sum V_n}{N}$, contradicting the inequality of averages.[14]

Q.E.D.

Proposition 3 is of economic interest when $\mathbf{b}^+$ is the only efficient profile. This would be the case if, for example, for any n and any $\tilde{\mathbf{b}}_{-n}$ there is a net collective gain when individual n takes action b^+:

$$a^* [\sigma(\mathbf{x} \mid \tilde{\mathbf{b}}_{-n}, b^+) - \sigma(\mathbf{x} \mid \tilde{\mathbf{b}}_{-n}, b^-)] > c(b^+) - c(b^-).$$

The proposition is stated in a more general form without this condition for expositional simplicity and to better highlight the role of influence in this analysis.

To give a numerical example, assume that $\epsilon = 0.10$ and that the per capita gain a^* is, say, three times the additional cost of higher effort, so that $\frac{l}{a^*} = 0.33$. Then using the values from the table in

Section 3.3, the profile $\mathbf{b}^*$ cannot be implemented for $N \geq 30$, regardless of any of the primitives of the model a^*, l, $\{\pi_n\}$, and σ.

Finally, with simple modifications, the model can be interpreted as one in which the production function σ represents externalities generated by the agents' actions. In this case, it would be natural to consider a fixed S, representing the way in which the externality affects the agents' payoffs. The proposition would then imply that the efficient profile cannot be an equilibrium, regardless of the complexity of the mechanism generating externality.

4.2. A Game with Production Complementarities

This section introduces equilibrium considerations by viewing the outcome function σ as a strategic choice of another player. The setting is that of a non-cooperative game with production complementarities between a Principal and N small agents. We use our results on influence, an iterated dominance argument, and the fact that strategy choices are common knowledge in equilibrium to restrict the set of Nash equilibria.

Agent n has a payoff function $u(a, b) = a - c_n(b)$ with $c_n(b^*) - c_n(b^-) \geq l > 0$, as in the partnership problem of Section 4.1. Note that we now interpret a as an outcome common to all agents, so $s_n = \frac{1}{N}$ for all N. The Principal chooses a contingent strategy $\sigma : X^N \to [0, 1]$ which maps observed signals into a probability of the action a^*. Let $d(\mathbf{b})$ denote the frequency of b^* in a realized vector of actions $\mathbf{b}$. We assume that the Principal has a payoff function of the form $v(a, d(\mathbf{b}))$. Note that this does not imply that the Principal's strategy (or the agents' conjectures about it in their equilibrium reasoning) is anonymous. We consider a non- cooperative game with the above described payoffs, where agents choose a profile $\tilde{\mathbf{b}}$ of actions and the Principal chooses a strategy σ simultaneously and independently. A pair $(\sigma, \tilde{\mathbf{b}})$ is a Nash equilibrium for the game if the usual definition applies.

The motivation for this description of the game is as follows. The Principal decides what action to take after observing the (payoff-irrelevant) vector of signals $\mathbf{x}$ but before the payoff-relevant choice of $\tilde{\mathbf{b}}$ is revealed. In this setting, the Principal conditions his action a on the agents' signals $\mathbf{x}$ to influence the choice of $\tilde{\mathbf{b}}$. For example, imagine the Principal as a teacher and the agents as students. The teacher's action may be choosing how much to invest in preparing good classes and the students choose whether or not to study hard. The Principal observes signals, *e.g.*, class participation, exam results, which he can use to determine whether it will be worth his while to make high effort. Production complementarties enters the model through the assumption that high effort is worthwhile for the Principal (agents) only if the agents (Principal) take high effort. If this coordination fails, then all players prefer to supply low effort.

To model such production complementarties, we assume that for every a, the function $v(a,d) : [0,1] \to I\!R$ is monotone, and that $v(a^+,1) > v(a^-,1)$ and $v(a^+,0) < v(a^-,0)$. Monotonicity implies that there is d^* such that $v(a^+,d) \leq v(a^-,d)$ if and only if $d \leq d^*$. Thus, high effort on the part of the Principal is worthwhile to him only if d exceeds the threshold d^*. For simplicity of exposition we assume that $a^- = 0$.

PROPOSITION 4: *Let* $\alpha = \frac{l}{a^+}$. *Then,*

i) *For any* N, *there can be at most* K^*_α *agents who play* b^+ *in any Nash equilibrium.*

ii) *For large enough* N, *the game has a unique Nash equilibrium in which* $b_n = b^-$ *for every* n *and* $\sigma(\mathbf{x}) = a^-$ *for every vector of signals* $\mathbf{x}$.

Proof: For part (i), let $(\sigma, \tilde{\mathbf{b}})$ be a Nash equilibrium of the game, so all agents have a common belief that the strategy σ will played. By (**), applied when $s_n = \frac{1}{N}$, only α-pivotal agents play b^+ with positive probability. By Theorem 2, there can be at most K^*_α such agents.

For part (ii), let $(\sigma, \tilde{\mathbf{b}})$ be any Nash equilibrium. From part (i) we know that $b_n = b^+$ with positive probability for at most K^*_α players. Thus, the ratio of agents playing b^+ cannot exceed $\frac{K^*_\alpha}{N}$. For $N > \frac{K^*_\alpha}{d^*}$, the Principal's best response is $\sigma(\mathbf{x}) = a^-$ for all $\mathbf{x}$. For such σ, every player's influence is zero, so every agent plays b^-.

Q.E.D.

4.3. Cooperation in a Noisy Prinsoner's Dilemma

In this section we consider a two stage game with random matching where the first stage is a prisoner's dilemma with unobserved actions and the second stage is a coordination game. Our goal is to examine the extent to which coordination in the second stage can be used to support cooperation in the prisoner's dilemma stage. Our main result states that the number of players who cooperate is bounded uniformly over N and all Nash equilibria.

