Review of Economic Studies (2005) **72**, 543–566 0034-6527/05/00220543$02.00
© 2005 The Review of Economic Studies Limited

Endogenous Games and Mechanisms: Side Payments Among Players

MATTHEW O. JACKSON and SIMON WILKIE
California Institute of Technology

First version received August 2002; *final version accepted May* 2004 (*Eds.*)

We characterize the outcomes of games when players may make binding offers of strategy contingent side payments before the game is played. This does not always lead to efficient outcomes, despite complete information and costless contracting. The characterizations are illustrated in a series of examples, including voluntary contribution public good games, Cournot and Bertrand oligopoly, principal–agent problems, and commons games, among others.

1. INTRODUCTION

Game theory and mechanism design are powerful tools that have become essential in the modelling of economic interactions. Generally, in modelling interactions from public goods contributions to imperfect competition among firms, the game being played or mechanism being designed is viewed in isolation. That is, we usually treat the game as being fixed from the players' perspective. The analysis of many games viewed in such isolation leads to a prediction of an inefficient outcome, since in many contexts there are externalities present. For instance voluntary public goods contributions games and commons games have well-known free rider problems and equilibria that are Pareto inefficient. Similar results hold for many other games, such as those with imperfect competition or production externalities such as pollution.

In practice, however, we often see players side contracting to improve efficiency. For instance, large donors often match the donations of other donors in contributions games. We see this in public radio and television station fundraising where one donor will agree to donate an amount equal to that donated by other donors in some time period (sometimes even subject to minimum or maximum donations, or subject to the donations exceeding some amount). This practice extends more generally and, for example, many employers offer to match their employees' contributions to any charities. On an intuitive level this type of side contracting can help overcome externalities and reduce inefficiencies. The promise to match donations increases the impact that a donation has and can essentially compensate for the externality—representing the value that the donation would have to others. Similar side contracting appears in the tragedy of the commons games in the form of international fishing and international pollution agreements, where often some promises of side payments are included. Again, the side payments can help promote efficiency by changing the incentives so that each party more fully sees the total impact or value that its actions generate.

While one can see an intuitive role for such side contracting, it is important to fully understand how such side contracting affects the outcome of the game. Which side contracts will agents write, and will the ability of agents to side contract lead to efficiency? These are the central

questions that we address in this paper. There is a widespread belief among economists in the efficiency properties of what may be called "Coasian Contracting". The simple but powerful idea put forth by Coase (1960) says that if property rights are well defined, and bargaining is costless, then rational agents faced with externalities should contract to come to an efficient outcome. Roughly speaking, with fully symmetric information and no transactions costs, agents should be able to come to an agreement that supports an efficient strategy profile as an equilibrium point of the game with side payments.[1] In this paper we hold this reasoning to a careful scrutiny, and find that the issue is surprisingly subtle. Side contracting does not always lead to efficiency even when there are no transactions costs, complete information, and binding contracts. In fact, even if we start with a game that has Pareto efficient Nash equilibria, side contracting on the part of players can change the equilibrium structure so that all equilibria are inefficient!

The perspective that we take here is to view a game as being embedded in a larger game where in a first stage players may engage in side contracting that can effectively rewrite pay-off functions and then play the eventual altered game in the second period. This takes the eventual game that is played to be *endogenous*. In particular, we examine the following scenario: a set of agents are to play a game with known pay-offs. Before playing the game, the agents can make enforceable offers of strategy contingent side payments to each other. So, players can make offers of the sort, "If actions x are played in the game that we are about to play, I will pay you an amount y". The offers that can be made can be contingent on the actions of more than one player and can differ depending on the profile of actions. Offers are publicly observed and legally enforceable, and actions taken in any subsequent play of the game are also observable to any third party such as a court. Such offers modify the net pay-offs in the game and this affects the equilibrium behaviour. From this point of view, the game has become endogenous. We explore how the ability to make such enforceable strategy contingent offers affects the equilibrium pay-offs of the game.

Our main results are a complete characterization of the set of supportable equilibrium pay-offs in endogenous games. We show that the equilibrium outcomes of a game with this costless stage of pre-play promises of side payments need not be efficient. Thus, we cannot rely on endogenous side payments to solve the inefficiency problem. Moreover, side contracting may introduce inefficiency where the equilibrium was efficient without side contracting. Our results provide a complete characterization of the supportable equilibrium outcomes, and how these depend on the structure of the game. Thus, we identify the class of games for which such endogenous side payments will result in efficient equilibria. This class includes some interesting examples such as some specifications of the tragedy of the commons and Bertrand games, but also excludes many interesting examples like voluntary contribution public good games and Cournot games.

In order to preview some of the intuition, and to give an idea of some of the main issues that arise, let us turn to an example.

Example 1. A prisoner's dilemma.

The intuitive argument for how side contracting might support efficient outcomes, and the reasoning of Coase, is roughly as follows. One agent can offer a second agent compensation as

1. One way to view property rights in a game theoretic setting is that the specification of actions embodies the specification of each agent's rights, and so property rights are built into the specification of the game. For instance, consider a classic example where an agent holds property rights and his or her approval is necessary in order for a firm to pollute. One can model this as a very simple game where the agent's actions are "allow" or "not allow", and the pollution only takes place if the agent plays "allow". More complicated interactions lead to richer games. The extension of the classic Coasian view is then that the game augmented with the possibility of side payments will result in an efficient equilibrium outcome.

a function of the second agent's action that effectively reflects any externality that the second agent's action has on the first agent. Essentially, it is as if the first agent says to the second agent "the benefit to me is x if you take action A rather than action B, and the cost to you of taking action A rather than B is only y where $x > y$, and so I will pay you z, where $x \geq z \geq y$, if you play A instead of B". Any z such that $x \geq z \geq y$ will provide sufficient incentives.

The more subtle issue arises when this is put into a richer setting, where more than one agent is taking an action at the same time. Then strategic factors come into play that make the analysis significantly more complicated.

To make things concrete, let us consider an example which has been well studied in the literature.

The pay-offs are represented as follows:

	C	D
C	$2, 2$	$-1, 4$
D	$4, -1$	$0, 0$

This has the classic form of a prisoner's dilemma, with the unique equilibrium being (D, D).

Now let us examine the intuition that efficient play should be supportable if players can make binding offers of action contingent side payments before the game is played. Consider the efficient situation where both players play (C, C). The column player would gain 2 by deviating. This deviation would hurt the row player by 3. So it is in the row player's interest to offer any payment of at least 2 and no more than 3 to the column player contingent on the column player playing C. The only such payment that makes sense from the row player's perspective is a payment of 2, since giving any more is simply a gratuitous transfer to the other player. The same logic works in reverse, so that the column player is willing to make a payment of 2 to the row player contingent on the row player playing C. Taking these two transfers into account, the net pay-offs to the two players looks as follows:

	C	D
C	$2, 2$	$1, 2$
D	$2, 1$	$0, 0$

The action contingent side payments have changed the game so that (C, C) is an equilibrium (and in this example in weakly dominant strategies). This insight is first due to Guttman (1978), and has been extended to a variety of voluntary contribution games and other games with externalities by Danziger and Schnytzer (1991), and Varian (1994*a*), among others.

This, however, is not the end of the story. We can ask whether these particular transfers are part of equilibrium play. For instance, if the column player is offering the row player a payment of 2 contingent on the row player playing C, is it in the row player's interest to offer a payment of 2 contingent on the column player playing C? The answer is no. Suppose that the row player deviates and offers to pay the column player $1 + \varepsilon$ for each play of C, by either player[2]

2. Note that we can view this game as a decision of whether or not to contribute to a public good, and the transfers as a form of matching contract (it can be shown that they have equivalent effects in public goods games). The reason that this does not contradict the analyses of Guttman (1978), Danziger and Schnytzer (1991), and Varian (1994*a*), is that they only consider a limited form of matching contracts, where matching or payments can only be made in proportion to the actions taken by the other agents (see also Qin (2002) who considers also payments made only contingent on own action). As we see here, either player would strictly gain by deviating and using a different sort of contract (arguably just as simple). In fact one observes matching offers of the form "I will match any contributions", rather than just "I will match the contributions made by other people". Such seemingly minor differences in contract specification have important implications for incentives, as illustrated in this example.

(where $\varepsilon > 0$).[3] The resulting game is as follows:

	C	D
C	$2-2\varepsilon, 2+2\varepsilon$	$-\varepsilon, 3+\varepsilon$
D	$3-\varepsilon, \varepsilon$	$0, 0$

This has a unique equilibrium which is inefficient, but better from the row player's perspective: the column player plays C and the row player plays D. Thus, it will not be an equilibrium for the players to offer the "efficient" promises of transfers. As will follow from the theorems we prove below, there is no equilibrium in this example that results in the efficient play. In fact, in the context of this game, any equilibrium must involve some mixing, as it is not only C, C that cannot be supported, but in fact any pure play (see the discussion of this example that follows Theorem 2).

The reason that efficiency is not always obtained, is as follows. Players can use transfers to ensure that other players internalize externalities. However, they can also use transfers to try to manipulate other players' behaviour more generally. Sometimes, these objectives are at odds with each other, and then it is impossible to support efficient outcomes in equilibrium. This is captured in our main results, which can roughly be characterized as follows. First let us consider the case of two players. Determine the pay-off that a given player can obtain by offering transfers to try to best manipulate the other player's behaviour. Now examine a particular set of (efficient) actions that one might like to support. The sum of the pay-offs obtained from this (efficient) profile of actions needs to be at least as large as the sum of the pay-offs that the players can obtain through their respective optimal transfers. This turns out to be a strong condition that rules out obtaining efficient outcomes in many but not all games. Once one turns to the case of three or more players, the analysis changes fairly dramatically. Effectively, the ability to make transfer payments to several agents simultaneously allows agents to commit themselves to following certain actions. This possibility of commitment leads to efficient outcomes. We defer a fuller discussion of the three or more player case until later in the paper.

Before presenting the model and results, let us discuss the relationship between our work and other work in additional detail.

As discussed above, our analysis is related to the study of matching games that have been analysed by Guttman (1978, 1987), Danziger and Schnytzer (1991), Guttman and Schnytzer (1992), and Varian (1994*a*,*b*), among others. They show that efficiency can be obtained when agents can undertake matching plans in the context of public goods and some other settings with externalities. In many contexts our results are at odds with the results from those papers. The reason behind the difference in results is that those papers limit the set of contracts that are available to agents so that they can only make particular types of transfers. As seen in the above example, if agents can choose from a richer set of transfers (and they have a strict incentive to) then efficiency no longer holds.

