Probabilistic Cheap Talk†

Bhaskar Chakravorti*

John P. Conley**

and

Bart Taub***

† **Published as** Bhaskar Chakravorti, John Conley and Bart Taub,"Probabilistic Cheap Talk" *Social Choice and Welfare*, 2002, Vol. 19, pp. 281-294. We are grateful to Dilip Abreu, Robert Aumann, Phil Reny, Mark Walker, and several seminar audiences for their comments.

* The Monitor Corporation

** Department of Economics, UIUC, j.p.conley@vanderbilt.edu

*** Department of Economics, UIUC

Abstract

We consider a model in which there is uncertainty over when a one-shot game will be played. We show how a mechanism designer can implement desirable outcomes in certain economic games by manipulating only the probability that the game is played in a given round while leaving all other aspects of the game unchanged. We also show that if there is no discounting, this uncertainty imparts a sequential structure that is almost mathematically equivalent to a repeated version of the game with discounting. In particular, a folk theorem applies to such games. Thus, games of probabilistic cheap provide a third interpretation of the repeated game framework with the additional feature that expected payoff is invariant to the probability of the game ending.

Keywords: Probabilistic cheap talk, implementation, preference revelation, prisoner's dilemma, folk theorem .

JEL: C72, D43, D62.

1. Introduction

Nash's (1950, 1951 1953) original program in game theory was a mixture of cooperative and noncooperative approaches. Most famously, he considered the outcome of one-shot simultaneous move games when agents were motivated only by personal gain. He noted, however, that in many cases this behavioral assumption made it impossible for agents to coordinate their strategies is ways which would serve their mutual interests. Nash's alternative was to suppose that agents were motivated instead by a sense of fairness as captured by a set of axioms on outcomes. Nash showed that when agents bargained over the division of spoils in a way that took their ethical beliefs into account, they could often overcome this problem and achieve efficient solutions.

Taken separately, neither of these approaches is fully satisfactory. On the one hand, everyday experience suggests that agents often do manage to cooperate a much greater degree than narrow self-interest would allow. For example, people give to charity and participate in the electoral process in very high proportions. The finding that agents do not behave as noncooperative game theory predicts is also strongly confirmed by experimental evidence.[1] On the other hand, economists are rightly loath to give up the notion that agents are rational maximizers of well defined objectives. This approach has generated many insights that have been empirically confirmed. Clearly, individuals do respond to incentives to a large degree, and certainly behave in the aggregate as if they were narrowly motivated in a wide variety of economic situations.

One way to reconcile these two competing views of behavior is to explore conditions under which mutually beneficial or ethically motivated behavior can also be rationalized as serving the personal interests of individuals. The most famous result in this spirit is probably the Folk Theorem. Fudenberg and Maskin (1986), for example, shows that for a broad class of one-shot games, if play is infinitely repeated and agents are patient enough, efficient outcomes become supportable as non-cooperative equilibria. One immediate problem with this approach is that in many real world situations, repeated

[1] See for example Guth *et al.*(1982).

play is simply not feasible. For example, awarding a defense contract for a particular fighter aircraft or collecting contributions from various levels of governments to fund the building of specific bridge are actions with are essentially one-shot in nature. An even deeper and more general problem, however, is that the repeated game structure generates an embarrassment of riches. In addition to the desirable outcomes, the Folk Theorem tells us that every other individually rational payoff can be supported as noncooperative equilibrium. Thus, the repeated game framework explains why agents might cooperate, but does not suggest why this outcome is any more likely than any other.

This lack of predictive power focused attention on situations in which fair, efficient, focal, or otherwise desirable outcomes were the only ones which could be supported by self-interested play. The result was a large literature on mechanism design and implementation theory. Famous early papers include Vickrey (1961), Clark (1971) and Groves and Ledyard (1971).[2] The difficulty with this approach is that the games which support desirable outcomes are typically quite complicated and often involve catastrophic punishments or extreme sensitivity to the actions of a single agent. Moreover, despite a great deal of time and effort spent developing such mechanisms, their impact on allocation decisions in the real world has been limited.[3]

The main focus of this paper is address this issue by exploring a class of sequential games which arise naturally. We view our approach as complementary to the standard implementation literature. The implementation tradition considers a mediator who is uninformed about the preferences and other private information known to the agents, but who has complete control over the institutions of the mechanism. In particular, the mediator is typically free to choose the strategy sets and payoff functions of the agents. Such freedom has given rise to a large number of mechanisms, many of which

[2] Also see Jackson and Moulin (1992) for a recent example and a nice survey of the literature.

