

Sustaining Cooperation Through Voluntarily Offering of “Hostages”*

Cheng-Zhong Qin
Department of Economics
University of California
Santa Barbara, CA 93106

December 16, 1997

Abstract

A hostage from a player is offered with respect to a particular action. It is defined here to be a quantity of an asset of the player, held in escrow by a third party, that will be given to the opponent if the player takes the action. For a prisoner’s dilemma game, it is shown that to support cooperation in subgame perfect equilibrium, the values of hostages must be great enough (bounded below) to induce each to cooperate, but small enough (bounded above) to prevent each either from inducing the other to defect or from making it too costly for the player to defect when the opponent defects. Necessary and sufficient conditions for the existence of subgame perfect equilibrium with cooperation are equivalent to the lower bound being less than or equal to the upper bound.

1 Introduction

Central in prisoner’s dilemma is that gains from cooperation are difficult to achieve. The standard approach for achieving cooperation is based on punishment. Punishment generally does not give rise to cooperation in one-shot

*Helpful comments by UCSB seminar participants, especially those by Ted Bergstrom, Rod Garratt, and Steve Trejo are gratefully acknowledged.

or finitely repeated prisoner's dilemma games,¹ but can support cooperation if the number of repetition is infinite. A problem with using infinite repetitions to support cooperation, however, is that the "folk theorem" then implies that the set of equilibria is large; that is, nearly any feasible and individually rational outcome can be supported in equilibrium.²

In this paper, I consider a different approach for achieving cooperation. The approach is based on the holding of hostages.³ The approach has been used in practice since antiquity and variants of the approach are widely used today. An institutional difference between ancient and modern applications is that hostages in ancient times were generally physical persons, whereas deposits of money and physical capital fill the economic role of hostages today.

To illustrate the flavor of the approach, consider the period of the Warring States in Chinese history between 475 and 221 B.C.⁴ During part of the period, state Chu and state Qin were the two most powerful amongst many states. Each of the two powerful states wanted to be dominant. They could form alliances either with each other or with other weaker states. Doing the former, they could avoid conflict, while doing the latter, they could each have the opportunity to become stronger than the opponent. The second strategy could also lead more likely to war. Mutual cooperation could make both Chu and Qin better off than mutual competition. However, each preferred

¹See Luce and Raiffa (1957, p. 97–102).

²See Aumann and Shapley (1976) and Rubinstein (1979) for the folk theorem for repeated games without discounting, and Fudenberg and Maskin (1986) for the folk theorem with discounting. Among other things, this approach to resolving the prisoner's dilemma is motivated by the common belief that cooperative outcomes might be sustained by the fear of punishments by other players for failing to cooperate. When the game is played repeatedly, players can observe opponents' past actions and can choose their actions so as to make defectors' future payoffs as small as they can enforce. A punishment of this kind that a player can carry out is known as a "trigger strategy": any deviation from cooperation causes the player to carry out a punitive action that lasts forever (see for example Fudenberg and Tirole 1991). For a recent evolutionary approach to resolving the prisoner's dilemma, see Bergstrom (1995).

³Schelling (1960, p. 135) suggests that the institution of hostages is an ancient technique that deserves to be studied by game theory.

⁴This is a period in which not only was China divided into several states but also various schools of thought including the Confucian school were formed in that period. See Liu (77-6 B.C.). I thank James Tong for reference to the English translation of Liu (77-6 B.C.) by Crump (1970).

to becoming stronger irrespective of the other one's choice. The situation can be described as a prisoner's dilemma game. Realizing the difficulties of achieving mutual cooperation, an advisor from state Qin suggested that the king of Chu offer his heir as a hostage to Qin and the same advisor suggested the king of Qin do likewise to induce the two states not to attack each other (see Liu 77-6 B.C., p. 244-247). Losing the heir is problematic for dynastic succession in ancient times. The hostages in this example are of great intrinsic values to the hostage-givers but not to the hostage-takers. This asymmetry in valuation made the hostages good safeguards for alliance between Chu and Qin.⁵

In modern times holding people as hostages is not allowed by law. This paper considers a hostage from a player as a quantity of an asset of that player. A hostage is offered by a player with respect to a particular action (usually defection), such that the hostage will be placed in the opponent's possession if the player takes the action. It is assumed that the required personal commodities are valuable to the opponents as well as to the owners themselves, but they value their own personal commodities at least as much as the opponents do. Economic examples of hostages of this kind include cash a buyer places on deposit or down payment for purchase of a property, bond provisions posted (Ricketts 1994, p. 150) or liquidated damages stipulated by parties to a contract (Cooter and Ulen 1997, p. 213-214), and transaction or relationship-specific assets that buyer and seller to a transaction invest in order to create a mutual reliance relation (Williamson 1983 and Palay 1996).⁶

