INTERNATIONAL ECONOMIC REVIEW
Vol. 32, No. 3, August 1991

OBSERVABLE CONTRACTS: STRATEGIC DELEGATION AND COOPERATION*

BY CHAIM FERSHTMAN, KENNETH L. JUDD, AND EHUD KALAI[1]

The role of commitments in noncooperative games is well acknowledged and documented. One way to achieve commitments is by letting delegates represent the players of a game. In this paper we study a delegation game in which the players can use agents strategically to play on their behalf and the contracts they sign with them are common knowledge. We show that in such cases every Pareto optimal outcome of the game can become the unique subgame perfect Nash equilibrium of the delegation game. We demonstrate this result by discussing the Cournot-type duopolistic game.

1. INTRODUCTION

It is a common observation in conflict situations that players are quite frequently represented by agents who play the game on their behalf. Lawyers often represent clients in negotiation, agents represent actors and sports players, managers represent owners, and elected officials represent voters. Thus, in discussing conflict situations in social science, specific attention must be paid to the possibility of hiring agents who participate in the game on behalf of the real players.

Besides considering the implications of such delegation, a fundamental question is the explanation of this phenomenon. Why does a player hire someone who will represent him in a game? Clearly one possible explanation is that there are games in which having special skills is essential. But can we argue that the only purpose in hiring a lawyer is always his superior knowledge of the law? Or is it possible that players can gain strategic advantage by having someone with different incentives play the game on their behalf?

The potential benefits of using delegates as credible commitment has already been emphasized by Schelling (1956, 1960). Considering, for example, the Nash bargaining problem, it is already known that distorting the player's utility function might benefit the player e.g., Kannai (1977), Crawford and Varian (1979), Kihlstrom, Roth and Schmeidler (1980), and Sobel (1981). This result can be discussed within a framework introduced by Kurz (1977, 1980) in which the original game is transformed to a noncooperative distortion game in which players' strategies consist of utility functions that may be distorted from their true utilities for strategic reasons. Clearly, sending an agent can be equivalent to *credibly* reporting a distorted utility function, providing, of course, that the agent indeed has such a utility function and his utility function is public information. The benefits from the

* Manuscript received September 1989; revised March 1990.

[1] The authors wish to thank the participants of the Game Theory Workshop held at Tel Aviv University and two anonymous referees for helpful comments. This research was partly supported by NSF Economics Grants Nos. SES-8409798 and SES-8606581.

use of such an agent should, however, exceed the cost of obtaining the agent's services. Similarly, if we consider the strategic bargaining game (Rubinstein 1982) it is clear that a player can benefit by sending a representative who is less impatient than he is. In an oligopoly framework it is already recognized that sending managers with distorted objective functions might benefit the owners of the firm and can be a part of an equilibrium behavior (e.g., Vickers 1985, Fershtman 1985, Fershtman and Judd 1986, 1987, and Sklivas 1987).

The main purpose of this paper is to analyze the extent to which the set of equilibria of a strategic game changes when agents are strategically allowed to represent the main players. The outcome of such an analysis depends crucially on the degree of commitment we allow in the game. Once the players are able to commit to certain contracts we are not entirely in the world of noncooperative games. The main question is, of course, whether, in order to achieve the collusive outcome, we should go all the way to cooperative games and allow principals to sign enforceable contracts one for the other. The main conclusion of this paper is that even in highly noncooperative games, cooperative outcomes emerge as equilibria in the game with delegation, providing that each principal is fully committed to the contract he signed with his agent and the contracts are fully observed.[2]

Our delegation game is described as follows. To every two person strategic game we associate a delegation game in which agents play the original game on behalf of the original players—their principals. More specifically, we formulate a two-stage game. In the first stage each principal provides his agent with a compensation scheme. These compensation schemes determine each agent's final reward as a function of the principal's payoffs. In the second stage, each agent, after learning both compensation schemes, plays the original game, choosing his principal's strategy so as to maximize his own final payoff. The principals then receive their payoffs in the original game, net of promised compensation to their agents.