[3] There are exceptions, of course. The auction literature is probably the most important. See McAfee and McMillan (1987) for example. Also, see Balinski and Sonmez (1999) for a description of a mechanism that is used to allocate places in Turkish higher education.

have unnatural strategy spaces and extremely complex payoff rules.[4] Our approach, on the other hand, involves trading off having a minimally informed designer in favor of having one who has minimal control over economic institutions. The mediator might know the aggregate benefit of public project, for example, but may be forced to use the basic voluntary contribution mechanism and be unable to modify this mechanism to reward or punish agents for choosing particular strategies. Our main result is to show that if the such a highly constrained designer is allowed the ability to randomly delay the arrival of payoffs as his sole instrument, then he can still uniquely implement desirable social choice rules.

At a more formal level, we investigate a class of games which lie on the boundary between repeated games and games with cheap talk. Agents simultaneously choose strategies in our model. With some probability, δ, the strategies are payoff relevant and the game ends. With probability $1-\delta$, the strategies are payoff irrelevant, and each agent observes the strategy choices of all the other agents, which ex post are cheap talk. Play then moves to a new round, and the process is repeated ad infinitum. The game ends and payoffs arrive with certainty if we aggregate these probabilities over the entire time horizon; however, payoffs are actually received once and only once. From the standpoint of economic resources used, therefore, the game is one-shot. We call this a game with probabilistic cheap talk (PCT).[5]

The following example illustrates PCT. Suppose your department is considering hiring a new assistant professor. The chairman comes into your office and asks you to serve on the hiring committee. You must accept or refuse this request without knowing who else has, or will, agree to serve on the committee. Thus, commitments to serve are

---

[4] The concern for more "natural" mechanisms has given rise to a new branch of the implementation literature where the objective is to find mechanisms which correspond to realistic institutions; see Dutta, Sen and Vohra (1993), Saijo, Tatamitani and Yamato (1993), and Thomson (1993).

[5] Note that in games with cheap talk there are preliminary rounds in which play is fictitious with certainty followed by a payoff relevant bound. In our case, there is a fixed probability that play is fictitious in each round up until the game is actually played. This is why we call the structure we explore in the paper "probabilistic cheap talk". Note that if we were to set the probability of fictitious play close to one for a number or rounds, and then change it to zero, this model would look very much like a game with cheap talk. Nevertheless, we wish to emphasize that the motivation for this paper and the results we obtain are quite different from those of Crawford and Sobel (1982), Farrell (1982) and others.

simultaneous and secret. This is a classic prisoners' dilemma situation. Department members are asked to contribute to the common good, while their own interests are best served by not cooperating and free-riding off the efforts of other. The twist here is that it is not a certainty that the dean will approve funding for the new position. There is only a probability δ that you will actually end up spending December reading the CV's of the candidates. If the dean does not approve funding then the commitment to serve is ex post cheap talk. The game is repeated in subsequent years until (we fervently hope) the dean finally gives authorization to make an offer. Clearly, this can be mapped on to any other situation in which public goods are voluntarily provided. This example also illustrates that it is not unnatural for a human agent (in this case, the dean) to control the probability of cheap talk ending. It is not as clear that it would make as much sense for an agent to control a discount factor. Thus, the PCT framework leads itself much more easily to mechanism design than the repeated game framework despite their mathematical similarity.

The plan of this paper is as follows. In section 2, we describe a PCT game formally and discuss our equilibrium concept. In section 3, we show how to use a PCT game to implement the proportional cost sharing rule in a public good provision problem. In section 4, we discuss the relationship between PCT games, and repeated games and make concluding remarks.

2. The model

We begin by giving a formal definition of an abstract game with PCT. Let N be the set of agents. Let M^i be the set of *moves* available to agent $i \in N$.[6] Let $v^i : M^1 \times \ldots \times M^n \equiv M \to \Re^n$ be the payoff function for $i \in N$. Let $G \equiv < N, M, v >$ denote the corresponding one-shot game. We use m to denote $(m^i)_{i \in N}$ and m^{-i} to

[6] We consider in the interests of simplicity, we consider only pure strategies here. We discuss this restriction further after Lemma 1 and 2.

denote $(m^j)_{j\in N\setminus\{i\}}$.

Let T be the set of time periods $\{1,2,\ldots\}$, and define the random variable $X \in \{\mathrm{CT},\mathrm{DL}\}$, where CT stands for *cheap talk* and DL stands for *deadline*. We denote the PCT extension of a game G as $\Gamma(G,\delta)$ where for each $t\in T$, DL is realized with probability $\delta\in(0,1]$, and CT is realized with probability $1-\delta$. At the start of each period $t\in T$, agents simultaneously choose moves. If DL is realized, the payoffs are generated for the moves the agents chose. If CT is realized, then the payoffs are not distributed, the chosen moves are treated as publicly known cheap talk, and the process is repeated in period $t+1$.