The analysis here requires a costless neutral third-party to hold hostages for the players so that hostages are returned or delivered to the opponents depending on players' actions. The need for a third-party is due to fact that a hostage from a player is valuable to the opponents as well as to the player himself. In the real estate market, the escrow company is a neutral agent appointed by both parties to a real estate transaction. The escrow holder's function is to hold funds and title in trust for buyer, seller and lender until

⁵For a discussion of conditions that qualify hostages as good safeguards for cooperation, see Cooter and Ulen (1997, p. 201).

⁶Such hostages are termed as *economic equivalents of hostages* by Williamson (1983) during his investigation of how to use hostages to support exchange. Williamson argues that not only are economic equivalents of hostages widely used to effect credible commitments, but failure to recognize the economic purposes served by hostages has been responsible for repeated policy error.

all of the conditions of the transaction have been completed.⁷

With the possibility of offering hostages admitted in a game, the game is extended so that a play takes two steps. Players decide independently what hostages to offer in the first step. Their hostages are then made known to the players and are followed by actual play of the game in the second step. An overall strategy of a player specifies not only what hostage to offer, but also which action to take in the game contingent upon hostages offered. The focus here is on the possibility of achieving cooperation by allowing players to offer hostages with respect to defection before the actual play of the prisoner's dilemma game.

To eliminate contingent actions that are not credible, subgame perfect equilibrium will be applied. I show that the use of hostages can induce cooperation in prisoner's dilemma games. Specifically, I prove that in terms of equilibrium actions and hence payoffs, there are at most two Nash equilibria when the possibility of hostages is admitted: mutual defection and mutual cooperation (theorem 1). To support cooperation in subgame perfect equilibrium, I prove that values of hostages must be great enough (bounded below) to induce each to cooperate, but small enough (bounded above) to prevent each either from inducing the opponent to defect by offering a sufficiently small hostage or from making it too costly to defect when the opponent defects (theorem 2). The lower bound measures the *gain* to a defector given that the other cooperates. The upper bound reflects the fact that a player's hostage cannot overly cover the *loss* incurred to the defector when his defection causes the player to defect as well, and at the same time, the player's hostage cannot make cooperation the *only* best response to the opponent's defection *even* when the opponent has offered a small hostage. Finally, I prove that necessary and sufficient conditions for the existence of subgame perfect equilibrium with cooperation are equivalent to the lower bound being less than or equal to the upper bound (theorem 3).

The next section specifies prisoner's dilemma with hostages. Results and proofs are in section 3. Section 4 contains a discussion of possible extensions to cases with more than two players and/or actions and applications to the study of contractual relations.

⁷For other roles that a neutral third-party plays, see Arrow, et al (1995, p. 22-24). Section 4 of this paper discusses briefly how to relax the requirement of a third-party.

2 Prisoner's Dilemma and Hostages

Consider a one-shot prisoner's dilemma game in which each player chooses either to cooperate (C) or to defect (D). Payoffs are as follows

		Player 2	
Player 1	C	D	
C	R, R	S, T	
D	T, S	P, P	

Table 1: The prisoner's dilemma game with $S < P < R < T$.

The first figure in each cell is the payoff to player 1 and the second is the payoff to player 2. Following Axelrod (1984, p. 8), P is the punishment for mutual defection, R is the reward for mutual cooperation, S is the sucker's payoff, and T is the temptation to defect. The pair (D, D) is the only Nash equilibrium for the prisoner's dilemma game, and yields each player a payoff of P . Each would obtain a higher payoff of R if both cooperate; that is if the pair (C, C) describes players' actions. In noncooperative games players by definition cannot enter binding agreement, so mutual cooperation cannot be achieved directly. Action D is always preferred by each player to action C regardless of the opponent's action. To induce cooperation, we introduce:

DEFINITION 1 *A hostage offered by a player with respect to an action is a quantity of some personal commodity of that player, which will be placed in the opponent's possession if the player takes the action.*

REMARK 1 *Because the personal commodity of a player is assumed to be valuable to players, a lost hostage represents a reduction in the hostage-giver's payoff and an increment in the hostage-receiver's payoff. In light of this role of hostages, an alternative but indirect definition can be given as follows.*

DEFINITION 2 *A hostage offered by a player with respect to an action is an amount of payoff that will be placed in the opponent's possession if the player takes the action.*