Our formulation differs from the delegation and distortion game models cited above in one important aspect. We assume that contracts between principals and their agents are fully observed and thus can be conditioned upon in the agents' game. Clearly, such an assumption implies that we are setting a higher degree of commitment than in previous works on delegation but what is interesting to see is that we need *exactly* this extra commitment in order to implement the collusive outcome by delegation. In other words, in order to achieve the cooperative outcome we do not have to go all the way to models in which contract between the main players can be signed and enforced. It is sufficient to have the possibility of hiring agents providing that the contracts signed with them are public information.

The result in this paper contrasts strongly with the literature in which it is argued that in the Cournot oligopoly case delegation equilibrium leads to a more competitive equilibrium (see, e.g., Vickers 1985, Fershtman 1985, Fershtman and Judd 1987, and Sklivas 1987). In these works, however, it is assumed that the agents' game strategies are not conditional upon the compensation scheme agents receive

[2] For the analysis of delegation games in which compensation schemes are private information see Myerson (1982) and Katz (1987).

from their principals. In the agents' game each agent chooses his strategy as the best response to the other agents' strategies independent of the compensation schemes other agents have. By not allowing strategies to be conditioned on compensation, agents do no have the ability to "punish" principals for giving the "wrong" compensation. Clearly, the fact that by letting agents condition their strategies on the observed compensation scheme we attain the collusive outcome illustrates the importance of such an assumption.

One can also contrast our work with previous results in game theory. It is already known that once the game is repeated, every individual rational feasible outcome of the underlying game is an equilibrium of the repeated game. (See, for example, Aumann 1981, Rubinstein 1979, Friedman 1985, and Benoit and Krishna 1985.) In our setting the game is played only once, but the strategic use of agents plus the ability to commit to them enable players to obtain the cooperative outcome.

We illustrate our result by considering a Cournot-type duopolistic game in which we allow firms to hire managers who run the firms on their behalf. The role of the managers in our setting is to choose the quantity produced by the firm. We show that even when we consider a one-shot game such that the two firms meet only once in the market, every collusive outcome can be obtained in such a duopoly game with delegation.

2. THE DELEGATION GAME

Our analysis of delegation begins with a game representing the basic conflict. The *underlying game* is a 2-player strategic form game with the set of players $P = (p_1, p_2)$. We let $S = S_1 \times S_2$ be the set of *strategy combinations* in this game. The payoffs of the players are described by a *utility function* $u = (u_1, u_2)\colon S \to \mathbb{R}^2$. We will use $G = (P, S, u)$ to denote this underlying game.

Nash equilibrium of this game is a strategy combination (s_1^*, s_2^*) with $u_1(s_1^*, s_2^*) \geq u_1(s_1, s_2^*)$ for every $s_1 \in S_1$, and similarly for player 2. We consider in this paper only games for which there is at least one (pure strategy) Nash equilibrium.

For such a game we define the associated *delegation game*, D, as follows: the set of players $N = (p_1, p_2, a_1, a_2)$ where p_1 and p_2 are called *principals* and a_i is called the *agent* or *delegate* of p_i.

The set of strategies of p_i is

$$C_i = \{c_i\colon \mathbb{R} \to \mathbb{R}_+ : c_i \text{ is weakly monotonically increasing}\}.$$

We refer to an element of C_i as a *compensation function* of agent a_i. Note that agent a_i's compensation depends only on his principal's gross payoff. We restrict the compensation functions to be weakly monotonic. Besides being intuitively appealing, there is a technical need for such a restriction. Subgame perfection cannot be obtained without the weak monotonicity assumption. It guarantees the existence of an equilibrium in every conceivable agents' game since it *preserves* the (pure strategy) equilibria of the original game.

We assume that the contracts signed between each principal and his agent are public information. Moreover, each principal is fully committed to this contract.

There is no hidden contract that the principal and his agent agreed upon. Thus, when an agent comes to make his choice of an action he is already informed about $(c_1, c_2) \in C \equiv C_1 \times C_2$. We therefore define the agents' strategies as follows.

The strategy set of every agent a_i is $R_i = \{r_i: C \rightarrow S_i\}$ and we let $R = R_1 \times R_2$. We call an element of R_i a *response function* of agent i.