Denote the number of rounds of ex post cheap talk by $t\in T$. The *history* of talk at t is denoted by h_t. Let $\mathcal{H}$ be the set of all possible histories over all t. Let $\mathcal{H}_t$ be the space of all possible histories at time t. We shall set $h_1=\emptyset$. A *strategy profile* for $i\in N$ is a set of mappings $s=\{s_t^i\}_{t=1}^{\infty}$ with $s_t^i:\mathcal{H}_t\to M^i$. Let S^i be the class of all possible strategy profiles for agent i and $S=\times_{i\in N}S^i$.

For any agent $i\in N$, participating in a strategy profile $s\in S$ yields the following expected payoff $x_t^i: S\times\mathcal{H}_t\times(0,1]\to R_+$ at time t, given history h_t:

$$x_t^i(s,h_t,\delta)\equiv\left\{\delta v^i(s_t(h_t))+\sum_{k=t+1}^{\infty}\delta(1-\delta)^{k-t}v^i(s_k(h_k))\right\}$$

where the histories after t are generated by equilibrium play of s given h_t.

A strategy profile $s\in S$ is *subgame perfect equilibrium (SPE) of the game* $\Gamma(G,\Delta)$ if

$$\forall\, t,\ \forall\, i\in N,\ \forall\, \bar{s}^i\in S^i,\ \text{and}\ \forall\, h_t\in\mathcal{H}_t,$$

$$x_t^i(s,h_t,\Delta)\geq x_t^i(s^1,\ldots,\bar{s}^i\ldots,s^n,h_t,\Delta).$$

In order to develop our equilibrium concept, we next develop definitions about stationarity. If for some $t'<t$ the agents have chosen $m\in M$ in round t', we shall say that the resulting history h_t *contains* m *at* t'. We shall write this as $m\in_{t'} h_t$.

DEFINITION: A history $h\in\mathcal{H}$ is *stationary* if

$$\exists\ \bar{m}\in M \text{ s.t. } h=(\bar{m},\bar{m},\ldots)$$

and $h_t \in H_t$, the restriction of h to the first t rounds, h_t, is also said to be stationary if

$$\exists\ \bar{m} \in M \text{ s.t. } h = (\bar{m}, \bar{m}, \ldots, \bar{m}).$$

DEFINITION: A strategy $s = \{s_t\}_{t=1}^{\infty}$ is said to be *partially stationary* if

$$\exists\ \bar{m} \in M \text{ s.t. } s(h_1) = \bar{m} \text{ and } h_t(\bar{m}, \bar{m}, \ldots, \bar{m}), \text{ then } s_t(h_t) = \bar{m}.$$

A partially stationary strategy thus generates a stationary history along the equilibrium path. Off the equilibrium path, however, the moves of agents can be arbitrarily complicated. In particular, the "punishment strategies" are in no way constrained to be stationary.

The equilibrium concept used in this paper is *Pareto-dominant partially stationary subgame perfect equilibrium*: the set of partially stationary SPE strategies of $\Gamma(G, \delta)$ whose payoffs Pareto dominate the payoffs of any other partially stationary SPE strategy. Restricting attention to Pareto dominant equilibrium payoffs is motivated by an appeal to axiomatic logic, but which we leave informal at this stage: if such an equilibrium exists all agents would agree that it is the most preferred outcome. The requirement that strategies be partially stationary is in the tradition of Rubinstein and Wolinsky's (1984) notion of *semi-stationarity* in a model of pairwise bargaining, and Green and Laffont's (1987) notion of *posterior implementable* equilibrium in a model of cheap talk. In both these cases, as in ours, agents receive no payoff until the last round of play. Thus, the past in any round of play involves only cheap talk, while the future, as viewed from any round, is identical in expectation. By forcing histories into just two paths, stationary and deviations from stationarity, in each round t the game is exactly identical to the game agents face in any other round t'. It can therefore be an equilibrium to respond with the same strategy choices.[7]

[7] Our equilibrium concept might seem restrictive. It is possible, however, to drop the Pareto dominance requirement at the cost of adding a "burning money" game on top of the existing structure, as in Ben-Porath and Dekel (1988), or Fudenberg and Tirole (1991), p.461.

We define $S_M^M \subset S$ to be the class of *trigger strategy profiles* in which all agents play a partially stationary strategy and respond to any deviation from the equilibrium move by going to a punishment move in all future rounds.

> DEFINITION: A strategy profile is a trigger strategy if $\exists\ \overline{m}, \tilde{m}$ such that $s^{\bar{m}}_{\tilde{m},t}(h_t) = \bar{m}$ if $t = 1$ or if $h_t = (\bar{m}, \bar{m}, \ldots, \bar{m})$, and $s^{\bar{m}}_{\tilde{m},t}(h_t) = \tilde{m}$ otherwise.