To see how the offering of hostages changes the payoff structure of the game, definition 2 will be applied. Because players are allowed to offer hostages, the prisoner's dilemma game is extended to contain two steps: the *hostage-offering* step and the *action-taking* step. Also, because hostages are made known to the players at the beginning of the action-taking step, each player's choice of action in the second step can be contingent upon hostages offered by both players. Thus, a player's strategy in the extended game specifies not only what hostage to offer, but also which action to take contingent upon each possible pair of hostages. That is, the player's choice of action induced by his strategy is a mapping that maps pairs of hostages into action C or D . That a player's strategy must specify, for each pair of hostages, which action to take follows from the definition of a strategy. A strategy of a player is a complete plan that specifies which move to make in each contingency that can possibly arise (see Shubik 1982, p. 34). Here a contingency that can possibly arise from a player's point of view is any offer of hostage from the opponent as well as from the player himself. Under definition 1, the hostage offered by a player will be delivered into the opponent's possession even when both the player and the opponent defect.⁸

For $i = 1, 2$, let $H_i \geq 0$ denote a hostage offered by player i and let \mathcal{H}_i denote the set of all possible hostages from player i . Then, $\mathcal{H}_i = [0, \infty]$.⁹ Set $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$. From the description given in the preceding paragraph, player i 's choice of action is a mapping Φ_i from \mathcal{H} to $\{C, D\}$. A strategy of player i is a pair $\sigma_i = (H_i, \Phi_i)$, where H_i and Φ_i are as above. Finally, let Σ_i denote the set of all strategies of player i . A hostage of H_j from player j is worth $\alpha_i H_j$ to player $i \neq j$, where α_i is a constant between 0 and 1. Hostages are asymmetrically or symmetrically valued by players depending on the values of α_i , for all i . Given $\sigma_i = (H_i, \Phi_i) \in \Sigma_i$, $i = 1, 2$, the payoff to player 1 associated with the strategy pair (σ_1, σ_2) is denoted by $U_1(\sigma_1, \sigma_2)$.

⁸A justification for this assumption is that mutual cooperation is socially optimal.

⁹For the results in section 3 of this paper to be valid, it is sufficient to require that each player has a personal commodity whose entire value to the player is no less than $T - R$. Furthermore, the personal commodities of the players, upon which hostages are created, do not have to be divisible (see the discussion given at the end of section 3).

By construction,

$$U_1(\sigma_1, \sigma_2) = \begin{cases} R & \text{if } \Phi_1(H) = \Phi_2(H) = C \\ S + \alpha_1 H_2 & \text{if } \Phi_1(H) = C \text{ and } \Phi_2(H) = D \\ T - H_1 & \text{if } \Phi_1(H) = D \text{ and } \Phi_2(H) = C \\ P - H_1 + \alpha_1 H_2 & \text{if } \Phi_1(H) = \Phi_2(H) = D \end{cases} \quad (1)$$

where $H = (H_1, H_2)$. Player 2's payoff $U_2(\sigma_1, \sigma_2)$ is similarly determined.

DEFINITION 3 A strategy pair (σ_1, σ_2) with $\sigma_1 = (H_1, \Phi_1)$ and $\sigma_2 = (H_2, \Phi_2)$ is said to support cooperation if and only if $\Phi_1(H) = \Phi_2(H) = C$.

Given strategies $\sigma_1 = (H_1, \Phi_1)$ and $\sigma_2 = (H_2, \Phi_2)$, $\Phi_1(H)$ and $\Phi_2(H)$ are the *realized actions* of player 1 and player 2, in the sense that they are the actions players take in the actual play of the prisoner's dilemma game following the pair $H = (H_1, H_2)$ of hostages. Thus definition 2 says that a strategy pair supports cooperation if and only if the realized actions of the players are C . Nash equilibrium is defined on $\Sigma = \Sigma_1 \times \Sigma_2$.

DEFINITION 4 A strategy pair (σ_1^*, σ_2^*) with $\sigma_1^* = (H_1^*, \Phi_1^*)$ and $\sigma_2^* = (H_2^*, \Phi_2^*)$ is a Nash equilibrium if $U_1(\sigma_1, \sigma_2^*) \leq U_1(\sigma_1^*, \sigma_2^*)$, for all $\sigma_1 \in \Sigma_1$, and $U_2(\sigma_1^*, \sigma_2) \leq U_2(\sigma_1^*, \sigma_2^*)$, for all $\sigma_2 \in \Sigma_2$.