The game is played between an even number of players N. A player's payoffs is the sum of his payoffs in the two stages described below. In the first stage players choose actions (Cooperate b^+ or Defect b^-) then matched uniformly and randomly with each other. We distinguish between the actions $\{b^+, b^-\}$ and the signals $\{x^+, x^-\}$, which we interpret as 'perceived' cooperation and defection. The first stage is noisy in the sense that $\pi(x^+ \mid b^+) = \pi(x^- \mid b^-) = 1 - \epsilon$. So if action b is played, the corresponding signal is observed with high probability, but the other signal is observed with probability ϵ. At the end of stage 1 the vector of signals of all participants $\mathbf{x}$ is made public. We assume, for simplicity, that the payoff of player n in stage 1 is obtained as a function of the actions.[15]

The second stage is played without noise. Again, players are randomly and uniformly matched. Player n observes the signals from the first stage, but not the actual actions. Thus, his strategy in the second stage is of the form $\sigma_n : X^N \to [0, 1]$, where $\sigma_n(\mathbf{x})$ is interpreted as the probability of playing a^+ in the second stage given a vector of observed signals $\mathbf{x}$. The profile of second period strategies is denoted by $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$.

	b^+	b^-
b^+	1, 1	$-l$, $1+g$
b^-	$1+g$, $-l$	0, 0

Stage 1: Noisy Prisoner's Dilemma
$g, l > 0$

	a^+	a^-
a^+	c^+, c^+	0, 0
a^-	0, 0	c^-, c^-

Stage 2: Coordination Game
$c^+ > c^- > 0$

A similar example of a two stage game without noise and with a continuum of players was presented in Fudenberg and Levine (1988). They show that there is an equilibrium profile in which every player cooperates in the first stage. The idea is to design a profile in which every player is pivotal in the sense that, given the equilibrium actions of all other players, a single defection in the first stage causes a switch to a worse continuation in the second stage. Random matching in an infinitely repeated noisy prisoner's dilemma was studied by Ellison (1994). An important difference with our specification of the prisoner's dilemma stage game concerns the way information is transmitted. Ellison focuses on a special case of an information structure in which a player conditions his play only on the signal of the player with whom he was matched. In our model, we give greater freedom for players to choose the set of signals used in determining their play in the second stage. Proposition 5 below shows that the degree of cooperation is bounded uniformly over all possible information structures.

PROPOSITION 5: *There is $\alpha > 0$ such that in any equilibrium $(\tilde{\mathbf{b}}, \boldsymbol{\sigma})$ the number of players who play b^* with positive probability is at most $K^*_\alpha - 1$.*

The intuition for the result is as follows. Second stage play is represented by N mechanisms $(\sigma_1, \ldots, \sigma_N)$, each depending on the N-vector of signals $\mathbf{x}$. Because of uniform random matching, each player cares only about his influence on the average behavior of the remaining players in the second stage. A necessary condition for cooperation is that this influence is sufficiently large to compensate for the forgone rewards from not defecting in the first stage. Think of the allocation of influence in the N mechanisms $(\sigma_1, \ldots, \sigma_N)$ in terms of an $N \times N$ matrix with generic entry V_{mn} denoting the influence of player m's signal on σ_n. There are many ways in which influence can be distributed. For simplicity, imagine that for each n there is a subset M_n of players such that σ_n ignores signals of players outside M_n. If the M_n's are large (*e.g.*, $M_n = N$ for each n), then each player has small influence on all mechanisms. Our bounds would then imply that the influence of each player on average second stage behavior is small. On the other hand, if each M_n is small (so only a few players matter relative to σ_n) and there is not much overlap between them, then a typical player has large influence relative to only a few mechanisms, so his influence on the average is again small. The proof essentially shows that the degree of cooperation is bounded by mechanisms that consolidate influence in a single common subset of players $M \subset N$. We can then apply Theorem 2 to derive a bound on how M can be.

Proof: Denote $\sigma_{-n}(\mathbf{x}) = \frac{1}{N-1}\sum_{m \neq n} \sigma_m(\mathbf{x})$. Since the players are matched uniformly and randomly in the second stage, each will care only about the expected frequency of the play of a^+. Thus, the best response of player n is of the form.

$$BR_n(\mathbf{x}) = \begin{cases} a^+ & \text{if } \sigma_{-n}(\mathbf{x}) > \bar{p} \\ a^- & \text{if } \sigma_{-n}(\mathbf{x}) < \bar{p} \\ a^+ \text{ or } a^- & \text{if } \sigma_{-n}(\mathbf{x}) = \bar{p} \end{cases}$$

for some threshold $\bar{p}$ determined by the payoffs of the coordination game. In deciding on whether to play b^+ or b^-, player n will take into account only his influence on σ_{-n}. Let $V_{mn} = V_m(\sigma_n)$, so the influence of player m on σ_{-m} is $V_m = \frac{1}{N-1}\sum_{\substack{n=1 \\ n \neq m}}^{N} V_{mn}$.

For a given profile, let p_n denote player n's random payoff in the second stage and z_n denote the probability of facing cooperation in the first stage. Player n plays b^+ with positive probability in the first stage if and only if

$$\begin{aligned} z_n g + (1 - z_n)l &\leq [(1-\epsilon)E(p_n \mid x^+) + \epsilon E(p_n \mid x^-)] - [\epsilon E(p_n \mid x^+) + (1-\epsilon)E(p_n \mid x^-)] \\ &= (1 - 2\epsilon)[E(p_n \mid x^+) - E(p_n \mid x^-)]. \end{aligned} \tag{+}$$

That is, the RHS represents the gain of player n playing b^- instead of b^+, while the RHS represents the effect of shifting the distribution on his signal by putting more weight on x^+ than before.