Our convention is to have the superscripted move be the one played along the equilibrium path, and the subscripted move be the punishment move. In this paper, we are particularly concerned with games that have a prisoners' dilemma flavor. Such games typically have a dominant strategy equilibrium which has the property that when all players save one use this strategy, the payoff to the remaining player is at a minimum. More precisely, we focus on punishment states in which all players punish a deviating player conditional on his deviating optimally. We capture this idea formally in the following definition.

> DEFINITION: Let $\underline{v}^i \equiv \max_{m^i} \min_{m^{-i}} v^i(m_i, m^{-i})$. A move $\tilde{m}$ is said to be a *minimizing dominant strategy equilibrium* (MDSE) of a game G if it is a dominant strategy equilibrium and
>
> $$v^i(\tilde{m}) = \underline{v}^i$$
>
> for each i.

Thus, $\tilde{m}^{-i}$ imposes the lowest payoff possible on agent i conditional on agent i's move.[8]

Lemma 1. *Suppose that $\tilde{m}$ is a MDSE of the game G, and let s be an arbitrary partially stationary strategy SPE of $\Gamma(G, \delta)$ with an equilibrium message of $\bar{m}$. Then the trigger strategy $s^{\bar{m}}_{\tilde{m}}$ is also an SPE of $\Gamma(G, \delta)$.*

Proof/

[8] We thank an anonymous referee for suggesting this way of defining MDSE.

Begin by noting that the stationary history of $\bar{m}$ each period is generated by both s and $s^{\bar{m}}_{\tilde{m}}$ in equilibrium. Since the aggregate probability that the game ends over the whole time horizon is one, playing $\bar{m}$ each round guarantees each agent $i \in N$ the same expected payoff of $v^i(\bar{m})$.

Now suppose agent i considers defecting from this stationary equilibrium pay. Under $s^{\bar{m}}_{\tilde{m}}$, such a deviation is punished by $\tilde{m}^{-i}$ which by construction gives the agent $\underline{v}^i$ in each period subsequent to this defection. It follows that the best he can do in the period he defects is to maximize his one period payoff given that the remaining agents play $\tilde{m}$. Formally the best he can to is to get the expected payoff:

$$\delta v^i(m^{i*}, \bar{m}^{-i})$$

where

$$m^{i*} \in \{m^i \in M^i \mid v^i(m^i, \bar{m}^{-i}) \geq v^i(\hat{m}^i, \bar{m}^{-i}) \; \forall \; \hat{m}^i \in M^i\}.$$

Now consider the optimal deviation for agent i under strategy s. Note that by sending the same optimal one-round defection message defined above, he can guarantee himself expected payoff of $\delta v^i(m^{i*}, \bar{m}^{-i})$ in the period he defects. Recall, however, that $\tilde{m}$ is an MDSE which is played under $s^{\bar{m}}_{\tilde{m}}$ in all subsequent rounds and that this results in at most the payoff $\underline{v}^i$ to agent i for any message he might send. It follows that optimal play in subsequent rounds under strategy s must give agent i at least as much expected payoff each period as he would get when he plays optimally under strategy $s^{\bar{m}}_{\tilde{m}}$.

We conclude that since the optimal defection yields a weakly smaller payoff to agent i under $s^{\bar{m}}_{\tilde{m}}$ than s, if s is a Nash equilibrium, then $s^{\bar{m}}_{\tilde{m}}$ must also be one.

It only remains to show that $s^{\bar{m}}_{\tilde{m}}$ is subgame perfect. We have already shown agents are following best responses in equilibrium. Since by definition, $\tilde{m}$ is a dominant strategy equilibrium, it is immediate that for any history which leads to agents playing this strategy, agents must be playing a best response as well.

∎

Lemma 1 says that if a payoff can be achieved by a partially stationary SPE strategy, then it can also be achieved by a trigger strategy which punishes nonstationary play by reverting to a stage-game MDSE if one exists. Thus, when such a one-shot MDSE exists, we can restrict attention to this class of trigger strategies without loss of generality. We also remark that we have considered only pure strategies in the interest of simplifying notation. Provided that agents can observe the actual probabilities or distributions that agents use in their mixed strategies, nothing in the argument above depends on this restriction. In fact, in Chakravorti, and Conley and Taub (1996) we explicitly consider mixed strategies.

3. An Application: Proportional Allocation of Costs of a Public Good

A standard institution for funding a discrete public-good project is to ask for voluntary contributions. Agents, whose valuations of the public good are hidden and heterogeneous, are asked to report the benefit they would receive, and the cost is shared out in proportion to these reports. The outcome is severe under-reporting of benefits and widespread free riding.