Thus for a strategy pair to be in Nash equilibrium, neither player can be better off by changing his hostage and/or his action plan. To eliminate action plans that are not credible (i.e., not self-enforcing contingent upon a pair of hostages), we consider subgame perfect equilibrium. Denote by $\Gamma(H)$ the subgame following the pair $H = (H_1, H_2) \in \mathcal{H}$. Then, by (1), $\Gamma(H)$ is given by

DEFINITION 5 A strategy pair (σ_1^*, σ_2^*) with $\sigma_1^* = (H_1^*, \Phi_1^*)$ and $\sigma_2^* = (H_2^*, \Phi_2^*)$ is a subgame perfect equilibrium if it is a Nash equilibrium and for every $H \in \mathcal{H}$, $(\Phi_1^*(H), \Phi_2^*(H))$ is a Nash equilibrium for the subgame $\Gamma(H)$.

For each pair $H = (H_1, H_2) \in \mathcal{H}$, the following lemma establishes a pure

		Player 2	
		C	D
Player 1	C	R, R	$S + \alpha_1 H_2, T - H_2$
D	$T - H_1, S + \alpha_2 H_1$	$P - H_1 + \alpha_1 H_2, P - H_2 + \alpha_2 H_1$	

Table 2: The subgame $\Gamma(H)$.

strategy Nash equilibrium for the subgame $\Gamma(H)$. This lemma will be applied in section 3. Its proof is trivial and is therefore omitted.

LEMMA 1 *Let $H = (H_1, H_2)$ be a pair of hostages in \mathcal{H} .*

(i) *If $T - R < P - S$, then $\Phi^*(H) = (\Phi_1^*(H), \Phi_2^*(H))$ is a Nash equilibrium for the subgame $\Gamma(H)$, where*

$$\Phi^*(H) = \begin{cases} (D, D) & \text{if } H_1 < T - R \text{ and } H_2 \leq P - S \\ & \text{or } H_1 \leq P - S \text{ and } H_2 < T - R \\ (D, C) & \text{if } H_1 < T - R \text{ and } H_2 > P - S \\ (C, D) & \text{if } H_1 > P - S \text{ and } H_2 < T - R \\ (C, C) & \text{if } H_1, H_2 \geq T - R \end{cases} \quad (2)$$

(ii) *If $T - R = R - S$, then $\Phi^*(H)$ is a Nash equilibrium for the subgame $\Gamma(H)$, where*

$$\Phi^*(H) = \begin{cases} (D, D) & \text{if } H_1, H_2 \leq T - R \text{ and } H \neq (T - R, T - R) \\ (D, C) & \text{if } H_1 < T - R \text{ and } H_2 > T - R \\ (C, D) & \text{if } H_1 > T - R \text{ and } H_2 < T - R \\ (C, C) & \text{if } H_1, H_2 \geq T - R \end{cases} \quad (3)$$

Note that when $T - R < H_i < P - S$, for $i = 1, 2$, mutual defection is another pure strategy Nash equilibrium in the subgame $\Gamma(H)$. Furthermore, under condition (i) of lemma 1, when $H_i > P - S$, action C is the dominant strategy for player i in the subgame $\Gamma(H)$ regardless of player j 's hostage.

3 Results

In this section we prove that the use of hostages is quite effective in inducing cooperation in a prisoner's dilemma game. First, we prove in theorem 1 that in any Nash equilibrium the realized actions of players must be identical. We then prove in theorem 2 that in any Nash equilibrium with cooperation (if it exists) the values of players' hostages are bounded by $T - R$ from below, and the values of player 1's hostages are bounded by $\min\{(R - P)/\alpha_2, P - S\}$ while those of player 2's hostages are bounded by $\min\{(R - P)/\alpha_1, P - S\}$ from above.¹⁰ From table 1, it follows that the lower bound $T - R$ measures the gain to a defector given the other cooperates. The difference $R - P$ measures the loss to the defector in the prisoner's dilemma when his defection causes the other to defect as well. And, from table 2, if $T - R \leq P - S$ and $H_j > P - S$, then cooperation is the dominant strategy for player j regardless of player i 's hostage of H_i . Thus, the upper bound reflects the fact that to support cooperation in subgame perfect equilibrium, a player's hostage cannot overly cover the opponent's loss incurred when the opponent's defection causes the player himself to defect as well, and at the same time, a player's hostage cannot make it too costly for the player to defect when the opponent defects. Finally, we prove in theorem 3 that a subgame perfect equilibrium with cooperation exists if and only if $T - R \leq \min\{(R - P)/\alpha_i, P - S\}$, for $i = 1, 2$.