Fix an equilibrium and let $\{1, \ldots, M\}$ denote the set of players who play b^+ with positive probability. Following an argument similar to the one used in the proof of Theorem 2, we may modify each σ_n into a new mechanism σ_n' that depends only on the signals of players $m \leq M$, $m \neq n$, and such that the sum of influences of these players increases: $\sum_{\substack{m=1 \\ m \neq n}}^{M} V_m(\sigma_n') \geq \sum_{\substack{m=1 \\ m \neq n}}^{M} V_{mn}$. Writing $V_{mn}' = V_m(\sigma_n')$, we have

$$\sum_{m=1}^{M} V_m' = \frac{1}{N-1}\sum_{m=1}^{M}\sum_{\substack{n=1 \\ n \neq m}}^{N} V_{mn}' = \frac{1}{N-1}\sum_{n=1}^{N}\sum_{\substack{m=1 \\ n \neq m}}^{M} V_{mn}' \geq \frac{1}{N-1}\sum_{n=1}^{N}\sum_{\substack{m=1 \\ n \neq m}}^{M} V_{mn} = \sum_{m=1}^{M} V_m.$$

From Theorem 1, there is a symmetric environment such that for every n there is a symmetric, ϵ-extremal profile and an anonymous mechanism $\hat{\sigma}_n$ depending only on the signals of players $m \leq$

M, $m \neq n$ such that: $\sum_{\substack{m=1 \\ m \neq n}}^{M} \hat{V}_{mn} \geq \sum_{\substack{m=1 \\ m \neq n}}^{M} V'_{m}$, where $\hat{V}_{mn} = V_m(\hat{\sigma}_n)$. Note that in the new environment the number of signals has changed; the 'profiles' $\hat{\sigma}_n$'s are used in computing bounds on influence and have no strategic meaning in terms of the players' actions.

Since $p_n \leq c^{-} + (c^{+} - c^{-})\sigma_{-n}$, (+) implies that a necessary condition for playing b^{+} with positive probability is

$$\min\{g, l\} \leq (1-2\epsilon)(c^{+} - c^{-})[E(\sigma_{-n} \mid x^{+}) - E(\sigma_{-n} \mid x^{-})] \leq (1-2\epsilon)(c^{+} - c^{-})V_n.$$

Setting $\alpha = \frac{\min\{g,l\}}{(1-2\epsilon)(c^{+}-c^{-})}$, for every $m \leq M$ we must have $V_m \geq \alpha$, and

$$\sum_{m=1}^{M} \hat{V}_m \geq \sum_{m=1}^{M} V'_m \geq \sum_{m=1}^{M} V_m \geq M\alpha.$$

Thus, there is $\bar{m} \leq M$ such that $\hat{V}_{\bar{m}} = \frac{1}{N-1} \sum_{\substack{n=1 \\ n \neq m}}^{N} \hat{V}_{\bar{m}n} \geq \alpha$. This in turn implies that there is $\bar{n}$ such that $\hat{V}_{\bar{m}\bar{n}} \geq \alpha$. The anonymity of $\hat{\sigma}_{\bar{n}}$ implies that $\hat{V}_{m\bar{n}} \geq \alpha$ for every $m \leq M$, $m \neq \bar{n}$. This means that every such player m is α-pivotal with respect to the mechanism $\hat{\sigma}_{\bar{n}}$, so by Theorem 2, M is bounded by $K^{*}_{\alpha} + 1$.

Q.E.D.

APPENDIX

For a player n let t_n^+ and t_n^- denote the pair of signals at which his maximum influence is achieved (that is, $V_n(F, \tilde{\mathbf{t}}) = V_n(F, \tilde{\mathbf{t}}; t_n^+, t_n^-)$). It is also convenient to define his *conditional influence* given the signal of player n' to be:

$$V_n(F, \tilde{\mathbf{t}} \mid t_{n'} = t_{n'}^m) = \sum_{\mathbf{t}_{-n}} P(\mathbf{t}_{-n} \mid t_{n'} = t_{n'}^m) \left[F(\mathbf{t}_{-n}, t_n = t_n^+) - F(\mathbf{t}_{-n}, t_n = t_n^-)\right]$$

With this notation, average influence $V(F, \tilde{\mathbf{t}})$ can be expressed in terms of the probability distribution of player n's signals as:

$$N\, V(F; \tilde{\mathbf{t}}) = V_n(F, \tilde{\mathbf{t}}; t_n^+, t_n^-) + \sum_{m=1}^{M_n} P(\tilde{t}_n = t_n^m) \underbrace{\sum_{n' \neq n}^{N} V_{n'}(F, \tilde{\mathbf{t}} \mid t_n = t_n^m)}_{A^m}.$$

The second part of this expression: $\sum_{m=1}^{M} P(\tilde{t}_n = t_n^m) A^m$ is a linear function of agent n's distribution $P(\tilde{t}_1 = t_1^m)$ and is therefore maximized at some signal which we denote $t_n^{\max}$.