This institution can be described as a one-shot game. The game has a dominant strategy outcome in which every agents' pledge is zero, and the good is not provided. The PCT extension of this one-shot game overcomes this incentive: it uniquely implements the proportional sharing rule, and since it induces agents to report their valuations truthfully, the costs are shared in accord with the agents' actual valuations. The designer, who chooses the termination probability, need only know the sum of the benefits over all agents in order to find the appropriate δ. Neither the individual components nor the true distribution of this aggregate across agents need be observable to him.

The formal description of the public goods game, $G^{pg} \equiv < N, M, v >$, is as follows. Let B^i be the private benefit that agent i receives if the public project is built, and C

be the cost of the project. We assume that for all $i \in N$, $C > B^i > 0$, so that no individual would choose to build the project by himself. For all $i \in N$, $M^i \equiv \Re_+$. An agent's message is interpreted as a report of the benefits he receives from the project. The payoff function for all $i \in N$ is:

$$v^i(m) = \begin{cases} B^i - C\frac{m^i}{\sum_{j=1}^n m^j} & \text{if } \exists\, j \in N \text{ s.t. } m^j > 0 \\ 0 & \text{if } \forall\, j \in N, m^j = 0. \end{cases}$$

and so the information of each agent, as well as that of the designer, includes the total benefits $\sum_{j=1}^n m^j$, obtained through anonymous demand surveys of samples of the population, statistical extrapolation, expert consultations, or through aggregate signals.

Let $\tilde{m} = (0, \ldots, 0)$. If the move $\tilde{m}$ is made, the project is not built, and all the agents pay zero. Since $C > B_i$ for all i, it is immediate that $\tilde{m}$, free riding, is an MDSE of the one-shot game G^{pg}, with $\underline{v}^i = 0$. We start with the following lemma.

Lemma 2. *Suppose $C \leq \sum_{j=1}^n B^j$. Then for $\delta^* = 1 - \frac{C}{\sum_{j=1}^n B^j}$, $s_{\tilde{m}}^{\bar{m}}$ is a partially stationary SPE of $\Gamma(G^{pg}, \delta^*)$, if and only if there is $k \geq 0$ such that $\bar{m} = (kB^1, \ldots, kB^n)$.*

Proof/

First suppose $k = 0$. Then by Lemma 1 $\bar{m} = \tilde{m} = (0, \ldots, 0)$. Clearly, $s_{\tilde{m}}^{\tilde{m}}$ is an SPE for any δ since $\tilde{m}$ is the only Nash equilibrium of G^{pg}.

Now suppose that $k > 0$. To see that $s_{\tilde{m}}^{\bar{m}}$ is an SPE consider any $i \in N$ and any $h \in \mathcal{H}$. This history could have evolved in one of two ways.

Suppose first that the game has not ended at any given t, that h_t is stationary, and that the other agents have been playing kB^{-i} each round. If i makes the optimal defection from $m^i = kB^i$, which is $m^i = 0$, then his expected payoff is $\delta^* B^i$. This is because he gets the benefit of free riding in round t if the game ends. If the game does not end at t then all agents report zero in the next round and all future rounds, resulting in a payoff of zero whenever the game happens to end. On the other hand, not defecting yields a payoff of $B^i - \frac{B^i C}{\sum_{j=1}^n B^j}$. But by construction,

$$\delta^* B^i = \left(1 - \frac{C}{\sum_{j=1}^n B^j}\right) B^i = B^i - \frac{B^i C}{\sum_{j=1}^n B^j}.$$

Thus, it is a best response for i to play kB^i since no additional benefit is gained by defecting.

Suppose on the other hand that at some t, some agent j were to play $\hat{m}^j \neq kB^j$. Since $\tilde{m}$ is a Nash equilibrium of G, and all agents besides i play $\tilde{m}^{-i}$ in all subsequent rounds, it is a best response for agent i to play $\tilde{m}^i$ in all subsequent rounds given this history.

Finally, suppose there is another trigger strategy $s_{\tilde{m}}^{\hat{m}}$, which is an SPE of $\Gamma(G, \delta)$ such that

$$\nexists k > 0 \ \text{ s.t. } \hat{m} = kB.$$

But then:

$$\exists \; i \in N, \text{ s.t. } \frac{\hat{m}^i}{\sum_{j=1}^n \hat{m}^j} > \frac{B^i}{\sum_{j=1}^n B^j}.$$

Thus:

$$\delta^* B^i = \left(1 - \frac{C}{\sum_{j=1}^n B^j}\right) B^i > B^i - \frac{\hat{m}^i C}{\sum_{j=1}^n \hat{m}^j},$$

and defecting has a higher expected payoff than abiding by the trigger strategy. This contradicts the hypothesis that $s_{\tilde{m}}^{\hat{m}}$ is an SPE of $\Gamma(G, \delta)$, and therefore no such SPE trigger strategy can exist.

