

Shino, Junnosuke

Working Paper

2x2 delegation games with implementability in weakly undominated SPNE

Working Paper, No. 2012-02

Provided in Cooperation with:

Department of Economics, Rutgers University

Suggested Citation: Shino, Junnosuke (2012) : 2x2 delegation games with implementability in weakly undominated SPNE, Working Paper, No. 2012-02, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/59487>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

2×2 Delegation Games With Implementability In Weakly Undominated SPNE

Junnosuke Shino ^{*†}

Abstract

In this paper we study delegation environments based on Fershtman, Judd, and Kalai ([3], hereafter FJK). By imposing a certain assumption on the notion of implementability, called *implementability with mutually rational agents*, they show that every efficient outcome can be fully implemented in subgame perfect Nash equilibrium (SPNE). For their analysis, we first argue that FJK's model can and should be interpreted as a problem in mechanism design. With this in mind, we first modify their model so that agents' participation decision is explicitly built in. Then, we argue that FJK uses a non-standard solution concepts in the mechanism design literature – in order to attain full implementability, they strengthen notion of implementability, instead of using a refined equilibrium concept. In response, we follow the standard mechanism design approach – employ a refined SPNE (weakly undominated SPNE, U-SPNE) as the equilibrium concept, while keeping the notion of the implementability unchanged. By applying U-SPNE, we show that in certain classes of 2×2 games – including prisoners' dilemma, coordination games, and battle of sexes – every efficient outcome is fully implementable in U-SPNE.

JEL classifications: C72, D04, L13

^{*}Department of Economics, Rutgers, The State University of New Jersey, NJ, USA and Financial Markets Department, Bank of Japan, Tokyo, Japan. E-mail: junnosuke.shino@gmail.com

[†]I thank to Prof. T.Sjostrom for helping me with numerous suggestions about the implementation theory and mechanism design, and to Prof. R.McLean and Prof. C.Campbel for very useful comments.

1 Introduction

Delegation is a frequently observed phenomenon in social, political and economic strategic interactions. In interstate relations, it is typically a diplomat or embassy representative who negotiates with another country's diplomat to maximize one's nation's self-interest, rather than the heads of ministries or politicians themselves. Similarly, a private firm's manager or CEO in competitions with rivals can be viewed as a delegated agent with task of maximizing shareholders' profit. The common structure underlying all such conflict situations can be characterized as a principal-agents problem accompanied with strategic interactions between agents.

It is therefore quite important to analyze delegation environments at a general level. Since Shelling [8], a great deal of attention has been paid to analyzing strategic delegation by using game-theoretic framework. Fershtman and Judd [2] studied a model of Cournot duopoly with delegation and unobservable efforts by agents. Persson and Tabellini [7] analyzed international monetary policy coordination by casting central banks as agents acting on behalf of the public. Persson *et al.* [6] also examined public finance with a delegation setup. In this paper, following the pioneering work of Fershtman, Judd, and Kalai ([3], hereafter FJK), we analyze a general strategic delegation situation in which principals' proposals of compensation schemes to their agents are commonly observable before the agents' interaction.

Specifically, the order of the play in the FJK delegation game is as follows: suppose two principals face a conflict represented by a strategic form game (called an *underlying game*). First, for the underlying game, each principal simultaneously proposes to her agent a compensation scheme, which is contingent on payoffs in the underlying game. Next, commonly observing the both schemes, each agent chooses one of actions from the strategy set in the underlying game. Finally, depending on an outcome derived by their interactions, the compensation is paid to each agent based on the scheme, and its principal gets the payoff in the underlying game minus the compensation she pays to her agent.

Here it is worth noting that this delegation game described above can be viewed through the lens of mechanism design or implementation theory. First, for each underlying game, we may specify a socially efficient outcome. Such a specification can be regarded as a (*social*)

choice function. Second, by applying a certain *equilibrium concept* to the second stage game induced by a compensation scheme, we can assign a predicted action profile to each underlying game. Lastly, an *outcome function* can be defined as a function from such a predicted action profile to an outcome in the underlying game.

Therefore, the delegation game thus lends itself naturally and aptly to the tools of mechanism design. With this in mind, we first slightly modify the FJK's delegation game so that agents' participation decision is explicitly built in. Next, we examine their methodology to deal with the delegation game. In the mechanism design literature, two building blocks (solution concepts) in considering whether an efficient outcome can be attained are *equilibrium* and *implementation*. Given a certain equilibrium, we say that a profile of compensation schemes proposed by principals *implements* an outcome in the equilibrium, when there exists an equilibrium in which the proposed compensation scheme attains the outcome via an agents' action. Furthermore, if there is a compensation scheme that always attains the outcome in the equilibrium, irrespective of the agents' actions, then principals can attain the optimal outcome *for sure*, by using such a scheme. In this case we say the scheme *fully implements* the outcome in the equilibrium concept.