PROPOSITION A.1: *Fix $\epsilon > 0$, F and $\tilde{\mathbf{t}} \in \Delta_\epsilon^N$. Then there is a restricted signal set $\hat{T}_1$ of three distinct signals for player 1, a distribution $\hat{t}_1 \in \text{ext}\, \Delta_\epsilon(\hat{T}_1)$ and a mechanism $\hat{F} : \hat{T}_1 \times T_{-n} \to [0, 1]$ such that*

$$V(F, \tilde{\mathbf{t}}) \leq V(\hat{F}, \hat{t}_1 \times \tilde{\mathbf{t}}_{-n})$$

Proof: Consider the subset of signals $\hat{T}_1' = \{t_1^+, t_1^-, t_1^{\max}\}$ in which some of the signals may be repeating (this will necessarily be the case if player 1 has only two signals). Consider the distribution $\hat{t}_1'$ which assigns each of t_1^+, t_1^- probability ϵ and probability $1 - 2\epsilon$ to $t_1^{\max}$. Define $\hat{F}' : \hat{T}_1' \times T_{-n} \to [0, 1]$ as the natural restriction of the original F to the restricted signals space $\hat{T}_1'$. Since individual 1's influence depends on the value taken by F at $\{t_1^+, t_1^-\}$ only, his contribution to total influence is unaffected by this change in the signal space and the distribution. From the definition of $t_1^{\max}$ we also have that shifting weight $1 - 2\epsilon$ to it (weakly) increases the total contribution of other players to total influence. Thus, $V(F, \tilde{\mathbf{t}}) \leq V(\hat{F}', \hat{t}_1' \times \tilde{\mathbf{t}}_{-n})$.

We now convert $\hat{T}_1'$ to a set of three distinct signals $\hat{T}_1$. For later use, it will be notationally convenient to choose $\hat{T}_1 = \{0, 1, 2\}$ to be a standard signal space common to all players. If the signals $\{t_1^+, t_1^-, t_1^{\max}\}$ are all distinct, then identify $t_1^{\max}$ with 1, t_1^+ with 2, and t_1^- with 0, and define $\hat{F} : \hat{T}_1 \times T_{-n} \to [0, 1]$ to coincide with $\hat{F}_1'$ under this identification of signals. We now turn to the various cases in which $\{t_1^+, t_1^-, t_1^{\max}\}$ fail to be distinct.

Assume first that $t_1^+ = t_1^{\max}$, so the probability of this combined signal is actually $1 - \epsilon$. In this case, identify $t_1^{\max}$ with 1, and t_1^- with 0. We then split from signal 1 a new signal 2 that carries probability ϵ (so signal 1 is now left with probability $1 - 2\epsilon$). Define the mechanism $\hat{F}_1$ so that signal 2 is redundant in the sense that $\hat{F}$ treats signal 2 in exactly the same way as signal 1. That is, the new mechanism $\hat{F} : \hat{T}_1 \times T_{-n} \to [0, 1]$ is defined by

$$\hat{F}(0, \mathbf{t}_{-1}) = F(t_1^-, \mathbf{t}_{-1})$$

$$\hat{F}(1, \mathbf{t}_{-1}) = F(t_1^+, \mathbf{t}_{-1})$$

$$\hat{F}(2, \mathbf{t}_{-1}) = F(t_1^+, \mathbf{t}_{-1}).$$

Note that this does not affect the influence of any individual. The proof for the case in which $t_1^- = t_1^{\max}$ is similar. Finally, if $t_1^+ = t_1^-$, then the mass of 2ϵ assigned to t_1^m can be split over two signals which F treats in the same manner, and this can be done without reducing influence.

Q.E.D.

Call $(\{\hat{T}_n\}_{n=1}^N\}, \tilde{\mathbf{t}})$ the *standard* environment if: (1) all agents have the same signal sets $\hat{T}_n = \{0, 1, 2\}$; (2) each agent has an extremal distribution such that signal 1 has probability $1 - 2\epsilon$. In such environment, call a mechanism F *regular* if each agent's maximum influence is achieved when his signal changes from 0 to 2.

PROPOSITION A.2: *Fix $\epsilon > 0$, F and $\tilde{\mathbf{t}} \in \Delta_\epsilon^N$. Then there is a regular mechanism $\hat{F}$ in the standard environment $(\{\hat{T}_n\}_{n=1}^N\}, \hat{\mathbf{t}})$ such that*

$$V(F, \tilde{\mathbf{t}}) \leq V(\hat{F}, \hat{\mathbf{t}}).$$

Proof: Apply Proposition A.1 relative to player 1 to obtain a new signal space $\hat{T}_1 \times T_2 \times \cdots \times T_N$, a vector of random signals $\tilde{\mathbf{t}}^1 = \hat{t}_1 \times \tilde{\mathbf{t}}_{-n}$, and a mechanism $\hat{F}_1$ so that $V(F, \tilde{\mathbf{t}}) \leq V(\hat{F}_1, \tilde{\mathbf{t}}^1)$. Repeating this process for player 2 relative to $\hat{F}_1$ and $\tilde{\mathbf{t}}^1$ yields a new mechanism $\hat{F}_2$ and profile $\tilde{\mathbf{t}}^2$ such that average influence does not decrease. Continuing in this manner for all remaining players, we obtain a sequence of pairs $(\hat{F}_n, \tilde{\mathbf{t}}^n)$ along which $V(\hat{F}_n, \tilde{\mathbf{t}}^n)$ is increasing. The claim is proved by setting $\hat{F} = \hat{F}_N$ and $\hat{\mathbf{t}} = \tilde{\mathbf{t}}^N$. The new distribution $\hat{\mathbf{t}}$ is symmetric and extremal by construction.

Q.E.D.