■

Lemma 2 establishes that the only SPE trigger strategies of $\Gamma(G, \delta)$ have all agents either reporting a profile of messages that are proportional to the true benefit profile or reporting zero. Thus, if any agent over or under reports his true benefit, it must be by the same percentage as the over or under reports of all of the other agents in equilibrium. Theorem 1 shows that even though there is an infinity of SPE, there are only two partially stationary SPE payoffs. Note also that allowing for mixed strategies would not materially change this result. This is because the logic of the proof is that at δ^*, agents are just balanced between defecting and cooperating. Thus, for a mixed strategy to be supportable as an equilibrium, it would have to give each agent exactly the same expected payoff as the pure strategies defined above.

Theorem 1. *Suppose $C \leq \sum_{j=1}^n B^j$. If $\delta^* = 1 - \frac{C}{\sum_{j=1}^n B^j}$, then the only partially stationary SPE payoffs of $\Gamma(G^{pg}, \delta^*)$ are*

$$\left\{(B^1 - \frac{CB^1}{\sum_{j=1}^n B^j}, \ldots, B^n - \frac{CB^n}{\sum_{j=1}^n B^j}), (0, \ldots, 0)\right\}$$

Proof/

By Lemma 2, $s_{\tilde{m}}^{\bar{m}}$ is an SPE trigger strategy if and only if for some $k \geq 0$, $\bar{m} = (kB^1, \ldots, kB^n)$. Since $\tilde{m} = (0, \ldots, 0)$ is an MDSE, by Lemma 1 the payoffs associated with these trigger strategies are the only partially stationary SPE payoffs. Since for all $k > 0$, the expected payoffs are:

$$\left(B^1 - \frac{CB^1}{\sum_{j=1}^n B^j}, \ldots, B^n - \frac{CB^n}{\sum_{j=1}^n B^j}\right),$$

and for $k = 0$, the expected payoffs are:

$$(0, \ldots 0),$$

these are the only partially stationary SPE payoffs.

■

Finally, we show that regardless of δ, if the sum of the benefits is less than the cost, building the project is never an equilibrium. This is important if the designer makes a mistake in estimating the total benefit. It means that bad projects will never be built regardless of these errors.

Theorem 2. *Suppose $C > \sum_{j=1}^n B^j$. Then for all $\delta > 0$, $s_{\tilde{m}}^{\tilde{m}}$ is the only SPE trigger strategy of $\Gamma(G^{pg}, \delta)$.*

Proof/

Clearly, $s_{\tilde{m}}^{\tilde{m}}$ is an SPE since $\tilde{m}$ is the only Nash equilibrium of G.

To see that there can be no other equilibrium note the following. Since $C > \sum_{j=1}^n B^j$, for all $\hat{m} \neq \tilde{m}$, it must be that for some agent $i \in N$

$$\frac{\hat{m}^i}{\sum_{j=1}^n \hat{m}^j} > \frac{B^i}{\sum_{j=1}^n B^j},$$

since either the reports are in proportion to benefits, or at least one agent's report is more than proportionate to his benefit. Also, for all $i \in N$,

$$C \frac{B^i}{\sum_{j=1}^n B^j} > B^i$$

Then for all $\delta > 0$,

$$\delta B^i > 0 > B^i - \frac{B^i C}{\sum_{j=1}^n B^j} > B^i - \frac{m^i C}{\sum_{j=1}^n m^j}.$$

Thus, for all $\delta > 0$, defecting yields positive expected payoff, while abiding by any trigger strategy other than the one enforcing $\tilde{m}$ yields a negative expected value.

■

Of the two partially stationary SPE payoffs,

$$\left(B^1 - \frac{CB^1}{\sum_{j=1}^n B^j}, \ldots, B^n - \frac{CB^n}{\sum_{j=1}^n B^j} \right)$$

strongly Pareto dominates $(0, \ldots 0)$. Thus, if the designer knows the sum of the benefits, he can choose δ such that it uniquely implements the sharing of costs in proportion to the benefits, in Pareto dominant partially stationary SPE.