FJK, however, do not use this standard definition of implementation. Instead, they impose a certain assumption on implementability, which they call *implementability with mutually rational agents*, and show that every socially efficient outcome can be fully implemented with mutually rational agents in SPNE. A reason why they adopt such an unorthodox approach is because, in delegation games, few significant results can be derived by using these relatively weak solution concepts – SPNE and (standard definition of) implementability. In particular, in deriving SPNE, multiple equilibria often emerge in an induced game among agents, and some of them may not be socially optimal. As such, it could be reasonable to use “stronger solution concepts” to obtain full implementability.

Yet despite the challenges to obtain full implementability, the modern mechanism design approach advocates the use of refined equilibrium concept, opposed to directly strengthening the notion of implementability. To this end, our objective in this paper is to reformulate FJK's analysis within the framework of mechanism design and to derive full implementability. In particular, we employ a modification of SPNE, called (*weakly*) *undominated SPNE* in Austen-

Smith and Banks [1] ¹, as a refined equilibrium concept, while keeping the notion of the implementation unchanged.

Under this setup, we show that, in certain classes of 2×2 game – including prisoners’ dilemma, general coordination games, and battle of sexes – every socially efficient outcome is fully implementable by using these standard solution concepts. We also show that, by using an example of 3×3 coordination game, not only does our standard procedure attain full implementability, our solution concepts could also solve problem of multiple equilibria, an issue where FJK’s solution concepts fall short.

Finally, it may worth noting here that depending on how one defines fullness of the implementability, different approach could be considered. For example, we could alternatively take uniqueness of (weakly undominated) SPNE into consideration, instead of uniqueness of an outcome. In this case, in addition to induced games among agents, even game among principals should have a unique equilibrium. However, if principals can communicate each other, they would coordinate on best equilibrium. Therefore, our weaker notion is reasonable to be scrutinized.

The organization of the rest of this paper is as follows. In Section 2, we give formal definition of the delegation game and the implementability by weakly undominated SPNE. Main results are shown in the next two sections. Section 3 examines 2×2 game. An example of the multiple equilibria problem in 3×3 is shown in Section 4. Some concluding remarks are made in Section 5.

2 The Model and The Solution Concept

2.1 The Model

We consider a situation where two principals face a conflict represented by a strategic form game $G = (P, \{S_i\}_{i=1,2}, \{u_i\}_{i=1,2})$. $P = \{p_1, p_2\}$ is the set of players, and p_i is called *principal i*. S_i is the set of strategies of p_i , and u_i is p_i ’s utility function, where $u_i : S \equiv S_1 \times S_2 \rightarrow \mathbf{R}$. We call G

¹In the context of mechanism design, Yu [10] applied this concept to public goods provision environment. Also, note that this is an extension of *implementation using undominated strategies*, which was initially proposed in Palfrey and Srivastava [5] and applied in Jackson *et al* [4], Sjostrom [9] and many others, to extensive form games.

an *underlying game*. Denote $u \equiv (u_1, u_2) : S \rightarrow \mathbf{R}^2$.

For an underlying game G , suppose principal i could delegate agent i and consider the following three-stage strategic delegation environment. At the first stage, each principal simultaneously proposes to her agent a “compensation scheme”. Principal i ’s compensation scheme gives a monetary reward to agent i , depending on which payoff in G is realized by agents’ interaction in the last stage. Next, in the second stage, each agent simultaneously decides whether to participate the game or not, after observing not only his own contract but also his opponent’s. If (at least) one of agents decides not to participate, the game ends at this stage. In this case, agents obtain (common) reservation wage, $\epsilon > 0$, while principals get a constant payoff normalized to zero by, for example, exerting “outside option.” This setup is natural to describe a situation where considerable expertise – any knowledge or experiences – is needed to enter the underlying game and only agents have such expertise. If, on the other hand, both of agents decides to participate, the game moves on to the final stage. In this stage, agent i chooses an action $s_i \in S_i$ in the underlying game. Then “total payoff” of $u_i(s_1, s_2)$ is realized and depending on this, monetary rewards are paid to agent i based on the compensation scheme proposed at the first stage. What principal i obtains is the total payoff, $u_i(s_1, s_2)$, minus the reward for agent i .

Formally, for a underlying game G , its *delegation game*, denoted by $D(G)$, is defined as

$$D(G) \equiv (N, \{C_i, L_i\}_{i=1,2}, \{U_i^p, U_i^a\}_{i=1,2}, \epsilon). \quad (1)$$

- $N \equiv P \cup A$ is the set of players where $P = \{p_1, p_2\}$ is the set of *principals* and $A = \{a_1, a_2\}$ is the set of *agents*.
- C_i is the set of strategies of p_i where $C_i = \{c_i \mid c_i : \mathbf{R} \rightarrow \mathbf{R}_+, \text{non-decreasing function}\}^2$. $c_i \in C_i$ is called a *compensation function* of p_i . Define $C \equiv C_1 \times C_2$.
- L_i is the set of strategies of a_i defined as follows. First, let P and NP be an agent choice of “Participate” and “Not Participate” respectively. Then agent i ’s *participate function* is defined as $D_i = \{d_i \mid d_i : C \rightarrow \{P, NP\}\}$. d_i specifies agent i ’s choice of participating or

²Throughout this study, we preserve the FJK’s assumption of non-decreasing compensation function. This assumption can be justified by, following FJK, “Besides being intuitively appealing, ... subgame perfection cannot be obtained without the weak monotonicity assumption.” on page 553 of [3]

not, contingent on his observation about C . Next, we define *response function* R_i where $R_i = \{r_i \mid r_i : C \rightarrow S_i\}$, which specifies agent i 's an action in the last stage. Finally, agent i 's strategy set L_i is defined as $L_i \equiv D_i \times R_i$, that is, i 's strategy set is the set of the profile of the participation function and the response function.