THEOREM 1 *Players' realized actions and hence payoffs are identical in any Nash equilibrium.*

Proof. We prove this theorem by contradiction. Let $\sigma_1^* = (H_1^*, \Phi_1^*) \in \Sigma_1$ and $\sigma_2^* = (H_2^*, \Phi_2^*) \in \Sigma_2$ be such that the pair (σ_1^*, σ_2^*) consists of a Nash equilibrium. Suppose $\Phi_1^*(H^*) \neq \Phi_2^*(H^*)$. Without loss of generality, we assume $\Phi_1^*(H^*) = C$ and $\Phi_2^*(H^*) = D$. Then, by table 2,

$$U_1(\sigma_1^*, \sigma_2^*) = S + \alpha_1 H_2^* \text{ and } U_2(\sigma_1^*, \sigma_2^*) = T - H_2^*. \quad (4)$$

Consider $\sigma_2 = (H_2^*, \Phi_2) \in \mathcal{H}_2$, where $\Phi_2(H^*) = C$ and $\Phi_2^*(H) = \Phi_2^*(H)$ for $H \in \mathcal{H}$ with $H \neq H^*$. Then, $U_2(\sigma_1^*, \sigma_2) = R$, and hence $U_2(\sigma_1^*, \sigma_2) \leq U_2(\sigma_1^*, \sigma_2^*)$ implies that $R \leq T - H_2^*$ or $H_2^* \leq T - R$. Consider now $\sigma_1 =$

¹⁰When $\alpha_i = 0$, the ratio $(R - P)/\alpha_i$ is understood to be ∞ .

$(H_1, \Phi_1) \in \mathcal{H}_1$ with $H_1 = 0$ and $\Phi_1(H) = D$, for all $H \in \mathcal{H}$. If $\Phi_2^*(0, H_2^*) = D$, then $U_1(\sigma_1, \sigma_2^*) = P + \alpha_1 H_2^*$. Since $S < P$, (4) implies that $U_1(\sigma_1^*, \sigma_2^*) < U_1(\sigma_1, \sigma_2^*)$, which contradicts to the assumption that (σ_1^*, σ_2^*) is a Nash equilibrium. If $\Phi_2^*(0, H_2^*) = C$, however, $U_1(\sigma_1, \sigma_2^*) = T$. Since $H_2^* \leq T - R$ and $0 \leq \alpha_1 \leq 1$, we have $S + \alpha_1 H_2^* \leq S + \alpha_1(T - R)$. The condition that $S < R < T$ then implies that $S + \alpha_1(T - R) < (1 - \alpha_1)R + \alpha_1 T < T$. Thus, $U_1(\sigma_1^*, \sigma_2^*) < U_1(\sigma_1, \sigma_2^*)$, which again contradicts to the assumption that (σ_1^*, σ_2^*) is a Nash equilibrium. ■

Given a Nash equilibrium, theorem 1 makes it possible to use the term of *equilibrium action* to mean players' identical actions realized in the equilibrium. The result implies that in terms of equilibrium actions and hence payoffs there can be at most two Nash equilibria. It is not difficult to see that defection is a Nash equilibrium action. Furthermore, defection is also a subgame perfect equilibrium action (see the discussion given at the end of section 3). Note also that the result claims nothing about the sizes of the associated hostages. However, if the equilibrium action is D , the associated equilibrium hostages of both players must be 0.

DEFINITION 6 *Let $\sigma_1^* = (H_1^*, \Phi_1^*)$ and $\sigma_2^* = (H_2^*, \Phi_2^*)$ be strategies of player 1 and player 2. The strategy pair (σ_1^*, σ_2^*) consists of a Nash (subgame perfect) equilibrium with cooperation if and only if the pair consists of a Nash (subgame perfect) equilibrium and supports cooperation.*

We now prove:

THEOREM 2 *Let $\sigma_1^* = (H_1^*, \Phi_1^*)$ and $\sigma_2^* = (H_2^*, \Phi_2^*)$ be strategies of player 1 and player 2 such that the pair (σ_1^*, σ_2^*) consists of a subgame perfect equilibrium with cooperation. Then, $T - R \leq H_i^* \leq \min\{(R - P)/\alpha_j, P - S\}$, for $i \neq j$.*

Proof. Let $i = 1$ and $j = 2$. The case where $i = 2$ and $j = 1$ can be similarly proved. Observe first that $U_1(\sigma_1^*, \sigma_2^*) = R$ and $U_2(\sigma_1^*, \sigma_2^*) = R$. Let Φ_1 be given by $\Phi_1(H) = D$, for $H \in \mathcal{H}$ with $H_1 = H_1^*$, and $\Phi_1(H) = \Phi_1^*(H)$, for all other pairs $H \in \mathcal{H}$. Then, $U_1(\sigma_1, \sigma_2^*) = T - H_1^*$, where $\sigma_1 = (H_1^*, \Phi_1)$. Thus, for $U_1(\sigma_1, \sigma_2^*) \leq U_1(\sigma_1^*, \sigma_2^*)$, we must have $T - H_1^* \leq R$. That is, we must have $H_1^* \geq T - R$.