4. Discussion and Conclusions.

Remark 1. The proportional-sharing scheme for the division of costs is what is recommenced by the Lindahlian tradition of the benefit theory of taxation. The division of benefits can also be given an axiomatic justification. It turns out that the equilibrium payoffs are exactly those suggested by the bargaining solution of by Kalai and Smorodinsky (1975). We construct the bargain problem as follows: If the project is worth building, them the total social surplus is $\sum_i B^i - C$. This surplus can be linearly transferred by altering the contributions each agent makes on a one for one basis. Agents can also refuse to contribute which provides then a payoff of zero. Therefore the disagreement point is $(0, \dots, 0)$ and the feasible set is given by simplex: $\sum_i u^i = \sum_i B^i - C$. The best a player can expect to do is contribute nothing and free ride off everybody else's contributions. Thus, the ideal point is B. All that remains is to note that the intersection of the chord between the 0 and B gives every agents a payoff of

$$(\sum_i B^i - C)\frac{B^i}{\sum_i B^i} = B^i - C\frac{B^i}{\sum_i B^i},$$

which agrees with this Lindahlian rule. Since this is payoff is the unique Pareto-dominant partially stationary SPE equilibrium of the PCT game we describe, we conclude that PCT *implements* the Kalai-Smorodinsky bargaining solution to this cost-sharing problem in this of noncooperative equilibrium concept.

Remark 2. There is a close relationship between PCT games and repeated games with discounting. Recall the expected payoff from a strategy for a PCT game. For any agent $i \in N$, participating in a strategy profile $s \in S$ yields the following expected payoff $x_t^i : S \times \mathcal{H}_t \times (0, 1] \to R_+$ at time t, given history h_t:

$$x_t^i(s, h_t, \delta) \equiv \left\{ \delta v^i(s_t(h_t)) + \sum_{k=t+1}^{\infty} \delta(1-\delta)^{k-t} v^i(s_k(h_k)) \right\}$$

where the histories after t are generated by equilibrium play of s given h_t. For given δ, this game is identical in structure to a standard repeated game, because the probabilistic payoff $\delta v^i(s_t(h_t))$ can just as easily be interpreted as the periodic payoffs of a repeated game, with $1 - \delta$ as the discount factor. It is immediate that we may invoke a standard folk theorem, which we adapt from Fudenberg and Tirole (1991). Previous folk theorem results have required the existence of enough resources to play the game more than once, and the payoffs that are feasible in any given period are exactly the same regardless of the discount factor. As a result, these existing theorems are not applicable to situations in which institutional constraints prevent the actual repetition of a game. When a PCT game is interpreted from the perspective of the folk theorem, on the other hand, the periodic payoff is $\delta v^i(s_t(h_t))$, which depends on δ; when agents play stationary strategies, the larger "discounting" effect of increasing δ is exactly offset by an increase in the implicit periodic payoffs.[9]Thus, the expected value of playing stationary strategies is invariant with respect to δ.

To summarize, in Chakravorti, Conley and Taub (1996) we explored probabilistic cheap talk extensions of the one-shot prisoner's dilemma game, focusing especially on how the value of the cheap talk probability affected the shape of the set of subgame perfect equilibria. We extended this work here in two directions. First, we showed how applying PCT to a standard example a of voluntary contribution public goods provision game allow us to implement the Lindahlian cost sharing rule in an extremely simple and natural way. The mechanism designer need only control the probability that payoffs are delayed, but otherwise accepts the institution of the voluntary contribution game as given. Many other examples are possible, and we treat some of these, mainly from public finance, in Chakravorti, Conley and Taub (1998). In that paper we show that it is possible to implement the equal sharing of costs rule in a variation of the voluntary contribution game described above, and also to implement equal sharing of monopoly profits in the standard Bertrand oligopoly game.

Second, we discuss how the folk theorem applies to PCT extensions of a broad class of games beyond prisoner's dilemmas. We also pointed out that this implies that PCT games provide a different interpretation of the infinitely repeated game with discounting.

Because of the equivalence of PCT games with repeated games with discounting and the corollary folk theorem, if we eschew the partial stationarity restriction on the set of strategies then the unique implementation results achieved here will then no longer hold. Stahl (1991), and van Damme (1992) demonstrate that for the infinitely repeated prisoners' dilemma there is no value of the discount factor for which the set of Pareto efficient equilibria is a singleton, and without the imposition of stationarity

[9] More concretely, in recursive form the expected payoff is $x^i_t(s, h_t, \delta) = \delta v^i(s_t(h_t)) + (1-\delta)x^i_{t+1}(s_{t+1}, h_{t+1}, \delta)$. Changes in δ alter the weighting of current versus future payoffs, but in such a way that the expected payoff is always identical to the set of one-shot payoffs.

PCT necessarily inherits this non-uniqueness.

It is possible to allow agents to use any strategies (instead of only the partially stationary ones), and still achieve the partially stationary outcome if we allow the designer slightly more power. Specifically, the designer must be able to observe when any agent changes his strategy choice from the previous round (although the strategy itself need not be observed) and to condition the termination probability on this observation. We explore this approach more thoroughly in our 1996 paper.