PROPOSITION A.3: *In the standard environment $(\{T_n\}_{n=1}^N\}, \tilde{\mathbf{t}})$, for any regular mechanism F there is an anonymous regular mechanism F' such that $V(F, \tilde{\mathbf{t}}) = V(F', \tilde{\mathbf{t}})$.*

Proof: Let σ be any permutation of the set of players' names, and $\tilde{\mathbf{t}}$ be any symmetric distribution (references to $\tilde{\mathbf{t}}$ are dropped for notational simplicity). Define the new mechanism F^σ by $F^\sigma(\mathbf{t}) = F(\sigma(\mathbf{t}))$. We show that $V_n(F^\sigma) = V_{\sigma^{-1}(n)}(F)$ for every n:

$$\begin{aligned} V_n(F^\sigma) &= E(F^\sigma \mid t_n = 2) - E(F^\sigma \mid t_n = 0) \\ &= E(F \mid t_{\sigma^{-1}(n)} = 2) - E(F \mid t_{\sigma^{-1}(n)} = 0) \\ &= V_{\sigma^{-1}(n)}(F). \end{aligned}$$

Thus,

$$\begin{aligned} V(F^\sigma) &= \frac{1}{N} \sum_{n=1}^{N} V_n(F^\sigma) \\ &= \frac{1}{N} \sum_{n=1}^{N} V_{\sigma^{-1}(n)}(F) \\ &= V(F). \end{aligned}$$

Let p be the uniform probability distribution on the set of all permutations σ over players' names. Define

$$F'(\mathbf{t}) = \sum_{\sigma} p(\sigma) F^\sigma(\mathbf{t}).$$

Obviously, $F'(\mathbf{t})$ is symmetric and

$$V(F') = \sum p(\sigma)V(F^{\sigma}) = \sum p(\sigma)V(F) = V(F).$$

Q.E.D.

PROPOSITION A.4: *Fix the standard environment $(\{\hat{T}_n\}_{n=1}^{N}\}, \tilde{\mathbf{t}})$ and let F_m denote the majority rule relative to signals 0 and 2. Let F be any regular mechanism F, then*

$$V(F, \tilde{\mathbf{t}}) \leq V(F_m, \tilde{\mathbf{t}}) \leq R_{\cdot,N}.$$

For the proof we will need some additional notation. As before, it is convenient to think of signals 0 and 2 as 'No' and 'Yes' respectively, and signal 1 as representing 'all-other-signals', or 'Abstain'. Fix any set of $N-1$ players, and let $K = 0, \ldots, N-1$ be the random variable denoting the number of non- abstaining players out of this set, and let $P_{N-1}(K)$ denote its probability. Since the signal sets and the profiles are symmetric, the identity of the $N-1$ players is irrelevant.

Proof: Let k be the random variable denoting the number of players saying 'Yes' (*i.e.*, whose signal is 2) out of $N-1$ players, and let $P_{N-1}(k; K)$ denote its conditional probability (or simply $P(k \mid K)$ when N is clear from the context).

It is convenient to express F as a function $F(k, K)$ of the number of Yes's k and the number of voting players K. For given K and k, if player n changes his signal from 0 to 2 (a change that gives him maximum influence), then there will be $K+1$ non-abstaining players and $k-1$ Yes's, and the outcome changes from $F(k, K+1)$ to $F(k+1, K+1)$. Note that while K and k represent the relevant uncertainty from the perspective of player n, as far as F is concerned the total number of non-abstaining players is $K+1$.

Fix F that satisfies the assumptions of the proposition (we drop references to $\tilde{\mathbf{t}}$ for notational simplicity): define

$$V_n^K = \sum_{k=0}^{K} P(k \mid K) \; [F(k+1, K+1) - F(k, K+1)];$$

and note that player n's influence is just:

$$V_n(F) = \sum_{K=0}^{N-1} P(K) \;\; V_n^K.$$

Thus, V_n^K represents player n's influence conditional on there being K non-abstaining players out of the remaining $N-1$ players.

It is convenient to rewrite V_n^K as:

$$\begin{aligned} V_n^K(F) = -P(0 \mid K)\; F(0, K+1) + {} & F(k=1, K+1)\; [P(k=0 \mid K) - P(k=1 \mid K)] \\ & + \dots \\ & + F(k, K+1)\; [P(k-1 \mid K) - P(k \mid K)] \\ & + \dots \\ & + F(K-1, K+1)\; [P(K-2 \mid K) - P(K-1 \mid K)] \\ & + F(K+1, K+1)\; P(K \mid K). \end{aligned}$$

We view $V_n^K(F)$ as a function of the values $F(k, K+1)$, $k = 0, \dots, K$ for fixed probability weights and with the constraints that $0 \leq F(k, K+1) \leq 1$, for $k = 0, \dots, K$. Since the constraint set is bounded, the maximum of V_n^K is achieved, and at which point the first order conditions must be satisfied. Since $P(0 \mid K)$ and $P(1 \mid K) > 0$, we must have $F(0, K+1) = 0$ and $F(k, K+1) = 1$.

For $k = 1, \dots, K$ we have

$$\begin{aligned} & \frac{P(k \mid K)}{P(k-1 \mid K)} > 1 \implies F(k, K+1) = 0 \\ \text{and} \qquad & \frac{P(k \mid K)}{P(k-1 \mid K)} < 1 \implies F(k, K+1) = 1. \end{aligned}$$

That is, for a fixed K, the highest value of V_n^K is achieved when, $F(\cdot, K+1)$ is zero as long as the binomial probability $P(k \mid K)$ is increasing in k, and 1 as long as $P(k \mid K)$ is decreasing in k.

Conditional of K, the signals of the K non-abstaining players are independently and identically distributed with probability 0.5 for 0 and 0.5 for 2. Thus, $P(k \mid K)$ is a Bernoulli distribution of K independent and identically distributed random variable with probability of success 0.5. This is a symmetric distribution whose maximum value is $\rho_{K,0.5}$ by definition.