We also note that the reason that inefficiency arises in our setting is different from that in the contracting literature. Here inefficiency arises from when each agent is attempting to offer transfers that manipulate other agents' behaviour to his or her advantage, and not necessarily to what is socially desirable. This is a different contracting failure from those that have been the primary focus of the recent contracting literature, such as imperfections related to costs of contracting, asymmetric information, limited enforcement of contracts, and non-verifiability of actions or information.[4] Some of the existing contracting literature can be embedded in our framework as special cases with added restrictions on the admissible contracts. In that regard,

3. If $\varepsilon = 0$ there are two equilibria to the game, but both are still inefficient and involve the row player playing D.

4. For a recent overview of this extensive literature, see MacLeod (2002). Anderlini and Felli (2001) provide a nice discussion of the relationship of that literature to failures of the Coase theorem.

our results provide a robustness check on these papers, and perhaps a rationale for the emergence of contracting restrictions. For example the common agency literature beginning with Bernheim and Whinston (1986), is a special case of our model where only the players labelled "principal" are allowed to make offers, and the admissible transfers can only depend on the actions of the agents. Those limitations, can result in different predictions (*e.g.* see Prat and Rustichini, 2003).[5]

One exception to viewing a game as fixed is delegation games (*e.g.* see Fershtman, Judd and Kalai (1991) and a recent application in Miller and Pazgal (2001)).[6] In delegation games players may hire another player to play the game for them. This effectively allows a player to change their own incentives and thus can change the outcomes of a game. In delegation games a player can only change their own pay-off structure, and cannot make promised payments to other players to induce them to change their strategies, and so the results are not very closely related to those here.

2. DEFINITIONS

A set $N = \{1, \ldots, n\}$ of players interact in two stages.

First, let us offer an informal description of the process.

Stage 1. Players simultaneously announce transfer functions. That is, each player announces a profile of functions indicating the payments that they promise to make to each other player as a function of the full profile of actions chosen in the second-stage game.

Stage 2. Players choose actions in the game.

Pay-offs. The pay-off that player i receives is his or her pay-off in the game plus all transfers that other players have promised to i conditional on the actions played in the game minus the transfers player i promised to make to other players conditional on the actions played in the game.

The transfer functions that are announced in Stage 1 are binding. There are many ways in which this could be enforced, ranging from reputation, posting a bond with a third party, to having legal enforcement of contracts.[7]

We also point out that players can effectively refuse any part of another player's promised transfers by announcing a transfer that returns the other player's transfer.[8] We point the reader to the discussion section for a fuller discussion of the importance of being able to refuse transfers.

5. Another application is the contracting externalities literature. For example in Aghion and Bolton (1987) there are three players, the incumbent seller, a customer and an entrant. They show that the customer and incumbent may contract to an inefficient outcome that deters entry. In their framework the entrant is not allowed to make pre-game offers to the incumbent or the customer. Segal (1999) shows how many contracting papers can be unified by the concept of a contracting externality. Again our results provide insight into the role played by the restrictions on the class of contracts used in these papers. Segal and Whinston (2003) also provide insight into how allowing for rich contracts can matter by changing the information revealed in the context of contracting between a principal and multiple agents.

6. A few other exceptions are the analysis of choices of mechanisms by competing sellers (*e.g.* see McAfee, 1993), choice of voting rules (*e.g.* see Barbera and Jackson, 2004), flexibility on the part of the planner (*e.g.* Baliga and Sjöström, 1995), mechanism selection more generally (*e.g.* see Lagunoff (1992), Allen (2001)), as well as earlier work by Kalai and Samet (1985), who looked at players trying to come to a unanimous and binding agreement as to a social state that is to be chosen (see also Kalai (1981), Bensaid and Gary-Bobo (1996)). For recent introductions to the implementation and mechanism design literatures, and some additional discussion of endogenous mechanisms, see Jackson (2001, 2003).

7. For an alternative framework where unanimity is required to enforce another's offer see Ray and Vohra (1997) or Qin (2002).

8. In terms of the legal enforcement of contracts, one might worry that some of the promises of transfers in our model lack what is called "consideration". Contracts where one player makes a promise contingent only on his or her own action are sometimes not enforceable because of the lack of "adequate consideration" by the other player—*i.e.* the other player did not do anything. However, these promises are easily approximated by promises that vary in some way on the other players' actions, or can be enforced by other things outside of our model such as reputation.

We now provide formal definitions.

The underlying (second-stage) game

The second-stage game consists of a finite pure strategy space X_i, with $X = \times_i X_i$. Let $\Delta(X_i)$ denote the set of mixed strategies for player i, and let $\Delta = \times_i \Delta(X_i)$. We denote by x_i, x, μ_i and μ generic elements of X_i, X, $\Delta(X_i)$, and Δ, respectively. In some cases we use x_i and x to denote elements in $\Delta(X_i)$ and Δ, respectively, that place probability one on x_i and x.

The restriction to finite strategy spaces provides for a simple presentation of the results, avoiding some technical details. Nevertheless, games with a continuum of actions are important, and we provide results for the case of games with continuous action spaces in the Appendix. These results are a straightforward extension of the finite case.

Pay-offs in the second-stage game are given by a von Neumann–Morgenstern utility function $v_i : X \to \mathbb{R}$.

The first-stage transfer functions

The transfer functions that player i announces in the first period are given the vector of functions $t_i = (t_{i1}, \ldots, t_{in})$, where $t_{ij} : X \to \mathbb{R}_+$ represents the promises to player j as a function of the actions that are played in the second-period game. So, if x is played in the second period, then i transfers $t_{ij}(x)$ to player j.

Let $t = (t_1, \ldots, t_n)$. Also, denote by t_i^0 the degenerate transfers such that $t_{ij}^0(x) = 0$ for all $x \in X$, and let $t^0 = (t_1^0, \ldots, t_n^0)$.

The pay-offs

The pay-off to player i given a profile of transfer functions t and a play x in the second-period game is then[9]

$$U_i(x, t) = v_i(x) + \sum\nolimits_{j \neq i} (t_{ji}(x) - t_{ij}(x)).$$

So, given a profile of transfer functions t and a mixed strategy μ played in the second-period game, the expected utility to player i is

$$\mathrm{EU}_i(\mu, t) = \sum\nolimits_x \times_i \mu_i(x_i) \left[v_i(x) + \sum\nolimits_{j \neq i} (t_{ji}(x) - t_{ij}(x)) \right].$$

Let $\mathrm{NE}(t)$ denote the set of (pure and mixed) Nash equilibria of the second-stage game where pay-offs are given as above. So this is the set of Nash equilibria taking a profile of transfer functions t as given, and only varying the strategies in the second-period game.

Supportable strategies and pay-offs

A pure strategy profile $x \in X$ of the second-stage game together with a vector of pay-offs $\overline{u} \in \mathbb{R}^n$ such that $\sum_i \overline{u}_i = \sum_i v_i(x)$ is *supportable* if there exists a subgame perfect equilibrium of the two stage game where some t is played in the first stage and x is played in the second stage (on the equilibrium path), and $U_i(x, t) = \overline{u}_i$.

Supportability is a condition that applies to a combination of a strategy profile and a set of pay-offs. We refer to both since in some cases transfers must be made on the equilibrium path to

9. This assumes transferable utility, and it would be interesting to see how this extends to situations where private goods transfer at different rates across players.

support x as part of an equilibrium. In such cases the pay-offs including transfers differ from the original underlying pay-offs without transfers.

The definition supportability looks at pure strategies in terms of what is played on the equilibrium path. In many games (in fact generically), there is a unique x that is efficient. Thus, it makes sense to focus on pure strategy equilibria, at least in terms of the second period. The focus on pure strategies in terms of t's is for technical convenience, as the space of mixed strategies over all such transfer functions is a complicated animal (measures over functions).[10]

Surviving equilibria

In addition to understanding supportability in the two stage process, we are also interested in the following question. When does an equilibrium of the original underlying game survive to be supportable when the two stage process is considered?

Consider a pure strategy profile $x \in X$ of the second-stage game that is an equilibrium of the second stage when no transfers are possible ($x \in \text{NE}(t^0)$). Such an equilibrium *survives* if there exists a subgame perfect equilibrium of the two stage game where some t is played in the first stage and x is played in the second stage (on the equilibrium path), with net pay-offs being $U_i(x, t) = v_i(x)$.

Note that x survives if and only if x is a Nash equilibrium of the second-stage game and $(x, v(x))$ is supportable in the two stage process. Together, the notions of supportability and surviving then give us an idea of how the set of equilibrium and equilibrium pay-offs change when players can make binding transfer commitments.

Existence

Our characterization results provide conditions for existence of pure strategy equilibria of the overall two stage game. As we shall see, equilibria always exist in games with three or more players, and exist under some conditions in two player games. Example 4 shows that there are two player games where equilibrium must involve some mixing over transfer functions. While we have not found an example of non-existence of equilibrium (in mixed strategies with two players), that is an open question.[11]

3. TWO PLAYER GAMES

The results for two player games and games with more than two players differ significantly and so we treat them separately. We start with an analysis of two player games.

The following notion plays an important role in the characterization results that follow.

Solo transfers

Suppose that only one player were allowed to propose transfers in the first stage. We can consider the transfers that would be best from this player's perspective.

10. In many of the examples where efficiency turns out not to be supportable in pure strategies, allowing for mixed strategies would not help. We are not sure whether this is always the case. However, it is important to consider mixed strategies off the equilibrium path in the second-period game, and we explicitly account for this.

11. One can ensure existence of a perfect equilibrium in the two stage game by bounding the possible transfer functions to provide a compact strategy space and then applying a theorem by Harris (1985). However, bounding the transfer functions is a bit problematic in our context because it limits the ability of agents to undo the other players' transfers. For instance, if a player wishes to refuse the transfer of another player which is at the maximal level, then that player could not offer to make any additional transfers.

Let[12]

$$u_i^s = \sup_{t_i}\left[\min_{\mu \in \mathrm{NE}(t_{-i}^0, t_i)} \mathrm{EU}_i(\mu, t_{-i}^0, t_i)\right].$$

So, a player's "solo" pay-off is the one obtained when the player is allowed to announce any transfer function that he or she likes and other players cannot make any transfers. As there may be several equilibria in the second-stage game that result from any given transfer function, we must have some idea of which one will be relevant. This definition imagines the worst continuation equilibrium for i once t is in place. This turns out to be the correct definition for the characterizations that follow.

To get some intuition for the definition above, consider the prisoners' dilemma game from Section 1:

	C	D
C	$2, 2$	$-1, 4$
D	$4, -1$	$0, 0$

Here the solo pay-off for the row player (and similarly the column player) is 3. By making a payment of $1 + \varepsilon$ conditional on the column player playing C, the new matrix becomes

	C	D
C	$1 - \varepsilon, 3 + \varepsilon$	$-1, 4$
D	$3 - \varepsilon, \varepsilon$	$0, 0$

This has a unique equilibrium leading to a pay-off of $3 - \varepsilon$ for the row player. Taking the sup over such payments leads to a pay-off of 3 for the row player, which is the solo pay-off.

