ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic
Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 80, No. 1 (January, 2012), 363–411

DEFINABLE AND CONTRACTIBLE CONTRACTS

MICHAEL PETERS
University of British Columbia, Vancouver, BC V6T 1Z1, Canada

BALÁZS SZENTES
London School of Economics, London, WC2A 2AE, United Kingdom

Econometrica, Vol. 80, No. 1 (January, 2012), 363–411

DEFINABLE AND CONTRACTIBLE CONTRACTS

BY MICHAEL PETERS AND BALÁZS SZENTES[1]

This paper analyzes Bayesian normal form games in which players write contracts that condition their actions on the contracts of other players. These contracts are required to be representable in a formal language. This is accomplished by constructing contracts which are definable functions of the Godel code of every other player's contract. We provide a complete characterization of the set of allocations supportable as pure-strategy Bayesian equilibria of this contracting game. When information is complete, this characterization provides a folk theorem. In general, the set of supportable allocations is smaller than the set supportable by a centralized mechanism designer.

KEYWORDS: Definability, contract theory, folk theorem.

1. SELF-REFERENTIAL STRATEGIES AND RECIPROCITY IN STATIC GAMES

IN THIS PAPER, we characterize the allocation rules attainable by players in a Bayesian game when they have the ability to commit themselves by writing contracts that condition on other players' contracts.

The idea that contracts might condition on other contracts is not new in economics. The most commonly known expression of this idea is well known in the industrial organization literature (e.g., Salop (1986)) as the "meet the competition" clause in which one firm commits itself to lower its price when any of its competitors does. A similar idea appears in trade theory as the principle of *reciprocity* (Bagwell and Staiger (2001)). Countries enact trade legislation in which they agree to abide by a trade agreement like GATT. Such legislation commits the country to lowering tariffs in response to trade legislation by another country that lowers tariffs, provided this other country's legislation agrees to abide by GATT. Finally, tax treaties sometimes have this flavor, for example, out of state residents who work in Pennsylvania are exempt from Pennsylvania tax as long as they live in a state that has a reciprocal agreement that exempts out of state residents (presumably from Pennsylvania) from state taxes.

In all of these examples, commitments are made that are conditional on commitments of others, and are used to support cooperative outcomes. The literature treats these situations as static games of complete information. Additionally, the contracts that are used to support equilibrium are idiosyncratic, so that only the simplest most stylized problems are amenable to analysis. For example, in the meet the competition argument, firm A offers to sell at a high price provided its competitor, firm B, also sets a high price. If instead, B offers any price below the highest price, A commits itself to sell at its marginal cost. If B believes this commitment, then one best reply for B is to set the highest price. In the trade and taxation treaties mentioned above, a state cooperates

[1] We would like to thank the editor and three referees for many useful suggestions that we ended up incorporating into the paper.

DOI: 10.3982/ECTA8375

by offering a reciprocal contract that cooperates if and only if the other state does the same.

Tennenholtz (2004) suggested a way to model mutually dependent commitment devices. His players compete using computer programs that condition their actions on other programs. He showed that all individually rational outcomes in complete information games can be supported as equilibria.[2] Basically, two programs implement cooperative actions if they know that they have the same syntax; otherwise, they punish each other. As he was only concerned with showing that individually rational outcomes can be supported as program equilibrium, he did not give a complete description of what the set of possible programs looks like. Kalai, Kalai, Lehrer, and Samet (2010), whose method we illustrate below, specifically constructed a set of commitment devices for a two player game of complete information with the property that outcome functions are supportable as equilibria in these commitment devices if and only if they are individually rational.

This paper considers a two-stage contracting game in a Bayesian environment. At the first stage, players offer contracts. A contract restricts the action spaces of a player as a function of the other contracts. At the second stage, players take actions from their restricted action spaces. Our objective in this paper is twofold. First, we identify two properties of any abstract contract space that lead to a complete characterization of supportable outcomes. We refer to these properties as cross-referentiality and invariant punishment. *Cross-referentiality* is a generalization of the Tennenholz idea that contracts can recognize each other. *Invariant punishment* means that if a player wants to, he can write contracts that commit him to any of his pure actions, while inducing the same reaction from the other players each time. It is this property that allows us to show that players' payoffs cannot be held below their individually rational level in games of complete information.[3] Second, we show that there exists a contract space which satisfy these properties. We assume that the set of feasible contracts is a set that consists of finite texts written in some language. Conditional on the language, this description of contracts is natural and has the advantage that the set of contracts is universal in the sense that the same set of feasible contracts can be used to model competition in every environment. This set is at once rich, descriptive, and must, within limits set by the language in which contracts are written, be robust to the introduction of new contracts. We explain how these contracts can be written so that they specify actions that are conditional on the contracts of other players.

[2] By "individually rational outcome" in a game of complete information, we mean an outcome in which each player receives at least his minmax payoff.

[3] More specifically, if players can react to the outcomes of a deviation by appropriately crafting their contract, then they could conceivably respond to the action that a deviating player takes. This would make it possible to keep some players' payoffs at their maxmin level, instead of just their minmax.

We use our methods to show how to extend the static complete information folk theorem results of papers like Tennenholtz (2004) or Kalai et al. (2010) to games of incomplete information. We provide a complete characterization of outcomes that can be supported as (pure-strategy) equilibria in finite contracting games of incomplete information. Although there is no agreed definition in the literature of what a folk theorem would say in games with incomplete information, our results do indicate that contracts are more restrictive in such games. In the complete information case, it is known that outcomes in which every player receives more than his or her minmax payoff can be supported as an equilibrium in contracts with the appropriate choice of commitment devices. One way to think of such outcomes is that they are the outcomes that could be supported by a centralized mechanism designer, who can enforce actions of all players who agree to participate in his mechanism and who minmaxes players who unilaterally decide not to participate. In the incomplete information case, we show that there are outcome functions that a centralized mechanism designer like this can support, but which cannot be supported with contracts.

The reason that a folk theorem like result does not hold with incomplete information has to do with "participation." A player who does not play along with some cooperative agreement enforced by contracts will still observe any information that the contracts themselves convey about the types of the other players. In the play of the ensuing game, the deviator can make use of this information when choosing his actions. To capture this, we show that outcome functions are supportable if and only if they are supportable by a mechanism designer who can condition actions only on publicly observable messages.

1.1. *Contractible Contracts*

When trying to describe a broad set of feasible contracts, it is easy to get lost in complexities associated with the infinite regress that arises when a contract specifies a commitment that depends on whether another contract specifies a commitment that depends on whether the first contract specifies.... One way to get around this infinite regress is simply to impose ad hoc restrictions on the set of feasible contracts to ensure that the infinite regression does not arise. This is the approach developed in Kalai et al. (2010). We explain this approach briefly to motivate the broad approach that we adopt.

We can apply the argument in Kalai et al. (2010) to a simple two player prisoner's dilemma. We want to construct a set of commitment devices that will support cooperation. Define a contract called *This Contract* which works as follows

$$\text{This Contract} = \begin{cases} C, & \text{if other player's contract} = \text{This Contract}, \\ D, & \text{otherwise.} \end{cases}$$

If both players in the prisoner's dilemma offer This Contract, then they are unambiguously obliged to cooperate. If one of the players offers something else, then the other is unambiguously required to defect. As long as the deviator's contract specifies some unambiguous outcome against This Contract, cooperation is a Nash equilibrium. Depending on what other contracts are described by the word "otherwise" in the sentence above, there may be many other possible equilibrium outcomes as well.

To complete the description of the contracting game, let Θ be a set of feasible contracts defined in such a way that $\theta \in \Theta \Longrightarrow \theta : \Theta \to A$. We are about to judiciously construct Θ. We have just described a reciprocal contract θ^* such that

$$\theta^*(\theta') = \begin{cases} C, & \text{if } \theta' = \theta^*, \\ D, & \text{otherwise.} \end{cases}$$

Suppose we simply add to this a pair of "constant" contracts $\theta_c \in \Theta$ such that $\theta_c(\theta') \equiv C$ and a contract $\theta_d \in \Theta$ such that $\theta_d(\theta') \equiv d$. The constant contracts could be used to support the Nash equilibrium in the prisoner's dilemma in the obvious way. Since the outcomes (C, C) and (D, D) are the only two outcomes for which all players receive at least their minmax payoffs, the collection consisting of the contract θ^* and the two constant contracts already supports all outcomes for which each player receives at least her minmax payoff. Furthermore, if $\Theta = \{\theta^*, \theta_c, \theta_d\}$, it is pretty clear that no other (pure) outcomes can be supported as equilibrium in contracts. In other words, now that we have the set Θ, we have a complete characterization of the set of equilibrium outcomes. This characterization amounts to a folk theorem.

One desirable feature of contracts that this approach lacks is that they be robust to contractual innovation. Absent such a property, one is never sure whether economic properties that emerge from a contractual model are not just artifacts of the way the contracts are modelled. Ideally, if we want commitment devices that can condition on the devices used by others, it would be desirable that the set of feasible contracts or commitment devices would include *all* functions from itself into the set of feasible actions. This would ensure that any new contract we could dream up would already be a feasible contract. Unfortunately, this is impossible because the cardinality of the set of functions with a given domain is larger than the cardinality of that domain (Cantor's theorem). Our approach is instead to describe the largest set of contracts that can be written in a finite set of characters using first-order logic.

This is not simply a theoretical issue. Observe that in the formulation above, the set of contracts made available seems much too restrictive. Contracts depend on other contracts, but in a very limited way. If a player offers the contract that he is "supposed" to offer, then things go well; otherwise, something bad happens. Yet this bad outcome cannot depend in any way on what the deviator actually does. In other words, the punishment imposed on a deviator only depends on the deviator's identity and not on the actual deviation. This is not at all a natural property of contracts that depend on other contracts. In Kalai et al. (2010) and Tennenholtz (2004), this is true by assumption.

Relaxing this constraint on contracts creates a difficulty when attempting to provide a full characterization of supportable outcomes, because natural bounds on outcomes like the minmax payoff themselves rely on this property. When players minmax a deviating player, their actions do not vary with the action the deviator takes. If contracts allow reactions to depend on deviations, then, in principle, it might be possible to support outcomes in which some players' payoffs are below their minmax payoffs. In the complete information environment, there is an easy fix to this problem, which is to allow players to offer contracts which do not specify a single action, but rather specify a subset of their action spaces. With such a contract space, a deviator can always offer a contract which does not restrict his action space at all and then he can best respond to the action profile of the others. Then the worst possible punishment players can impose on a deviator is indeed an action profile which minmaxes him. Therefore, restricting attention to punishments which are independent of the actual deviation is without loss of generality. Unfortunately, this same approach fails in games with incomplete information, because players who punish a deviator might not restrict their action spaces to a single action precisely because they might not want to reveal too much information to a deviator. Therefore, following a deviation, even the nondeviators will be forced to choose their actions strategically. Hence, a deviator might actually benefit from restricting his action space.

Furthermore, the set of feasible contracts described above is specially tailored to the economic problem to which it is applied. The commitment devices we just described obviously will not be much use in a game with more than two players or more than two actions. Indeed, if we simply relabel the actions so that C stands for defection and D stands for cooperation, then these commitment devices will support only the noncooperative outcome. We can create new sets of commitment devices to handle these changes without much problem. What we would like to do instead is to provide a set of commitment devices that always works.

Perhaps more important, the contracts are obviously chosen because we know we want to support a specific pair of actions. If we perturb the payoffs in a way that changes the essential economics of the problem (for example, by making the socially desirable outcome (C, D)), then we have to reconstruct the set of contracts to get the result we want. We provide a set of contracts that can be described independently of the (payoffs in the) game to which they are applied. So there are two ways in which our contracts are universal. First, any particular set of contracts, like the contracts we described above, can be rewritten in our language so that they are embedded in the larger set of contracts we describe. Second, the set of contracts we describe can be used to model contractual competition no matter what are the actual payoffs in the game.

Perhaps the main contribution of our approach is to show that even when contracts are universal in the sense that we have described, we can still understand the economic logic of contractual situations using a modified version of the "minmax a deviator" logic that is described above.

1.2. *How Definability Works*

We accomplish this universality by allowing players to write finite texts in a first-order arithmetic language that describes their commitments. This allows players to punish deviators in potentially very complex ways and involves no ad hoc restrictions like those above. The way we do this is to observe that the texts associated with contracts can be reinterpreted as definable functions of the Gödel codes of texts written by other players. Since Gödel codes and definable functions are unfamiliar to most economists, we give an informal description of the method below.

We start by endowing each player with a formal language and require each contract to be a text written in this language. A text is a finite string of symbols. It is well known that there are bijections from the set of texts into the set of integers. One such mapping is called Gödel coding. This implies that any contract uniquely corresponds to an integer. A contract of a player is a mapping from contract profiles to subsets of his action space. Since the contracts correspond to integers, one can think of such a contract as a description of an arithmetic correspondence from the codes of contracts to the codes of the names of the actions. There is a well known set of arithmetic correspondences, called the *definable correspondences*, which can be precisely described in formal language by using finitely many characters. (We formally define this set later.) Hence, one can think of the contract space as the set of definable functions from $\mathbb{N}^m \to 2^{\mathbb{N}}$, where m is the number of players. The domain of these functions is the vectors of the codes of the players' contracts and the range of these functions is the subsets of the codes of the names of the actions. We identify the contract space of a player with the set of definable correspondences.

To see how our approach works, return to the simple prisoner's dilemma game. Let $[c]$ denote the Gödel code of the contract c and refer to $[c]$ as the encoding of c. For any pair of contracts c_1 and c_2, the action (C or D) taken by player 1 is $c_1([c_2])$ and similarly for player 2. Since every pair of actions determines a payoff, this procedure associates a unique payoff with every pair of contracts.

There are many things that are not definable contracts that also have Gödel codes. We want to make use of some of these other things. In particular, we want to use definable functions with *free variables.* Interpreting n as the encoding of the other player's contract, here is a definable contract with a free variable:

$$\gamma_x(n) = \begin{cases} C, & n = x, \\ D, & \text{otherwise.} \end{cases}$$

A free variable has the natural interpretation that x can take on any integer value. Definable contracts with free variables also have Gödel codes. The contract with free variable that we want is a slight modification of the one above, in particular,

$$c_x(n) = \begin{cases} C, & n = \left[\langle x\rangle^{(x)}\right], \\ D, & \text{otherwise}, \end{cases} \tag{1}$$

is also a contract with a free variable. The mapping $\langle x\rangle^{(x)}$ is the composition of two functions. First, the function $\langle x\rangle$ is the inverse operation to the Gödel coding; that is, $\langle n\rangle$ is the text whose Gödel code is n. Second, if ϕ is a text with one free variable, then $\phi^{(n)}$ is the same text where the value of the free variable is set to be n. Hence, if n is a Gödel code of a definable contract with one free variable, then $\langle n\rangle^{(n)}$ is itself a definable contract (without a free variable). $[\langle n\rangle^{(n)}]$ is just the Gödel code of whatever this definable contract happens to be.

We want to create a contract by fixing a very specific value for x in (1). In particular, the value of x we are interested in is $[c_x]$. Since $[c_x]$ is the Gödel code of a contract with a free variable, the right-hand side of (1) requires that we decode $[c_x]$ to get c_x, then fix x at $[c_x]$ to get the contract $c_{[c_x]}$. Putting all this together gives

$$c_{[c_x]}(n) = \begin{cases} C, & n = \left[c_{[c_x]}\right], \\ D, & \text{otherwise}, \end{cases}$$

so

$$c_{[c_x]}([c_2]) = \begin{cases} C, & [c_2] = \left[c_{[c_x]}\right], \\ D, & \text{otherwise}. \end{cases}$$

This is the contract which corresponds to the one we called This Contract or θ^* in our discussion above. The difference is that This Contract now reacts to a much broader set of contracts than to what θ^* did. The contract θ^* could only respond to itself and to the two constant contracts. The contract $c_{[c_x]}$ responds to any definable function (in fact, it specifies an action for every finite text).

To press the analogy with θ^* in the problem above, if player 2 also uses strategy $c_{[c_x]}$, then $[c_2] = [c_{[c_x]}]$, which evidently triggers the cooperative action by player 1. The same argument applies for player 2. Player 2 can deviate to any alternative definable strategy c' that she likes. Since every definable strategy has a Gödel code, the reaction of player 1 and, consequently, both players' payoffs are well defined. As the Gödel coding is injective, $c' \neq c_{[c_x]}$ implies the Gödel code of c' is not equal to $[c_{[c_x]}]$, and the deviation by player 2 induces player 1 to respond by switching from C to D.

What this argument illustrates is that our contract space is large enough that we can always find a contract that corresponds to θ^* in our existing set of contracts, without having to construct it explicitly from the details of the game. This is the property that makes our contracts universal in the sense that exactly the same set of contracts can be used to characterize equilibrium outcomes in all games.

The introduction of Gödel coding into our model requires some explanation. Take it as given that contracts must be expressed in a formal language. Then a contract of a player must give *precise instructions* on how to restrict his action space as a function of the texts submitted by the other players. To describe the contracting game, one must carefully define the notion of precise instructions and the set of those texts which give these instructions. Any such definition would lead to a definition of a set of arithmetic correspondences which can be described as finite texts. To see this, suppose that there is a text which gives instructions on how to pin down a subset of a player's action space as a function of the other texts. Then there is also a text that gives the same instructions as a function of the Gödel codes of texts of other players instead of their texts. This is because the Gödel coding and its inverse are definable functions, that is, they can be described as texts.[4] This implies that this new text describes an arithmetic correspondence. In the paper, we adopt the definition of definable functions from number theory instead of introducing a new definition.

Since the set of definable functions is the largest set of arithmetic functions which can be described in a first-order language, our contract space is the largest given the restriction to contracts which can be expressed as texts. Implicitly, our approach makes it possible for players to offer any finite text as a contract. We simply identify the original text with the corresponding definable mapping.

In the prisoner's dilemma example above, we assume that a contract of player 1 is a definable function, say $c(x)$ and that player 1 can take action a ($\in \{C, D\}$) if and only if the code of a is in the set $c([c_2])$, where c_2 is the code of player 2's contract. If we allow players to write any text as contracts, then they could write down the following text:

> My contract can be described by the following definable function: $c(x)$, where the interpretation of $c(x)$ is the following. If the code of the text of the contract of player 2 is x, then I can only take action a if the Gödel code of a is in $c(x)$. Finally, the Gödel code of a text is defined as follows: ....