Given a 4-tuple of strategies (c_1, c_2, r_1, r_2) we define the utilities of the four players in D as follows:

$$U_i^P(c_1, c_2, r_1, r_2) = u_i(r(c)) - c_i(u_i(r(c))),$$

and

$$U_i^a(c_1, c_2, r_1, r_2) = c_i(u_i(r(c))).$$

Given the strategies (c_1, c_2, r_1, r_2) the agents' choice of actions is $r(c) = (r_1(c), r_2(c))$. The ith principal thus gets the game payoffs $u_i(r(c))$ minus the compensation he pays to his agent, i.e., $c_i(u_i(r(c)))$.

DEFINITION 1. $(r_1^*, r_2^*) \in R$ *is a* Nash equilibrium *in the agents' game induced by* $(c_1, c_2) \in C$ *if* $U_1^a(c_1, c_2, r_1^*, r_2^*) \geq U_1^a(c_1, c_2, r_1, r_2^*)$ *for every* $r_1 \in R_1$, *and similarly for agent* (2).

We let $EA(c) = \{(r_1^*(c), r_2^*(c)) \in S$: (r_1^*, r_2^*) *is an equilibrium in the agents' game induced by* $c\}$.

DEFINITION 2. $(c_1^*, c_2^*, r_1^*, r_2^*)$ *is a* subgame perfect *Nash equilibrium of D if:*

i) $U_1^P(c_1^*, c_2^*, r_1^*, r_2^*) \geq U_1^P(c_1, c_2^*, r_1^*, r_2^*)$ *for every* $c_I \in C_I$ *and the same for principal 2; and*

ii) *for every pair of compensation functions* $c \in C$, (r_1^*, r_2^*) is a Nash equilibrium of the agents' game induced by c, i.e., $r^*(c) \in EA(c)$.

Notice that the weak monotonicity of the compensation functions guarantees that every Nash equilibrium s^* of the underlying game G is in $EA(c)$ for every c. However, $EA(c)$ will typically contain other equilibria as well. Moreover, without monotonicity or other similar restrictions, obtaining subgame perfection is impossible. It is always possible to construct "wild" compensation schemes for which there is no equilibrium in the induced agents' game.

DEFINITION 3. c implements $u \in \mathbb{R}^2$ *via* $r \in \mathbb{R}$ *if* (c, r) *is a subgame perfect equilibrium of D with* $u(r(c)) = u$.

Under this definition, when u is implemented, u_i is shared by principal i and his agent. Given the above definition, one can prove trivially that if s^* is a Nash equilibrium of the underlying game, every feasible payoff pair $u \geq u(s^*)$ is implementable in the delegation game.[3]

The above definition of implementation might, however, be intuitively unattractive. It disregards the usual difficulty of dealing with multiple equilibria as well as

[3] The proofs are trivial and can be found in Fershtman, Judd and Kalai (1987).

a choice of Pareto dominated equilibrium in the agents' game. In cases of multiplicity the principals have to count on the agents choosing a particular equilibrium. It would, however, be naive on the part of the principals to expect the agents to choose the exact actions when the agents are indifferent between several actions or when $EA(c)$ is not a singleton and in particular when there is an element in $EA(c)$ that from the agents' point of view, Pareto dominates the equilibrium the principals wish them to choose.

Game theory, however, does not yet have the proper equilibrium concept to deal with such conceptual difficulties. We thus reconstruct our formulation of the delegation game and the implementation concept in order to exclude some unattractive phenomena of this type.

Multiplying all the payoffs of a player in a strategic game by a positive constant results in a new game, isomorphic to the original one. For our delegation game, an implication of this observation is important. It is reasonable to assume that if the principals change a compensation scheme (c_1, c_2), say, to $(.5c_1, c_2)$, the strategies of the agents will stay the same, yielding principals one and two the same payoffs from the underlying game. Yet, principal one would have to spend half as much on compensating his agents. Since this argument can be applied repeatedly, no equilibrium with positive payments to agents exists. As in the standard principal-agent literature, we resolve this problem by assuming that there is a $\varepsilon > 0$ which is the smallest amount of money that the agents are willing to accept in order to participate in the game.