Our first result is a characterization of the Nash equilibria of a game that survive when transfers are introduced.

Theorem 1. *If $n = 2$, then a Nash equilibrium x of the underlying game survives if and only if $v_i(x) \geq u_i^s$ for each i. Moreover, if x survives then there is an equilibrium in the overall process where no transfers are made in the first stage and x is played in the second stage.*

The formal proof of Theorem 1 uses the proof of Theorem 3 and appears in the Appendix. However, the intuition is fairly simple and we explain it here.

First, let us show that this condition is sufficient to have x survive. Consider a Nash equilibrium x such that $v_i(x) \geq u_i^s$. On the equilibrium path let players make no transfers in the first stage (play t^0) and then play x in the second stage. Off the equilibrium path, if some player offers transfers in the first period, then identify the worst equilibrium for that player in the resulting second-stage game and have that be played in the continuation (and if more than one player offers transfers in the first period then play any equilibrium in the second stage). This is easily seen to be a subgame perfect equilibrium of the overall game: the best pay-off a player can get by deviating in the first stage is no more than their solo pay-off, which is not improving; and given that no transfers are made in the first stage, x will be an equilibrium in the second stage.

Next, consider a Nash equilibrium x that survives. We argue that $v_i(x) \geq u_i^s$. Let t be any transfer function that is made in the first stage as part of an equilibrium where x is played in the second stage. Suppose to the contrary that $v_i(x) < u_i^s$. The definition of solo pay-off then implies that there exists a transfer function $\bar{t}_i$, so that the pay-off to i under worst continuation equilibrium under $\bar{t}_i, t_j^0$ is higher than $v_i(x)$. So, let i do the following: make a transfer that

12. Note that the min in this expression are well defined since the set of Nash equilibria of a finite game is compact. The sup is necessary as there may be no maximizer. For instance, consider a game where $X_1 = \{x_1\}$ while $X_2 = \{x_2, x_2'\}$. Let player 2 be indifferent between the actions and player 1 prefer that 2 play x_2. Any positive transfer from 1 to 2 leads to a unique equilibrium of x_2, but a 0 transfer leads to a minimizing equilibrium of x_2'.

cancels out the transfers under t_j and then adds $\bar{t}_i$ on top. That is, let i announce $\widehat{t_i} = \bar{t}_i + t_j$. Note that the pair of transfers $\widehat{t_i}, t_j$ leads to exactly the same second-stage pay-offs as $\bar{t}_i, t_j^0$. Thus, from the definition of $\bar{t}_i$, it follows that if i deviates to $\widehat{t_i}$ while j plays t_j then even the worst continuation equilibrium in the second stage will result in a pay-off which is higher than $v_i(x)$. This contradicts the fact that t was part of an equilibrium where x was played in the second stage.

To get an idea of the impact of Theorem 1, we illustrate it in the context of several examples.

Example 2. Only the efficient equilibrium survives.

	A	B
A	2, 2	0, 0
B	0, 0	1, 1

In this pure coordination game, there are two equilibria (A, A) and (B, B). The solo pay-offs are 2 for each player[13] and so it follows from Theorem 1 that the only equilibrium that survives once transfers are allowed is (A, A).

Example 3. The efficient equilibrium *does not* survive.

Consider the following game, which has an efficient equilibrium of U, L leading to pay-offs of 2, 2. It is easily checked that the solo pay-offs are 3 to each player.

	L	C	R
U	2, 2	0, 0	0, 0
M	0, 0	3, 0	0, 0
D	0, 0	0, 0	0, 3

So, in this game the efficient equilibrium does not survive. In fact, no equilibrium survives, and the only equilibria of the two stage process (if they exist) must involve mixing over transfer functions announced in the first stage.

Next, we provide an example which shows why it was necessary to consider mixed strategies in the definition of solo pay-offs.

Example 4. Mixed strategies in the solo definition.

Consider the following game.

	L	C	R
U	1, 10	0, 0	0, 0
M	0, 0	3, 0	0, 10
D	0, 0	0, 10	3, 0

Let us check to see if the strategies $x = (U, L)$ with pay-offs $\bar{u} = (1, 10)$ survive. The highest pay-off in the entire matrix for the column player is 10, and so the column player's solo pay-off is no more than 10 and that part of the characterization is satisfied. So, we need only check that the pay-off of 1 is at least as large as the row player's solo pay-off. If we use a pure strategy solo definition, then the best pay-off that the row player can induce is 1. This would suggest that U, L would survive. However, this is not the case, and we can see where mixed strategies affect

13. For instance, if the row player makes the transfer of 2 to the column player conditional on (B, A) being played, then the unique equilibrium becomes (A, A).

the outcome. Suppose that the row player pays the column player 2 conditional on U being played. That leads to the following matrix of pay-offs.

	L	C	R
U	$-1, 11$	$-1, 1$	$-1, 1$
M	$0, 0$	$3, 0$	$0, 10$
D	$0, 0$	$0, 10$	$3, 0$

This game has a unique Nash equilibrium which is a mixed strategy equilibrium (equal mixing on M and D by row, and C and R by column) leading to a pay-off of 1·5 to the row player. This means that the threat point for the row player is indeed 1·5, as whatever transfers are made by the column player, the row player can always return those, add a payment of 2 conditional on U being played, and expect at least 1·5 in the continuation. That is what is captured in the definition of solo pay-offs. In fact, given that the solo pay-offs of the players are 1·5 and 10, respectively, Theorem 2 below will show that the no action-pay-off pair is supportable in this example. What that means is that any equilibrium in the two stage process will involve mixing over the transfers functions announced in the first stage.[14]

Note that the reasoning behind Theorem 1 did not rely on the fact that x was a Nash equilibrium of the second-stage game to begin with. So, in fact we have just argued a necessary condition for supportability as well as survivorship. This is stated in the following theorem.

Theorem 2. *If $n = 2$, then $(x, \overline{u})$ is supportable only if $\overline{u} \geq u_i^s$ for each i.*

While only presenting necessary conditions, Theorem 2 is a still useful result, since $\overline{u} \geq u_i^s$ is a demanding condition that fails in some games. That is true of Example 1 in the introduction, as one can directly check that in that example the solo pay-off of each player is 3, and there is no pair of actions possible that could lead to a pay-off of at least 3 to each player at the same time. Note that since no pure strategy profile is supportable, and moreover since no mixed strategy profile in the second stage leads to at least 3 to both players, any equilibrium must involve some mixing over transfer functions.

A full characterization of supportability for two players

While necessary for supportability, the condition that $\overline{u} \geq u_i^s$ for each i is not always enough to ensure that $(x, \overline{u})$ is supportable. We now present the full necessary and sufficient condition for supportability.

We begin by deducing some additional necessary conditions by thinking about the minimal transfers that are necessary to support some $(x, \overline{u})$. We remark that we may need to allow $\overline{u} \neq v(x)$ if we wish to support some x as part of an equilibrium, as x may not be a Nash equilibrium in the absence of any transfers and so some side payments may be necessary. Thus, in order to characterize supportability it will be important to have some idea of what transfers are (minimally) necessary.

Minimal transfers

The *minimal transfer function profile* $t^{x,\overline{u}}$ for a pair $x, \overline{u}$ is defined by:

$$t_{ij}^{x,\overline{u}}(\widehat{x}) = \begin{cases} \max[v_i(\widehat{x}) - \overline{u}_i, 0] & \text{if } \widehat{x}_j = x_j \\ 0 & \text{otherwise.} \end{cases}$$

14. Since the total of the solo pay-offs is 11·5, any pure strategies of transfers in the first stage must lead to an equilibrium in the second stage that is below the solo pay-off of one of the two players, who could thus benefit by deviating in the first stage.

The idea of a minimal transfer function is straightforward. If we want x to be a Nash equilibrium with pay-off $\overline{u}$, then this will impose some minimal necessary conditions on transfers. First, if i's pay-off at x, $v_i(x)$, is larger than $\overline{u}_i$, then it must be that i transfers the excess $(v_i(x) - \overline{u}_i)$ to the other player or else i's pay-off would not be $\overline{u}_i$ at x. Second, in order for this to be a Nash equilibrium, no player should obtain a higher pay-off if they deviate to some other strategy $\widehat{x}_i$. Thus, they would have to transfer $\max[v_i(\widehat{x}_i, x_{-i}) - \overline{u}_i, 0]$ to the other player.

Minimal transfers are illustrated in the following example.

Example 5. Minimal transfer function

Consider the following game.

	L	R
U	4, 4	0, 6
D	5, 0	0, 6

Consider supporting $x = (\mathrm{U}, \mathrm{L})$ with pay-offs $\overline{u} = (2, 6)$.[15] In order to have (U, L) result in these pay-offs, it would have to be that the row player transfers at least 2 to the column player conditional on (U, L).[16] The row player would also have to transfer at least 3 to the column player conditional on (D, L), as otherwise (U, L) could not be an equilibrium. So, these transfers are the minimal transfers to support $x = (\mathrm{U}, \mathrm{L})$ and $\overline{u} = (2, 6)$.

We now define the solo pay-offs noting that these minimal transfers (or some larger transfers) would have to be in place in order to lead to $x, \overline{u}$ as part of an equilibrium outcome in the two stage process. So, this definition is similar to that of solo pay-offs, except that now the other player(s) are assumed to play at least the minimum transfers instead of not playing any transfers.

Modified solo pay-offs

$$u_i^{ms}(x, \overline{u}) = \sup_{t_i}\left[\min_{\mu \in \mathrm{NE}(t_{-i}^{x,\overline{u}}, t_i)} \mathrm{EU}_i(\mu, t_{-i}^{x,\overline{u}}, t_i)\right].$$

The modified solo pay-offs represent the best possible pay-off a player can guarantee himself, assuming the worst possible continuation pay-off and that the other player plays the minimal transfer functions.