Since the Gödel coding is a recursive function, the ellipses (...) can be replaced by a precise description of this coding. All we do is to identify the text above with $c(x)$.

2. LITERATURE

As we mentioned in the Introduction, our paper is not the first to show how contractual devices can be used to support cooperative play. Much of the lit-

[4] In particular, this implies that players do not need to agree to use the Godel code of other contracts. They can use the Godel code unilaterally, and the implications of the contract will be understood by the others provided they agree on the underlying language in which contracts are written.

erature in this area follows an idea developed by Fershtman and Judd (1987) in which actions are delegated to an agent who is given the appropriate incentives to carry out actions that might not otherwise be part of a noncooperative equilibrium. This idea was developed by Katz (2006), who used it to prove a folk theorem for a very specialized environment.

The idea that agents might report deviations provides the basis for the menu theorems in common agency, like Martimort and Stole (1998), Peters (2001), Han (2006), and Martimort and Moreira (2010). Recently Yamashita (2010) suggested a method that can be used to extend the common agency approach to games in which there are many agents.[5]

The idea that principals could learn about deviations by communicating with agents is developed in Epstein and Peters (1999). They showed that for every environment, there is a universal set of mechanisms for that environment such that any set of indirect mechanisms used to model competition between principals can be embedded in that universal set. Each element of this universal set is described by a sequence of payoffs. Since each indirect mechanism that a principal can offer corresponds to some sequence of payoffs, agents can report the mechanisms being used by other principals by reporting the sequences that correspond to each of the other principals' mechanisms. In this sense, the agent's type corresponds with his usual payoff type along with a sequence representing the mechanisms of the other principals. Every equilibrium in a competing mechanism game can then be represented as an equilibrium in a game where principals offer universal mechanisms and agents truthfully report both their payoff type and their market information.

There are a couple of important differences between our paper and theirs. First of all, our formulation makes it possible to provide a characterization of the outcome functions that are supportable as equilibria. This is a major advantage over Epstein and Peters (1999), which only provides a set of contracts that might be used to support equilibrium. Second, players offer contracts that condition directly on contracts of other players instead of asking agents to describe these contracts. In our model, there is no communication at all between players after contracts have been announced, so there are really no agents at all. Indeed, we illustrate that despite this limit on players' ability to communicate, contracts support a rich set of type contingent outcome functions. Our main theorem shows that contracts are equivalent to a mechanism in which players communicate their type information publicly. In the private value case, contracts support all the outcome functions that can be supported by a centralized mechanism designer, illustrating the flexibility of this approach.

However, the primary objective of Epstein and Peters (1999) and our paper is the same—to find a language that makes it possible to describe contracts that depend on other contracts. The essential conceptual difference is that Epstein

[5] See Peters and Troncoso-Valverde (2009) for a full characterization of supportable outcomes using his method.

and Peters (1999) created a language to describe contracts that uses sequences of payoffs from the game to describe contracts. In this sense, the approach in Epstein and Peters (1999) resembles the approach in Kalai et al. (2010) that we described above, except that the contracts in Epstein and Peters (1999) are immune to contractual innovation. So the contracts that work in one game will not work in a new game with a different payoff structure. As we mentioned, the contracts that we describe work for all games. The cost of our universality is that the set of contracts we describe is countable for the finite environments we are interested in. The universal set of contracts in Epstein and Peters (1999) is finite in finite environments because the set of payoffs is finite.

Our approach is closely related to ideas in the computer science literature. One paper we have already mentioned that uses this approach is Tennenholtz (2004). He has players writing programs that determine their actions. Using an idea due to von Neumann, he allows these programs to use other programs as data, which has the effect of making the output of each player's program depend on the other players' programs. We illustrated the idea with our description of This Contract in the introduction. Tennenholz does not give a complete description of the set of feasible contracts. We explain below how our approach differs from the assumption that players choose Turing machines to play against one another.

The paper by Kalai et al. (2010) gives a complete characterization of equilibrium outcomes in two-player complete information games. They constructed a set of (game specific) commitment devices which can be used to support *correlated* strategies in which all players' payoffs exceed their minmax payoffs. Specifically, in some games their devices support outcomes in which all players receive payoffs that exceed their best payoffs with Tennenholz's programs. This is accomplished by constructing commitment devices that allow players to correlate their actions while using independent randomizing devices. We extend part of their argument to games of incomplete information. We do not deal with correlation or any other form of randomization simply because we are trying to keep a notationally and technically demanding problem relatively simple.

Finally, the problem we model is one in which privately informed players offer contracts which will depend on their types in most Bayesian equilibria of the contracting game. Since our interest is in contracts rather than mechanism design, we do not allow players to communicate privately after agreeing to the contract us is done in an informed principal problem Myerson (1983). Despite this, we are able to show that contract equilibria support a rich set of type contingent outcomes. Indeed, this is a major advantage of our approach over the simple "cooperate or be minmaxed" approach in the computer science literature, since contracts have to respond to other players' contracts in a much more sophisticated way to make it possible for one player's action to depend on another player's type without any explicit communication. We could not have shown this as effectively if we had allowed private communication.

Of course, the results we have about the limits of contract equilibrium arise from this restriction. We could support a larger class of outcome functions if we allowed private communication.

3. THE MODEL

There are m players indexed by $i \in \{1, \dots, m\}$. Player i has a finite action space A_i, and let A denote $\bigtimes_{i=1}^{m} A_i$. Player i has a type t_i drawn from a finite set T_i, and let T denote $\bigtimes_{i=1}^{m} T_i$. Each type profile has a strictly positive probability and the joint distribution of types is common knowledge. The payoff of player i is $u_i : A \times T \to \mathbb{R}$.

Each player i is endowed with a contract space C_i and let C denote $\bigtimes_{i=1}^{m} C_i$. Each element of C_i defines a mapping from C to $2^{A_i} \setminus \{\emptyset\}$. Let $\widetilde{c}_i$ denote the mapping induced by c_i. That is, a contract of player i specifies a nonempty subset of his action space for each contract profile. It is important to note that we do not identify a contract with the mapping it defines. This allows for the possibility that two different contracts induce the same mapping, that is, $\widetilde{c}_i(c) = \widetilde{c}'_i(c)$ for all $c \in C$, but $c_i \neq c'_i$.

The contracting game takes place in two stages. In the first stage, each player submits a contract from his contract space simultaneously. Let $c_i \in C_i$ denote the contract submitted by player i. For each player i, consider the following subset of A_i determined by the contract profile $c = (c_1, \dots, c_n)$:

$$S_i(c) = \{a_i : a_i \in \widetilde{c}_i(c)\}.$$

In the second stage, player i takes action from $S_i(c)$ simultaneously. As always, $S(c) = \bigtimes_i S_i(c)$.

In what follows we restrict attention to pure strategies. A strategy, of player i consists of a mapping from his type space to his contract space and a mapping from his types and first-stage contract profiles to his action space. Let Γ_i denote the first stage strategies of player i, that is,

$$\Gamma_i = \{\gamma_i : \gamma_i \in C_i^{T_i}\},$$

where $C_i^{T_i}$ denotes the set of functions with domain T_i and range C_i. Similarly let $\mathcal{A}_i$ denote the set of second-stage strategies of player i, that is,

$$\mathcal{A}_i = \{\alpha_i : \alpha_i \in A_i^{T_i \times C} \text{ and } \forall t_i \in T_i, \forall c \in C, \alpha_i(t_i, c) \in S_i(c)\}.$$

Let $\gamma(t)$ denote $(\gamma_1(t_1), \dots, \gamma_m(t_m))$ and let $\alpha(t, c) = (\alpha_1(t_1, c), \dots, \alpha_m(t_m, c))$ for all $t \in T$ and $c \in C$. The strategy profile $(\gamma^*, \alpha^*) \in (\bigtimes_i \Gamma_i) \times (\bigtimes_i \mathcal{A}_i)$ constitutes a Bayesian equilibrium if and only if for all $i \in \{1, \dots, m\}$, $t_i \in T_i$, $\gamma_i \in \Gamma_i$, and $\alpha_i \in \mathcal{A}_i$,

$$(2) \qquad \begin{aligned} &E_{t_{-i}}\big(u_i\big(\alpha^*((t_i, t_{-i}), \gamma^*(t_i, t_{-i})), (t_i, t_{-i})\big) \mid t_i\big) \\ &\quad \geq E_{t_{-i}}\big(u_i\big(\bar{\alpha}((t_i, t_{-i}), \bar{\gamma}(t_i, t_{-i})), (t_i, t_{-i})\big) \mid t_i\big), \end{aligned}$$

where $\bar{\alpha} = (\alpha_i, \alpha^*_{-i})$ and $\bar{\gamma} = (\gamma_i, \gamma^*_{-i})$.

A deterministic outcome function in our model is a mapping from T to A. We say that an outcome function $s : T \to A$ is supportable as a Bayesian equilibrium in the contracting game if there is a Bayesian equilibrium (γ^*, α^*) such that $\alpha^*(t, \gamma^*(t)) = s(t)$ for all t.

Bayesian equilibrium imposes no restriction at all on the second-stage actions $\alpha^*_{-i}(t_{-i}, (\gamma_i, \gamma^*_{-i}(t_{-i})))$ that player i anticipates when he deviates, apart from $\alpha^*_{-i}(t_{-i}, (\gamma_i, \gamma^*_{-i}(t_{-i}))) \in S_{-i}(\gamma_i, \gamma^*_{-i}(t_{-i}))$. For example, $\alpha^*_j(t_j, (\gamma_i, \gamma^*_{-i}(t_{-i})))$ could be strictly dominated for player j with type t_j by some other action in $S_j(\gamma_i, \gamma^*_{-i}(t_{-i}))$. For this reason, it may be that refinements of equilibrium are necessary in applications to rule out this kind of second-stage behavior.

Refinements are completely incidental to our formalism, although they are obviously going to make a difference to the precise set of outcome functions that are supportable as equilibria. The less controversial refinements, like sequential equilibrium and perfect Bayesian equilibrium do not fit easily into our environment since our contract spaces are not necessarily finite and player types can be correlated. The detailed formalism we need to modify these concepts takes us well beyond our main purpose in this paper. For these reasons, we restrict attention to equilibria where players do not take strictly dominated actions in the continuation games generated by a contract profile. In the Appendix, we provide a more abstract description that describes refinements in a manner that is independent of the particular game that is being played (see Section A.1).

For each $t_{-i} \in T_{-i}$ and $\bar{A} = \times_{i=1}^m \bar{A}_i \subset A$, define $\mathcal{R}_i(\bar{A}, t_{-i})$ as the set of action profiles $a_{-i} \in A_{-i}$ such that for each $j \neq i$, a_j is not strictly dominated[6] for player j when his type is t_j, given that the players are constrained to choose actions in $\bar{A}$. We say that (γ^*, α^*) is an $\mathcal{R}$ equilibrium of the contracting game if (2) holds, and, in addition, for every i, $\gamma_i \in C_i$, and $t_{-i} \in T_{-i}$,

$$\alpha^*_{-i}(t_{-i}, \bar{\gamma}(t_{-i})) \in \mathcal{R}_i\big(S(\bar{\gamma}(t_{-i})), t_{-i}\big), \tag{3}$$

where $\bar{\gamma}(t_{-i}) = (\gamma_i, \gamma^*_{-i}(t_{-i}))$.

3.1. *Properties of the Contract Space*

Our main characterization theorem relies on two properties, which we describe formally in this section. We have already mentioned these properties: cross-referentiality, and invariant punishment. Cross-referentiality is intended

[6] Here, strictly dominated means "no matter what he believes about the types of his opponents." This is intended to be a very weak refinement. In particular, here we ignore any information j might have learned from the contracting phase when evaluating whether or not a particular action is strictly dominated.

to generalize the term This Contract that we used to support cooperation in the prisoner's dilemma game in the introduction. The complication is that, generally, a contract has to respond to many players who have many different types. For each such player and type, a contract needs to recognize the particular contract for that player and the type which reciprocally recognizes the player's own contract.

CROSS-REFERENTIAL PROPERTY: For all $(N_1, \dots, N_m) \in \mathbb{N}^m$, $N_i \geq 1$, for all functions $r_i : \mathbb{N}^m \to 2^{A_i} \setminus \{\emptyset\}$, $p_i^j : \mathbb{N}^{m-1} \to 2^{A_i} \setminus \{\emptyset\}$ $(i, j \in \{1, \dots, m\}, i \neq j)$, there exists a set of contracts for all $i \in \{1, \dots, m\}$, $\{c_i^{n_i}\}_{n_i \in \{1, \dots, N_i\}} \subset C_i$, such that $c_i^{n_i} \neq c_i^{n_i'}$ if $n_i \neq n_i'$ and

$$\widetilde{c}_i^{n_i}(c_1, \dots, c_m) = \begin{cases} r_i(n_i, n_{-i}), & \text{if } \forall k \neq i \; c_k^{n_k} = c_k, \\ p_i^j(n_i, n_{-ij}), & \text{if } \forall k \neq i, j c_k^{n_k} = c_k \text{ and } \nexists n_j \text{ s.t. } c_j^{n_j} = c_j, \\ A_i, & \text{otherwise.} \end{cases} \tag{4}$$

Notice that the conditions on the right-hand side of (4) explicitly depend on (cross-referential contracts) $c_j^{n_j}$ and, vice versa, the mapping determined by $c_j^{n_j}$ explicitly depends on $c_i^{n_i}$. If each player j offers a contract $c_j^{n_j}$ for some $n_j \leq N_j$, then $c_i^{n_i}$ restricts the action space of player i to be $r_i(n)$, where $n = (n_1, \dots, n_m)$. If each player offers such a contract except player j, then $c_i^{n_i}$ restricts the action space of player i to be $p_i^j(n_{-j})$. In any other cases, $c_i^{n_i}$ imposes no restrictions on the actions of player i. Intuitively, $\{c_i^{n_i}\}_{n_i}$ can be thought of as the set of contracts from which player i is supposed to choose in an equilibrium. If each player chooses from these sets, then the second-stage restrictions are defined by the functions $\{r_i\}_i$. If a single player, say player j, offers a contract which is not in $\{c_j^{n_j}\}_{n_j}$, then the contracts of the other players restrict their action spaces according to the functions $\{p_i^j\}_i$. One can think about these functions as the *contractual punishments* imposed on deviators.

In the construction given above, the reaction of each player when one of the others fails to use one of the cross-referential contracts is independent of what contract this player actually offers. This is similar to the approach we used in the prisoner's dilemma problem. The next property is used to ensure that the logic associated with fixed punishments like this will still be valid in richer contract spaces.

This property requires, that no matter what contracts the other players are offering, the contract space is rich enough that the remaining player is able to write contracts that will commit him to any subset of his actions that he wants while inducing the same commitments from his opponents in response to each different subset. Of course, any sensible contract space will let a player commit himself to an arbitrary subset of his actions. Yet the player must make these commitments by announcing different contracts. The other players' contracts make commitments that depend on the contract this player offers, so the other players' commitments will generally change as the remaining player's commitment changes. We want the contract space to be rich enough that the player can write the contracts in such a way that the different commitments he makes induce exactly the same commitments from his opponents.

We do not need to know what these contracts are or what commitments they elicit from the other players. All we require is that these contracts exist so that there is *some* fixed commitment a player can elicit from the others. What this property delivers is the fact that in any equilibrium, a player must do at least as well as he does against this fixed commitment. This ensures that all equilibrium outcomes can be implemented with the cross referential contracts we described above. Formally

Invariant Punishment Property: For all $(N_1, \dots, N_m) \in \mathbb{N}^m$, $N_i \geq 1$, for all sets of contracts $\{c_j^{n_j}\}_{j, n_j \in \{1,\dots,N_j\}}$ $(c_j^{n_j} \in C_j$, $c_j^{n'_j} \neq c_j^{n_j}$ if $n'_j \neq n_j)$, and for every i, there are functions $p_k^i : \mathbb{N}^{m-1} \to 2^{A_k} \setminus \{\emptyset\}$ $(k \neq i)$, such that for any function $f_i : \mathbb{N}^{m-1} \to 2^{A_i} \setminus \{\emptyset\}$, there is a contract $c_i^* \in C_i$ such that for all $n_{-i} \in \bigtimes_{j \neq i} \{1, \dots, N_j\}$,

$$\tilde{c}_i^*(c_i^*, (c_j^{n_j})_{j \neq i}) = f_i(n_{-i}) \tag{5}$$

and for all $k \neq i$,

$$\tilde{c}_k^{n_k}(c_i^*, (c_j^{n_j})_{j \neq i}) = p_k^i(n_{-i}). \tag{6}$$

Again, the set $\{c_i^{n_i}\}_{n_i}$ can be thought of as the collection of those contracts from which player i chooses, depending on his type in an equilibrium. Given the strategies of the others, each alternative contract that he offers induces a commitment correspondence $f_i(n_{-i})$ and elicits some kind of response by the others. The Invariant Punishment Property guarantees that there must exist some collection of punishment correspondences, $\{p_j^i\}_{j \neq i}$, such that for *any* commitment correspondence f_i that player i wants, he can write his own contract in such a way that the others respond with exactly the same punishment $\{p_j^i\}_{j \neq i}$.

In Section 5, we specify a contract space which satisfies both of these properties. This space is going to be the set of definable functions. There are other spaces which contain cross-referential objects. One such space is the set of Turing machines which is often used in game theoretic analysis. One can even think about the programs in Tenneholz (2004) as Turing machines: he used this space to model contracts in a complete information environment. If we modelled the contracts by Turing machines, then the input of a player's machine would be the descriptions of the machines submitted by the other players, and the output would be a subset of the player's action space.

This space would satisfy the Cross-Referential Property, but not the Invariant Punishment Property. This is because players could submit *universal machines* that would simulate the machine of a deviator. Once the simulation is completed, these machines could recommend an action profile which is worse for the deviator. This action profile can depend on the result of the simulation, that is, on the actual deviation. This suggests that players can push the payoff of a deviator below his minmax value. Of course, a deviator could also submit a universal machine that simulates the machines of the others, and then best responds to their outputs. The problem is that, in general, these universal machines will not halt on each other. Indeed, it is not clear how one can properly define the game because of this halting problem.