DEFINITION 4. *c implements u via r* with mutual rational agents *if in addition to implementing u, r satisfies the following property: for every $\bar{c} \in C$, $\bar{c}(u(r(\bar{c}))) \geq (\varepsilon, \varepsilon)$ whenever there is $s \in EA(\bar{c})$ with $\bar{c}(u(s)) \geq (\varepsilon, \varepsilon)$.*

Intuitively, the above definition requires that if there is an equilibrium in the agent game in which both agents get at least ε they will choose such an equilibrium. Moreover, such a choice is made at every subgame, i.e., for every pair of compensation schemes. The mutual rationality condition is an assumption on the agents' selection among multiple Nash equilibria. Agents will coordinate their actions in order to avoid zero payoffs if possible. Once the agent's equilibrium payoff is less than ε he does not participate, which implies that the principal has to make his choice of action. In this case, we are back to our original game in which s^* is an equilibrium.

DEFINITION 5. *c* uniquely implements *u with mutual rationale agents if there is an $r \in R$ such that c implements u via r with mutual rational agents and if for some $\bar{u}$ and $\bar{r} \in R$, c implements $\bar{u}$ via $\bar{r}$ with mutual rational agents then $\bar{u} = u$.*

Unique implementations are attractive since they guarantee the principals the vector u without depending on a specific choice of equilibrium by their agents. Thus, once we require unique implementation, the problem of multiplicity is resolved. There still might be multiple equilibria in the agents' game, but all these equilibria yield the same payoffs.

DEFINITION 6. *We say that T_i is a* target compensation function *if for every u_i, $T_i(u_i) \in \{0, \varepsilon\}$.*

Target compensation functions pay nothing unless a minimal level of utility is obtained for the principal and pay ε if that target level is obtained or exceeded. We are now ready to state our main result.

THEOREM 1. (Folk theorem in delegation games.) *If u is (strictly) Pareto optimal in G and for $i = 1, 2$, $u_i - u_i(s^*) > \varepsilon$ for some Nash equilibrium s^* of G, then u is uniquely implementable with mutual rational agents. Moreover, the implementation can be done by target compensation functions.*

PROOF. Let $T_{u_i} \in C_i$ be a target compensation function defined by

$$T_{u_i}(\bar{u}_i) = \begin{cases} \varepsilon, & \text{if } \bar{u}_1 \geq u_i \\ 0, & \text{otherwise,} \end{cases}$$

and define $r(\bar{c})$ as follows:

i) if there is an $s \in EA(\bar{c})$ with $\bar{c}(u(s)) \geq (\varepsilon, \varepsilon)$ then let $r(\bar{c})$ be such an s;

ii) otherwise, let $r(\bar{c}) = s^*$.

We will show that (T_u, r) uniquely implements u with mutual rational agents. It is clear that $u(r(T_u)) = u$. By playing the above strategies the agents' payoffs is $(\varepsilon, \varepsilon)$ and by the fact that u is strictly Pareto optimal in G it is the only outcome such that $T_u(u) \geq (\varepsilon, \varepsilon)$. By construction, $r(c) \in EA(c)$ for every $c \in C$ and it is mutually rational. Now we check that T_{u_1} is indeed a best response strategy to (T_{u_2}, r_1, r_2).

Suppose that principal one deviates and plays $c_1 \in C_1$. One of the following two cases must hold.

a) There exists $s \in EA(c_1, T_{u_2})$ such that $(c_1(u_1(s)), T_{u_2}(u_2(s))) \geq (\varepsilon, \varepsilon)$. By the Pareto optimality of u it follows that principal one cannot be better off from this deviation.

b) Otherwise, the outcome of the game is s^* and $u_1(s^*) < u_1 - \varepsilon$.

T_u uniquely implements u with mutual rational agents since every $s \in EA(T_u)$ with $c(u(s)) \geq (\varepsilon, \varepsilon)$ has $u(s) = u$. □

The reason that the delegation game "works" is that through their agents (and their compensation schemes) the principals can commit to playing cooperative strategies (with enough flexibility to allow them to retaliate against deviations). When the principals cannot commit directly to each other (such as in a one-shot game and when binding agreements are not permissible) the ability to commit to their agents is very beneficial as was illustrated in the theorem. One could question the ability of the principals to commit to their agents and actually pay them according to the promised compensation schemes. This commitment, however, is credible when binding agreements between them are possible and legally enforceable. Also, a long-term relationship between the principal and an agent would enhance these commitment possibilities even when the principal is using the agent in a sequence of games with different opponents.