Suppose that $H_1^* > P - S$. Then, in the subgame $\Gamma(H_1^*, 0)$ (see table 2), action D is the dominant strategy for player 2. Thus, $\Phi_2^*(H_1^*, 0) = D$. Since $H_1^* > P - S$, the only best response of player 1 to player 2's action D in $\Gamma(H_1^*, 0)$ is action C . Hence, $(\Phi_1^*(H_1^*, 0), \Phi_2^*(H_1^*, 0)) = (C, D)$ and $U_2(\sigma_1^*, \sigma_2) = T$, where $\sigma_2 = (H_2, \Phi_2^*)$ with $H_2 = 0$. This shows that $U_2(\sigma_1^*, \sigma_2) > U_2(\sigma_1^*, \sigma_2^*)$, which contradicts to the assumption that (σ_1^*, σ_2^*) is a subgame perfect equilibrium. Thus, $H_1^* \leq P - S$.

To prove that $H_1 \leq (R - P)/\alpha_2$, consider $\sigma_2 = (H_2, \Phi_2) \in \mathcal{H}_2$ with $H_2 = 0$ and $\Phi_2(H) = D$ for all $H \in \mathcal{H}$. If $\Phi_1^*(H_1^*, 0) = C$, then $U_2(\sigma_1^*, \sigma_2) = T$, and hence $U_2(\sigma_1^*, \sigma_2) > U_2(\sigma_1^*, \sigma_2^*)$. We thus conclude that $\Phi_1^*(H_1^*, 0) = D$. In this case, $U_2(\sigma_1^*, \sigma_2) = P + \alpha_2 H_1^*$. Thus, for $U_2(\sigma_1^*, \sigma_2) \leq U_2(\sigma_1^*, \sigma_2^*)$, it must be true that $P + \alpha_2 H_1^* \leq R$ or $H_1^* \leq (R - S)/\alpha_2$. ■

The first part of the above proof shows that to induce player 2 to cooperate, player 1's hostage must be great enough (i.e., bounded from below by $T - R$). The logic can be seen as follows. If player 2 chooses to cooperate contingent upon small hostages offered by player 1, then player 1 would prefer offering such a small hostage and defecting himself. Thus, player 2 cannot be induced to cooperate by player 1 unless player 1 offers a great enough hostage. The second part of the proof shows that to prevent player 2 from inducing player 1 to defect (e.g., by offering a hostage of 0), player 1's hostage must be bounded by $(R - P)/\alpha_2$ from above. At the same time, to prevent player 2 from defecting, player 1's hostage must not make cooperation in the subsequent subgame the only best response to player 2's defection. Thus, player 1's hostage must be also bounded by $P - S$ from above.

REMARK 2 *In the case where $\alpha_1 = \alpha_2 = 1$, to support cooperation in Nash equilibrium, the common upper bound of players' hostages has shown to be $R - P$ while lower bound remains to be $T - R$ in Qin (1996). For α_1, α_2 in $[0, 1]$, a similar logic will show that to support cooperation in Nash equilibrium, the upper bound becomes $(R - P)/\alpha_2$ for player 1's hostages and $(R - P)/\alpha_1$ for player 2's hostages. Thus, when $\alpha_i = 0$ (i.e., player j 's hostages have no intrinsic value to player i), player j 's hostages are not bounded from above in Nash equilibrium.*

We now apply theorem 2 to prove our last result of this section.

THEOREM 3 *A subgame perfect equilibrium with cooperation exists if and only if $T - R \leq \min\{(R - P)/\alpha_i, P - S\}$, for $i = 1, 2$.*

Proof. By theorem 2, for $i \neq j$, player j 's hostage offered in any subgame perfect equilibrium with cooperation must be bounded by $T - R$ from below and by $\min\{(R - P)/\alpha_i, P - S\}$ from above. Thus, $T - R \leq \min\{(R - P)/\alpha_i, P - S\}$, for $i = 1, 2$.