4. THE CHARACTERIZATION THEOREM

There are several ways to state our characterization theorem. One of our objectives is to compare the set of equilibrium outcomes of the contracting game to the set of outcome functions implementable by a centralized mechanism designer who can control the actions of all the players who agree to participate in his mechanism. To help illustrate the relationship, we define a class of mechanisms called *public message mechanisms* (PMM) and show that the set of equilibrium outcomes relative to these mechanisms is identical to the set of equilibrium outcomes in the contracting game.

The following class of two-stage mechanisms are called public message mechanisms. In the first stage, players simultaneously decide whether or not to participate in the mechanism. Players who participate send public messages from a countable message space. At the same time, a player who does not participate publicly submits a commitment device, which imposes a restriction on his action space as a function of the messages sent by the participants.[7] In the second stage, the mechanism restricts the action spaces of the participating players as a function of the messages of the participants. These restrictions, however, cannot depend on the functions submitted by the nonparticipants. Finally, players simultaneously take actions from their restricted action spaces.

Next, we define the public message mechanism formally.

DEFINITION 1: Let N_i be a countable message space for each i and let $N = \bigtimes_{i=1}^{m} N_i$. Suppose that $\emptyset \notin \bigcup_{i=1}^{m} N_i$ and let $\bar{N}_i = N_i \cup \{\emptyset\}$ for each i.[8] In addition, let

$$\varrho_i = \left\{\rho_i : \rho_i \in (2^{A_i} \setminus \{\emptyset\})^{\bigtimes_{i=1}^{m} \bar{N}_i}\right\}.$$

[7] A restriction is a nonempty subset of the action space.

[8] The symbol $\emptyset$ can be interpreted as the message of a player who does not participate in the mechanism.

For each $\rho^* = (\rho_1^*, \ldots, \rho_m^*)$, $\rho_i^* \in \varrho_i$, consider the following two-stage game. In the first stage, players take actions simultaneously. The first-stage action space of player i is $N_i \cup \varrho_i$. In the second-stage, after observing the first-stage action profile, players take actions simultaneously. Suppose that the first-stage action of player j is δ_j. Let $n_j = \delta_j$ if $\delta_j \in N_j$ and $\emptyset$ otherwise. Then the second-stage action space of player i is $\rho_i^*(n_1, \ldots, n_m)$ if $\delta_i \in N_i$ and $\rho_i(n_1, \ldots, n_m)$ if $\delta_i = \rho_i \in \varrho_i$. We call this game the public message mechanism defined by (N, ρ^*).

The main difference between a PMM and a standard direct mechanism is that the reports of the players are publicly observable. This has several consequences. First, a nonparticipant player can learn about the types of the participants through their messages and can make his action contingent on these messages. As a result, nonparticipation is more profitable in a PMM than it would be in a comparable mechanism when messages are privately conveyed to the mechanism designer. Second, to prevent players from refusing to participate so that they can make use of this information, the mechanism might not induce participants to fully reveal their types in the first stage. Finally, since the messages do not necessarily coincide with the types, the mechanism designer might not want to specify a single action for every player in response to some public messages. If, instead, he allows the player to choose from a subset of his actions, then he makes it possible for the player's action to depend on his private information which was not revealed through his message. Furthermore, he can exploit a nonparticipant's uncertainty about the types of others when implementing a punishment.

Intuitively, the relationship between a PMM and the contracting game can be explained as follows. In a contracting equilibrium, a player with different types offers different contracts. Since the contracts are publicly observable, players learn about each other's types from the contracts. The contracts' information content about the types corresponds to the public messages in a PMM. An equilibrium contract profile specifies restrictions on the action spaces of the players. These restrictions correspond to the second-stage restrictions of a PMM if each player participates. The Invariant Punishment Property corresponds to the property of a PMM which says that the restrictions imposed on participants do not depend on the commitment devices submitted by the nonparticipants. The reader should think about a nonparticipant in a PMM as a deviator in the contracting game and think of the commitment device of a nonparticipant as the deviator's contract. The Invariant Punishment Property implies that it is without loss of generality to assume that uncooperative behavior by one player in the contracting game provokes a punishing contractual response from the others that does not depend on how the deviator goes about being uncooperative.

We restrict attention to deterministic mechanisms and pure strategies. Our main theorem can be stated as follows.

THEOREM 1: *An outcome function is implementable as an $\mathcal{R}$ equilibrium in the contracting game if and only if it is implementable as an $\mathcal{R}$ equilibrium by a public message mechanism.*

A *simple public message mechanism* is a public message mechanism in which each player's message space is a partition of his type space. The mechanism is incentive compatible if each player prefers to report the partition element that contains his true type. It is individually rational if every player, irrespective of his type, would prefer to participate in the mechanism than unilaterally commit himself to a subset of his actions that depends on the partition elements reported by the other players. By standard arguments in mechanism design, an outcome function can be supported as an $\mathcal{R}$ equilibrium in a public message mechanism if and only if there is an incentive compatible and individually rational simple public message mechanism that implements the same outcome function as an $\mathcal{R}$ equilibrium. For this reason, we restrict attention to simple public message mechanisms which are incentive compatible and individually rational.

Next we characterize the set of implementable outcome functions with constraints. Let $\tau_i : T_i \to 2^{T_i} \setminus \{\emptyset\}$ be the partition from which player i must choose his report. Let τ, τ_{-i}, and τ_{-ij} denote $\bigtimes_{i=1}^{n} \tau_i$, $\bigtimes_{j\neq i} \tau_j$, and $\bigtimes_{k\neq i,j} \tau_k$, respectively. Let $r_i(t) \in 2^{A_i} \setminus \{\emptyset\}$ denote the restricted action space of player i if each player participates and the message sent by player j is $\tau_j(t_j)$. Since the restrictions can depend only on the partition elements that each player reports, r_i must be measurable with respect to τ, that is, $r_i(t) = r_i(t')$ whenever $\tau(t) = \tau(t')$. Furthermore, let $p_i^j(t_{-j}) \in 2^{A_i} \setminus \{\emptyset\}$ denote the restriction on the action space of player i if all players but player j participate and the message sent by player q is $\tau_q(t_q)$ for all $q \neq j$. The function $p_i^j(t_{-j})$ is measurable with respect to τ_{-ij}. A simple public message mechanism is given by $(\tau, r, p) = (\{\tau_i\}_{i=1}^m, \{r_i\}_{i=1}^m, \{p_i^j\}_{i,j=1}^m)$.

A public message mechanism only constrains players to subsets of their action spaces, so we need to describe what happens at the second stage. We start with the equilibrium path. Let s_i denote the strategy of player i at the second stage if each player participates; that is, $s_i : T \to A_i$, such that $s_i(t) \in r_i(t)$ for all t, and s_i is measurable with respect to τ_{-i}. Note that since player i knows his own type, s_i does not have to be measurable with respect to τ_i. Next, we describe the strategies of the players following a deviation. Let

$$F_i^\tau = \{f_i : f_i \in (2^{A_i} \setminus \{\emptyset\})^{T_{-i}}, f_i \text{ is } \tau_{-i} \text{ measurable}\}.$$

The set F_i^τ is the action space of player i in the first stage if he does not participate in the mechanism. If player i submits f_i ($\in F_i^\tau$) and player j reports $\tau_j(t_j)$ for $j \neq i$, then player i's restricted action space is $f_i(t_{-i})$ ($\subset A_i$) in the second stage. Let s_i^j denote the second-stage strategy of player i if all players but player j participate; that is, $s_i^j : T_{-j} \times F_j^\tau \to A_i$ such that $s_i^j(t_{-j}, f_j) \in p_i^j(t_{-j})$, and s_i^j is measurable with respect to τ_{-ij}.

An outcome function $s = (s_1, \dots, s_m)$ is supportable as an equilibrium in the public message game (or alternatively, is implementable by a simple public message mechanism) if there is a simple public message mechanism (τ, r, p) such that the following inequalities hold. The first one guarantees that each player sends a truthful message in the first stage of the game. For each $i = 1, \dots, m$ and for all $t_i, t_i' \in T_i$

$$\begin{aligned}(7)\qquad & E_{t_{-i}}\big(u_i(s(t), t)|t_i\big) \\ & \geq E_{t_{-i}}\Big(\max_{a_i \in r_i(t_i', t_{-i})} E_{t_{-i}'}\big(u_i(a_i, s_{-i}(t_i', t_{-i}'), (t_i, t_{-i}'))|t_{-i}' \in \tau_{-i}(t_{-i})\big)|t_i\Big).\end{aligned}$$

The max operator on the right-hand side implies that the player has to choose a best reply from his restricted action space given his posterior belief. Taken together, these constraints for all the players require that the play in the second stage constitutes a Bayesian equilibrium of the game in which each player chooses an action from his restricted set of actions, given posterior beliefs about other players' types.

To deal with deviations at the first stage of the PMM, we require that, for each $t_i \in T_i$,

$$\begin{aligned}(8)\qquad & E_{t_{-i}}\big(u_i(s(t), t)|t_i\big) \\ & \geq \max_{f_i \in F_i^\tau} E_{t_{-i}}\Big(\max_{a_i \in f_i(t_{-i})} E_{t_{-i}'}\big(u_i(a_i, s_{-i}^i(t_{-i}', f_i), (t_i, t_{-i}'), t)| \\ & \quad t_{-i}' \in \tau_{-i}(t_{-i})\big)|t_i\Big).\end{aligned}$$

This inequality says that even if the deviator chooses a best reply from the set of actions to which he is restricted, he cannot gain by deviating.

We say that an outcome function is supported as an $\mathcal{R}$ equilibrium of the public message mechanisms (alternatively, is $\mathcal{R}$-implementable by a simple public message mechanism) if (7) and (8) hold, and, in addition, for every i and $f_i \in F_i^\tau$,

$$(9)\qquad s_{-i}^i(t_{-i}, f_i) \in \mathcal{R}_i(f_i \times p_{-i}^i, t_{-i}).$$

With this formalism, we can restate our theorem as follows:

COROLLARY 1: *The outcome function s is implementable as an $\mathcal{R}$ equilibrium in the contracting game if and only if there is a simple public message mechanism (τ, r, p) for which* (7), (8), *and* (9) *hold.*

4.1. *The Proof of Theorem 1*

PROOF OF THE "IF" PART OF THEOREM 1: Suppose that there exists a PMM which $\mathcal{R}$-implements the outcome function $s = (s_1, \dots, s_m) : T \to A$. According to the arguments in the previous section, this implies that for all i and j, there exists a partition of the type space $\tau_i : T_i \to 2^{T_i} \setminus \{\emptyset\}$, on-path restriction $r_i : T \to 2^{A_i} \setminus \{\emptyset\}$, off-path restrictions $p_i^j : T_{-j} \to 2^{A_i} \setminus \{\emptyset\}$, and off-path strategies $s_i^j : T_{-j} \times F_j^\tau \to A_i$ such that s_i and r_i are measurable with respect to τ_{-i}, p_i^j and s_i^j are measurable with respect to τ_{-ij}, $s_i(t) \in r_i(t)$, $s_i^j(t_{-j}, f_j) \in p_i^j(t_{-j})$, and (7)–(9) are satisfied. In what follows, we construct an $\mathcal{R}$ equilibrium in the contracting game which implements the outcome function s.

For each i, let N_i denote the number of elements in the partition generated by τ_i. Then there exists a τ_i-measurable surjection $\sigma_i : T_i \to \{1, \dots, N_i\}$. Define $\bar{r}_i : \mathbb{N}^m \to 2^{A_i} \setminus \{\emptyset\}$ such that $\bar{r}_i(\sigma_1(t_1), \dots, \sigma_m(t_m)) = r_i(t_1, \dots, t_m)$ and define $\bar{p}_i^j : \mathbb{N}^{m-1} \to 2^{A_i} \setminus \{\emptyset\}$ such that $\bar{p}_i^j((\sigma_k(t_k))_{k\neq j}) = p_i^j((t_k)_{k\neq j})$. This is possible because σ_i and r_i are τ_i-measurable and p_i^j is τ_{-ij}-measurable. The Cross-Referential Property guarantees that there exists a set of contracts for all i, $\{c_i^{n_i}\}_{n_i\in\{1,\dots,N_i\}} \subset C_i$, such that $c_i^{n_i} \neq c_i^{n_i'}$ for $n_i \neq n_i'$ and

$$\widetilde{c}_i^{n_i}(c_1, \dots, c_m) = \begin{cases} \bar{r}_i(n_i, n_{-i}), & \text{if } \forall k \neq i \; c_k^{n_k} = c_k, \\ \bar{p}_i^j(n_{-j}), & \text{if } \forall k \neq i, j \; c_k^{n_k} = c_k \text{ and } \nexists n_j \text{ s.t. } c_j^{n_j} = c_j, \\ A_i, & \text{otherwise.} \end{cases}$$

Now define $c_i^{t_i}$ to be $c_i^{\sigma_i(t_i)}$. The first-stage strategy of player i, $\gamma_i : T_i \to C_i$, is given by $\gamma_i(t_i) = c_i^{t_i}$. Since σ_i is τ_i-measurable, γ_i is also τ_i-measurable. Using the definitions of $\bar{r}_i$ and $\bar{p}_i^j$, $\widetilde{c}_i^{t_i}$ can be written as

$$\text{(10)} \qquad \widetilde{c}_i^{t_i}(c_1, \dots, c_m) = \begin{cases} r_i(t_i, t_{-i}), & \text{if } \forall k \neq i \; c_k^{t_k} = c_k, \\ p_i^j(t_{-j}), & \text{if } \forall k \neq i, j \; c_k^{t_k} = c_k \text{ and } \nexists t_j \text{ s.t. } c_j^{t_j} = c_j, \\ A_i, & \text{otherwise.} \end{cases}$$

It remains to specify the second-stage strategy of player i, $\alpha_i : T_i \times C \to A_i$ for each i. If, for all j, there is a $t_j \in T_j$ such that player j offered a contract $\gamma_j(t_j) = c_j^{t_j}$, then $\alpha_i(t_i, \gamma(t)) = s_i(t)$. This strategy is well defined because $s_i(t_i, t_{-i}) = s_i(t_i, t'_{-i})$ whenever $\alpha_i(t_i, \gamma(t_i, t_{-i})) = \alpha_i(t_i, \gamma(t_i, t'_{-i}))$.[9] Suppose now that one player deviated, say player j, and offered a contract c_j, and player k offered $c_k^{t_k}$ for some $t_k \in T_k$ for all $k \neq j$. Define $f_j^{c_j} : T_{-j} \to 2^{A_j}$ as

$$\text{(11)} \qquad f_j^{c_j}(t_{-j}) = \widetilde{c}_j(c_j, (c_k^{t_k})_{k\neq j}) = \widetilde{c}_j(c_j, \gamma_{-j}(t_{-j})).$$

Notice that $f_j^{c_j} \in F_j^\tau$ because γ_{-j} is τ_{-j}-measurable. Define $\alpha_i(t_i, (c_j, (c_k^{t_k})_{k\neq j}))$ to be $s_i^j(t_{-j}, f_j^{c_j})$. These strategies are well defined because s_i and γ_{-i} are τ_{-i}-

[9] This follows from σ_j being a surjection for each j.

measurable, s_i^j is τ_{-ij}-measurable, and σ_j is a surjection. In addition, they are consistent with the restrictions imposed by the contracts defined by (10), that is, $s_i(t) \in r_i(t)$ and $s_i^j(t_{-j}, f_j) \in p_i^j(t_{-j})$.

Next, we argue that the strategies described above constitute an $\mathcal{R}$ equilibrium in the contracting game. First, we show that the strategies $\{s_i\}_{i=1}^m$ are optimal in the second stage. The constraint (7) with $t_i = t_i'$ requires $s_i(t_i, t_{-i})$ to be a best response of player i with type t_i to the strategies of the other players. It remains to show that players do not have an incentive to deviate at the contracting stage. Suppose that player j with type t_j offers a contract $c_j \neq c_j^{t_j}$. We consider two cases.

Case 1: There exists a $t_j' \in T_j$ such that $c_j = c_j^{t_j'}$. Then, by (7), this deviation is not profitable no matter what the second-stage strategy of player j is.

Case 2: $c_j \neq c_j^{t_j'}$ for all $t_j' \in T_j$. Such a deviation induces player i ($i \neq j$) with type t_i to take action $s_i^j(t_{-j}, f_j^{c_j})$, where $f_j^{c_j}$ is defined by (11). Hence, by (8), such a deviation cannot be profitable.

Moreover, these strategies satisfy the constraints imposed by the refinement concept (3). This is because

$$\begin{aligned}\alpha_i\bigl(t_i, (c_j, \gamma_{-j}(t_{-j}))\bigr) &= s_{-j}^j(t_{-j}, f_j^{c_j}) \in \mathcal{R}_j(f_j^{c_j} \times p_{-j}^j, t_{-j}) \\ &= \mathcal{R}_j\bigl(S(c_j, \gamma_{-j}(t_{-j})), t_{-j}\bigr),\end{aligned}$$

where the equalities are satisfied by the definitions of α_i, $f_j^{c_j}$, and γ, and the middle of the chain is just (9).

Finally, since $\alpha_i(t_i, \gamma(t)) = s_i(t)$ for all i and t, these strategies indeed implement the outcome function s. *Q.E.D.*

PROOF OF THE "ONLY IF" PART OF THEOREM 1: Fix an $\mathcal{R}$ equilibrium in the contracting game which implements the outcome function $s = (s_1, \dots, s_m) : T \to A$. For all i and j, we construct the objects $\tau_i : T_i \to 2^{T_i}$, $r_i : T \to 2^{A_i} \setminus \{\emptyset\}$, $p_i^j : T_{-j} \to 2^{A_i} \setminus \{\emptyset\}$, and $s_i^j : T_{-j} \times F_j^\tau$ such that s_i and r_i are measurable with respect to τ_{-i}, p_i^j and s_i^j are measurable with respect to τ_{-ij}, and $s_i(t) \in r_i(t)$, $s_i^j(t_{-j}, f_j) \in p_i^j(t_{-j})$. Then we show that (7), (8), and (9) are satisfied.