One could conceive of defining an agent compensation scheme on other parameters of the game. For example, an agent's compensation could depend on the difference in the payoff of his principal minus the payoff of his opponent's principal, or even on combinations of payoffs and actions in the underlying game. These modifications are possible but they only serve to complicate a simple idea. They may also come at a cost of losing the subgame perfection property of the implementations, since they may create a large number of unreasonable subgames.

3. EXAMPLE: DUOPOLY GAME

To highlight the main feature of our analysis we next discuss the Cournot duopoly game. Consider a duopolistic market in which the inverse demand function is given by $p = a - b(q_1 + q_2)$ where p is the market price and q_i is the ith firm's output. Assume that the cost function is given by $TC_i(q_i) = mq_i$ where $a > m \geq 0$ is the constant marginal cost. The ith firm's profit function is thus $\pi_i(q_1, q_2) = q_i(a - b(q_1 + q_2)) - mq_i$. The unique equilibrium of such a one-shot game is $q_1^* = q_2^* = (a - m)/3b$ and the equilibrium payoffs are $\pi_i(q_1^*, q_2^*) = (a - m)^2/9b$.

Now let us consider the delegation game in which the owner of each firm hires a manager who will make the choice of q_i on his behalf. The owner and the manager sign a contract that specifies the manager's compensation as a function of the firm's performance.

Let $\hat{q}_1 = \hat{q}_2 = (a - m)/4b$ and $\hat{\pi}_i = (a - m)^2/8b$ be the maximum symmetric collusive output levels and the collusive payoffs, respectively.

PROPOSITION 1. *In the above duopolistic game the collusive outcome,* $(\hat{q}_1, \hat{q}_2)$, *can be implemented uniquely with mutual rational agents if* $\hat{\pi}_i - \pi_i(q_1^*, q_2^*) > \varepsilon$.

Proposition 1 is an immediate corollary of Theorem 1. However, in order to get some intuition on the type of strategies and contracts that give rise to such an equilibrium we will specify the players's strategies.

Let $T_{\hat{\pi}_i}$ be the following target compensation function

$$T_{\hat{\pi}_i}(\pi_i) = \begin{cases} \varepsilon, & \text{if } \pi_i \geq \hat{\pi}_i \\ 0, & \text{otherwise} \end{cases} \qquad i = 1, 2.$$

Let the managers' strategies for a specific pair of compensation schemes c be as follows: when possible, the managers will choose a pair of output rates q which are a Nash equilibrium in the mangers' game induced by c and $c(\pi(q)) \geq (\varepsilon, \varepsilon)$; otherwise they will choose the Cournot production rates q_i^*.

Clearly, the above construction implies that once the compensation scheme $T_{\hat{\pi}_i}$ are given the managers will choose the only production rate that yields the payoffs $(\varepsilon, \varepsilon)$ which are the collusive production rate $(\hat{q}_1, \hat{q}_2)$. Also, given the response of the managers to various incentive schemes, the owners can do no better than use these target schemes. Subgame perfection of the above equilibrium means that the owners can be convinced that if they deviate from their compensation scheme the managers will indeed respond as stated.

It is interesting to observe that the principals' commitment to cooperate through their agents is done by sending out agents who are less "hungry" than themselves. This is done by flattening the agents' compensation from some critical level of the principals' utility.

This Cournot analysis should be compared with previous analyses of competing principal-agent pairs. Fershtman and Judd (1987), Vickers (1985), and Sklivas (1987) examined the same game but restricted contracts to be linear in profits and sales; in this case the owners' profits were less than in the underlying game's Nash equilibrium. As was stated previously, the difference between these works and ours is that we allow agents to condition their strategy on the two compensation schemes which are assumed to be public information. Proposition 1 is not correct without such an assumption.

In Fershtman and Judd (1986) it was assumed that the manager's effort was unobserved, creating a moral hazard problem; again, the outcome was not a cooperative one for the principals. From the latter analysis it is clear that the folk theorem of our analysis may break down when contracts must deal with moral hazard problems with managers as well as coordinate cooperation with opposing principals. The value of delegation in incomplete information games is an open and interesting question.