To prove the condition is also sufficient, we will construct a strategy pair and show, under the condition of the theorem, that the pair consists of a subgame perfect equilibrium with cooperation. To this end, consider $\sigma_1^* = (H_1^*, \Phi_1^*)$ and $\sigma_2^* = (H_2^*, \Phi_2^*)$, where Φ_i^* is as in lemma 1 and H_i^* is given by

$$T - R \leq H_i^* \leq \min\{(R - P)/\alpha_j, P - S\}, \quad (5)$$

for $i \neq j$. Under the condition of the theorem, H_1^* and H_2^* are well defined, and by (2) and (3),

$$\Phi_1^*(H^*) = \Phi_2^*(H^*) = C. \quad (6)$$

Let $\sigma_1 \in \Sigma_1$ be any strategy of player 1. Then, by (2), (3), and table 2,

$$U_1(\sigma_1, \sigma_2^*) = \begin{cases} S + \alpha_1 H_2^* & \text{if } H_1 < T - R \text{ and } \Phi_1(H_1, H_2^*) = C \\ P - H_1 + \alpha_1 H_2^* & \text{if } H_1 < T - R \text{ and } \Phi_1(H_1, H_2^*) = D \\ R & \text{if } H_1 \geq T - R \text{ and } \Phi_1(H_1, H_2^*) = C \\ T - H_1 & \text{if } H_1 \geq T - R \text{ and } \Phi_1(H_1, H_2^*) = D \end{cases} \quad (7)$$

Since $T - R \leq H_2^* \leq \min\{(R - P)/\alpha_1, P - S\}$ and $H_1 \geq 0$, (7) implies $S + \alpha_1 H_2^* \leq R$, $P - H_1 + \alpha_1 H_2^* \leq P + \alpha_1 H_2^* \leq R$, and $T - H_1 \leq R$. Thus, $U_1(\sigma_1, \sigma_2^*) \leq U_1(\sigma_1^*, \sigma_2^*)$. Similarly, given σ_1^* , $U_2(\sigma_1^*, \sigma_2) \leq U_2(\sigma_1^*, \sigma_2^*)$, for all $\sigma_2 \in \Sigma_2$. We conclude that (σ_1^*, σ_2^*) is a Nash equilibrium. By lemma 1, $(\Phi_1^*(H), \Phi_2^*(H))$ is a Nash equilibrium for the subgame $\Gamma(H)$, for all $H \in \mathcal{H}$. Thus, by (6), (σ_1^*, σ_2^*) is a subgame perfect equilibrium with cooperation. ■

Note that defection is also a subgame perfect equilibrium action. To see this, consider $\sigma_i^* = (H_i^*, \Phi_i^*)$ where $H_i^* = 0$ and Φ_i^* is as in lemma 1, for $i = 1, 2$. Then, it follows from (2) and (3) that $\Phi_1^*(0, H_2) = D$ for all $H_2 \in \mathcal{H}_2$ and $\Phi_2^*(H_1, 0) = D$ for all $H_1 \in \mathcal{H}_1$. Thus, $\Phi_1^*(H^*) = \Phi_2^*(H^*) = D$, and hence

$U_1(\sigma_1^*, \sigma_2^*) = P$ and $U_2(\sigma_1^*, \sigma_2^*) = P$. Moreover, for any $\sigma_1 = (H_1, \Phi_1) \in \Sigma_1$, $U_1(\sigma_1, \sigma_2^*) = S$ if $\Phi_1(H_1, 0) = C$ and $U_1(\sigma_1, \sigma_2^*) = P - H_1$ if $\Phi_1(H_1, 0) = D$. Therefore, $U_1(\sigma_1, \sigma_2^*) \leq U_1(\sigma_1^*, \sigma_2^*)$ in either case. Similarly, $U_2(\sigma_1^*, \sigma_2) \leq U_2(\sigma_1^*, \sigma_2^*)$ for all $\sigma_2 \in \Sigma_2$. This shows that (σ_1^*, σ_2^*) is a Nash equilibrium and, by lemma 1, (σ_1^*, σ_2^*) is also a subgame perfect equilibrium. Finally, to support cooperation in subgame perfect equilibrium, players' personal commodities do not have to be divisible so long as their entire values to the owners satisfy (5).

Taking theorems 2 and 3 together, to support cooperation in subgame perfect equilibrium the idea here is, through the offering of hostages, to make the net gain to a defector negative given the other cooperates, and to make the net loss to the defector positive when his defection causes the other to defect as well. And, a player's hostage cannot "tie" the player's hands too strongly, in the sense that the hostage cannot make it too costly for the player to defect when the opponent defects. Since $S < P < R < S$, the necessary and sufficient condition of theorem 3 covers many prisoner's dilemma games.