Denote the equilibrium contract of player i with type t_i by $c_i^{t_i}$. Define the partition τ as

$$\tau_i(t_i) = \bigl\{t_i' \in T_i : c_i^{t_i} = c_i^{t_i'}\bigr\}.$$

For all $i \in \{1, \dots, m\}$, let

$$r_i(t) = \widetilde{c}_i^{t_i}\bigl(c_i^{t_i}, (c_j^{t_j})_{j \neq i}\bigr). \tag{12}$$

Notice that r_i is measurable with respect to τ_{-i} by the definition of $\tau = \bigtimes_{i=1}^{m} \tau_i$.

Let $\alpha_i : T_i \times C \to A_i$ denote the second-stage strategy of player i. Observe that

$$\alpha_i(t_i, (c_j^{t_j})_j) \in \tilde{c}_i^{t_i}(c_i^{t_i}, (c_j^{t_j})_{j \neq i}) \tag{13}$$

by the rules of the contracting game. Since the contracting equilibrium implements $s = (s_1, \dots, s_m)$, it follows that $\alpha_i(t_i, (c_j^{t_j})_j) = s_i(t_1, \dots, t_m)$. Notice that $s_i(t)$ is measurable with respect to τ_{-i}. In addition, $s_i(t) \in r_i(t)$ by (12) and (13).

We are ready to show that the triple $(\{\tau_i\}_{i=1}^m, \{r_i\}_{i=1}^m, s)$ satisfies (7). First, consider this constraint with $t_i' = t_i$. This constraint requires $\alpha_i(t_i, (c_j^{t_j})_j)$ to be a best response of player i with type t_i to the strategies of the other players on the equilibrium path. Since α_i was an equilibrium strategy, it has to be a best response and hence (7) must be satisfied. Second, consider (7) with $t_i' \neq t_i$. Then this constraint says that player i with type t_i is better off offering the contract $c_i^{t_i}$ than offering $c_i^{t_i'}$. Indeed, the left-hand side is just his expected equilibrium payoff and the right-hand side is the maximum payoff of player i with type t_i if he offered $c_i^{t_i'}$. Since $c_i^{t_i}$ was an equilibrium contract, player i is better off offering $c_i^{t_i}$ than any other contract, hence, (7) is satisfied.

It remains to construct p_i^j and s_i^j for all i, j $(i \neq j)$ and to show that (8) is also satisfied. For each i, let N_i denote the number of elements in the partition of T_i generated by τ_i. For each i, fix a τ_i-measurable surjection $\sigma_i : T_i \to \{1, \dots, N_i\}$. For each $f_j \in F_j^\tau$ and for all $n_{-j} \in \bigtimes_{k \neq j}\{1, \dots, N_k\}$, define $\bar{f}_j(n_{-j})$ to be $f_j(t_{-j})$ if $n_{-j} = \sigma_{-j}(t_{-j})$. The function $\bar{f}_j$ is well defined because f_j is τ_{-j}-measurable and σ_{-j} is a τ_{-j}-measurable surjection. For all $n_i \in \{1, \dots, N_i\}$, define $c_i^{n_i}$ to be $c_i^{t_i}$ if $\sigma_i(t_i) = n_i$. The Invariant Punishment Property guarantees that there are functions $\bar{p}_k^j : \mathbb{N}^{m-1} \to 2^{A_k} \setminus \{\emptyset\}$ $(k \neq j)$ such that for each $f_j \in F_j^\tau$, there is a contract $c_j^{f_j} \in C_j$ which satisfies

$$\tilde{c}_j^{f_j}(c_j^{f_j}, (c_i^{n_i})_{i \neq j}) = \bar{f}_j(n_{-j}) \quad \text{and} \quad \tilde{c}_k^{n_k}(c_j^{f_j}, (c_i^{n_i})_{i \neq j}) = \bar{p}_k^j(n_{-j}) \tag{14}$$

for all $n_{-j} \in \bigtimes_{i \neq j}\{1, \dots, N_i\}$ and $k \neq j$. For all $t_{-j} \in T_{-j}$, define $p_k^j(t_{-j})$ to be $\bar{p}_k^j(\sigma_{-j}(t_{-j}))$. Notice that p_k^j is τ_{-j}-measurable. Using this notation and the definition of $\bar{f}_j$, (14) can be rewritten as

$$\tilde{c}_j^{f_j}(c_j^{f_j}, (c_i^{t_i})_{i \neq j}) = f_j(t_{-j}) \quad \text{and} \quad \tilde{c}_k^{t_k}(c_j^{f_j}, (c_i^{t_i})_{i \neq j}) = p_k^j(t_{-j})$$

for all $t_{-j} \in T_{-j}$ and $k \neq j$. For each $f_j \in F_j^\tau$ and $k \neq j$, define $s_k^j(t_{-j}, f_j)$ to be $\alpha_k(t_k, (c_j^{f_j}, (c_i^{t_i})_{i \neq j}))$. The function s_i^j is obviously measurable with respect to τ_{-ij}. Given these notations, (8) requires that player j cannot profitably deviate by offering $c_j^{f_j}$. Therefore, this constraint is satisfied. Finally, since the strategies in the contracting game satisfly the refinement,

$$
\begin{aligned}
s^j_{-j}(t_{-j}, f_j) &= \alpha_{-j}\big(t_{-j}, (c_j^{f_j}, (c_k^{t_k})_{k\neq j})\big) \\
&\in \mathcal{R}_j\big(\widetilde{c}_j^{f_j}(c_j^{f_j}, (c_k^{t_k})_{k\neq j}), \big(\widetilde{c}_i^{t_i}(c_j^{f_j}, (c_k^{t_k})_{k\neq j})\big)_{i\neq j}, t_{-j}\big) \\
&= \mathcal{R}_j(f_j(t_{-j}) \times p^j_{-j}(t_{-j}), t_{-j})
\end{aligned}
$$

and, hence, (9) is satisfied. *Q.E.D.*

5. DEFINABLE CONTRACTS

5.1. *The Language and the Gödel Coding*

We consider a formal language that is sufficiently rich to allow its user to state propositions in arithmetic. (The Appendix provides a formal definition of a first-order language; see Section A.2.) Furthermore, the set of statements in this language is closed under the finite applications of the Boolean operations $\neg$, $\vee$, and $\wedge$. In addition, the language contains variable symbols, such as x, y, which enable one to express, for example, Fermat's last theorem:

$$\forall n, x, y, z\big\{[(n \geq 3) \wedge (x \neq 0) \wedge (y \neq 0) \wedge (z \neq 0)] \rightarrow (x^n + y^n \neq z^n)\big\}.$$

In fact, one can also express statements in the language that involve any finite number of free variables. For example, "x is a prime number" is a statement in the language. The symbol x is a free variable in the statement. Another example for a predicate that has one free variable is "$x < 4$." One can substitute any integer into x and then the predicate is either true or false. This particular one is true if $x = 0, 1, 2, 3$ and is false otherwise.

A *text* is a finite string of symbols. Let $\mathfrak{L}$ be the set of all texts of the formal language. It is well known that one can *construct* a one-to-one function $\mathfrak{L} \rightarrow \mathbb{N}$. Let $[\varphi]$ be the value of this function at $\varphi \in \mathfrak{L}$; call it the Gödel code of the text φ.

In what follows, we define a class of functions which can be represented by finitely many characters in our formal language.

DEFINITION 2: The function $f : \mathbb{N}^k \rightarrow 2^{\mathbb{N}}$ is said to be *definable* if there exists a first-order arithmetic statement ϕ in $k + 1$ free variables such that for all $a \in \mathbb{N}^k$, $b \in f(a)$ if and only if $\phi(a, b)$ is true.

We provide a formal definition of first-order arithmetic statement in the Appendix. We illustrate the previous definition with an example.

EXAMPLE: Consider the following function defined on $\mathbb{N}$:

$$f(a) = \begin{cases} 0, & \text{if } a \text{ is an even number,} \\ 1, & \text{if } a \text{ is an odd number.} \end{cases}$$

We show that this function is definable by constructing the corresponding predicate ϕ:

$$\phi(x, y) \equiv \{\{y = 1\} \vee \{y = 0\}\} \wedge \{\exists z : 2z = y + x\}.$$

Notice that ϕ indeed has two free variables. (The variable z is not free because there is a quantifier in front of it.) The first part of ϕ states that y is either 1 or 0. The second part says that $x + y$ is divisible by 2. Clearly, $f(a) = 0$ if and only if $\phi(a, 0)$ is true and $f(a) = 1$ if and only if $\phi(a, 1)$ is true.

If the statement ϕ defines the function f and if θ is true, then $\phi \wedge \theta$ also defines f. We make use of this observation to construct different but computationally equivalent contracts in the next section.

DEFINITION 3: Suppose that f_n is a function mapping from $\mathbb{N}^k$ to $2^{\mathbb{N}}$ for all $n \in \mathbb{N}^q$. Suppose that there exists a first-order arithmetic statement ϕ in $q + k + 1$ free variables such that for all $n \in \mathbb{N}^q$ and $a \in \mathbb{N}^k$, $b \in f_n(a)$ if and only if $\phi(n, a, b)$ is true. Then we call the expression f_x a *definable function with q free variables*, where $x = (x_1, \dots, x_q)$ is a vector of variable symbols.

A definable function from $\mathbb{N}^k$ to $2^{\mathbb{N}}$ with q free variables is essentially a definable function from $\mathbb{N}^{k+q}$ into $2^{\mathbb{N}}$. Therefore, many properties of definable functions are also properties of definable functions with free variables.

We can now describe some properties of definable functions that we need in our proofs. We need two pieces of notation. First, recall from the introduction that if $n \in \mathbb{N}$, then $\langle n \rangle$ denotes the text whose Gödel code is n, that is, $[\langle n \rangle] = n$. Let g be a function from $\{1, \dots, q\}$ to the set of variable symbols such that $g(i) \neq g(j)$ if $i \neq j$; that is $(g(1), \dots, g(q))$ is a finite vector of variable symbols. Then, for any text φ and $(n_1, \dots, n_q) \in \mathbb{N}^q$, let $\varphi^{(n_1, \dots, n_q)}$ denote the text where if the symbol $g(k)$ stands for a free variable in φ, then $g(k)$ is replaced by n_k in φ for $k = 1, \dots, q$.[10] For example, if $g(1) = x_1$ and $g(2) = x_2$, φ is $x_1 < x_2$, $n_1 = 1$ and $n_2 = 2$, then $\varphi^{(n_1, n_2)}$, is $1 < 2$.[11]

Suppose that $g(i) = x_i$ for $i = 1, \dots, q$. Consider the following text in n free variables: $\langle x_k \rangle^{(x_1, \dots, x_q)}$, where $k \leq q$. Since the Gödel coding is a bijection, $\langle n_k \rangle$ is

[10] The text $\varphi^{(n_1, \dots, n_q)}$ depends on g, which specifies the list of free variables that are to be replaced. We suppress this dependence in the notation just to make it simpler.

[11] Of course, it is possible that the text φ does not contain some of the symbols $\{g(1), \dots, g(q)\}$. In that case, there is no substitution for the missing symbols in $\varphi^{(n_1, \dots, n_q)}$. For example, $g(1) = x_1$ and $g(2) = x_2$ and φ is $x_2 > 2$, then $\varphi^{(3,4)}$ is $4 > 2$, because x_1 does not appear in φ.

a text for each $n_k \in \mathbb{N}$. Since $\varphi^{(n_1,\dots,n_q)}$ is defined for all φ and $(n_1,\dots,n_q)\in\mathbb{N}^q$, $\langle n_k\rangle^{(n_1,\dots,n_q)}$ is a text for all $(n_1,\dots,n_q)\in\mathbb{N}^q$. The following lemma is a well known result in mathematical logic.

LEMMA 1: *Let $(g(1),\dots,g(q))$ be a vector of distinct free variables and, for all $k\in\{1,\dots,q\}$, let $f_k(n_1,\dots,n_q)=[\langle n_k\rangle^{(n_1,\dots,n_q)}]$. Then f_k is a definable function for all $k\in\{1,\dots,q\}$.*

This basic result is central to the construction of cross-referential contracts. The next result is used to show that the contracts we construct to support various kinds of equilibrium are definable.

LEMMA 2: *For any set A, let χ_A denote the characteristic function of A.*

(i) *If $A\subset\mathbb{N}^k$ is finite, then χ_A is definable.*

(ii) *Let $A, B\subset\mathbb{N}^k$. If χ_A and χ_B are definable, then $\chi_{A\cap B}$, $\chi_{A\cup B}$, and $\chi_{A\setminus B}$ are definable.*

(iii) *Let $A_1,\dots,A_m\subset\mathbb{N}^k$ and $B_1,\dots,B_{m+1}\subset\mathbb{N}$. If $\bigcap_{i=1}^m A_i=\{\emptyset\}$, $\chi_{A_1},\dots,\chi_{A_m}$ are definable and $B_1,\dots,B_m$ are finite, then the following correspondence, $f:\mathbb{N}^k\to\mathbb{N}$, is definable:*

$$f(n)=\begin{cases} B_1, & \text{if } n\in A_1,\\ \vdots & \vdots\\ B_m, & \text{if } n\in A_m,\\ B_{m+1}, & \text{otherwise.}\end{cases}$$

(iv) *Suppose that $D_1,\dots,D_k\subset\mathbb{N}$ are finite sets, $g:\bigtimes_{i=1}^k D_i\to 2^{\mathbb{N}}$, and $B_1,\dots,B_{k+1}\subset\mathbb{N}$, are finite sets. For all $n\in\mathbb{N}^k$, let n_i denote the ith coordinate of n. Then the following correspondence is definable:*

$$f(n)=\begin{cases} g(n), & \text{if } n\in\bigtimes_{i=1}^k D_i,\\ B_1, & \text{if } \{i:n_i\notin D_i\}=\{1\},\\ \vdots & \\ B_k, & \text{if } \{i:n_i\notin D_i\}=\{k\},\\ B_{k+1}, & \text{otherwise.}\end{cases}$$

We point out that all the contracts we use to construct equilibrium in the paper are in the form of Lemma 2(iv).

See the Appendix for the proof.

5.2. *The Contract Space*

With a slight abuse of notation, we refer to $[a_i]$ as the Gödel code of the text describing action a_i.[12] In addition, for each $\bar{A}_i \subset A_i$, let $[\bar{A}_i] = \{[a_i] : a_i \in A_i\}$. We define the contract space of player i, C_i, as the set of all arithmetic statements defining functions from $\mathbb{N}^m$ to $2^{[A_i]} \setminus \{\emptyset\}$ in the sense of Definition 2.

For all $c_i \in C_i$, define the correspondence $\widetilde{c}_i : C \to 2^{A_i} \setminus \{\emptyset\}$ induced by c_i as

$$a_i \in \widetilde{c}_i(c_1, \dots, c_m) \quad \Leftrightarrow \quad c_i([c_1], \dots, [c_m], [a_i]) \text{ is true.} \tag{15}$$

That is, given the contract profile $(c_1, \dots, c_m)$, player i can take action a_i if and only if the Gödel code of this action is an element of the function defined by c_i evaluated at the vector $([c_1], \dots, [c_m])$.

To show that the statement of Theorem 1 is valid with this contract space, we have to show that both the Cross-Referential Property and the Invariant Punishment Property hold.

5.2.1. *The Cross-Referential Property*

Let $N_i \geq 1$, $r_i : \mathbb{N}^m \to 2^{A_i} \setminus \{\emptyset\}$, and $p_i^j : \mathbb{N}^{m-1} \to 2^{A_i} \setminus \{\emptyset\}$ for $i, j \in \{1, \dots, m\}$, $i \neq j$. Let x_j denote the vector $(x_j^1, \dots, x_j^{N_j})$ and let $x = (x_1, \dots, x_m)$, where $x_j^{n_j}$ is a variable symbol for all j and n_j. For each i and $n_i \in \{1, \dots, N_i\}$, let $h_i^{n_i}(x)$ denote $[\langle x_i^{n_i} \rangle^{(x)}]$ and consider the following in $\sum_{i=1}^m N_i$ free variables

$$f_x^{i,n_i}([c_1], \dots, [c_m]) = \begin{cases} [r_i(n_i, n_{-i})], & \text{if } \forall k \neq i \; h_k^{n_k}(x) = [c_k] \text{ and} \\ & h_k^{n_k}(x) \neq h_k^{n'_k}(x) \text{ if } n_k \neq n'_k, \\ [p_i^j(n_i, n_{-ij})], & \text{if } \forall k \neq i, j \; h_k^{n_k}(x) = [c_k] \text{ and } h_k^{n_k}(x) \neq h_k^{n'_k}(x) \\ & \text{if } n_k \neq n'_k \text{ and } \nexists n_j \text{ s.t. } h_j^{n_j}(x) = [c_j], \\ [A_i], & \text{otherwise.} \end{cases} \tag{16}$$

We prove in the Appendix that this function is definable (see Lemma 6). Let $\varphi^{i,n_i}(x, y, z)$ be a statement which defines the function f_x^{i,n_i}, where y is an m-dimensional vector of variable symbols and z is a variable symbol. Define $c_x^{i,n_i}(y, z)$ to be $\varphi^{i,n_i}(x, y, z) \wedge (n_i + 1 > n_i)$ and let $\gamma_i^{n_i}$ denote its Gödel code.[13] In addition, let $\gamma_i = (\gamma_i^1, \dots, \gamma_i^{N_i})$ and $\gamma = (\gamma_1, \dots, \gamma_m)$. Observe that the con-

[12] In other words, we identify an action with a text which describes it.