It is fairly easy to see that it will be necessary that each player's pay-off exceed the modified solo pay-offs in order to have some $(x, \overline{u}_i)$ be supportable. To see this, suppose that we can support $(x, \overline{u}_i)$ as part of a two stage equilibrium. Let t be the transfers that are played in the first stage. We know that each t_i must be at least as large (as a function of each action) as the minimal transfer functions, or else x would not be part of an equilibrium play. Now suppose to the contrary that the modified solo pay-offs are larger than $\overline{u}_i$ for some agent i. This means that player i has some transfer function, denoted $\widehat{t}_i$, which when played against the other player's minimal transfer function is such that all equilibria in the second stage have a pay-off to i above $\overline{u}_i$. Let player i deviate by announcing $\widehat{t}_i$ plus the difference between whatever the other player has announced and what the other player's minimal transfer function is, $t_{-i} - t_{-i}^{x,\overline{u}}$. Thus the total transfers are the same as if i played $\widehat{t}_i$ and $-i$ played the minimal transfer function $t_{-i}^{x,\overline{u}}$. This means that regardless of which equilibrium is played in the subgame that follows, the resulting pay-off to i

15. As the column player can get a pay-off of at least 6 simply by not announcing any transfers and then playing R, supporting any set of strategies will require a pay-off of at least 6 to the column player.

16. We say "at least" as it is conceivable that the row player transfers more and the column player transfers some back.

will be above $\overline{u}_i$ and so this deviation is improving. This is a contradiction, and so it follows that $\overline{u}_i$ must be at least as large as the modified solo pay-offs for each player.

It turns out that the necessary condition that we have just outlined is also sufficient for supportability. Thus, we have the following complete characterization of supportable action-pay-off pairs.

Theorem 3. *If $n = 2$, then $(x, \overline{u})$ is supportable if and only if $\overline{u}_i \geq u_i^{ms}(x, \overline{u})$ for each i. Moreover, if $(x, \overline{u})$ is supportable it is supportable with the minimal transfer function profile $t^{x,\overline{u}}$.*

As we have already argued, it is necessary that the modified solo pay-offs be no higher than $\overline{u}_i$ for each i in order to support $(x, \overline{u})$. Let us argue that this condition is also sufficient, and that then supportability can be achieved by the minimal transfers. It is clear from the definition of minimal transfers that x will be an equilibrium in the second stage with corresponding pay-offs $\overline{u}$. So, we need only specify what happens if some player deviates and announces another transfer function. If one player deviates, then in the following subgame play the worst equilibrium for that player in the subgame that follows (and if both players deviate from announcing the minimal transfers then play any equilibrium). We need only check that no player can gain by deviating to announce some other transfer function in the first stage. If a player deviates, the worst equilibrium for that player will be played in the subgame that follows. By the definition of modified solo pay-offs, the pay-off for the deviating player i will be no more condition than that player's modified solo pay-offs. Since the player's modified solo pay-offs are no larger than $\overline{u}_i$, the deviation cannot be improving.

We can now illustrate our results by applying them to some important applications.

4. APPLICATIONS

To see the implications of the results above, let us examine some common settings.

One sided externalities

Consider a classic one sided externality, such as Coase's example of a steel mill affecting a laundry. Let x_1 denote the output of the steel mill and x_2 denote the output of the laundry. Take these to fall in some finite sets (and see the Appendix for a treatment of continuum action spaces). The utility functions are $v_1(x_1)$ and $v_2(x_1, x_2)$, so that the steel mills production affects the laundry's pay-off. Let there be a unique Nash equilibrium x_1^n, x_2^n, and a unique efficient point x_1^*, x_2^* under transferability, and that these result in different pay-offs. Thus, $v_1(x^n) > v_1(x^*)$ and $v_2(x^*) > v_2(x^n)$, so that the steel mill is not accounting for the externality it imposes on the laundry.

Let us consider supporting the efficient solution together with pay-offs where the steel mill gets the pay-off it would have under the Nash equilibrium $\overline{u}_1 = v_1(x^n)$, and the laundry gets the pay-off it gets from the efficient solution after compensating the steel mill for playing the efficient action: $\overline{u}_2 = v_2(x^*) - (v_1(x^n) - v_1(x^*))$.[17]

We need to determine what the minimal transfer functions and corresponding modified solo pay-offs are. Since the steel mill's Nash pay-off is the highest possible under the second-stage game, its minimal transfer function is simply t_1^0. By definition, the laundry's minimal transfer function satisfies $t_2(x) = \max[v_2(x) - \overline{u}_2, 0]$ if $x_1 = x_1^*$, and 0 otherwise. This implies that $t_2(x^*) = v_1(x_1^n) - v_1(x_1^*)$.

17. As the steel mill can get at least this pay-off by not offering any transfers and choosing an optimal action in the second-period game (given whatever transfers are made by the laundry), this is the minimal possible supportable pay-off for the steel mill.

We know from Theorem 3 that $x^*, \overline{u}$ will be supportable if and only if the modified solo pay-offs are no higher than $\overline{u}$. Let us check that this is the case. First, let us examine the steel mill's modified solo pay-offs. Since the laundry's action does not affect the steel mill's pay-off, the steel mill's highest pay-off will come when it makes no transfers. The steel mill can earn its Nash pay-off by either playing the Nash strategy, or the efficient strategy (given the laundry's minimal transfer). It cannot earn more. Thus the steel mill's modified solo pay-off is its Nash pay-off. Second, let us examine the laundry's modified solo pay-off. Given that the steel mill's minimal transfer function is 0, the laundry's modified solo pay-off is just its solo pay-off. Since the steel mill can always play its Nash strategy, x_1^n, the most that the laundry could ever hope to get at some x would be the total utility of both players from x, less a transfer to make sure that the steel mill gets at least its Nash payment. This takes its maximum value at $x = x^*$ (by the definition of x^*) and then the laundry must make a transfer that is exactly the minimal transfer to support the efficient play. Hence, by Theorem 3 the efficient point is supportable and Coase's claim that the polluter and victim can reach an efficient outcome is verified in our explicit model of transfer payments.

Public goods

We now move to a problem where the externalities are two sided, and see that supporting the efficient outcome is no longer always possible.

Consider a two person game of voluntary contributions to a public good. Let $x_i \in \mathbb{R}_+$ be player i's contribution and her utility be $v_i(x_1, x_2) = 2\theta_i(x_1 + x_2)^{\frac{1}{2}} - x_i$. Suppose that $\sum \theta_j = 1$ and $\theta_1 > \theta_2 > 0$. This ensures a unique Nash equilibrium in the contribution game (in the absence of any transfers), such that $x_1^n = (\theta_1)^2, x_2^n = 0$. The associated utilities are $v_1(x_1^n, x_2^n) = (\theta_1)^2$ and $v_2(x_1^n, x_2^n) = 2\theta_2\theta_1$. The efficient contribution level is any pair such that $\sum x_i = 1$. Moreover, the net utilities at any efficient allocation sum to 1, and so any pair $(x, \overline{u})$ that we might think of supporting where x is efficient must have $\sum \overline{u}_i = 1$.

Let us show that it follows from Theorem 2 that no efficient outcome is supportable. We do this by showing that the sum of the solo pay-offs are more than 1; and so it cannot be that each $\overline{u}_i$ is as large as the solo pay-offs, which is a necessary condition by Theorem 2.

First, consider 1's solo pay-off u_1^s. Consider the transfer function defined by $t_1(x) = \theta_1 x_2$. If this offer is made then in any equilibrium of the second stage that follows, it will be that $x_2 = 1$ (and $x_1 = 0$). Thus, $u_1^s \geq 2\theta_1 - \theta_1 = \theta_1$. Second, consider 2's solo pay-off u_2^s. Set $t_2(x) = (\theta_1)^2 - (2\theta_1 - 1)$ if $x_1 = 1$ and $t_2(x) = 0$ otherwise.[18] If this offer is made then in the second-stage equilibrium that follows $x_1 = 1$ (and $x_2 = 0$). Thus, $u_2^s = 2\theta_2 - [(\theta_1)^2 - (2\theta_1 - 1)] = 1 - (\theta_1)^2$. Putting these two solo pay-offs together we find that $\sum u_i^s \geq \theta_1 + (1 - (\theta_1)^2) > 1$, as claimed.[19]

Bertrand competition

Consider the case of two firms competing in a Bertrand market. Let each firm have a linear cost function $c(q_i) = cq_i$ as a function of their production quantity q_i, and the demand function be described by $Q(p)$ where $Q = \sum q_i$ and p is the lowest price offered by any firm. Here the

18. In the case of continuum action spaces, this can be substituted for by a carefully constructed continuous function and still give exactly the same incentives.

19. We note that tragedy of the commons problems, where a group of individuals share a common resource, have results that are similar to those of public goods, with inefficiency being pervasive. This can be seen as a variation of a voluntary public good contribution game, where one translates usage of the common resource into the negative of contribution to a public good. Holding back on usage is similar to contributing to a public good.

strategic variable of each firm is a price $p_i \in \mathbb{R}_+$,[20] and we can write their profits as $\pi_i(p_1, p_2)$. Follow the textbook Bertrand rule that firms charging the lowest price split the market evenly and that firms with higher prices sell zero. The Nash equilibrium pay-off for each firm in the underlying Bertrand game is zero. Let π^m denote the industry pay-off if both firms charge the monopoly price. Can we support the strategy $p_i = p^m$ and pay-offs $\overline{u}_i = \frac{\pi^m}{2}$ for each i?

Again, let us apply Theorem 3 by checking that the modified solo pay-offs do not exceed $\overline{u}$. Here, the corresponding minimal transfer functions $\overline{t}$ are defined by $t_i(p_1, p_2) = \max[\pi_i(p_1, p_2) - \frac{\pi^m}{2}, 0]$ if $p_{-i} = p^m$ and 0 otherwise. Let us then consider the modified solo pay-off to firm i, u_i^{ms}. First, consider any equilibrium that might result in the second stage following a first stage where i offers some transfer function t_i and player $-i$ offers the minimal transfer function. If the second stage does not involve i playing p^m, then there are no transfers from $-i$ to i. Also, for any price that i announces above c, player $-i$ can get arbitrarily close to the full market by slightly undercutting and so the pay-offs to i in such an equilibrium cannot exceed $\frac{\pi^m}{2}$. Now consider an equilibrium where i ends up playing p^m. Player $-i$, can get at least half the monopoly profits by playing p^m too, and so again i's pay-off cannot exceed $\frac{\pi^m}{2}$. Thus, the modified solo pay-offs do not exceed $\overline{u}$ and so by Theorem 3 the efficient (collusive) outcome is supportable.

Cournot duopoly

We now turn to a Cournot duopoly and see that we find a contrast with the Bertrand conclusions. Under Bertrand competition, the firms could support monopoly profits through appropriate transfers, as their solo pay-offs were no higher than half of the monopoly profits. In contrast, under Cournot competition we shall see that the solo pay-offs are higher than half of the monopoly profits and so the collusive monopoly outcome is not supportable in a classic linear Cournot world.