We show that the optimal mechanism F_m has the form of a simple majority rule. If $K = 0$, then $V_n^K = F(1,1) - F(0,1)$, a number which is maximized at the 'majority rule' $F(1,1) = 1$ and $F(0,1) = 0$. For $K \geq 1$, note that

$$\frac{P(k \mid K)}{P(k-1 \mid K)} = \frac{\binom{K}{k}}{\binom{K}{k-1}} = \frac{K-k+1}{k}.$$

Assume first that $K = 2L$ for some non-negative integer L (*i.e.*, K is even). In this case, $P(k \mid K)$ achieves its unique maximum at $k = L$, so F_m must be of the form

$$k \leq L \implies F_m(k, K+1) = 0$$

$$k > L \implies F_m(k, K+1) = 1$$

which is a majority rule. If $K = 2L + 1$ for some integer L (*i.e.*, K is odd), then $P(k \mid K)$ achieves its maximum at both $k = L$ and $k = L + 1$. In this case, we can set F_m to be the majority rule:

$$k \leq L + 1 \implies F_m(k, K+1) = 0$$

$$k > L + 1 \implies F_m(k, K+1) = 1$$

which again has the form of a majority rule.

We have therefore shown that conditional on K, the mechanism that maximizes influence is a simple majority rule for which influence is $\rho_{K,0.5}$. Averaging over $K = 0, \ldots, N-1$ with the probability distribution $P(K)$ yields the desired result.

Q.E.D.

Proof of Theorem 1: The proof follows by combining propositions A.1 through A.4. Using Proposition A.1 and A.2 we can reduce any general problem to one in a standard environment and a regular F without reducing average influence. Propositions A.3 then shows that any such F is equivalent (*i.e.*, yields the same influence for each player) to a mechanism that is anonymous. Finally, Proposition A.4 shows that simple majority rule is the anonymous rule with the maximum influence and computes the bound.

Q.E.D.

Proof of Theorem 2: The bound is clearly achieved as described in the statement of the theorem. It is also clear that no anonymous mechanism on a symmetric environment can exceed this bound. Suppose, by way of contradiction, that there is $N > K_\alpha^*$ such that there is a profile $\tilde{\mathbf{t}}$ and a mechanism F with $L > K_\alpha^*$ α-pivotal players. For convenience, reorder the players so that the α-pivotal players are the first L and let $V^L(F, \tilde{\mathbf{t}}) \geq \alpha$ denote their average influence. Also, write any vector of signals $\mathbf{t}$ as $(\mathbf{t}_L, \mathbf{t}_{-L})$. Now $V^L(F, \tilde{\mathbf{t}})$ is maximized at some vector $\bar{\mathbf{t}}_{-L}$ of signals of players outside L. Define the new mechanism F_L by setting $F_L(\mathbf{t}) = F(\mathbf{t}_L, \bar{\mathbf{t}}_{-L})$ for every $\mathbf{t}$. Clearly, $V^L(F_L, \tilde{\mathbf{t}}) \geq V^L(F, \tilde{\mathbf{t}})$. That is, F_L ignores what players outside L do, yet still increases the average influence of players in L (note, however, that there may be less than L α-pivotal players under F_L). We can now interpret F_L as a mechanism in a problem with $L > K_\alpha^*$ players. For this problem, we know from the proof of Theorem 1 that there is an anonymous F' and a symmetric $\tilde{\mathbf{t}}'$ in a symmetric environment with L players such that $V(F', \tilde{\mathbf{t}}') \geq V(F_L, \tilde{\mathbf{t}}) \geq \alpha$. This together with the symmetry of $\tilde{\mathbf{t}}'$ and anonymity of F' imply that there are L α-pivotal players, which contradicts the definition of K_α^* and the assumption that $L > K_\alpha^*$.

Q.E.D.

Proof of Theorem 3:

$$NV(F) = N\sum_{\mathbf{t}} P(\mathbf{t})V(F:\mathbf{t}) = \sum_{\mathbf{t}} P(\mathbf{t}) \sum_{n} V_n(F:t_n)$$

$$= \sum_{n}\sum_{t_n}\sum_{\mathbf{t}_{-n}} P(\mathbf{t}_{-n}:t_n)V_n(F:t_n)$$

$$= \sum_{n}\sum_{t_n} V_n(F:t_n)\sum_{\mathbf{t}_{-n}} P(\mathbf{t}_{-n}:t_n)$$

$$= \sum_{n}\sum_{t_n} P(t_n)V_n(F:t_n).$$

Let $t^*_{t_n}$ and t_{*t_n} denote the signals of player n at which his influence is achieved when his actual signal is t_n. Also, let t^*_θ and $t_{*\theta}$ be the signals at which player n maximizes his influence if he knew that the aggregate state is θ. That is,

$$\max_{t_n\in T_n} E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n},t_n)\,|\,\theta) - \min_{t_n\in T_n} E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n},t_n)\,|\,\theta) = E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n},t^*_\theta)\,|\,\theta) - E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n},t_{*\theta})\,|\,\theta).$$

Then,

$$\sum_{n=1}^{N}\sum_{t_n} P(t_n)V_n(F:t_n) = \sum_{n}\sum_{t_n} P(t_n)\,[E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t^*_{t_n})\,|\,t_n) - E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t_{*t_n})\,|\,t_n)]$$

$$= \sum_{n}\sum_{t_n} P(t_n)\sum_{\theta} P(\theta\,|\,t_n)[E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t^*_{t_n})\,|\,t_n,\theta) - E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t_{*t_n})\,|\,t_n,\theta)]$$

$$\leq \sum_{n}\sum_{t_n} P(t_n)\sum_{\theta} P(\theta\,|\,t_n)[E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t^*_\theta)\,|\,t_n,\theta) - E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t_{*\theta})\,|\,t_n,\theta)]$$

$$= \sum_{n}\sum_{t_n} P(t_n)\sum_{\theta} P(\theta\,|\,t_n)[E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t^*_\theta)\,|\,\theta) - E_{\mathbf{t}_{-n}}(F(\mathbf{t}_{-n}:t_{*\theta})\,|\,\theta)] \qquad (*)$$

$$= \sum_{n}\sum_{t_n} P(t_n)\sum_{\theta} P(\theta\,|\,t_n)V_n(F:\theta)$$

$$= \sum_n \sum_\theta V_n(F:\theta) \sum_{t_n} P(\theta \mid t_n) P(t_n)$$

$$= \sum_\theta P(\theta) \sum_n V_n(F:\theta).$$

where the conditional independence assumption was used to conclude that taking conditional expectation relative to t_n in the expression in square brackets in (*) is superfluous when θ is known.