[13] Notice that $n_i + 1 > n_i$ is always true, and, hence, $c_x^{i,n_i}(y, z)$ and $\varphi^{i,n_i}(x, y, z)$ define the same function. Such a statement, however, makes it possible that a player with two different types offers two different but computationally equivalent contracts.

tract c_γ^{i,n_i} defines the function

$$\begin{aligned} & f_\gamma^{i,n_i}([c_1],\ldots,[c_m]) \\ & \quad = \begin{cases} [r_i(n_i,n_{-i})], & \text{if } \forall k \neq i \; h_k^{n_k}(\gamma) = [c_k] \text{ and} \\ & h_k^{n_k}(\gamma) \neq h_k^{n'_k}(\gamma) \text{ if } n_k \neq n'_k \\ [p_i^j(n_{-j})], & \text{if } \forall k \neq i, j \; h_k^{n_k}(\gamma) = [c_k] \text{ and } h_k^{n_k}(\gamma) \neq h_k^{n'_k}(\gamma) \\ & \text{if } n_k \neq n'_k \text{ and } \nexists n_j \text{ s.t. } h_j^{n_j}(\gamma) = [c_j], \\ [A_i], & \text{otherwise.} \end{cases} \end{aligned}$$

Recall that $h_q^{n_q}(x) = [\langle x_q^{n_q}\rangle^{(x)}]$ and $\gamma_q^{n_q} = [c_x^{q,n_q}]$. Hence $h_q^{n_q}(\gamma) = (c_x^{q,n_q})^{(\gamma)} = c_\gamma^{q,n_q}$. Therefore, the previous function can be rewritten as[14]

$$f_\gamma^{i,n_i}([c_1],\ldots,[c_m]) = \begin{cases} [r_i(n_i,n_{-i})], & \text{if } \forall k \neq i \; [c_\gamma^{k,n_k}] = [c_k], \\ [p_i^j(n_{-j})], & \text{if } \forall k \neq i, j \; [c_\gamma^{k,n_k}] = [c_k] \text{ and} \\ & \nexists n_j \text{ s.t. } [c_\gamma^{j,n_j}] = [c_j], \\ [A_i], & \text{otherwise.} \end{cases} \tag{17}$$

Define $c_i^{n_i}$ to be c_γ^{i,n_i}. Therefore, by (15),

$$\begin{aligned} & \widetilde{c}_i^{n_i}(c_1,\ldots,c_m) \\ & \quad = \begin{cases} r_i(n_i,n_{-i}), & \text{if } \forall k \neq i \; c_k^{n_k} = c_k, \\ p_i^j(n_{-j}), & \text{if } \forall k \neq i, j \; c_k^{n_k} = c_k \text{ and } \nexists n_j \text{ s.t. } c_j^{n_j} = c_j, \\ A_i, & \text{otherwise,} \end{cases} \end{aligned}$$

which is just (4). In addition, since $c_i^{n_i} = \varphi^{i,n_i}(\gamma, y, z) \wedge (n_i + 1 > n_i)$, then $c_i^{n_i} \neq c_i^{n'_i}$ if $n_i \neq n'_i$ and, hence, the Cross-Referential Property is satisfied.

5.2.2. *The Invariant Punishment Property*

To prove that our contract space satisfies the Invariant Punishment Property, it is enough to prove the following lemma.

LEMMA 3: *For all $(N_1,\ldots,N_m) \in \mathbb{N}^m$, $N_i \geq 1$, for all sets of contracts $\{c_j^{n_j}\}_{j,n_j\in\{1,\ldots,N_j\}}$ ($c_j^{n_j} \in C_j$, $c_j^{n_j} \neq c_j^{n'_j}$ if $n_j \neq n'_j$), and for every i, there are functions $p_k^i : N^{m-1} \to 2^{A_k} \setminus \{\emptyset\}$ ($k \neq i$) such that for any function $f_i : \mathbb{N}^{m-1} \to 2^{A_i} \setminus \{\emptyset\}$, there is a contract $c_i^* \in C_i$ such that for all $n_{-i} \in \bigtimes_{j\neq i}\{1,\ldots,N_j\}$,*

$$\widetilde{c}_i^*(c_i^*, c_{-i}^{n_{-i}}) = f(n_{-i}), \tag{18}$$

[14] Notice that $c_\gamma^{k,n_k} \neq c_\gamma^{k,n'_k}$ if $n_k \neq n'_k$; hence, these conditions became irrelevant in the definition of f_γ^{i,n_i} (see the previous footnote).

and for all $k \neq i$

$$\widetilde{c}_k^{n_k}(c_i^*, c_{-i}^{n_{-i}}) = p_k^i(n_{-i}), \tag{19}$$

where $\widetilde{c}_j$ *is defined by* (15) *for all* j *and* $c_j \in C_j$.

First, we reformulate the statement of the lemma. Let $\mathcal{N}_i$ denote $\{1, \dots, N_i\}$ and let $\mathcal{N} = \bigtimes_{i=1}^m \mathcal{N}_i$. Let $(2^{A_i})^{|\mathcal{N}_{-i}|}$ denote the set of $\bigtimes_{j\neq i} N_j$-dimensional vector of nonempty subsets of A_i. For all $(A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \in (2^{A_i})^{|\mathcal{N}_{-i}|}$, define $S((A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}})$ as

$$\big\{(A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} : \forall n_{-i} \in \mathcal{N}_{-i},\ \exists c_i \text{ s.t. } \widetilde{c}_i(c_i, c_{-i}^{n_{-i}}) = A_i^{n_{-i}},$$
$$\widetilde{c}_{-i}^{n_{-i}}(c_i, c_{-i}^{n_{-i}}) = A_{-i}^{n_{-i}}\big\}.$$

$(A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \in S((A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_j})$ implies that there exists a contract c_i, available for player i such that if he offers c_i and the contract profile of the other players is $c_{-i}^{n_{-i}}$, then the contract profile $(c_i, c_{-i}^{n_{-i}})$ restricts the action space of player i to be $A_i^{n_{-i}}$ and the action spaces of the other players to be $A_{-i}^{n_{-i}}$. We claim the following lemma holds.

LEMMA 4: *The statement of Lemma* 3 *is equivalent to the following statement. For all* $(N_1, \dots, N_m) \in \mathbb{N}^m$, $N_i \geq 1$, *for all sets of contracts* $\{c_j^{n_j}\}_{j,n_j\in\{1,\dots,N_j\}}$ $(c_j^{n_j} \in C_j, c_j^{n_j} \neq c_j^{n'_j}$ *if* $n_j \neq n'_j)$, *and for every* i,

$$\bigcap_{(A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}}} S\big((A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}}\big) \neq \{\emptyset\}. \tag{20}$$

PROOF: Suppose first that (20) is true and that $(A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}}$ is an element of the intersection. Define $p_{-i}^i(n_{-i}) = (p_k^i(n_{-i}))_{k\neq i}$ to be $A_{-i}^{n_{-i}}$ for all n_{-i}. Fix a function $f_i : \mathcal{N}_{-i} \to 2^{A_i} \setminus \{\emptyset\}$ and consider $S((f_i(n_{-i}))_{n_{-i}\in\mathcal{N}_{-i}})$. Then there exists a c_i^* such that equations (18) and (19) are satisfied because $(p_{-i}^i(n_{-i}))_{n_{-i}} \in S((f_i(n_{-i}))_{n_{-i}})$.

Conversely, suppose that (20) is false. Then for all $p_{-i}^i(n_{-i}) = (p_k^i(n_{-i}))_{k\neq i} : \mathcal{N}_{-i} \to A_{-i}$ there exists $(A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \in (2^{A_i})^{|\mathcal{N}_{-i}|}$ such that

$$(p_{-i}^i(n_{-i}))_{n_{-i}\in\mathcal{N}_{-i}} \notin S\big((A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}}\big).$$

Define $f_i(n_{-i})$ to be $A_i^{n_{-i}}$ for all $n_{-i} \in \mathcal{N}_{-i}$. Then, by the definition of S, there does not exist a contract c_i^* such that (18) and (19) are satisfied. *Q.E.D.*

By the previous lemma, to show Lemma 3, we only have to prove (20). We have relegated this proof to the Appendix. Here, we sketch the proof for the case of two players, where $N_1 = N_2 = 1$. Let c_2 denote the contract of player 2. For all $(\{\emptyset\} \neq)\ B_1 \subset A_1$,

$$S(B_1) = \{B_2 : \exists c_1 \text{ s.t. } \widetilde{c}_1(c_1, c_2) = B_1,\ \widetilde{c}_2(c_1, c_2) = B_2\}.$$

We have to show that $\bigcap_{\{\emptyset\} \neq B_1 \subset A_1} S(B_1) \neq \{\emptyset\}$. Suppose that, by contradiction, $\bigcap_{B_1} S(B_1) = \{\emptyset\}$. Then for all $B_2 \subset A_2$, there exists a $(\{\emptyset\} \neq)\ B_1 \subset A_1$ such that $B_2 \notin S(B_1)$. Therefore, one can construct a function, $g : 2^{A_2} \setminus \{\emptyset\} \to 2^{A_1} \setminus \{\emptyset\}$ such that

$$\forall B_2 \subset A_2 : B_2 \notin S(g(B_2)).$$

Let f^{c_2} denote the function defined by c_2. Define the function in one free variable, f_x, as

$$f_x([c_1], [c_2]) = \bigl[g\bigl(\bigl\langle f^{c_2}\bigl(\bigl[\langle x\rangle^{(x)}\bigr], [c_2]\bigr)\bigr\rangle\bigr)\bigr].$$

We show (see Lemma 7 in the Appendix) that since g has a finite domain, f_x is a definable function in one free variable. Let c_x denote a statement which defines f_x and let γ denote its Gödel code. Notice that $\langle \gamma \rangle^{(\gamma)} = c_\gamma$. Hence,

$$f_\gamma([c_1], [c_2]) = \bigl[g\bigl(\bigl\langle f^{c_2}([c_\gamma], [c_2])\bigr\rangle\bigr)\bigr]. \tag{21}$$

Notice that

$$\widetilde{c}_2(c_\gamma, c_2) \in S(\widetilde{c}_\gamma(c_1, c_2))$$

by the definition of S. Alternatively,

$$\widetilde{c}_2(c_\gamma, c_2) \notin S\bigl(g(\widetilde{c}_2(c_\gamma, c_2))\bigr) = S(\widetilde{c}_\gamma(c_\gamma, c_2)),$$

where the exclusion follows from the construction of g, and the equality follows from (15) and (21). The previous two displayed statements contradict each other and, hence, $\bigcap_{B_1 \subset A_1} S(B_1) \neq \{\emptyset\}$.

6. APPLICATIONS AND EXAMPLES

This section accomplishes two goals. First, we illustrate some properties of the contracting equilibrium by two examples. Second, we compare the set of outcome functions implementable by a centralized mechanism with those implementable by our contracting game.

6.1. *Example 1*

Suppose that there are two players. The action space of player 1 is $\{a, b\}$ and the action space of player 2 is $\{l, m, r\}$. The type space of player 1 is $D = \{-1, 1\}$

and the type space of player 2 is $T \times D = \{3, -3\} \times \{-1, 1\}$. Any realization of the type of each player is equally likely and the types of the players are independently distributed. If the type of player 1 is d_1 and the type of player 2 is (t, d_2), then the payoffs to the players are defined by the matrix

$$\begin{array}{c|c|c|c|}
 & l & m & r \\
\hline
a & 6, t & 2+d_1, 2+d_2 & 0, 0 \\
\hline
b & 0, -t & 0, 0 & 2, 2 \\
\hline
\end{array}$$

We want to show that the following outcome function can be supported as an $\mathcal{R}$ equilibrium in the contracting game:

$$s(d_1, t, d_2) = \begin{cases} (a, l), & \text{if } t = 3, \\ (b, l), & \text{if } t = -3. \end{cases}$$

In this outcome function the action of player 1 varies with the type of player 2. Therefore, if the players played this game without the contracting stage, the outcome function s could not be implemented as a Bayesian equilibrium. Moreover, player 2 always takes action l in the outcome function s. Given l, player 1 would like to deviate and take action a no matter what the types are. Below we show how to construct a contract of player 2 which prevents such a deviation.

In what follows, c denotes the equilibrium contract of player 1 and c^t denotes the equilibrium contract of player 2 if the second coordinate of his type is t. Define the mappings determined by these contracts as

$$\widetilde{c}(c_1, c_2) = \begin{cases} a, & \text{if } c_2 = c^3, \\ b, & \text{if } c_2 = c^{-3}, \\ \{a, b\}, & \text{otherwise}, \end{cases}$$

and for $t = 3, -3$,

$$\widetilde{c}^t(c_1, c_2) = \begin{cases} l, & \text{if } c_1 = c, \\ r, & \text{otherwise}, \end{cases} \tag{22}$$

such that $c^3 \neq c^{-3}$. The Cross-Referential Property guarantees that these contracts lie in the contract space.

Notice that these contracts refer to each other and they implement the outcome function s. Indeed, player 1's contract, c, prescribes taking action a if the contract of player 2 is c^3 and action b if player 2's contract is c^{-3}. Similarly, player 2's contracts, c^3 and c^{-3}, prescribe taking action l if the contract of player 1 is c. We only have to show that players cannot profitably deviate.

The contracts always constrain the players to a specific action so they cannot deviate in the second stage of the game. We only need to check deviations in contracts. The payoff of player 2 is maximized by the outcome function, so we only have to show it for player 1. Any deviation of player 1 triggers action r by player 2. The best response of player 1 to this action is b and it provides him with a payoff of 2, which is smaller than his expected equilibrium payoff, 3.

Next, we discuss a number of features of this contract equilibrium that will be of interest in the general model.

INFORMATION CONTENT OF THE CONTRACTS: The contracts offered by player 2 for each of the realizations of the first coordinate of his types are different. In this sense, equilibrium play reveals something about player 2's type. In fact, it is precisely this feature that makes it possible for player 1 to take an action on the equilibrium path that depends on player 2's type. However, there is a limit to this in the sense that player 1's action can only vary with player 2's type to the extent that type information is revealed through the equilibrium contract. In this example, the action of player 1 cannot depend on d_2 because d_2 is not revealed through the equilibrium contracts.

The two contracts of player 2, c^3 and c^{-3}, are computationally equivalent, that is, they determine the same actions as a function of the contract profile. However, $c^3 \neq c^{-3}$, and player 2 uses his contract to communicate his type to player 1. This communication is not cheap talk because the contract c^t is also used by player 2 as a commitment to punish player 1 unless he credibly promises to make his action contingent on t by offering c.[15]

Alternatively, since contracts reveal information about types, a deviating player can make his second-stage action contingent on the types of the other players. This limits the set of outcome functions which are implementable by contractible contracts. We show that a centralized mechanism designer can implement more outcome functions than our contracting game because he can prevent a nonparticipant player from learning something about the others' types.

INVARIANT PUNISHMENTS: Any deviation of a player at the contracting stage triggers a restriction on the action space of the other player. These restrictions can be viewed as a punishment for a deviation. Observe that our equilibrium is supported by "punishments" that do not vary with the transgression. For example, player 2 simply commits himself to choose r whenever player 1 offers a contract that is different from his equilibrium contract. The Invariant Punishment property implies that assuming that the contractual punishment is invariant to the deviation is without loss of generality. (Of course, the punishment usually depends on the type of the punisher.)

We further explain the significance of this property. Since the punishment for any contractual deviation of player 1 is the same, he can best respond to the equilibrium contract of player 2. That is, the most profitable deviation of

[15] In fact, the allocation s cannot be implemented by introducing a cheap talk stage instead of the contracting stage because player 1 would not take action b if player 2 takes action l.

player 1 specifies a restriction which is a best response to the punishment of player 2 given the second-stage strategies. This allows us to use a logic that is similar to the minmax logic in games of complete information. The following table summarizes the best responses of player 1 as a function of the restrictions of player 2:[16]

	Restrictions of player 2		
	$\{m\}$	$\{r\}$	$\{m, r\}$
Best response of player 1	$\{a\}$	$\{b\}$	$\{a\}$ if $d_1 = 1$ $\{b\}$ otherwise
Expected payoff of player 1	2	2	2.5

Consider, for example, the last column of the table. The punishment of player 2 is $\{m, r\}$. What is the best response of player 1? If $d_1 = 1$, he can restrict his action space to be $\{a\}$. Player 2 observes this restriction and is forced to take his strictly dominant action m in the second stage. Similarly, player 1 can restrict his action space to be $\{b\}$ if $d_1 = -1$ and the best response of player 2 is r. Player 1's expected payoff is 2.5. It is clear from the table that player 1 can always achieve a payoff of 2 no matter what the punishment of player 2 is.

Consider now a modification of the original game such that the action profile (a, l) generates a payoff of 3 to player 1 instead of 6, but the payoffs are the same otherwise. In this case, our target outcome function generates a payoff of 1.5 to player 1. This payoff is lower than 2 and hence, the outcome function cannot be implemented as an $\mathcal{R}$ equilibrium in the contracting game. Next, we argue that our outcome function could be implemented even in the modified game if the restriction of player 2 could depend on the restriction of player 1.

Suppose now, that the restriction of player 2 can depend on the restriction of player 1 generated by a deviation. Next, we present another table which summarizes the most effective punishment of player 2 if the punishment can be contingent on the restriction implied by the deviation:

	Restrictions of player 1		
	$\{a\}$	$\{b\}$	$\{a, b\}$
Best punishment of player 2	$\{r\}$	$\{m\}$	$\{m, r\}$
Expected payoff of player 1	0	0	1.25

[16] It is never optimal to punish player 1 by taking action l. Therefore, we left this possibility out of the table.

Therefore, if the restriction of player 2 could depend on the restriction of player 1, player 1 can only achieve a payoff of 1.25. Hence, the target outcome function could be implemented even in the modified game.

EQUILIBRIUM REFINEMENT: The outcome functions that are supportable as equilibria generally depend on the equilibrium refinement concept. As we explained above, an important feature of the contract equilibrium is the restriction on the action space triggered by a deviation. Since deviations are off the equilibrium path, players never have to choose from these restricted sets of actions on the equilibrium path. Different refinement concepts impose different restrictions on these off-equilibrium choices.

To see how refinement matters consider again the modified version of our example and suppose that player 2 restricts his action space to $\{m, r\}$ whenever player 1 deviates. Suppose that one is interested in every Bayesian equilibrium. This concept does not impose any restriction on off the equilibrium play. Hence, if player 1 restricts his action space to be $\{a\}$, player 2 can still take action r although it is strictly dominated. Similarly, if player 1 restricts his action space to be $\{b\}$, player 2 can take action m. In both of these cases, the payoff of player 1 is zero. If player 1 restricts his action space to be $\{a, b\}$, then player 2 can play m if $d_2 = 1$ and play r otherwise. This provides player 1 with an expected payoff of at most 1.25. Hence, the outcome function s can be implemented as a Bayesian equilibrium even in the modified example.