Consider a Cournot duopoly where the action $x_i \in \mathbb{R}_+$ is quantity choice of firm i, inverse demand is linear where the market price is equal to $a - \sum_i x_i$, and costs of production are zero. The pay-off function to firm i is $v_i(x_1, x_2) = (a - \sum x_i)x_i$. In this case, the Cournot equilibrium quantities are $x_i^n = a/3$, and the resulting pay-offs are $v_i(x^n) = a^2/9$. If the firms were to collude to maximize their joint profits, they would choose the monopoly output, $x_1 + x_2 = a/2$ and split the monopoly profits.

Let us check that no such pair of strategies and split of monopoly profits (symmetric or asymmetric) is supportable. Again we apply Theorem 2 and check that the sum of the solo pay-offs exceeds the sum of the maximal possible profits. Here the relevant $\sum \overline{u}_i$ is equal to the monopoly profits $a^2/4$.

Consider player 1's solo pay-off, u_1^s. Fix any x_1' and consider transfers $t_1(x_1', 0) = (a - x_1)^2/4$, and $t_1(x) = a^2$ if $x_1 \neq x_1'$, and $t_1(x) = 0$ otherwise. In any equilibrium following these transfers, the play will be $x_1', 0$. Thus, these transfers are such that player 1 commits to play x_1' and to pay player 2 to stay out. To calculate a lower bound on player 1's solo pay-off u_1^s we can then solve:

$$\max V = (a - x_1)x_1 - (a - x_1)^2/4.$$

The solution is $x_1 = \frac{3a}{5}$ and so $u_1^s \geq \frac{a^2}{5}$. However the symmetric argument applies to player 2 and so $\sum u_i^s = \frac{2a^2}{5} > a^2/4$. Thus, by Theorem 2 no equilibrium with the duopoly earning monopoly profits is supportable.

20. This is a game that is usually analysed with a continuous action space. One can either approximate action spaces with some discrete grid and apply the results as is, or else see the Appendix where we show that the results extend to continuous action spaces.

Cardinal pay-offs and strategic structure

As we have seen through Theorem 3 and the applications above, the supportability of an efficient action profile depends on the specifics of the pay-off structure of the game, and that simple variations in a game can change its properties. We wish to emphasize that even variations in cardinal pay-offs can change supportability conclusions, regardless of whether the strategic structure of the game has changed. This can be seen directly through the following prisoners' dilemma examples, where games with the same basic strategic properties (a unique equilibrium in strictly dominant strategies, and it being Pareto dominated by the other pair of actions), but different cardinal pay-offs have different supportability characteristics. This means that the supportability characterization will not translate into some sort of characterization of the strategic properties of a game, but really relies on the cardinal structure of the game.

For the prisoners' dilemma below, the efficient actions of cooperation are not supportable, as we saw in Example 1.

	C	D
C	$2, 2$	$-1, 4$
D	$4, -1$	$0, 0$

However, the following variation on the same game, which has the same strategic properties has different supportability properties. Here, it is straightforward to see that the modified solo pay-offs are 3 for each player, and so now C, C is supportable.

	C	D
C	$3, 3$	$0, 4$
D	$4, 0$	$1, 1$

The reason that cardinal pay-offs play such an important role, is that the interplay between the two players' pay-offs is what determines what each player can expect when optimally structuring their transfer functions (their modified solo pay-offs). The cardinal structure determines how much one player has to pay another to sustain given actions, which are critical in determining what transfers are needed to sustain the efficient actions, and what alternative pay-offs the players can expect when they manipulate transfers to their best advantage.

5. THREE OR MORE PLAYERS

In the case of three or more players, it is relatively easier to support outcomes in the two stage process. That is captured in the following theorems, the first of which addresses the issue of survivability.

Theorem 4. *If $n \geq 3$, then every pure strategy Nash equilibrium of the underlying game survives.*

The reason for the much more positive outcomes with three or more players, and for instance the contrast between Theorems 1 and 4, is with more than two players it is possible for players to effectively commit themselves not to play certain strategies through the use of transfer functions. For example, consider a player 1 who would like to be able to commit not to playing an action x_1. Player 1 could simply say that he or she will pay some large amount, say M (which is higher than the maximum pay-off to any player in the matrix) to each other player if player 1 were to play x_1. In a two person game, player 2 can undo this by simply committing to pay M back to player 1 if player 1 plays x_1. However, in a three person game, player 2 would have to pay 2M back to player 1, and is only getting M from player 1, and so now it is prohibitively costly for player

2 to try to undo player 1's commitment. This type of commitment possibility makes supporting desired strategy-pay-off combinations much easier. The importance of commitment to strategies dates (at least) to Schelling (1960). In our analysis of three or more player games, any player can essentially become one who holds a bond (via promised transfers contingent on undesired actions being played) thus committing some other players to play certain strategies.

Note that if one wishes to introduce a third party to hold a bond in a two player game, this can be modelled simply by modelling the third party as a third player in the game who has no actions in the game and no pay-off other than transfers received (or made). We discuss this in more detail below.

Let us now turn to the question of supportability. As the full characterization involves conditions that are more difficult to apply, we first provide a simple set of sufficient conditions. The following theorem illustrates how easy it is to support action and pay-off profiles with three or more players.

Theorem 5. *If $n \geq 3$ and x is a strategy profile and there exists a Nash equilibrium $\widehat{x}$ such that $v_i(x) \geq v_i(\widehat{x})$ for all i, then $(x, v(x))$ is supportable.*

Theorem 5 states that a strategy profile that offers a Pareto improvement over (or is equivalent to) some Nash equilibrium pay-offs, is supportable. The proof appears in the Appendix. The intuition is that it is possible to use the Nash equilibrium as a threat point to which players revert if some player does not make the correct supporting offer of transfers in the first stage.

The following example illustrates the power of Theorem 5. It is also of interest since it shows how seemingly small restrictions in the set of admissible transfer functions can be critical. In particular we show how the analysis of common agency of Prat and Rustichini (2003) contrasts with what Theorem 5 predicts for a common agency example, and how this hinges on the set of admissible transfer functions.

Example 6. Efficiency in a common agency example.

Consider a setting with two principals and two agents. The agents are the only players who take actions. Let us label these as players 1 and 2. The principals are the only ones whose pay-offs depend on the play of the game.

	L	R
U	0, 0, 3, 0	0, 0, 0, 2
D	0, 0, 0, 2	0, 0, 2, 0

In this setting player 1 (an agent) takes an action up or down, while player 2 (also an agent) takes an action either left or right. The agents' pay-offs are always 0. Player 3 is a principal and would rather that the agents play (U, L) or (D, R), and player 4 is a principal who would rather that the agents would play (U, R) or (D, L). Thus, the principals have conflicting interests.

Theorem 5 shows that the efficient strategy pair (U, L) can be supported together with pay-offs (0, 0, 3, 0), since this is in fact an equilibrium of the game with no transfers.[21]

For the above example, Prat and Rustichini's (2003) results conclude that efficiency is not an equilibrium outcome and that the principals would play mixed strategies in the contracts they offer as Prat and Rustichini show in their matching pennies example. The key difference is that Prat and Rustichini only consider contracts between principals and agents, but not between different principals or between different agents. In the contracts that support efficiency in this

21. In fact, Theorem 4 could also be applied and note that all equilibria survive here.

example and underlie the proof of Theorem 5, there are transfers made off the equilibrium path between agents and/or principals, as a variety of transfer functions work.

As a simple, but useful corollary of Theorem 5, note that in any symmetric game that has a pure strategy Nash equilibrium, a symmetric efficient strategy profile will be supportable.

Corollary 1. *If $n \geq 3$ and the game has a pure strategy Nash equilibrium with symmetric pay-offs, then any efficient strategy that results in symmetric pay-offs is supportable.*

This corollary applies to the symmetric public goods game, commons games, and Cournot games for which such efficient outcomes were not supportable when $n = 2$.

As is clear from our results, there are differences between the consequences of the results for three or more players and those for two players. Let us offer two important observations in this regard.

Dummy players and bonding

Let us discuss how Theorem 5 shows that two players might use a third player as a bonding agent.

Consider a two person game (X_1, X_2, v_1, v_2). Let us say that we add a *dummy player* if we add a player with a degenerate singleton action space $X_3 = \{x_3\}$ and with $v_3(x) = 0$ for all $x \in X$.

Corollary 2. *If in a two person game there exists an efficient action pair (x_1, x_2) and a Nash equilibrium $(\widehat{x}_1, \widehat{x}_2)$ such that $v_i(x) \geq v_i(\widehat{x})$ for $i \in \{1, 2\}$, then if a dummy player is added to the game, x_1, x_2, x_3 is supportable together with $(\overline{u}_1, \overline{u}_2, 0) = (v_1(x), v_2(x), v_3(x))$ in the three person game.*

Note that the use of the third player in the corollary could also be viewed as placing deposits in escrow to be conditionally returned depending on the actions taken.[22]

A complete characterization of supportability

We now offer a complete characterization of supportability. Let

$$u_i^{ms}(\bar{t}) = \sup_{t_i}\left[\min_{\mu \in \mathrm{NE}(\bar{t}_{-i}, t_i)} \mathrm{EU}_i(\mu, \bar{t}_{-i}, t_i)\right].$$

This is similar to the definition of modified solo pay-offs that we had in the two player case, except that the transfer functions of the players other than i are fixed to some $\bar{t}_{-i}$ rather than to the minimal transfer functions. This difference is due to the fact that the minimal transfer functions are no longer uniquely tied down with three or more players.[23]

Theorem 6. *$(x, \overline{u})$ is supportable if and only if there exists $\bar{t}$ such that $x \in \mathrm{NE}(\bar{t})$, and $U_i(x, \bar{t}) = \overline{u}_i$ and $\overline{u}_i \geq u_i^{ms}(\bar{t})$ for each i.*

The proof of Theorem 6 is a straightforward variation on the proof of Theorem 3. See the proof of Theorem A1 in the Appendix for details.

The necessary and sufficient condition in Theorem 6 is more difficult to check than the corresponding condition in Theorem 3, as Theorem 3 shows that with $n = 2$ one only needs to

22. Such a device is discussed by Dutta and Radner (2001) as a means of partly solving a commons problem associated with investing in technological development related to slowing global warming.

23. For instance, a player's pay-off ties down how much they must be giving away to others, but does not tie down to whom the transfers might be made.

check the condition with respect to the uniquely defined minimal supporting $\bar{t}$. That is no longer the case with more than two players.