Q.E.D.

Proof of Theorem 4: For each n, fix a version of the conditional expectation $E(F \mid \mathcal{G}_n)(t)$. The proof consists of constructing a sequence of mechanisms $F = \hat{F}_0, \hat{F}_1, \ldots, \hat{F}_N = \hat{F}$ along which average influence increases, and such that under $\hat{F}$ all players have just three signals. We then apply the bound obtained in Theorem 1 for the finite signal case to $\hat{F}$.

Fix $0 < \eta < \epsilon$. For $n = 1, \ldots, N$ and any mechanism F, define the functions

$$W_{-n}(F, t_n = t) = \sum_{m<n}^{N} V_m(F, \eta \mid t_n = t) + \sum_{m>n}^{N} V_m(F, \epsilon \mid t_n = t)$$
$$W_{-n}(F) = \sum_{m<n}^{N} V_m(F, \eta) + \sum_{m>n}^{N} V_m(F, \epsilon).$$

That is, W_{-n} represents the sum of the η-influences of players before player n and the ϵ-influences of players after n.

For $1 \le n \le N$ let $\hat{F}_{n-1}$ be the mechanism constructed in the previous step (for $n = 1$, this is just the original mechanism F). Let A^+ be a set of measure η such that $E(F \mid \mathcal{G}_n)(t) \ge E(F \mid \mathcal{G}_n)(t')$ for any $t \in A^+$ and $t' \notin A^+$. Such set A^+ exists since P is non-atomic by assumption. Similarly, define A^- to be a set of measure η for which $E(F \mid \mathcal{G}_n)(t) \le E(F \mid \mathcal{G}_n)(t')$ for any $t \in A^-$ and $t' \notin A^-$. Let $t_n^* \in \operatorname{argmax}_{t \in [0,1]} W_{-n}(\hat{F}_{n-1}, t)$, $t_n^+ \in \operatorname{argmax}_{t \in A^+} W_{-n}(\hat{F}_{n-1}, t)$ and $t_n^- \in \operatorname{argmax}_{t \in A^-} W_{-n}(\hat{F}_{n-1}, t)$. If the maximum in these definitions is not achieved, then the same argument would go through by taking a sequence of signals for which $E(F \mid \mathcal{G}_n)$ converges to the supremum without changing the basic idea of the proof.

Define a new mechanism:

$$\hat{F}_n(t_n, \mathbf{t}_{-n}) = \begin{cases} F(t_n^+, \mathbf{t}_{-n}) & \text{if } t_n \in A^+ \\ F(t_n^-, \mathbf{t}_{-n}) & \text{if } t_n \in A^- \\ F(t_n^*, \mathbf{t}_{-n}) & \text{otherwise.} \end{cases}$$

Note that:

$$E(F \mid \mathcal{G}_n)(t_n^+) \geq \sup_{t \in A^+} E(F \mid \mathcal{G}_n)(t) \geq \inf_{\{A : \lambda(A) < \epsilon\}} \sup_{t \notin A} E(F \mid \mathcal{G}_n)(t)$$

and

$$E(F \mid \mathcal{G}_n)(t_n^-) \leq \inf_{t \in A^-} E(F \mid \mathcal{G}_n)(t) \leq \sup_{\{A : \lambda(A) < \epsilon\}} \inf_{t \notin A} E(F \mid \mathcal{G}_n)(t).$$

Thus,

$$\begin{aligned} V_n(F, \epsilon) &\leq E(F \mid \mathcal{G}_n)(t_n^+) - E(F \mid \mathcal{G}_n)(t_n^-) \\ &= E(\hat{F}_n \mid \mathcal{G}_n)(t_n^+) - E(\hat{F}_n \mid \mathcal{G}_n)(t_n^-) \\ &= V_n(\hat{F}_n, \eta). \end{aligned}$$

By the choice of t_n^+, t_n^- and t_n^*, we have $W_{-n}(\hat{F}_{n-1}) \leq W_{-n}(\hat{F}_n)$, and by the above argument we also have $V_n(\hat{F}_{n-1}, \epsilon) \leq V_n(\hat{F}_n, \eta)$. We therefore conclude that:

$$W_{-n}(\hat{F}_{n-1}) + V_n(\hat{F}_{n-1}, \epsilon) \leq W_{-n}(\hat{F}_n) + V_n(\hat{F}_n, \eta).$$

Continuing this process, we have:

$$N\,V(F, \epsilon) = W_{-1}(F) + V_1(F, \epsilon) \leq W_{-N}(\hat{F}_N) + V_N(\hat{F}_N, \eta) = N\,V(\hat{F}, \eta).$$

Note that under $\hat{F}_n$, each player $m \leq n$ effectively has only three signals: t_m^+, t_m^- and t_m^* each with probability at least η. Therefore, Theorem 1 applies to $\hat{F}$, from which we conclude that

$$V(F, \epsilon) \leq V(\hat{F}, \eta) \leq R_{\eta, N}.$$

The conclusion of the theorem now follows by taking $\eta \to \epsilon$ and noting that $R_{\eta, N}$ is continuous in η.

Q.E.D.

ENDNOTES

1- To put these possibilities in perspective, suppose that there are N players each having three signals: $\{0, 1, 2\}$. Then, with $N = 10$ players there are 6 majority rules, 2^{66} anonymous rules, and $2^{3^{10}} = 2^{59,049}$ general non-anonymous rules.