6.2. *Example 2*

The next example illustrates two more properties of the contracting equilibria. First, we show that the informational partition is nontrivial in general. That is, players reveal some information about their types but do not reveal them fully. Second, we show that some outcome functions can only be implemented by a contract profile that does not restrict the action space of a player to a single action. That is, a player must still have some flexibility in choosing his action in the last stage of the game. In this sense, contracts are generally *incomplete*.

Suppose that $m = 2$, $T_1 = A_1 = \{-1, 1\}^2$, and $A_2 = \{-1, 1\}^2 \times \{\alpha, \beta\}$. (The type space of player 2 is degenerate.) If the type of player 1 is $t_1 = (t_1^1, t_1^2)$, the action of player 1 is $a_1 = (a_1^1, a_2^1)$, and the action of player 2 is (a_2^1, a_2^2, α), then the payoff of each player is

$$t_1^1(a_1^1 + a_2^1) + t_1^2(a_1^2 + a_2^2).$$

If the type of player 1 is (t_1^1, t_1^2) and the action of player 2 is (a_2^1, a_2^2, β), then the payoff of player 2 is $5t_1^2 a_2^2$ and the payoff of player 1 is 0.

Notice that if the third coordinate of the action of player 2 is α, this game is a kind of coordination game where both players want to match the type of player 1 with their actions coordinatewise. If both players do so, then each player receives a payoff of 4 conditional on player 2 taking α. The problem is that if player 2 knows the type of player 1, he prefers to take action β and match the second coordinate of player 1's type with the second coordinate of his action. This would provide him with a payoff of 5.

Consider the outcome function $s_1(t_1^1, t_1^2) = (t_1^1, t_1^2)$ and $s_2(t_1^1, t_1^2) = (t_1^1, 1, \alpha)$. In this outcome function, each player matches the first coordinate of the type of player 1 but only player 1 matches the second coordinate. This outcome function generates an expected payoff of 3 to each player. Next, we argue that this outcome function can be implemented by contractible contracts as an equilibrium. To see this, consider a PMM where player 1 reports t_1^1 at the first stage. The mechanism imposes no restriction on the action space of the players. In the second stage, player 1 takes action (t_1^1, t_1^2) and player 2 takes action $(t_1^1, 1, \alpha)$. Obviously, no player has incentive to deviate.

Notice that the information partition is described by $\tau_1(t_1) = t_1^1$ for all t_1. Observe that (s_1, s_2) cannot be implemented such that all the information is revealed about player 1's type. This is because if player 1 would fully reveal his type, player 2 would not participate in the mechanism and would take action $(1, t_1^2, \beta)$.

Also notice that (s_1, s_2) cannot be implemented by complete contracts. This is because player 1 has to match the second coordinate of his own type with the second coordinate of his action. Recall that he can only reveal the first coordinate of his type. Therefore, it must be the case that $r_1(t_1^1, t_1^2)$ contains both $(t_1^1, -1)$ and $(t_1^1, 1)$.

6.3. *Comparison With Centralized Mechanisms*

The set of implementable outcome functions in the contracting game is fairly large. However, as we mentioned before, contract equilibrium imposes a restriction on feasible outcome functions. When a player decides to deviate at the contracting stage, he knows that he will learn something about the types of the other players when he observes their contracts. Therefore, a deviator's action in the last stage of the game can depend on the information about the types of the other players that are revealed by their contracts. This suggests that there are outcome functions which are implementable by centralized mechanisms (where the messages are private) but not by contracts.

By the standard revelation principle, a centralized mechanism asks the players to report their types *privately*. Then the mechanism requires each participant to take an action as a function of the reported type profile of the participants. If a player does not participate in the mechanism, he can take any action he wants. It is without loss of generality to restrict attention to mechanism–

equilibrium pairs where each player participates and truth-telling constitutes a Bayesian equilibrium.[17]

A centralized mechanism has to specify the target outcome function $s : T \to A$ and what actions the others take if player i does not participate, $\underline{s}^i_{-i} : T_{-i} \to A_{-i}$ for each i. Then the outcome function s is implementable by a centralized mechanism if and only if the following two sets of constraints are satisfied: for all $i \in \{1, \dots, m\}$ and $t_i, t'_i \in T_i$,

$$E_{t_{-i}}\big(u_i(s(t), t)|t_i\big) \geq E_{t_{-i}}\big(u_i(s(t'_i, t_{-i}), t)|t_i\big) \tag{23}$$

and

$$E_{t_{-i}}\big(u_i(s(t), t)|t_i\big) \geq \max_{a_i \in A_i} E_{t_{-i}}\big(u_i(a_i, \underline{s}^i_{-i}(t_{-i}), t)|t_i\big). \tag{24}$$

The inequality (23) is the incentive compatibility constraint that guarantees that a participant player reports his type truthfully. The inequality (24) is the participation constraint that guarantees that each player prefers to participate irrespective of his type.

To show that one can implement more outcome functions with centralized mechanisms than with contracts, we revisit the example of the previous subsection. Consider the outcome function $s^* = (s^*_1, s^*_2) : T \to A$, such that $s^*_1(t^1_1, t^2_1) = (t^1_1, t^2_1)$ and $s^*_2(t^1_1, t^2_1) = (t^1_1, t^2_1, \alpha)$. Notice that this outcome function provides each player with a payoff of 4, and it maximizes the sum of the players' payoffs among all outcome functions. We show that this outcome function can be implemented by a centralized mechanism but cannot be implemented by a PMM as an $\mathcal{R}$ equilibrium.

To show that s^* can be implemented by a centralized mechanism, it is enough to construct $\underline{s}^2_1 : T_1 \to A_1$ and $\underline{s}^1_2 : T_2 \to A_2$ such that (23) and (24) are satisfied for $i = 1, 2$ with $s = s^*$. Define $\underline{s}^2_1(t^1_1, t^2_1) \equiv (1, 1)$, that is, if player 2 does not participate, the mechanism requires player 1 to take action $(1, 1)$. Similarly, define $\underline{s}^1_2 \equiv (1, 1, \beta)$. Notice that s^* maximizes the payoff of player 1 among all outcome functions. Hence, (23) and (24) are satisfied for $i = 1$. Since, the type space of player 2 is degenerate, we only have to show that player 2 prefers to participate, that is, (24) holds for $i = 2$. Notice that the right-hand side of (24) is 0 and the left-hand side is 4.

Now, we argue that s^* cannot be implemented by a PMM. Notice that to implement s^*, player 2 must know the type of player 1. Therefore, player 1 has to fully reveal his type at the first stage of the mechanism. If player 2 decides not to participate and player 1 reveals that his type is (t^1_1, t^2_1), player 2 can take action $(1, t^2_1, \beta)$ in the last stage of the mechanism. This would generate a payoff of 5, which is larger than the payoff generated by s^*.

[17] Notice that the mechanism restricts the action space of the participants to singletons, that is, participating players do not choose actions strategically at the second stage. Therefore, each Bayesian equilibrium is also an $\mathcal{R}$ equilibrium.

Next, we identify some environments where the set of outcome functions implementable by centralized mechanisms is the same as the set of outcome functions implementable by contracts.

ASSUMPTION 1: *For all i and $t_i \in T_i$, there exist $\bar{a}_i(t_i) \in A_i$, $\bar{a}^i_{-i} \in A_{-i}$, and $U_i : T \to \mathbb{R}$ such that for all $(t_i, t_{-i}) \in T$, the following inequalities hold*:

(i) *$u_i(\bar{a}_i(t_i), a_{-i}, t) \geq U_i(t)$ for all $a_{-i} \in A_{-i}$.*

(ii) *$u_i(a_i, \bar{a}^i_{-i}, t) \leq U_i(t)$ for all $a_i \in A_i$.*

Part (i) of Assumption 1 says that player i has an action for each of his types which provides him a payoff of at least $U_i(t)$ no matter what the action profile of the other players is. Part (ii) says that players other than player i can take an action profile which holds player i down to at most $U_i(t)$, no matter what action player i takes.

Assumption 1 is arguably a strong assumption, but is satisfied in many economic environments. One way to interpret part (i) is that player i can choose not to interact with the other players and take his outside option. The value $U_i(t)$ can be thought of as the value of the outside option. Similarly, the action profile $\bar{a}^i_{-i}$ can be thought of as a profile of the other players which forces player i to exit.

Consider, for example, an auction environment where a single seller is selling a single object to many bidders. The players in this environment are the bidders and the seller. The type of a player is his signal about the value of the object. The action space of a bidder is the amount of transfers to the seller and the action space of the seller is the set of players (to whom he can sell the object). In this environment, Assumption 1 is obviously satisfied because a bidder can choose not to pay and the seller can decide to keep the object.

PROPOSITION 1: *Suppose that Assumption 1 is satisfied and the allocation $s : T \to A$ can be implemented by a centralized mechanism. Then the allocation s can be implemented as an $\mathcal{R}$ equilibrium in the contractible contracting game.*

PROOF: Suppose that a centralized mechanism implements s, that is, there exists $\{\underline{s}^i_{-i}\}_i$ such that (23) and (24) are satisfied. Notice that by part (i) of Assumption 1,

$$\max_{a_i \in A_i} E_{t_{-i}}\big(u_i(a_i, \underline{s}^i_{-i}(t_{-i}), t) : t_i\big) \geq E_{t_{-i}}(U_i(t) : t_i).$$

Therefore, (24) and the previous inequality imply:

$$E_{t_{-i}}\big(u_i(s(t), t) : t_i\big) \geq E_{t_{-i}}(U_i(t) : t_i). \tag{25}$$

Next, we construct a PMM which implements the outcome function s. Consider the following PMM defined by $(s, \{\tau_i\}^m_{i=1}, \{r_i\}^m_{i=1}, \{p^j_i\}^m_{i,j=1})$, where $\tau_i(t_i) =$

$\{t_i\}$, $r_i(t) = \{s_i(t)\}$, and $p_i^j(t_{-j}) = \{\bar{a}_i^j\}$. That is, the information partition is the full information partition, the equilibrium restrictions are singletons corresponding to s, and if player j does not participate, player i's action space is restricted to be the singleton $\{\bar{a}_i^j\}$. Since $p_i^j(t_{-j}) = \{\bar{a}_i^j\}$, then $s_i^j(t_{-j})$ must be $\bar{a}_i^j$. We have to show that (7), (8), and (9) are satisfied. Since $p_i^j(t_{-j})$ is singleton for all i, j, and t_{-j}, (9) is satisfied. Since $r_i(t_i', t_{-i}) = \{s_i(t_i', t_{-i})\}$, (7) coincides with (23). Since $\tau_i(t_i) = \{t_i\}$ and $p_i^j(t_{-j}) = \{\bar{a}_i^j\}$, (8) can be rewritten as

$$E_{t_{-i}}\big(u_i(s(t), t)|t_i\big) \geq E_{t_{-i}}\Big(\max_{a_i \in A_i} u_i(a_i, \bar{a}_{-i}^i, t)|t_i\Big).$$

By parts (i) and (ii) of Assumption 1, the right-hand side of the previous inequality is $E_{t_{-i}}(U_i(t)|t_i)$. Hence, this inequality is just (25), which is indeed satisfied. *Q.E.D.*

Notice that the statement of the proposition holds for any refinement. This is because the outcome function s can be implemented by a PMM in which all the restrictions on the action spaces of the players are singletons. That is, even off the equilibrium path, players do not make strategic choices at the second stage.

6.4. *Complete Information Environment*

In this section, we characterize the set of pure-strategy subgame perfect Nash equilibrium (SPNE) in our model if the players do not possess private information. We prove a pure-strategy folk theorem for this environment. That is, we show that for each player there exists a value such that an outcome function is implementable as a SPNE if and only if the payoff of each player is larger than his value.

Define the value for player i as

$$\underline{u}_i = \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} u_i(a_i, a_{-i}).$$

We refer to $\underline{u}_i$ as the pure minmax value of player i.

THEOREM 2: *The action profile $a^* = (a_1^*, \dots, a_m^*) \in A$ is supportable as a pure-strategy SPNE outcome in the contracting game if and only if $u_i(a^*) \geq \underline{u}_i$ for each i.*

PROOF: Suppose first that $u_i(a^*) \geq \underline{u}_i$ for all $i \in \{1, \dots, m\}$. We construct a PMM which implements a^* as a SPNE. For each j, let us fix an action a_i^j for player i ($i \neq j$) such that

$$(a_1^j, \dots, a_m^j) \in \arg\min_{a_{-j} \in A_{-j}} \max_{a_j \in A_j} u_j(a_j, a_{-j})$$

and let

$$\underline{a}_j \in \arg\max u_j(a_j, a^j_{-j}).$$

For all j and i $(i \neq j)$, define $r_i = \{a^*_i\}$, $p^j_i = \{a^j_i\}$, and $s^j_i(f_j) = a^j_i$ for all $f_j \in F_j$. Notice that $s^j_i(f_j) \in p^j_i$ and the restriction imposed by subgame perfection is irrelevant because the action space of each player is restricted to a single action followed by any deviation. By Theorem 1, to prove that these strategies constitute a Nash equilibrium, we have to show that both (7) and (8) are satisfied. Notice that since players have no private information, (7) boils down to

$$u_i(a^*) \geq \max_{a_i \in r_i} u_i(a_i, a^*_{-i}).$$

Since $r_i = \{a^*_i\}$, this inequality is obviously satisfied. Again, since players have no types, (8) can be rewritten as

$$u_i(a^*) \geq \max_{f_i \in F_i} \max_{a_i \in f_i} u_i(a_i, s^i_{-i}(f_i)) = \max_{f_i \in F_i} \max_{a_i \in f_i} u_i(a_i, a^i_{-i}),$$

where the equality follows from $s^i_{-i}(f_i) = a^i_{-i}$. Notice that the right-hand side is just $\underline{u}_i$ and, hence, this inequality is indeed satisfied.

Suppose now that a PMM implements a^*. Then, by Theorem 1, there exist on-equilibrium restrictions $r_i \in 2^{A_i} \setminus \{\emptyset\}$, off-equilibrium restrictions, $p^j_i \in 2^{A_i} \setminus \{\emptyset\}$, and off-equilibrium strategies $s^j_i : F_j \to A_i$, $s^j_i(f_j) \in p^j_i$, such that (7) and (8) are satisfied. Hence,

$$u_i(a^*) \geq \max_{f_i \in F_i} \max_{a_i \in f_i} u_i(a, s^i_{-i}(f_i)) \geq \max_{a_i \in A_i} u_i(a_i, s^i_{-i}(A_i)) \geq \underline{u}_i.$$

The first inequality is just (8), the second inequality follows from setting $f_i = A_i$, and the third inequality follows from the definition of $\underline{u}_i$. We can conclude that (8) can only be satisfied if $u_i(a^*) \geq \underline{u}_i$ for all i. *Q.E.D.*

One of the implications of Theorem 2 is that for the complete information case, the set of outcome functions that are implementable by a centralized mechanism designer is identical to the set of outcome functions that can be supported by equilibrium in the contracting game. As we have shown (by example) that the same is not generally true for games of incomplete information, this result serves to highlight one of the uses of our characterization theorem.

7. CONCLUSION

This paper shows how the contracts on contracts approach can be extended to environments with incomplete information. Definable contracts constitute the largest class of arithmetic contracts which can be written as a finite text in a first-order language. In this sense, definable contracts embed most other interesting classes of feasible contracts as subsets.

In contrast to the complete information case, we show that the folk theorem does not generally hold in the following sense. A centralized mechanism designer can implement outcome functions that cannot be supported as equilibrium with contractible contracts. This limitation is not a consequence of the set of feasible contracts, but rather of the fact that public contracts reveal information about nondeviators' types. The restriction to definable contracts allows us to provide a complete characterization of equilibrium and to prove this result. One of the results we provide as a part of our main theorem illustrates the role that punishments play in a static contracting environment.

We emphasize that this paper does not intend to do mechanism design and that restricting attention to our two-stage game is with loss of generality. For example, if we allow further communication after the contracting stage, the set of implementable outcome functions becomes larger. Similarly, allowing players to offer contracts and take actions sequentially leads to a different characterization theorem.

APPENDIX

A.1. *Refinement*

This section shows how to extend our characterization theorems to more general equilibrium refinements. A refinement is a restriction on the strategy rules of the different types of nondeviating players in the second stage following a deviation by some player i in the first stage. In the main text, we only ruled out the possibility that players choose strictly dominated actions. Strict dominance is a notion that depends on feasible sets of actions and payoff functions, but not on the game in which these are embedded. Generally, refinements impose restrictions that can depend on the sets to which the players are constrained in the second stage, the information that has been revealed by the nondeviators' first period contracts, and the outcome that would have prevailed had there not been any deviation. Informally, the sets to which players are constrained when choosing their second period actions are used to determine whether or not some actions are dominated for certain player types. The information conveyed by first period play is used for refinements like perfect Bayesian equilibrium that require the use of Bayes rule for making inferences about nondeviating players. The original equilibrium outcome is used in refinements like the "intuitive criterion," which restrict beliefs that nondeviators can have about the deviating player based on what he might have expected to gain by deviating.

Formally, let $s: T \to A$ denote the equilibrium outcome function, that is, $s(t) = \alpha^*(t, \gamma^*(t))$ for each t. Let τ_i be the partition of T_i generated by i's equilibrium strategy or, in the context of our contract game, $\tau_i(t_i) = \{t'_i : \gamma^*_i(t'_i) =$

$\gamma_i^*(t_i)\}$. Let $\mathcal{F}^i$ be a τ_{-i}-measurable correspondence from T_{-i} into A representing the sets to which the players are constrained when choosing their actions following a deviation by player i. If player i deviates to contract γ_i, then this correspondence is $\mathcal{F}^i(t_{-i}) = S(\gamma_i, \gamma_{-i}^*(t_{-i}))$.