Coalitional considerations

We have seen results for three or more players show that the strategic aspects of side contracts are critically dependent on the number of players, and in particular, differ dramatically depending on whether there are two or at least three players. An important part of this distinction between two and three or more players can be seen with respect to the commitment power that transfers enable in the different settings. A player can try to commit to playing a certain strategy by offering to pay large transfers to other players if he or she deviates from the prescribed strategy. With only two players, the second player can undo that commitment simply by offering transfers that cancel the first player's transfers. However, with three or more players, a player can promise to make large transfers to several players if he or she deviates from the prescribed strategy. No single player can unilaterally undo this commitment—it would take a coordinated action by all of the other players to undo this. Thus, such commitment is possible to sustain as part of an equilibrium with three or more players, while it was not sustainable with only two players. Part of the reason for the difference is that we have not considered coalitional deviations. If instead of Nash and subgame perfect equilibrium, we considered strong equilibrium and strong perfect equilibrium, then the reasoning behind the three or more player case would look more like the two player case. That is, collectively any coalition of $n-1$ players could always undo the transfers of any other player and maximize its pay-offs subject to only controlling the remaining player through promised transfers. This would result in benchmarks that are similar to the solo pay-offs for each coalition of $n-1$ players. In many contexts, this would again lead to combinations of coalitional pay-offs that exceed the total efficient pay-off in the game.

6. DISCUSSION

We have characterized the outcomes of games that are supportable when players can commit to making strategy contingent side payments to other players. Some basic conclusions from the results can be summarized as follows.

- The incentives to use side payments to affect the strategic aspects of the game are subtle, and at times conflict with efficiency.
- In some cases, efficient strategies that are equilibria in a game without side payments do not survive when side payments are introduced.
- The solo pay-offs (where only one player can make transfers) are key benchmarks in understanding what outcomes are supportable in games with side payments.
- With three or more players side payments allow for a sort of commitment to strategies that makes supporting efficient strategies (and others) easier to support than with only two players.

Let us discuss some of the restrictions on the types of side payments we have considered and how robust the results should be to changes.

Refusing contracts

We have not considered the possibility of allowing players to make choices regarding accepting transfer contracts from other players. We note however, that any player can always return the transfers made by any other player through their own transfers. *Thus, in equilibrium no player*

is ever accepting any transfers that they would rather not accept. This holds regardless of the number of players.

The timing of this, however, is a bit tricky. If one can allow players to simultaneously commit, *before seeing the transfer functions*, to which transfer functions they would accept they can rule out all sorts of possible transfer functions by the other players. The important key to such a commitment ability is that they could commit to refusing transfer functions *off the equilibrium path* as well as on it. Recent work by Yamada (2003) shows that introducing such commitment ability into our model for the case of two players, changes the supportability results to make them more permissive. This makes it clear that we need to develop a deeper understanding about how the timing and commitment ability of refusing transfers affects equilibrium outcomes.

This also suggests considering bilateral contracting, where both players need to sign any contract before it becomes active. Note, however, that some of the intuition we have developed here already has some implications for such a bilateral bargaining setting. After a bilateral contract is signed, an agent may still have an incentive to make a unilateral offer that effectively undoes important aspects of contract and pushes things in (inefficient) directions that are to his or her advantage. Completely eliminating this problem could be done by allowing agents to come together and write a contract that says "no other contracts involving these parties are possible". Our analysis suggests that such exclusionary contracts would be helpful in reaching efficiency, as otherwise agents might make unilateral promises undoing aspects of bilateral contracts.

In any case, our results may be thought of as showing that it is critical to consider more complicated forms of bargaining and contracting in order to support efficient outcomes. This provides a rich agenda for further analysis.

Timing

In our analysis, we have considered only the simultaneous determination of transfer contracts. Let us argue that this is largely inconsequential. Suppose instead, for instance, that players alternate in announcing transfer functions, and that the game does not end until two periods with no moves, and that they may modify their transfer functions in any way in a given period. This would allow a player to respond to the others' contracts, and so the (modified) solo pay-offs are still relevant. Thus, if we end at any equilibrium of such a game, it must be that each player is still receiving at least their modified solo pay-offs. This leads to a direct extension of our results. This type of reasoning would apply to any sequential structure, so long as the agents could modify their transfers to react to the other player.

Thus, in order for timing to really be an issue it must either be that some players are restricted not to be able to respond to the contracts of others or else there must be some frictions in timing, for instance in the form of time discounting and some time or effort cost to writing contracts. But note that neither of these situations should generally improve efficiency, and in some cases might harm it.

Negative transfers

In our analysis players cannot make threats of violence (perhaps at a cost to all players) or steal from or tax other players.[24] Threats might be useful in reaching efficiency in some cases.

Let us make an important observation about the robustness of "positive" transfers vs. "negative" ones. The positive transfer contracts that we have considered here are immune to renegotiation since these contracts only involve transfers from one player to others. In contrast,

24. See Schelling (1960) for some interesting discussion of the role of such threats.

violence (and in some cases even stealing) will generally be costly for the player inflicting the negative transfer, and so *ex post* it may be that all players can benefit from renegotiation. In short, allowing for threats of violence, stealing, punishments, etc., might be a useful additional tool for supporting efficient outcomes, but further study is needed and this will involve some attention to *ex post* renegotiation that was not needed in the analysis here.

Contracts on contracts

There are two other aspects of the contracting that deserve further attention.

First, the contracts that we have considered are not contingent on the contracts offered by other players.[25] Allowing for such contingencies presents substantial technical hurdles in modelling, as when each contract is contingent upon the form of the other it results in a self-referential problem. This was first pointed out in the competing mechanisms literature (see McAfee (1993), Peck (1997), Epstein and Peters (1999)). Considering the impact of such contingent contracts is an important open and difficult problem in many contexts. As one can see from Epstein and Peters (1999), it has been a challenge even to prove that problems involving such contingencies are well-posed! A reasonable conjecture (based in part on the understanding of modelling that comes from Epstein and Peters, 1999) is that we might consider contracting on a game with an augmented action space (some $M \times X$, where M is derived endogenously and incorporates some aspects of the contracting but is pay-off irrelevant in the second-stage game). In that case, the basic results we have here would still go through, as the solo pay-offs would be unchanged. While this seems to be a reasonable conjecture, it appears to be difficult to prove.

The second issue related to contracts on contracts is viewing additional contracting stages before the larger game we have examined here. That is, one might also think of the two stage process that we have considered here as a game, and then consider contracting before it, and so on.[26]

Looking to mechanism design and implementation

We close by noting that our results also have important implications for the mechanism design and implementation literatures. Our results on the survivability of equilibria show that if the mechanism designer cannot control the side contracting of agents, then even if the mechanism is implementing efficient outcomes (when no side contracting is considered), the agents will have incentives to alter the workings of the mechanism through side contracts. Understanding the implementation problem in this broader context could provide very different conditions for implementability. It also raises questions such as which sorts of mechanisms are least susceptible to being undone by side payments. As such side contracting is available (and observed) in many situations, our results here suggest that this is an essential next step in the mechanism design and implementation literatures.[27]

APPENDIX

Proof of Theorem 1. Asking whether x survives is equivalent to asking whether $(x, v(x))$ is supportable (where $v(x)$ is the vector with i-th entry $v_i(x)$). Since x is a Nash equilibrium of the second-stage game, it follows from the definition of $t^{x,\bar{u}}$ that $t^{x,v(x)} = t^0$. This implies that $u^{ms}(x, v(x)) = u^s$, and then Theorem 1 follows from Theorem 3. ‖

25. Having contracts be contingent on which contracts are accepted by other players can also matter, as shown in Spiegler (2000).

26. See Lagunoff (1992) for such an approach in the context of selecting mechanisms.

27. This echoes a theme of Hurwicz (1994), who offers compelling arguments for viewing mechanisms in a larger natural context. He points out that we need to better understand a variety of factors, ranging from the enforceability of the outcomes, to the impact natural actions that are available to agents outside of those of the mechanism.

Proof of Theorem 2. We show that

$$u_i^{ms}(x,\overline{u}) \geq u_i^s \tag{A.1}$$

for any i and $x, \overline{u}$. Given (A.1), the theorem then follows from Theorem 3.

So let us now show (A.1). Consider any t_i. Let $\widehat{t}_i = t_i + t_{ji}^{x,\overline{u}}$. It follows that

$$t_{ij}(x') - t_{ji}^0(x') = \widehat{t}_{ij}(x') - t_{ji}^{x,\overline{u}}(x')$$

for every x'. This implies that the net transfers across players are identical under (t_{-i}^0, t_i) and $(t_{-i}^{x,\overline{u}}, \widehat{t}_i)$ and so $\mathrm{NE}(t_{-i}^0, t_i) = \mathrm{NE}(t_{-i}^{x,\overline{u}}, \widehat{t}_i)$. Thus, for each t_i there exists $\widehat{t}_i$ such that

$$\min_{\mu \in \mathrm{NE}(t_{-i}^{x,\overline{u}}, \widehat{t}_i)} \mathrm{EU}_i(\mu, t_{-i}^{x,\overline{u}}, \widehat{t}_i) = \min_{\mu \in \mathrm{NE}(t_{-i}^0, t_i)} \mathrm{EU}_i(\mu, t_{-i}^{x,\overline{u}}, t_i).$$

Since this is true for any t_i, it follows that

$$\sup_{t_i}\left[\min_{\mu \in \mathrm{NE}(t_{-i}^{x,\overline{u}}, t_i)} \mathrm{EU}_i(\mu, t_{-i}^{x,\overline{u}}, t_i)\right] \geq \sup_{t_i}\left[\min_{\mu \in \mathrm{NE}(t_{-i}^0, t_i)} \mathrm{EU}_i(\mu, t_{-i}^{x,\overline{u}}, t_i)\right],$$

which establishes (A.1). ‖

Proof of Theorem 3. Let us first show that if $(x, \overline{u})$ is supportable, then $\overline{u}_i \geq u_i^{ms}(x, \overline{u})$ for each i.

Suppose to the contrary that $\overline{u}_i < u_i^{ms}(x, \overline{u})$ for some i and $(x, \overline{u})$ is supportable. It follows that there exists some t_i such that

$$\overline{u}_i < \min_{\mu \in \mathrm{NE}(t_{-i}^{x,\overline{u}}, t_i)} \mathrm{EU}_i(\mu, t_{-i}^{x,\overline{u}}, t_i). \tag{A.2}$$

Let $\overline{t}$ be any set of transfers for which $(x, \overline{u})$ is supported. Note that, as argued in the text, it must be that $\overline{t}_j \geq t_j^{x,\overline{u}}$. Let $\widehat{t}_i = t_i + \overline{t}_j - t_j^{x,\overline{u}}$. It follows that

$$t_i(x') - t_j^{x,\overline{u}}(x') = \widehat{t}_i(x') - \overline{t}_j(x')$$

for every x'. This implies that the net transfers across players are identical under $(\overline{t}_{-i}, \widehat{t}_i)$ and $(t_{-i}^{x,\overline{u}}, t_i)$ and so $\mathrm{NE}(\overline{t}_{-i}, \widehat{t}_i) = \mathrm{NE}(t_{-i}^{x,\overline{u}}, t_i)$. Thus, from (A.2) it follows that

$$\overline{u}_i < \min_{\mu \in \mathrm{NE}(\overline{t}_{-i}, \widehat{t}_i)} \mathrm{EU}_i(\mu, t_{-i}^{x,\overline{u}}, \widehat{t}_i).$$

Let i deviate from $\overline{t}$ and announce $\widehat{t}_i$ in the first stage. It follows from the inequality above that the worst possible continuation pay-off in the subgame that follows is better than the expected continuation under $\overline{t}$. This contradicts the fact that $\overline{t}$ was played in the first stage of an equilibrium that supports $(x, \overline{u})$.