Because this construction works, we conjecture that partially stationary PCT is empirically realistic. Consider a closed-door labor-management negotiation in which each side has received instructions from their respective principals. A mediator could set a random termination probability. The mediator cannot typically observe the strategy of each negotiator, because it is never fully revealed. However, the mediator might threaten to impose the deadline immediately if either of the negotiators leaves the room, thereby enabling themselves to receive new instructions—in essence, deviating from their stationary strategies. Such deviation from stationarity would initiate a punishment sequence, namely a reversion to a one-shot equilibrium and its consequent inefficient outcome. By the correct choice of the termination probability along with this strategy of punishing deviations from stationarity in this fashion, a unique cooperative outcome can be implemented. As with our public-goods model, the correct termination probability is determined by the payoff structure of the underlying one-shot game. If players have a high degree of bargaining power, something we can quantify in the same way we quantified eagerness in the public goods game, the termination probability is small. This translates into the expectation of protracted play, as one expects in actual negotiations between "difficult" bargainers.

The agreement of the parties to this structure of negotiations is thus an a priori agreement to implement a unique bargaining solution. Public-good projects and multi-party private investment projects are, correspondingly, often the product of closed-door negotiation of this sort. We entertain the hope that a more rigorous empirical examination of PCT in such situations will move bargaining and mechanism design from the realm of theory to the realm of description.

References

Balinski, M., and T. Sonmez (1999): "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory,* v. 84 pp. 73-94 .

Ben-Porath, E., and E. Dekel (1988): "Coordination and the Potential for Self-Sacrifice," working paper.

Chakravorti, B., J. Conley, and B. Taub (1998): "Economic Applications of Probabilistic Cheap Talk," *Topics in Mathematical Economics and Game Theory,* M. Wooders, editor, Fields Institute Communications, Toronto, forthcoming.

Chakravorti, B., J. Conley, and B. Taub (1996): "On Uniquely Implementing Cooperation in the One-Shot Prisoners' Dilemma," *Economic Theory,* v. 8, pp. 347-66.

Chakravorti, B., J. Conley, and B. Taub (1997): "On Uniquely Implementing Cooperation in the Prisoner's Dilemma: Corrigendum," *Economic Theory,* v. 9 pp. 377-8.

Crawford, V. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica,* v. 50, pp. 1431-1450.

Dutta, B., A. Sen, and R. Vohra (1993): "Nash Implementation through Elementary Mechanisms in Economic Environments," manuscript, ISI Delhi.

Farrell, J (1982): "Communication in Games," manuscript, MIT.

Fudenberg, D. and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting and Incomplete Information," *Econometrica,* v. 54, pp. 533-554.

Fudenberg, D. and J. Tirole (1991): *Game Theory.* MIT Press.

Green, J. and J. J. Laffont (1987): "Posterior Implementability in a Two-Person Decision Problem," *Econometrica,* v. 55, pp. 69-94.

Groves, T. and J. Ledyard (1977): "Optimal Allocation of Public Goods: A Solution to the 'Free Rider' Problem," *Econometrica,* Vol. 45, 783-809.

Guth, W., R. Schmittberger, and B. Schwarze (1982): "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization,* v. 3 pp. 367-88.

Jackson, M., and H. Moulin (1992): "Implementing a Public Project and Distributing its Cost," *Journal of Economic Theory,* Vol. 57, pp 125-140.

Kalai, E. and M. Smorodinsky (1975): "Other Solutions to Nash's Bargaining Problem," *Econometrica,* pp. 513-8.

McAfee, P., and J. McMillan (1987): "Auctions and Bidding," *Journal of Economic Literature,* v. 25 pp. 699-738.

Nash, J. (1950): "The Bargaining Problem," *Econometrica,* v. 18, pp. 155-62.

Nash, J. (1951): "Noncooperative Games," *Annals of Mathematics,* v. 54, pp. 289-95.

Nash, J. (1953): "Two-Person Cooperative Games," *Econometrica,* v. 21, pp. 128-40.

Rubinstein, A. and A. Wolinsky (1985): "Equilibrium in a Market with Sequential Bargaining," *Econometrica,* v. 53, pp. 1133-50.

Saijo, T., Y. Tatamitani, and T. Yamato (1993): "Toward Natural Implementation," manuscript, University of Tsukuba.

Stahl, D. (1991): "The Graph of the Prisoner's Dilemma Supergame Payoffs as a Function of the Discount Factor," *Games and Economic Behavior,* v. 3 pp. 368-84.

Thomson, W. (1993): "'Divide and Permute' and the Implementation of Solutions to the Problem of Fair Division," manuscript, University of Rochester.

van Damme, E. (1991): *Stability and Perfection of Nash Equilibria.* Berlin New York, Springer-Verlag.

Vickrey, W. (1961): "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance,* v. 16 pp. 8-37.