A *refinement* specifies for every deviator i, every $\mathcal{F}^i$ measurable with respect to some information partition τ_{-i}, and every status quo outcome function s, a nonempty set of action profiles for the nondeviators that the refinement allows for each profile of their types. Let $\mathcal{R}_i(s, \mathcal{F}^i, \tau, t_{-i}) \in 2^{A_{-i}} \setminus \{\emptyset\}$ describe this correspondence.[18] If an equilibrium has the nondeviators using strategy rules $\alpha_{-i}^*(t_{-i}, (\gamma_i, \gamma_{-i}^*(t_{-i})))$ in response to a deviation to contract γ_i by player i, then the equilibrium satisfies the refinement if

$$\alpha_{-i}^*\big(t_{-i}, (\gamma_i, \gamma_{-i}^*(t_{-i}))\big) \in \mathcal{R}_i\big(s, S(\gamma_i, \gamma_{-i}^*(t_{-i})), \tau, t_{-i}\big)$$

for each t_{-i}.

The properties to be built into the refinement correspondence $\mathcal{R}_i$ are going to depend on the particular application. For example, if types are independently distributed, then perfect Bayesian equilibrium is well defined and $\mathcal{R}_i(s, S(\gamma_i, \gamma_{-i}^*(t_{-i})), \tau, t_{-i})$ would, for each t_{-i}, consist of all action profiles for the nondeviators that constitute actions these types would jointly take in some Bayesian equilibrium of the game with action spaces $S(\gamma_i, \gamma_{-i}^*(t_{-i}))$ and beliefs given by posterior beliefs conditional on nondeviators types lying in $\tau_{-i}(t_{-i})$.[19] In a game of complete information with equilibrium contracts (c_i^*, c_{-i}^*) and equilibrium outcome a^*, a deviation to c' by player i supports the collection of action profiles $\mathcal{F}^i = S(c', c_{-i}^*)$, which are just the actions to which players are constrained in the second stage by their first period contracts. The subgame perfection refinement of Nash equilibrium would specify $\mathcal{R}_i(s, \mathcal{F}^i) \subset \mathcal{N}(\mathcal{F}^i)$, where $\mathcal{N}(\mathcal{F}^i)$ is just the set of action profiles $a_{-i} \in A_{-i}$ for which there exists an action $a_i \in S_{-i}(c', c_{-i}^*)$ such that (a_i, a_{-i}) constitutes a Nash equilibrium of the game with action spaces $S(c', c_{-i}^*)$.

Finally, if we simply want to describe Bayesian equilibrium, then we could do so by having $\mathcal{R}_i(s, \mathcal{F}^i, \tau, t_{-i}) = A_{-i}$. We refer to the collection of restrictions $\mathcal{R} = \{\mathcal{R}_i\}_{i=1,\dots,m}$ as a *refinement.*

Fix a refinement $\mathcal{R}$. Let τ be the information partition induced by the equilibrium strategies γ^* (that is, $\tau(t) = \{t' \in T : \gamma^*(t') = \gamma^*(t)\}$). We say that

[18] It is reasonable to require $\mathcal{R}_i$ to be measurable with respect to τ_{-i}, but we do not need this property for our formalism.

[19] Perfect Bayesian equilibrium does not have a well accepted definition when types are correlated. To see why, observe that when player i deviates, nondeviators have to make *some* inference about his type. The on path choices of the nondeviators reveal their types to be in some subset. The distribution of nondeviators' types within this subset depends on the deviating player's type. So either inferences about nondeviating players depend on actions of the deviator or the common prior assumption has to be abandoned.

(γ^*, α^*) is an $\mathcal{R}$ equilibrium of the contracting game if (2) holds and, in addition,

$$\alpha^*_{-i}(t_{-i}, \bar{\alpha}) \in \mathcal{R}_i\big(\alpha^*(\cdot, \gamma^*(\cdot)), S(\gamma_i, \gamma^*_{-i}(t_{-i})), \tau, t_{-i}\big)$$

for every i and $\gamma_i \in C_i$ (recall the notation $\bar{\alpha} = (\alpha_i, \alpha^*_{-i})$).

Recall from Section 4 that an outcome function is implementable by a PMM if and only if (7), (8), and (9) hold. Using the more general refinement described above, we only have to modify (9) as follows. For every i, $t_{-i} \in T_{-i}$, and $f_i \in F_i$,

$$s^i(t_{-i}, f_i) \in \mathcal{R}_i(s, f_i \times p^i_{-i}, \tau, t_{-i}). \tag{26}$$

Then an outcome function is $\mathcal{R}$-implementable by a public message mechanism if and only if (7), (8), and (9) hold.

Finally, the statement of Theorem 1 is valid even if $\mathcal{R}$ denotes restrictions imposed by a more general refinement concept. The proof is essentially identical to the one in Section 4.

A.2. *Definability*

Our goal here is to provide formal definitions for *arithmetic statement* and arithmetic statements with free variables. We define statements for any first-order logic and explain what is specific about number theory.

Each formal language has a set of symbols. The symbols of a first-order language are divided into two disjoint sets: the logic symbols, and the nonlogic symbols. The logic symbols include: (,), ∀, ∃, ¬, =, and infinitely many *variable symbols* $x_0, x_1, \ldots$. The nonlogic symbols include function symbols and relation symbols.

DEFINITION 4: $t = \langle F, R, \tau \rangle$ is a similarity type, where F is a set of function symbols, R is a set of relation symbols, and $\tau : F \cup R \to \mathbb{N}$ such that $\tau(r) > 0$ if $r \in R$.

The function τ tells how many variables the functions and the relations have. If $\tau(f) = 0$, then f is referred to as a *constant symbol*.

EXAMPLE: One of the similarity types corresponding to the Peano arithmetics, denoted by $q = \langle F, R, \tau \rangle$, is $F = \{0, 1, +, *\}$, $R = \{<\}$, $\tau(0) = \tau(1) = 0$, $\tau(+) = \tau(*) = \tau(<) = 2$. Notice that the 0 and the 1 are considered as functions with zero variables, that is, constant symbols. (We point out that the similarity type of arithmetics can be defined without the relation $<$. This relation can be then defined recursively.)

DEFINITION 5: Let $t = \langle F, R, \tau \rangle$ be a similarity type. Then the set of *expressions* of type t, denoted by $K(t)$, is the smallest set for which the following statements hold:

(i) $x \in K(t)$ for all variable symbols.

(ii) For all $f \in F$, if $\tau(f) = 0$, then $f \in K(t)$.

(iii) For all $f \in F$, if $\tau(f) = n$ and $k_1, \dots, k_n \in K(t)$, then $f(k_1, \dots, k_n) \in K(t)$.

Suppose that $t = q$. Then the following string of symbols are expressions in arithmetics: x, 0, 1, $x + 1$, $((x + 1) * (y + 1) + 1)$, and so forth.

Finally, we are ready to define the set of statements that correspond to a similarity type.

DEFINITION 6: Let $t = \langle F, R, \tau \rangle$ be a similarity type. Then the set of *statements* of type t, denoted by $F(t)$, is the smallest set for which the following statements hold:

(i) If $r \in R$, $\tau(r) = n$, and $k_1, \dots, k_n \in K(t)$, then $r(k_1, \dots, k_n) \in F(t)$.

(ii) If $k_1, k_2 \in K(t)$, then $k_1 = k_2 \in F(t)$.

(iii) If $\phi, \eta \in F(t)$, then $(\phi) \vee (\eta) \in F(t)$, $\neg(\phi) \in F(t)$, and $\exists x(\phi) \in F(t)$.

The set of arithmetic statements is defined according to the previous definition with $t = q$. Then the following string of symbols are statements in arithmetics: $x = y$, $\neg \exists x \exists y\ (y = x + 1)$, and so forth.

For each statement, one can enumerate the number of different variable symbols appearing in the statement. A variable is called a *free variable* in a statement if it does not appear right behind a quantifier. For example, the statement $\neg \exists x \exists y\ ((y = x + 1) \vee (z = 1))$ has three variable symbols: x, y, and z. However, both the x and the y appear behind a quantifier. Hence, the only free variable of this statement is z.

A.3. *Proofs*

PROOF OF LEMMA 2: (i) Suppose that $A = \{n^1, \dots, n^q\}$, where $n^i = (n^i_1, \dots, n^i_k) \in \mathbb{N}^k$ for $i = 1, \dots, q$. The characteristic function of the set A can be defined by the following statement in $k + 1$ free variables:

$$\phi(x_1, \dots, x_k, y) \equiv \left(\bigvee_{i=1}^{q} \left(\bigwedge_{j=1}^{k} (x_j = n^i_j) \right) \wedge (y = 1) \right) \vee \left(\neg \left(\bigvee_{i=1}^{q} \left(\bigwedge_{j=1}^{k} (x_j = n^i_j) \right) \right) \wedge (y = 0) \right).$$

(ii) Suppose that φ_A defines χ_A and φ_B defines χ_B, and let x denote $(x_1, \dots, x_k)$. Then $\chi_{A\cap B}$, $\chi_{A\cup B}$, and $\chi_{A\smallsetminus B}$ are defined by

$$\begin{aligned}
&[\varphi_A(x,1) \wedge \varphi_B(x,1) \wedge (y=1)]\\
&\quad \vee \big[\neg(\varphi_A(x,1) \wedge \varphi_B(x,1)) \wedge (y=0)\big],\\
&\big[(\varphi_A(x,1) \vee \varphi_B(x,1)) \wedge (y=1)\big]\\
&\quad \vee \big[\neg(\varphi_A(x,1) \vee \varphi_B(x,1)) \wedge (y=0)\big],\\
&[\varphi_A(x,1) \wedge \varphi_B(x,0) \wedge (y=1)]\\
&\quad \vee \big[\neg(\varphi_A(x,1) \wedge \varphi_B(x,0)) \wedge (y=0)\big],
\end{aligned}$$

respectively.

(iii) Let ϕ_{A_i} denote a statement defining χ_{A_i} for each i. Furthermore, let $B_i = \{b^{i1}, \dots, b^{iq_i}\}$. Then the following statement obviously defines the function f:

$$\begin{aligned}
&\left(\bigvee_{i=1}^{m}\left(\phi_{A_i}(x_1,\dots,x_k,1) \wedge \left(\bigvee_{l=1}^{q_i}(y=b^{il})\right)\right)\right)\\
&\quad \vee \left(\bigwedge_{i=1}^{m}\phi_{A_i}(x_1,\dots,x_k,0) \wedge \left(\bigvee_{l=1}^{q_{m+1}}\big(y=b^{(m+1)l}\big)\right)\right).
\end{aligned}$$

(iv) Let ϕ_D define the characteristic function of $D = \bigtimes_{i=1}^{k} D_i$, let ϕ_{D_j} define the characteristic function of D_j, and let φ_n define the characteristic function of $g(n)$ for $n \in D$. The statements ϕ_D, ϕ_{D_j}, and φ_n exist because of part (i) of this lemma. Furthermore, let $A_i = \{n \in \mathbb{N}^k : \{i : n_i \notin D_i\} = i\}$ for $i \in \{1, \dots, k\}$. The characteristic function of A_i for $i \in \{1, \dots, k\}$ is defined by

$$\begin{aligned}
&\psi_i(x_1,\dots,x_k,z)\\
&\quad = \left(\left(\bigwedge_{\substack{j=1\\ j\neq i}}^{k}\phi_{D_j}(x_j,1)\right) \wedge \phi_{D_i}(x_i,0) \wedge (z=1)\right)\\
&\qquad \vee \left(\neg\left(\left(\bigwedge_{\substack{j=1\\ j\neq i}}^{k}\phi_{D_j}(x_j,1)\right) \wedge \phi_{D_i}(x_i,0)\right) \wedge (z=0)\right).
\end{aligned}$$

Finally, let ξ_j define the characteristic function of B_j for $j = 1, \dots, k+1$. Then the function f is defined by the statement

$$\begin{aligned}
&\left(\phi_D(x,1) \wedge \left(\bigvee_{n\in D}(\varphi_n(y,1) \wedge (x=n))\right)\right) \\
&\bigvee_{i=1}^{k}(\phi_D(x,0) \wedge \psi_i(x,1) \wedge \xi_i(y,1)) \\
&\vee\left(\neg\left(\left(\phi_D(x,1) \wedge \left(\bigvee_{n\in D}(\varphi_n(y,1) \wedge (x=n))\right)\right)\right.\right. \\
&\left.\left.\bigvee_{i=1}^{k}(\phi_D(x,0) \wedge \psi_i(x,1) \wedge \xi_i(y,1))\right) \wedge \xi_{k+1}(y,1)\right).
\end{aligned}$$

Q.E.D.

LEMMA 5: *Let $N_i \geq 1$ for all $i \in \{1, \dots, m\}$. Suppose that $h_j^{n_j} : \mathbb{N}^{\times_j^m N_j} \to 2^{\mathbb{N}}$, is definable and that $|h_j^{n_j}(q)| = 1$ for all $j \in \{1, \dots, m\}$, $n_j \in \mathcal{N}_j$ and q. Suppose that $\bar{r} : \mathcal{N} \to 2^{\mathbb{N}} \setminus \{\emptyset\}$ and that $\bar{p}_i^j : \mathcal{N}_{-j} \to 2^{\mathbb{N}} \setminus \{\emptyset\}$ for each $i, j \in \{1, \dots, m\}$ ($i \neq j$) such that $\bar{r}(n)$ and $\bar{p}_i^j(n_{-i})$ are finite for each $n \in \mathcal{N}$, $n_{-i} \in \mathcal{N}_{-i}$, and i. In addition, let ($\{\emptyset\} \neq$) $\bar{A}_i \subset \mathbb{N}$ be a finite set. For all $(q_1, \dots, q_m) \in \mathbb{N}^{\times_i^m N_i}$ (where $q_i = (q_i^1, \dots, q_i^{N_i}) \in \mathbb{N}^{N_i}$) and $l = (l_1, \dots, l_m) \in \mathbb{N}^m$, define*

$$(27) \qquad f^{i,n_i}(q,l)$$

$$= \begin{cases} \bar{r}(n_i, n_{-i}), & \text{if } \forall k \neq i \ h_k^{n_k}(q) = l_k \text{ and} \\ & h_k^{n_k}(q) \neq h_k^{n'_k}(q) \text{ if } n_k \neq n'_k, \\ \bar{p}_i^j(n_i, n_{-ij}), & \text{if } \forall k \neq i, j \ h_k^{n_k}(q) = l_k, \text{ and } h_k^{n_k}(q) \neq h_k^{n'_k}(q) \\ & \text{if } n_k \neq n'_k \text{ and } \nexists n_j \text{ s.t. } h_j^{n_j}(q) = l_j, \\ \bar{A}_i, & \text{otherwise.} \end{cases}$$

Then the function $f^{i,n_i} : \mathbb{N}^{\times_i^m N_i + m} \to 2^{\mathbb{N}}$ is definable.

PROOF: Let $\mathcal{N}_i$ denote $\{1, \dots, N_i\}$ and let $\mathcal{N} = \times_{i=1}^m \mathcal{N}_i$. Let $\varphi_i^{n_i}(x, y_i)$ define $h_i^{n_i}$, where x is an $|\mathcal{N}|$-dimensional vector of variable symbols and y_i is a variable symbol. Let $\theta_r^n(z, v)$ define the characteristic function of the set $\bar{r}(n)$, let $\theta^n_{p_i^j}(z, v)$ define the characteristic function of $\bar{p}_i^j$, and let $\theta_{A_i}(z, v)$ define the characteristic function of $\bar{A}_i$. The letters z and v are variable symbols.

For all $n = (n_1, \dots, n_m) \in \mathbb{N}^m$, $n_k \in \{1, \dots, N_k\}$, define

$$\psi^n(x, y) \equiv \bigwedge_{k \neq i}\left(\varphi_k^{n_k}(x, y_k) \bigwedge_{\substack{n'_k \in \{1, \dots, N_k\} \\ n'_k \neq n_k}} (\neg\varphi_k^{n'_k}(x, y_k))\right).$$

Notice that $\psi^n(q,l)$ is true if and only if $h_k^{n_k}(q)=l_k$ for all $k\neq i$ and $h_k^{n_k}(q)\neq h_k^{n'_k}(q)$ whenever $n_k\neq n'_k$. That is, ψ^n corresponds to the condition in the first line of (27).

Similarly,

$$\begin{aligned}&\psi_j^{n_{-j}}(x,y_{-j})\\&\quad\equiv\bigwedge_{k\neq i,j}\Biggl(\varphi_k^{n_k}(x,y_k)\bigwedge_{\substack{n'_k\in\mathcal{N}_k\\ n'_k\neq n_k}}\bigl(\neg\varphi_k^{n'_k}(x,y_k)\bigr)\Biggr)\bigwedge_{n_j\in\mathcal{N}_j}\bigl(\neg\varphi_j^{n_j}(x,y_j)\bigr).\end{aligned}$$

Notice that $\psi_j^{n_{-j}}(q,l_{-j})$ is true if and only if $h_k^{n_k}(q)=l_k$, $h_k^{n_k}(q)\neq h_k^{n'_k}(q)$ whenever $n_k\neq n'_k$ for all $k\neq i,j$, and there is no n_j $(\in\mathcal{N}_j)$ s.t. $h_j^{n_j}(q)=l_j$. That is, $\psi_j^{n_{-j}}(x,y_{-j})$ corresponds to the second line of (27).

We are ready to construct a statement which defines the f^{i,n_i}:

$$\begin{aligned}&\bigvee_{n\in\mathcal{N}}[\psi^n(x,y)\wedge\theta_r^n(z,1)]\bigvee_{j,n_{-j}\in\mathcal{N}_{-j}}\bigl[\psi_j^n(x,y_{-j})\wedge\theta^n_{p_i^j}(z,1)\bigr]\\&\quad\vee\Biggl(\neg\Biggl(\bigvee_{n\in\mathcal{N}}[\psi^n(x,y)\wedge\theta_r^n(z,1)]\bigvee_{j,n_{-j}\in\mathcal{N}_{-j}}\bigl[\psi_j^n(x,y_{-j})\wedge\theta^n_{p_i^j}(z,1)\bigr]\Biggr)\\&\quad\wedge\theta_{A_i}(z,1)\Biggr).\end{aligned}$$

Q.E.D.