2- The motivation for this work is Al-Najjar (1996) where the results on influence are used to study the role of free-riding among small opponents in challenging the reputation of a central player in a repeated game setting.

3- Our results go through (with appropriate re-scaling of the bounds) if the interval $[0,1]$ is replaced with any bounded subset of real numbers.

4- Formally, we consider permutations $\sigma : \{1, \ldots, N\} \to \{1, \ldots, N\}$ such that $\sigma(n) = n$ for all $n \notin K$.

5- The definition is implicit in the interim incentive compatibility constraint in Mailath and Postlewaite (1990) where player n's influence is $\rho_n(v) - \rho_n(v')$ (p. 353).

6- In some applications, this definition is stronger than necessary because it might ignore useful model-specific information which could rule out some changes as being clearly suboptimal for the player. The definition can be modified accordingly by defining the max on a more restricted set of signal pairs.

7- Although our exposition is restricted to binary outcomes, one can easily extend the analysis to a finite number of L outcomes. Observe that the distance in the L-dimension simplex, which is contained R^{L-1}, is bounded by the sum of the distances along the axes. Thus, the influence for L outcomes is bounded by $L - 1$ times the bound for the binary case.

8- The error in Stirling's formula is bounded by a function which converges very rapidly to zero. Thus, the error in our approximation of ρ also improves rapidly in N. For example, for $N = 10$, the error in Stirling's formula is no more than 0.8%, so our approximation is inaccurate by no more than: $\max\left\{\left|\frac{1-0.008}{(1+0.008)^2} - 1\right|, \left|\frac{1+0.008}{(1+0.008)^2} - 1\right|\right\} \leq 0.025$, or less than 2.5%. Similarly, for $N = 100$, the error in our estimate of ρ is no more than 0.2%.

9- Without the conditional independence assumption, our bounds on influence (Theorem 3) do not hold. For example, suppose that there are $2N$ players each with two signals (0 or 1). Consider the distribution P obtained by first selecting at random a subset of N players, then assign to members of that set signal 1 and 0 otherwise. The only aggregate state in this case is the realization itself, so the conditional distribution $P(\cdot|\theta)$ does not belong to Δ_ϵ^N and assumption A2 fails. Consider now the mechanism $F(\mathbf{t}) = 1$ if the signals of exactly N players are 1 and 0 otherwise. Then every player is fully pivotal, and average influence is equal to 1 regardless of how large N is.

10- To see this, let $x(t)$ and $y(t)$ be two versions and let E be the set of measure zero on which they disagree. If $\inf_{\{A:P(A)<\epsilon\}} \sup_{t\notin A} y(t) > \inf_{\{A:P(A)<\epsilon\}} \sup_{t\notin A} x(t)$, then there must be a set B, with $P(B) < \epsilon$ such that $\inf_{\{A:P(A)<\epsilon\}} \sup_{t\notin A} y(t) > \sup_{t\in B^c} x(t)$. Since E has measure zero, $P(B\cup E) < \epsilon$, so $\inf_{\{A:P(A)<\epsilon\}} \sup_{t\notin A} y(t) \leq \sup_{t\in B^c\cap E^c} y(t) = \sup_{t\in B^c\cap E^c} x(t) \leq \sup_{t\in B^c} x(t)$, which is a contradiction.

11- The case of $t_n^- < 0$ requires a more careful consideration of the property right structure. This is discussed in Section 3.4.

12- It is easy to see that the proof yields a somewhat sharper bound $\frac{\max_n Et_n}{\beta} R_{\eta,N} + \frac{t^-}{\beta}\eta$. The weaker bound in the statement of the proposition is maintained for simplicity of exposition.

13- This conclusion is also obtained by Chari and Jones (1994) in a different model, where they also impose ex post individual rationality.

14- The inequality of averages states that for any finite sequence of numbers V_n, the harmonic average does not exceed the geometric average which does not exceed the arithmetic average. Formally, $\frac{N}{\sum_n \frac{1}{V_n}} \leq (\prod_{n=1}^N V_n)^{1/N} \leq \frac{\sum V_n}{N}$.

15- It may be more natural to assume that payoffs are determined using the signals. This has no qualitative impact on the result, but slightly complicates notation. The current assumption may be justified by viewing the payoff in stage 1 as accruing after both stages have been already played.

REFERENCES

Al-Najjar, N.I. (1996): "The Public-Good Nature of Reputation Building," MEDS Department, Kellogg GSM, Northwestern University.

Chari, V.V. and L. Jones (1994): "A Reconsideration of the Problem of Social Costs: Free Riders and Monopolists," Northwestern University.

Cremer, J. and R. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, **56**, 1247-57.

Ellison, G. (1994): "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies*, **61**, 567-88.

Fudenberg, D. and D. Levine (1988): "Open-Loop and Closed-Loop Equilibria in Dynamic Games with Many Players," *Journal of Economic Theory*, **44**, 1-18.

Fudenberg, D., D. Levine, and W. Pesendorfer (1995): "When are Non-Anonymous Players Negligible?," CMSEMS discussion paper no. 1114, Northwestern University.

Klibanoff, P. and J. Morduch (1995): "Decentralization, Externalities, and Efficiency," *Review of Economic Studies*, **62**, 223-47.

Mailath, G.J. and A. Postlewaite (1990): "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies*, **57**, 351-67.

Myerson, R. and M. A. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, **29**, 265-81.

Neeman, Z. (1994): "Property Rights and Efficiency of Public-Good Mechanisms Under Asymmetric Information," MEDS Department, Kellogg GSM, CMSEMS discussion paper no. 1092, Northwestern University.

Rob, R. (1989): "Pollution Claim Settlements under Private Information," *Journal of Economic Theory*, **47**, 307-333.