4 Discussion

This paper considers hostages of a player as fractions of some personal commodity of that player. Personal commodities are assumed to be valuable to the opponents as well as to owners themselves, but they value their own commodities at least as much as the opponents do. The offering of hostages changes the payoff structure of a game but not players' strategies in the game. The possibility of sustaining cooperation through voluntarily offering of hostages is analyzed for the 2-person one-shot prisoner's dilemma games. Our definition of hostage can be adapted to be applicable to games in which a player has more than two strategies or in which there are more than two players. When a player has more than two strategies in a game, a hostage of the player can be offered with respect to a subset of strategies, such that the hostage will be placed in the opponents' possession if the player plays any strategy in that subset. If there are more than two players, hostages may be offered bilaterally or multilaterally. In the latter case, a player's offer consists of a vector of hostages one for each opponent player. One way to relax the requirement of a neutral third-party to hold hostages is to consider

reputation effects or moralistic aggressions (for an application of the latter to IO problems, see Comanor and Frech 1987). Further research along these lines will be carried out in a framework of repeated game in the future.

Bond-posting is an incentive mechanism that induces parties to a contract to honor the contract (Ricketts 1994, p. 150). The law of contracts does not normally require parties to a contract actually to perform. Contracting parties are given the option of performance or breach. But if a contracting party breaches his deal, he is required to pay damages (Hirsch 1988, p. 142). As mentioned before, bond provisions posted or liquidated damages stipulated by parties to a contract can be considered as hostages of the kind studied in this paper. For efficiency purposes, incentive mechanisms inducing parties to honor a contract are important. Our results and their extensions can be useful for the study of such mechanisms. I hope to report on such work in the future.

References

- [1] Arrow, Kenneth J, Mnookin, Robert H, Ross, Lee, Tversky, Amos, and Wilson, Robert. *Barriers to Conflict Resolution*. New York: W. W. Norton & Company, 1995.
- [2] Aumann, Robert and Shapley, Lloyd S. “Long-Term Competition: A Game Theoretic Analysis.” mimeo, 1976.
- [3] Axelrod, Robert. *The Evolution of Cooperation*. Now York: Basic Books, Inc., Publishers, 1984.
- [4] Bergstrom, Theodore C. “On the Evolution of Altruistic Ethical Rules for Siblings.” *American Economic Review*, 85 (1995), 58–81.
- [5] Comanor, William S. and Frech III, H. E. “The Competitive Effects of Vertical Agreements: Reply.” *American Economic Review*, 77 (1987), 1069–1072.
- [6] Cooter, Robert and Ulen, Thomas. *Law and Economics*. Harper Collins Publishers, 2nd edition, 1997.

- [7] Fudenberg, Drew and Maskin, Eric. “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information.” *Econometrica*, 54 (1986), 533–554.
- [8] Fudenberg, Drew and Tirole, Jean. *Game Theory*. Cambridge, Massachusetts: The MIT Press, 1991.
- [9] Hirsch, Werner Z. *Law and Economics: An Introductory Analysis*. Academic Press, Second Edition, 1988.
- [10] Liu Xiang. *Zhan Guo Ce*. 77–6 B.C. Translated from the Chinese by James I Crump. Oxford University Press, 1970.
- [11] Luce, Duncan R and Raiffa, Howard. *Games and Decisions*. New York: John Wiley and Sons, Inc., 1957.
- [12] Palay, Thomas M. “Comparative Institutional Economics: The Governance of Rail Freight Contracting.” In *Case Studies in Contracting and Organization*. Edited by Scott E Masten. Oxford: Oxford University Press, 1996.
- [13] Ricketts, Martin. *The Economics of Business Enterprise*. New York: Harvester Wheatsheaf, 1994.
- [14] Qin, Cheng-Zhong. “Hostages” and Cooperation in the Prisoner’s Dilemma. UCSB Working Paper #8-96.
- [15] Rubinstein, Ariel. “Equilibrium in Supergames with the overtaking Criterion.” *JET*, 21 (1979), 1–9.
- [16] Schelling, Thomas. *The Strategy of Conflict*. Cambridge, Massachusetts: Harvard University Press, 1960.
- [17] Shubik, Martin. *Game Theory in the Social Sciences*. Cambridge, Massachusetts: The MIT Press, 1982.
- [18] Williamson, Oliver E. “Credible Commitments: Using Hostages to Support Exchange.” *AER*, 73 (1983), 519–540.