Next, let us show that if $\overline{u}_i \geq u_i^{ms}(x, \overline{u})$ for each i, then $(x, \overline{u})$ is supportable, and by $t^{x,\overline{u}}$.

Let us specify equilibrium strategies. In the first stage $t^{x,\overline{u}}$ is played and x is played in the second stage. A full specification of the equilibrium strategies includes specification of what happens off the equilibrium path as follows. If in the first stage player i plays $t_i^{x,\overline{u}}$ and player j plays $t_j \neq t_j^{x,\overline{u}}$, then in the second stage that follows the play is $\mu \in \mathrm{NE}(t_i^{x,\overline{u}}, t_j)$ that minimizes $\mathrm{EU}_j(\mu, t_i^{x,\overline{u}}, t_j)$ over $\mu \in \mathrm{NE}(t_i^{x,\overline{u}}, t_j)$. In a subgame following play of t such that $t_i \neq t_i^{x,\overline{u}}$ and $t_j \neq t_j^{x,\overline{u}}$, select any $\mu \in \mathrm{NE}(t)$. To see that this forms a subgame perfect equilibrium, note that by the definition of $t^{x,\overline{u}}$ it follows that if $t^{x,\overline{u}}$ is played in the first stage, then it is an equilibrium to play x in the second stage. So we need only show that there is no deviation away from $t^{x,\overline{u}}$ to $t_j \neq t_j^{x,\overline{u}}$ by some j. It follows from the definition of $u^{ms}(x, \overline{u})$ and our specification of off the equilibrium path behaviour that if any player j deviates from announcing $t_j^{x,\overline{u}}$ in the first stage then player j's pay-off will be no more than $u_j^{ms}(x, \overline{u})$. Since $\overline{u}_j \geq u_j^{ms}(x, \overline{u})$, it follows that this cannot be an improving deviation. ‖

Proof of Theorem 4. Let $M = 1 + \max_{i,x',x''}[v_i(x') - v_i(x'')]$. Fix a pure strategy Nash equilibrium x of the underlying game. Consider the transfer functions

$$t_{ij}(\widetilde{x}) = \begin{cases} 2\mathrm{M} & \text{if } \widetilde{x}_i \neq x_i \\ 0 & \text{otherwise.} \end{cases}$$

Under the above transfer functions it is a strictly dominant strategy for each player i to play x_i, and so x is a unique Nash equilibrium in the second-period game. Specify this behaviour on the equilibrium path, and off the equilibrium path choose any Nash equilibrium in the second stage. We need only show that a deviation to some $\widehat{t}_i$ by a player i is not

profitable for i. Such a deviation can only be improving if it leads to play of something other than x_{-i} by other players. (If only i changed actions, then i cannot do better given that x was a Nash equilibrium and $t_{ji}(\widehat{x}) = 0$ when $\widehat{x}_j = x_j$.) First, consider the case where a pure strategy Nash equilibrium $\widehat{x}$ is played in the second stage where $\widehat{x}_j \neq x_j$ for some $j \neq i$. Let there be $k \geq 1$ players $j \neq i$ such that $\widehat{x}_j \neq x_j$, and consider some such j. By playing $\widehat{x}$ player j's pay-off is

$$v_j(\widehat{x}) - (n-1)2\mathrm{M} + 2\mathrm{M}(k-1) + \widehat{t}_{ij}(\widehat{x}).$$

If j plays x_j instead, then j's pay-off is

$$v_j(\widehat{x}_{-j}, x_j) + 2\mathrm{M}(k-1) + \widehat{t}_{ij}(\widehat{x}_{-j}, x_j).$$

For $\widehat{x}_j$ to be a Nash equilibrium conditional on $\widehat{t}$, this implies that

$$\widehat{t}_{ij}(\widehat{x}) - \widehat{t}_{ij}(\widehat{x}_{-j}, x_j) \geq v_j(\widehat{x}_{-j}, x_j) - v_j(\widehat{x}) + (n-1)2\mathrm{M}.$$

Given our definition of M and the fact that $n - 1 \geq 2$, it follows that

$$\widehat{t}_{ij}(\widehat{x}) - \widehat{t}_{ij}(\widehat{x}_{-j}, x_j) > 3\mathrm{M}.$$

This implies that $\widehat{t}_{ij}(\widehat{x}) > 3\mathrm{M}$. This is true for any j with $\widehat{x}_j \neq x_j$. So, player i's utility in the new equilibrium is at most

$$v_i(\widehat{x}) - k3\mathrm{M} + k2\mathrm{M}.$$

For $k \geq 1$, the definition of M implies that this expression is less than $v_i(x)$. Thus, the deviation cannot be improving. ‖

Proof of Theorem 5. Consider x and a Nash equilibrium $\widehat{x}$ such that $v_i(x) \geq v_i(\widehat{x})$ for each i.

Set t as follows:

$$t_{ij}(\widetilde{x}) = \begin{cases} 2\mathrm{M} & \text{if } \widetilde{x}_{-i} = x_{-i} \text{ and } \widetilde{x}_i \neq x_i \\ 2\mathrm{M} & \text{if } \widetilde{x}_{-i} \neq x_{-i} \text{ and } \widetilde{x}_i \neq \widehat{x}_i \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify that $x \in \mathrm{NE}(t)$, as if i deviates then i pays M to each other player. To support $(x, v(x))$ have the strategies of the players be to play t in the first stage and x in the second stage. Specify off the equilibrium path strategies as follows. Conditional on a single player i deviating from t to some $\widehat{t}_i$ in the first stage, then play $\widehat{x}$ in the second period if $\widehat{x} \in \mathrm{NE}(t_{-i}, \widehat{t}_i)$ and otherwise play the worst Nash equilibrium for i out of $\mathrm{NE}(t_{-i}, \widehat{t}_i)$. Conditional on more than one player deviating from t in the first stage, play any Nash equilibrium in the resulting subgame.

To complete the proof of the theorem, we need only check that no player i can benefit by deviating to some $\widehat{t}_i$ in the first period. If $\widehat{x} \in \mathrm{NE}(t_{-i}, \widehat{t}_i)$, then the resulting play will be $\widehat{x}$ and so $t_{ji}(\widehat{x}) = 0$ for all $j \neq i$. Thus, the pay-off to i will be $v_i(\widehat{x}) - \sum_{j \neq i} \widehat{t}_{ij}(\widehat{x})$. Since this is less than $v_i(\widehat{x})$, it is less than $v_i(x)$ and cannot be a beneficial deviation. Thus, consider the case where $\widehat{x} \notin \mathrm{NE}(t_{-i}, \widehat{t}_i)$ but there is some pure strategy $\widetilde{x} \in \mathrm{NE}(t_{-i}, \widehat{t}_i)$. If $\widetilde{x} = x$ then it cannot be a beneficial deviation since $v_i(x) \geq v_i(x) - \sum_{j \neq i} \widehat{t}_{ij}(x)$.

We are left with the case where $\widetilde{x} \neq x$ and $\widetilde{x} \neq \widehat{x}$. Let us first show that it must be that $\widetilde{x}_k \neq x_k$ for at least two players k and j, with the possibility that $k = i$. To see this, suppose to the contrary that $\widetilde{x}_k \neq x_k$ for just one k. Given the definition of t_j for each $j \neq i$, it must be that i is paying at least $(2n-3)\mathrm{M}$ to each $j \notin \{i, k\}$ for whom $\widetilde{x}_j \neq \widehat{x}_j$ as otherwise j would rather play $\widehat{x}_j$. The transfers to i from each such j amount to at most M and are 0 from any other j. i also gets at most M from k. Thus, by the definition of M, this cannot be a beneficial deviation for i unless $x_{-i,k} = \widehat{x}_{-i,k}$. If $k = i$, then it must be that $x_{-i} = \widehat{x}_{-i}$ and $t_j(\widetilde{x}) = 0$ for all $j \neq i$. Since $\widehat{x}_i$ is a best response to $\widehat{x}_{-i}$ it follows that $v_i(\widehat{x}) \geq v_i(\widetilde{x})$, and so $v_i(\widehat{x}) \geq v_i(\widetilde{x}) - \sum_{j \neq i} \widehat{t}_{ij}(\widetilde{x})$, which implies that this could not be a profitable deviation. Therefore, the only such k must be some $k \neq i$, and thus $\widetilde{x}_{-k} = x_{-k}$. Thus, by the structure t_k for this to be a best response i must pay k at least $(2n-3)\mathrm{M}$ and gets M from k and 0 from other j's (for whom $\widetilde{x}_j = \widehat{x}_j$ as shown above). This cannot be profitable for i.

Thus we know that $\widetilde{x}_k \neq x_k$ for at least two distinct players, with the possibility that $k = i$. This means that $\widetilde{x}_{-j} \neq x_{-j}$ for each $j \neq i$ and so by the argument above we know that $\widetilde{x}_j = \widehat{x}_j$ for each $j \neq i$ in order for this to be a profitable deviation for i. This means that $t_j(\widetilde{x}) = 0$ for each $j \neq i$. However, then since $\widehat{x}_i$ is a best response to $\widehat{x}_{-i}$ it follows that $v_i(\widehat{x}) \geq v_i(\widetilde{x})$, and so $v_i(\widehat{x}) \geq v_i(\widetilde{x}) - \sum_{j \neq i} \widehat{t}_{ij}(\widetilde{x})$, which implies that this could not be a profitable deviation.

The extension to the case where in place of $\widetilde{x}$ there is a mixed strategy equilibrium is a straightforward extension of the above reasoning, working on the payments that are made in each realization of the support of the Nash equilibrium. ‖

Games with continuum actions

The major technical hurdle faced when the second-period game has infinite (pure) strategy spaces is finding the existence of a subgame perfect equilibrium in the two stage game. If discontinuous transfer functions are allowed (even off the equilibrium path) then there will be some subsequent subgames where no equilibrium exists. This presents a difficulty,

as even restricting attention to continuous transfer functions is then a problem, as it will not be a closed space. One must limit attention to some compact and convex set of transfer functions, for which there always exist second stage equilibria. With this approach, we describe here how the characterization theorems presented above hold in the continuum case.