LEMMA 6: *The function described by* (16) *is a definable function with* $|\times_{i=1}^m N_i|$ *free variables*.

PROOF: By Definition 3, we have to show that $f^{i,n_i}:\mathbb{N}^{\times_1^m N_j}\times\mathbb{N}^m\rightarrow 2^{\mathbb{N}}$ is definable, where $f^{i,n_i}(q,l)$ is defined by

$$\begin{cases}[r_i(n_i,n_{-i})], & \text{if } \forall k\neq i\ h_k^{n_k}(q)=l_k \text{ and}\\ & h_k^{n_k}(q)\neq h_k^{n'_k}(q) \text{ if } n_k\neq n'_k,\\ [p_i^j(n_i,n_{-ij})], & \text{if } \forall k\neq i,j\ h_k^{n_k}(q)=l_k,\ \text{and } h_k^{n_k}(q)\neq h_k^{n'_k}(q)\\ & \text{if } n_k\neq n'_k \text{ and } \nexists n_j \text{ s.t. } h_j^{n_j}(q)=l_j,\\ [A_i], & \text{otherwise}\end{cases}$$

for all $l=(l_1,\dots,l_m)\in\mathbb{N}^m$, $q_j=(q_j^1,\dots,q_j^{N_j})\in\mathbb{N}^{N_j}$ and $q=(q_1,\dots,q_m)$.

Notice that $|h_k^{n_k}(q)|=1$ for all q and that $h_k^{n_k}$ is definable by Lemma 1. Define $\bar{r}_i(n_i,n_{-i})$ to be $[r_i(n_i,n_{-i})]$ and $\bar{p}_i^j(n_i,n_{-ij})$ to be $[p_i^j(n_i,n_{-ij})]$. Finally let

$\bar{A}_i = [A_i]$. Notice that $\bar{r}_i(n_i, n_{-i})$, $\bar{p}_i^j(n_i, n_{-ij})$, and $\bar{A}_i$ are finite sets of $\mathbb{N}$. Hence, the statement of the lemma follows from Lemma 5. *Q.E.D.*

LEMMA 7: *Suppose that* $g : \mathbb{N}^k \to 2^B \setminus \{\emptyset\}$ *is a definable function where* B ($\subset \mathbb{N}^q$) *is finite. Let* $f : 2^B \to 2^D$ *an arbitrary function where* D ($\subset \mathbb{N}^l$) *is finite. Then* $f \circ g : \mathbb{N}^k \to 2^D$ *is a definable function.*

PROOF: Suppose that the statement ϕ in $k + q$ free variables defines the function g. That is, $\phi(a_1, \dots, a_k, b_1, \dots, b_q)$ is true if and only if $(b_1, \dots, b_q) \in g(a_1, \dots, a_k)$. First, we construct a statement φ in $k + q|B|$ free variables such that

$$\text{(28)} \qquad \varphi(a, b_1, \dots, b_{|B|}) \text{ is true} \quad \Leftrightarrow \quad g(a) = \{b_1, \dots, b_{|B|}\},$$

where b_i is a q-dimensional integer vector. (Notice that we do not assume that $b_i \neq b_j$ if $i \neq j$ in the previous equivalence.) To this end, let $x = (x^1, \dots, x^k)$, $y_i = (y_i^1, \dots, y_i^q)$ for all $i \in \{1, \dots, |B|\}$, and $z = (z^1, \dots, z^q)$. Define φ as

$$\varphi(x, y_1, \dots, y_{|B|}) = \left[\bigwedge_{i=1}^{|B|} \phi(x, y_i)\right] \wedge \left[\nexists z \left(\bigwedge_{i=1}^{|B|} z \neq y_i\right) \wedge \phi(x, y_i)\right].$$

This φ obviously satisfies (28). Next, we construct a statement ψ in $ql|B||D|$ free variables such that

$$\text{(29)} \qquad \psi(b_1, \dots, b_{|B|}, d_1, \dots, d_{|D|}) \text{ is true}$$
$$\Leftrightarrow \quad f(\{b_1, \dots, b_{|B|}\}) = \{d_1, \dots, d_{|D|}\},$$

where b_i is a q-dimensional integer vector for all i and d_i is an l-dimensional integer vector for all i. Suppose that $B_i \subset B$, $B_i = \{b_i^1, \dots, b_i^{|B|}\}$, and $f(B_i) = D_i = \{d_i^1, \dots, d_i^{|D|}\}$. Consider

$$\psi_{B_i}(y_1, \dots, y_{|B|}, z_1, \dots, z_{|D|}) = \left(\bigvee_{n=1}^{|B|}(y_j = b_i^n)\right) \bigwedge_{n=1}^{|B|} \left(\bigvee_{j=1}^{|B|}(y_j = b_i^n)\right)$$
$$\wedge \left(\bigvee_{n=1}^{|D|}(z_j = d_i^n)\right) \bigwedge_{n=1}^{|D|} \left(\bigvee_{j=1}^{|D|}(z_j = d_i^n)\right).$$

The first part in the first line says that $y_j \in \{b_1, \dots, b_{|B|}\}$ and the second part of the first line requires that for all b_i^n, there exists a y_j such that $y_j = b_i^n$. Similarly, the first part in the second line says that $z_j \in \{d_1, \dots, d_{|D|}\}$ and the second part of the second line requires that for all d_i^n, there exists a z_j

such that $z_j = d_i^n$. Obviously, $\psi_{B_i}(b^1, \dots, b^{|B|}, d^1, \dots, d^{|D|})$ is true if and only if $\{b^1, \dots, b^{|B|}\} = \{b_i^1, \dots, b_i^{|B|}\}$ and $\{d^1, \dots, d^{|D|}\} = \{d_i^1, \dots, d_i^{|D|}\}$. Let $2^B \setminus \{\emptyset\} = \{B_1, \dots, B_i, \dots, B_{2^{|B|}-1}\}$. We are ready to define ψ:

$$\psi(y_1, \dots, y_{|B|}, z_1, \dots, z_{|D|}) = \bigvee_{i=1}^{2^{|B|}-1} \psi_{B_i}(y_1, \dots, y_{|B|}, z_1, \dots, z_{|D|}).$$

This statement ψ obviously satisfies (29).

Finally, we are ready to construct the statement ζ which defines $g \circ f$.

$$\begin{aligned}\zeta(x, z) = {} & \exists y_1, \dots, y_{|B|}, z_1, \dots, z_{|D|} \\ & \big(\varphi(x, y_1, \dots, y_{|B|}) \wedge \psi(y_1, \dots, y_{|B|}, z_1, \dots, z_{|D|}) \\ & \wedge (z = z_1)\big). \end{aligned}$$

Q.E.D.

LEMMA 8: *Suppose that $f : \mathbb{N}^k \to 2^{\mathbb{N}}$ is definable, $g : \mathbb{N} \to 2^{\mathbb{N}} \setminus \{\emptyset\}$ is definable, and $|g(n)| = 1$ for all n. For each $n \in \mathbb{N}$, define*

$$h(n_1, \dots, n_k) = f(g(n_1), n_{-1}).$$

Then the function $h : \mathbb{N}^k \to 2^{\mathbb{N}}$ is definable.

PROOF: Suppose that $\theta(x, y)$ defines f, where $x = (x_1, \dots, x_k)$ is a vector of variable symbols and y is a variable symbol. Suppose that φ defines $g(z, v)$, where both z and v are variable symbols. Then the following statement ϑ obviously defines h:

$$\vartheta(x, y) \equiv \theta(z, x_{-1}, y) \wedge \varphi(z, x_1).$$

Q.E.D.

LEMMA 9: *Let $(N_1, \dots, N_m) \in \mathbb{N}^m$ $(N_i \geq 1)$ and $\mathcal{N}_i = \{1, \dots, N_i\}$. Suppose that $h_{n_{-i}} : \mathbb{N} \to 2^{\mathbb{N}}$ is definable for all $n_{-i} \in \mathcal{N}_{-i}$. Let $\{\emptyset\} \neq \bar{A}_i \subset \mathbb{N}$ be a finite set. Let $g = (g_1, \dots, g_{m-1}) : \mathcal{N}_{-i} \to \mathbb{N}^{m-1}$ be an injection. For each $l \in \mathbb{N}$, and $q = (q_1, \dots, q_m) \in \mathbb{N}^m$, define $f(l, q)$ as*

$$f(l, q_i, q_{-i}) = \begin{cases} h_{n_{-i}}(l), & \text{if } q_{-i} = g(n_{-i}), \\ \bar{A}_i, & \text{otherwise.} \end{cases}$$

Then the function f is definable.

PROOF: Let $\theta_{n_{-i}}(v, y)$ be a statement which defines $h_{n_{-i}}$ and let $\theta_{A_i}(y, z)$ denote a statement which defines the characteristic function of $\bar{A}_i$. The letters y, z, and v are variable symbols. Let $x = (x_1, \dots, x_m)$ be a vector of variable symbols. Then the following statement ϑ obviously defines f:

$$\begin{aligned}
&\vartheta(v, x, y) \\
&\quad = \bigvee_{n_{-i}\in\mathcal{N}_{-i}} \big(\theta_{n_{-i}}(v, y) \wedge (g(n_{-i}) = x_{-i})\big) \\
&\qquad \vee \Bigg(\neg\bigg(\bigvee_{n_{-i}\in\mathcal{N}_{-i}} \big(\theta_{n_{-i}}(v, y) \wedge (g(n_{-i}) = x_{-i})\big)\bigg) \wedge \theta_{A_i}(y, 1)\Bigg).
\end{aligned}$$

Q.E.D.

PROOF OF LEMMA 3: By Lemma 4, we only have to show that (20) holds for all $i \in \{1, \dots, m\}$ and any contract profile $\{c_j^{n_j}\}_{j,\, n_j}$. Suppose by contradiction that there exists an $i \in \{1, \dots, m\}$ and a contract profile $\{c_j^{n_j}\}_{j,\, n_j}$ such that $\bigcap_{(A_i^{n_{-i}})_{n_{-i}}} S((A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}}) = \{\emptyset\}$. Then, for all $(A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \in (2^{A_{-i}})^{|\mathcal{N}_{-i}|}$, there exists an $(A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \in (2^{A_i})^{|\mathcal{N}_{-i}|}$ such that $(A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \notin S((A_i^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}})$. Let us fix a function $f = (f_{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} : (2^{A_{-i}})^{|\mathcal{N}_{-i}|} \to (2^{A_i})^{|\mathcal{N}_{-i}|}$ such that for all $(A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \in (2^{A_{-i}})^{|\mathcal{N}_{-i}|}$,

$$(A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}} \notin S\big(f\big((A_{-i}^{n_{-i}})_{n_{-i}\in\mathcal{N}_{-i}}\big)\big).$$

Let $f^{c_{-i}^{n_{-i}}}$ denote the function defined by $c_{-i}^{n_{-i}}$. Define a function with one free variable, f_x, as

$$\begin{aligned}
&f_x([c_i], [c_{-i}]) \\
&\quad = \begin{cases} \Big[f_{n'_{-i}}\big(\big(\big\langle\big(f^{c_{-i}^{n_{-i}}}\big(\big[\langle x\rangle^{(x)}\big], [c_{-i}^{n_{-i}}]\big)\big)\big\rangle\big)_{n_{-i}\in\mathcal{N}_{-i}}\big)\Big], & \text{if } c_{-i} = c_{-i}^{n'_{-i}}, \\ [A_i], & \text{otherwise.} \end{cases}
\end{aligned} \tag{30}$$

We prove that f_x is a definable function in one free variable; see Lemma 10. Let c_x be an arithmetic statement defining f_x and let γ denote its Gödel code. Then

$$f_\gamma([c_i], [c_{-i}]) = \begin{cases} \big[f_{n'_{-i}}\big((\widetilde{c}_{-i}^{n_{-i}}(c_\gamma, c_{-i}^{n_{-i}}))_{n_{-i}\in\mathcal{N}_{-i}}\big)\big], & \text{if } c_{-i} = c_{-i}^{n'_{-i}}, \\ [A_i], & \text{otherwise.} \end{cases}$$

Notice that

$$(\widetilde{c}_{-i}^{n_{-i}}(c_\gamma, c_{-i}^{n_{-i}}))_{n_{-i}\in\mathcal{N}_{-i}} \in S\big((\widetilde{c}_\gamma(c_\gamma, c_i^{n_i}))_{n_{-i}\in\mathcal{N}_{-i}}\big) \tag{31}$$

by the definition of S. Alternatively,

$$\begin{aligned}
(\widetilde{c}_\gamma(c_\gamma, c_{-i}^{n_{-i}}))_{n_{-i}\in\mathcal{N}_{-i}} &= \big(f_{n_{-i}}\big(\big(\widetilde{c}_{-i}^{n'_{-i}}\big(c_\gamma, c_{-i}^{n'_{-i}}\big)\big)_{n'_{-i}\in\mathcal{N}_{-i}}\big)\big)_{n_{-i}\in\mathcal{N}_{-i}} \\
&= f\big((\widetilde{c}_{-i}^{n_{-i}}(c_\gamma, c_{-i}^{n_{-i}}))_{n_{-i}\in\mathcal{N}_{-i}}\big)
\end{aligned}$$

by the definition of c_γ. Therefore,

$$\big((\widetilde{c}^{\,n_{-i}}_{-i}(c_\gamma, c^{n_{-i}}_{-i}))_{n_{-i}\in\mathcal{N}_{-i}}\big) \notin S\big((\widetilde{c}_\gamma(c_\gamma, c^{n_{-i}}_{-i}))_{n_{-i}\in\mathcal{N}_{-i}}\big) \tag{32}$$

by the definition of f. Notice that (31) and (32) contradict to each other and, hence, (20) holds. *Q.E.D.*

LEMMA 10: *The function defined by* (30) *is a definable function with one free variable.*

PROOF: By Definition 3, we have to prove that $f : N^{m+1} \to 2^N$ is a definable function if

$$f(l, q_1, \dots, q_m) = \begin{cases} \Big[f_{n'_{-i}}\big(\big(\big\langle\big(f^{c^{n_{-i}}_{-i}}\big(\big[\langle l\rangle^{(l)}\big], [c^{n_{-i}}_{-i}]\big)\big)\big\rangle\big)_{n_{-i}\in\mathcal{N}_{-i}}\big)\Big], & \text{if } q_{-i} = [c^{n'_{-i}}_{-i}], \\ [A_i], & \text{otherwise} \end{cases}$$

for all $l \in \mathbb{N}$ and $q = (q_1, \dots, q_m) \in \mathbb{N}^m$. The function $f^{c^{n_{-i}}_{-i}}([\langle l\rangle^{(l)}], q_{-i})$ is definable ($\mathbb{N}^m \to 2^{\mathbb{N}}$) by Lemmas 1 and 8. Define $h_{n'_{-i}}(l)$ for all $l \in \mathbb{N}$ as

$$h_{n'_{-i}}(l) = \Big[f_{n'_{-i}}\big(\big(\big\langle\big(f^{c^{n_{-i}}_{-i}}\big(\big[\langle l\rangle^{(l)}\big], [c^{n_{-i}}_{-i}]\big)\big)\big\rangle\big)_{n_{-i}\in\mathcal{N}_{-i}}\big)\Big].$$

Then the function $h_{n'_{-i}}$ is definable for all $n'_{-i} \in \mathcal{N}_{-i}$ by Lemma 7. Finally, define $g(n_{-i}) = [c^{n_{-i}}_{-i}]$ for all $n_{-i} \in \mathcal{N}_{-i}$ and apply Lemma 9 to conclude that f is definable. *Q.E.D.*

REFERENCES

BAGWELL, K., AND R. W. STAIGER (2001): “Reciprocity, Non-Discrimination and Preferential Agreements in the Multilateral Trading System,” *European Journal of Political Economy*, 17, 281–325. [363]

EPSTEIN, L., AND M. PETERS (1999): “A Revelation Principle for Competing Mechanisms,” *Journal of Economic Theory*, 88, 119–160. [371,372]

FERSHTMAN, C., AND K. JUDD (1987): “Equilibrium Incentives in Oligopoly,” *American Economic Review*, 77, 927–940. [371]

HAN, S. (2006): “Menu Theorems for Bilateral Contracting,” *Journal of Economic Theory*, 131, 157–178. available at http://ideas.repec.org/a/eee/jetheo/v131y2006i1p157-178.html. [371]

KALAI, A. T., E. KALAI, E. LEHRER, AND D. SAMET (2010): “A Commitment Folk Theorem,” *Games and Economic Behavior*, 69, 127–137. [364-366,372]

KATZ, M. (2006): “Observable Contracts as Commitments: Interdependent Contracts and Moral Hazard,” *Journal of Economics and Management Strategy*, 15, 685–706. [371]

MARTIMORT, D., AND H. MOREIRA (2010): “Common Agency and Public Good Provision Under Asymmetric Information,” *Theoretical Economics*, 5, 159–213. [371]

MARTIMORT, D., AND L. STOLE (1998): "Communications Spaces, Equilibria Sets and the Revelation Principle Under Common Agency," Unpublished Manuscript, University of Chicago. [371]

MYERSON, R. (1983): "Mechanism Design by an Informed Principal," *Econometrica*, 51, 1767–1797. [372]

PETERS, M. (2001): "Common Agency and the Revelation Principle," *Econometrica*, 69, 1349–1372. [371]

PETERS, M., AND C. TRONCOSO-VALVERDE (2009): "A Folk Theorem for Competing Mechanisms," Unpublished Manuscript, University of British Columbia. [371]

SALOP, S. (1986): *Practises That Credibly Facilitate Ologopoly Coordination*. Cambridge: MIT Press. [363]

TENNENHOLTZ, M. (2004): "Program Equilibrium," *Games and Economic Behavior*, 49, 363–373. [364-366,372,376]

YAMASHITA, T. (2010): "Mechanism Games With Multiple Principals and Three or More Agents," *Econometrica*, 78, 791–801. [371]

Dept. of Economics, University of British Columbia, 497-1873 East Mall, Vancouver, BC V6T 1Z1, Canada; peters@econ.ubc.ca

and

Dept. of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom; b.szentes@lse.ac.uk.

Manuscript received January, 2009; final revision received April, 2011.