4. CONCLUDING REMARKS

In this paper we have demonstrated that once principals are able to sign contracts with agents who choose an action on their behalf and that these contracts are fully observable by all agents, then in such extended games the principals can obtain a Pareto outcome as an equilibrium. Having the ability to have such a commitment implies that the game does not fall exactly into the category of noncooperative games and should be classified in between cooperative and noncooperative games. Since casual observation indicates that delegates are occasionally used in daily life there is clearly a need for a detailed investigation of such a class of games. In this paper we did not discuss all the possible complex relationships that the existence of delegates introduces. The agents in our model are used solely for the purpose of strategic delegation. They do not have any expertise, information, or any other advantage vis-à-vis their principals. We thus concentrate on one specific and unexplored aspect of the delegation issue. Clearly, possible generalizations of our analysis should integrate strategic delegation with the standard principal-agent theory.

Tel Aviv University, Israel, and Kellogg Graduate School of Management, Northwestern University, U.S.A.

The Hoover Institution, Stanford University, U.S.A.

Kellogg Graduate School of Management, Northwestern University, U.S.A.

REFERENCES

AUMANN, R. J. "Survey on Repeated Games," in *Essays in Game Theory and Mathematical Economics* (Mannheim: Bibliographsiches Institut, 1981), 11–42.

BENOIT, J. P. AND V. KRISHNA, "Finitely Repeated Games," *Econometrica* 53 (1985), 905–922.

CRAWFORD, V. P. AND H. R. VARIAN, "Distortion of Preferences and the Nash Theory of Bargaining," *Economics Letters* 3 (1979), 203–206.

FERSHTMAN, C., "Managerial Incentives as a Strategic Variable in a Duopolistic Environment," *International Journal of Industrial Organization* 3 (1985), 245–253.

——— AND K. JUDD, "Strategic Incentive in Manipulation in a Rivalrous Agency," Technical Report No. 496, IMSSS, Stanford University, 1986.

——— AND ———, "Incentive Equilibrium in Oligopoly," *American Economic Review* 77 (1987), 927–940.

———, ——— AND E. KALAI, "Cooperation Via Delegation," Research Report No. 163, Hebrew University of Jerusalem, 1987.

FRIEDMAN, J. W., "Cooperative Equilibria in Finite Horizon Noncooperative Supergames," *Journal of Economic Theory* 35 (1985), 390–398.

KANNAI, Y., "Concavifiability and Construction of Concave Utility Functions," *Journal of Mathematical Economics* 4 (1977), 1–56.

KATZ, MICHAEL L., "Game-Playing Agents: Contracts as Pre-Commitments," manuscript, University of California, Berkeley, 1987.

KIHLSTROM, R. E., A. E. ROTH AND D. SCHMEIDLER, "Risk Aversion and Nash's Solution to the Bargaining Problem," in O. Moeschlin and D. Pallaschke, eds., *Game Theory and Mathematical Economics* (Amsterdam: North Holland, 1981).

KURZ, M., "Distortion of Preferences, Income Distribution and the Case for Linear Income Tax," *Journal of Economic Theory* 14 (1977), 291–298.

———, "Income Distribution and Distortion of Preferences: The l-Commodity Case," *Journal of Economic Theory* 22 (1980), 99–106.

MYERSON, R., "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics* 10 (1982), 67–81.

RUBINSTEIN, A., "Equilibrium in Supergames with the Overtaking Criterion," *Journal of Economic Theory* 21 (1979), 1–9.

———, "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50 (1982), 97–109.

SCHELLING, T. C., "An Essay on Bargaining," *American Economic Review* 46 (1956), 281–306.

———, *The Strategy of Conflict* (New York: Oxford University Press, 1960).

SKLIVAS, S. D., "The Strategic Choice of Management Incentives," *The Rand Journal of Economics* 18 (1987), 452–458.

SOBEL, J., "Distortion of Utilities and the Bargaining Problem," *Econometrica* 49 (1981), 597–617.

VICKERS, J., "Delegation and the Theory of the Firm," *Economic Journal* (supplement) 95 (1985), 138–147.