Consider a game where X_i is a compact metric space and let $\Delta_i(X_i)$ denote the Borel measures on X_i. Let v_i be continuous on X for each i. Consider the set of continuous transfer functions $T = \times_i T_i$.[28,29]

Thus, $\mathrm{NE}(t)$ is non-empty and compact for each t.[30]

As in the finite case, define

$$u_i^{ms}(\bar{t}) = \sup_{t_i \in T_i}\left[\min_{\mu \in \mathrm{NE}(\bar{t}_{-i}, t_i)} \mathrm{EU}_i(\mu, \bar{t}_{-i}, t_i)\right].$$

Note that $\min_{\mu \in \mathrm{NE}(\bar{t}_{-i}, t_i)} \mathrm{EU}_i(\mu, \bar{t}_{-i}, t_i)$ is well defined since $\mathrm{EU}_i(\mu, \bar{t}_{-i}, t_i)$ is continuous and linear in μ, and $\mathrm{NE}(\bar{t}_{-i}, t_i)$ is non-empty and compact.

Say that $\bar{t} \in T$ supports $(x, \bar{u})$ if

- $x \in \mathrm{NE}(\bar{t})$ and
- $U_i(x, \bar{t}) = \bar{u}_i$ for all i.

We find the following theorem that covers any n.

Theorem A1. *$(x, \bar{u})$ is supportable if and only if there exists a supporting $\bar{t} \in T$ such that $\bar{u}_i \geq u_i^{ms}(\bar{t})$ for each i.*

Proof of Theorem A1. Let us first show that if $(x, \bar{u})$ is part of a subgame perfect equilibrium with supporting $\bar{t}$, then $\bar{u}_i \geq u_i^{ms}(\bar{t})$ for each i.

Suppose to the contrary that $\bar{u}_i < u_i^{ms}(\bar{t})$ for some i. It follows that there exists some t_i such that

$$\bar{u}_i < \min_{\mu \in \mathrm{NE}(\bar{t}_{-i}, t_i)} \mathrm{EU}_i(\mu, \bar{t}_{-i}, t_i). \tag{A.3}$$

If player i deviates to play t_i, then for any μ that follows in the subgame, i will benefit. This contradicts the fact that $(x, \bar{u})$ is supported by $\bar{t}$.

Next, let us show that if $\bar{u}_i \geq u_i^{ms}(\bar{t})$ for each i, then $(x, \bar{u})$ is supportable.

Let us specify equilibrium strategies. In the first stage $\bar{t}$ is played and x is played in the second stage. If in the first stage some player i plays $t_i \neq \bar{t}_i$, then in the subgame that follows the play is $\mu \in \mathrm{NE}(\bar{t}_{-i}, t_i)$ that minimizes $\mathrm{EU}_i(\mu, \bar{t}_{-i}, t_i)$. In any other subgame select any μ. To see that this forms a subgame perfect equilibrium, note that by the support of $(x, \bar{u})$ by $\bar{t}$ it follows that if $\bar{t}$ is played in the first stage, then it is an equilibrium to play x in the second stage. So we need only show that there is no deviation away from $\bar{t}$ to $t_i \neq \bar{t}_i$ by some i. It follows from the definition of $u^{ms}(\bar{t})$ and our specification of off the equilibrium path behaviour that if any player i deviates from announcing $\bar{t}_i$ in the first stage, then player i's pay-off will be no more than $u_i^{ms}(\bar{t})$. Since $\bar{u}_i \geq u_j^{ms}(\bar{t})$, it follows that this cannot be an improving deviation. ‖

Acknowledgements. We thank Ken Hendricks, Philippe Jéhiel, Ehud Kalai, Roger Lagunoff, Bentley MacLeod, Nolan Miller, Hakan Orbay, Mike Peters, Michael Whinston, and seminar participants at the University of Arizona, Caltech, University of Texas, University of Toronto, U.B.C., USC, and the Decentralization Conference for helpful comments. We also thank the editor and anonymous referees for helpful suggestions. Financial support under NSF grants SES-9986190, SES-9986676, and SES-0316493 is gratefully acknowledged.

REFERENCES

AGHION, P. and BOLTON, P. (1987), "Contracts as Barriers to Entry", *American Economic Review*, **77** (3), 388–401.
ALLEN, B. (2001), "Supermechanisms" (Mimeo, University of Minnesota).
ANDERLINI, L. and FELLI, L. (2001), "Costly Bargaining and Renegotiation", *Econometrica*, **69**, 377–412.

28. One could use a more general space. Any space T for which $\mathrm{NE}(t)$ is non-empty and closed will work. In that case one may need to replace $\min_{\mu \in \mathrm{NE}(t)} \mathrm{EU}_i(\mu, t)$ in the definition of u_i^{ms} with inf, and make some corresponding adjustments in the proof of Theorem A1.

29. Note that we are not making assumptions on T that guarantee existence of an equilibrium in the overall two stage game, as for instance T need not be compact. We are simply making assumptions that will be enough to prove Theorem A1. This will be enough to guarantee that equilibrium will exist when the necessary conditions are satisfied, which then makes the necessary conditions necessary and sufficient and so we will get existence in that way.

30. In that case, $U_i(x, t)$ is continuous in x for each i, and then $\mathrm{EU}_i(\mu, t)$ is continuous and quasi-concave (in fact linear) in μ. Then by a theorem of Debreu, Fan, and Glicksberg (*e.g.* see Fudenberg and Tirole, 1991) there exists a Nash equilibrium of the game with t fixed. Closure of the set of Nash equilibria (using weak convergence) then follows easily from the continuity of $U_i(x, t)$ in x.

BALIGA, S. and SJÖSTRÖM, T. (1995), "Interactive Implementation", *Games and Economic Behavior*, **27**, 38–63.
BARBERA, S. and JACKSON, M. O. (2004), "Choosing How to Choose: Self-Stable Majority Rules and Constitutions", *Quarterly Journal of Economics*, **119** (3), 1011–1048.
BENSAID, B. and GARY-BOBO, R. J. (1996), "An Exact Formula for the Lion's Share: A Model of Pre-Play Negotiation", *Games and Economic Behavior*, **14**, 44–89.
BERNHEIM, B. D. and WHINSTON, M. D. (1986), "Menu Auctions, Resource Allocation, and Economic Influence", *Quarterly Journal of Economics*, **101**, 1–31.
COASE, R. H. (1960), "The Problem of Social Cost", *The Journal of Law and Economics*, **3**, 1–44.
DANZIGER, L. and SCHNYTZER, A. (1991), "Implementing the Lindahl Voluntary-Exchange System", *European Journal of Political Economy*, **7**, 55–64.
DUTTA, P. and RADNER, R. (2001), "Global Warming and Technological Change" (Mimeo, NYU).
EPSTEIN, L. and PETERS, M. (1999), "A Revelation Principle for Competing Mechanisms", *Journal of Economic Theory*, **88**, 119–160.
FERSHTMAN, C., JUDD, K. and KALAI, E. (1991), "Observable Contracts, Strategic Delegation, and Cooperation", *International Economic Review*, **32**, 551–559.
FUDENBERG, D. and TIROLE, J. (1991) *Game Theory* (Cambridge, MA: MIT Press).
GUTTMAN, J. M. (1978), "Understanding Collective Action: Matching Behavior", *American Economic Review*, **68**, 251–255.
GUTTMAN, J. M. (1987), "A Non-Cournot Model of Voluntary Collective Action", *Economica*, **54**, 1–19.
GUTTMAN, J. M. and SCHNYTZER, A. (1992), "A Solution of the Externality Problem Using Strategic Matching", *Social Choice and Welfare*, **9**, 73–88.
HARRIS, C. (1985), "Existence and Characterization of Perfect Equilibrium in Games of Perfect Information", *Econometrica*, **53**, 613–628.
HURWICZ, L. (1994), "Economic Design, Adjustment Processes, Mechanisms and Institutions", *Economic Design*, **1**, 1–14.
JACKSON, M. O. (2001), "A Crash Course in Implementation Theory", *Social Choice and Welfare*, **18**, 655–708.
JACKSON, M. O. (2003), "Mechanism Theory", in U. Derigs (ed.) *Optimization and Operations Research, Encyclopedia of Life Support Systems* (Oxford, U.K.: EOLSS) `http://www.eolss.net`.
KALAI, E. (1981), "Preplay Negotiations and the Prisoners' Dilemma", *Mathematical Social Sciences*, **1**, 375–379.
KALAI, E. and SAMET, D. (1985), "Unanimity Games and Pareto Optimality", *International Journal of Game Theory*, **14**, 41–50.
LAGUNOFF, R. (1992), "Fully Endogenous Mechanism Selection on Finite Outcome Sets", *Economic Theory*, **2**, 462–480.
MACLEOD, W. B. (2002), "Complexity and Contract", in E. Brousseau and J.-M. Glachant (eds.) *The Economics of Contract in Prospect and Retrospect* (Cambridge: Cambridge University Press).
MCAFEE, P. (1993), "Mechanism Design by Competing Sellers", *Econometrica*, **61**, 1281–1312.
MILLER, N. H. and PAZGAL, A. I. (2001), "The Equivalence of Price and Quantity Competition with Delegation", *RAND Journal of Economics*, **32**, 284–301.
PECK, J. (1997), "Competing Mechanisms and the Revelation Principle" (Unpublished Manuscript, Ohio State University).
PRAT, A. and RUSTICHINI, A. (2003), "Games Played Through Agents", *Econometrica*, **71**, 989–1026.
QIN, C.-Z. (2002), "Penalties and Rewards as Inducements to Cooperate" (Mimeo, U.C. Santa Barbara).
RAY, D. and VOHRA, R. (1997), "Equilibrium Binding Agreements", *Journal of Economic Theory*, **73**, 30–78.
SCHELLING, T. (1960) *The Strategy of Conflict* (Cambridge, MA: Harvard University Press).
SEGAL, I. (1999), "Contracting with Externalities", *Quarterly Journal of Economics*, **114** (2), 337–388.
SEGAL, I. and WHINSTON, M. (2003), "Robust Predictions for Bilateral Contracting with Externalities", *Econometrica*, **71**, 757–791.
SPIEGLER, R. (2000), "Extracting Interaction-Created Surplus", *Games and Economic Behavior*, **30**, 142–162.
VARIAN, H. R. (1994*a*), "Sequential Provision of Public Goods", *Journal of Public Economics*, **53**, 165–186.
VARIAN, H. R. (1994*b*), "A Solution to the Problem of Externalities when Agents are Well-Informed", *American Economic Review*, **84**, 1278–1293.
YAMADA, A. (2003), "Efficient Equilibrium Side Contracts", *Economics Bulletin*, **3** (6), 1–7.