- U_i^p and U_i^a are principal and agent i 's utility functions in $D(G)$ defined as:

$$U_i^p(c_1, c_2, l_1, l_2) = \begin{cases} u_i(r_1(c), r_2(c)) - c_i(u_i(r_1(c), r_2(c))) & \text{if } d_1(c) = d_2(c) = P \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$U_i^a(c_1, c_2, l_1, l_2) = \begin{cases} c_i(u_i(r_1(c), r_2(c))) & \text{if } d_1(c) = d_2(c) = P \\ \epsilon & \text{otherwise.} \end{cases} \quad (3)$$

For the following analysis, we denote a strategy profile of all players (agents and principals) by $(c, l) \equiv (c_1, c_2, l_1, l_2) \equiv (c_1, c_2, (d_1, r_1), (d_2, r_2))$.

For a compensation scheme profile $c \in C$, its subsequent game played at the last stage in the case where both agents participate can be specified. We call it *induced game by c* and denote by $G(c)$. For $c \in C$, we say $(s_1^*, s_2^*) \in S$ is a *Nash equilibrium (NE) in $G(c)$* iff $c_i(u_i(s_i^*, s_j^*)) \geq c_i(u_i(s_i, s_j^*))$ for all $s_i \in S_i$ and for $i \in \{1, 2\}$.

The following example may be useful to understand the model.

Example 2.1 (Coordination Game)

Suppose the underlying game G is a coordination game given by Fig.1.

G	s_{21}	s_{22}
s_{11}	2, 2	0, 0
s_{12}	0, 0	1, 1

Figure 1: Coordination Game

and consider the following two kinds of compensation profiles, $\hat{c} \equiv (\hat{c}_1, \hat{c}_2)$ and $\tilde{c} \equiv (\tilde{c}_1, \tilde{c}_2)$.

$$\hat{c}_i(0) = \hat{c}_i(1) = 0, \hat{c}_i(2) = 2\epsilon \quad \tilde{c}_i(0) = \tilde{c}_i(1) = 0, \tilde{c}_i(2) = \epsilon \quad (i = 1, 2) \quad (4)$$

Then for each compensation profile, its induced game, $G(\hat{c})$ or $G(\tilde{c})$, is described by the followings:

$G(\hat{c})$	s_{21}	s_{22}
s_{11}	$2\epsilon, 2\epsilon$	$0, 0$
s_{12}	$0, 0$	$0, 0$

$G(\tilde{c})$	s_{21}	s_{22}
s_{11}	ϵ, ϵ	$0, 0$
s_{12}	$0, 0$	$0, 0$

Figure 2: (Left) induced game $G(\hat{c})$ (Right) induced game $G(\tilde{c})$

NEs in $G(\hat{c})$ or $G(\tilde{c})$ are (s_{11}, s_{21}) and (s_{12}, s_{22}) . ■

Our model builds on FJK, but differs in that an agent's decision of participating is explicitly considered. With this setup, we can explicitly examine Participate Constraint (PC).

2.2 The Solution Concept

In this section, we define the solution concept applied to the delegation game. As we will define, the solution concept, called *undominated subgame perfect Nash equilibrium (U-SPNE)*, is a strategy profile $(c_1^*, c_2^*, l_1^*, l_2^*)$ satisfying the conditions of (7), (8), and (9).

To begin with, note that the delegation game $D(G)$ is an extensive form game with complete information. Therefore, it is natural to use "backward induction". As a first step, consider the last stage game played among agents. For a given compensation profile c , denote set of NEs in $G(c)$ by

$$EA(c) = \{(s_1, s_2) \in S \mid (s_1, s_2) \text{ is an NE in } G(c)\}. \quad (5)$$

The notion of the "subgame perfect Nash equilibrium" (SPNE) requires that response function r satisfies $[r(c) \in EA(c) \forall c]$. However, as stated in FJK, this condition does not work in the delegation game setup. Especially, unique implementation (defined later) is hard to obtain. To see this, here consider Example 2.1 and the compensation scheme \tilde{c} in Fig.2. Roughly speaking, we look for a compensation scheme such that (s_{11}, s_{21}) is uniquely attained as an "equilibrium" because $u(s_{11}, s_{21})$ is Pareto efficient in the underlying game. However, to employ the notion of $EA(c)$ can not exclude agents playing (s_{12}, s_{22}) in $G(\tilde{c})$.

Now we define $UEA(c)$ as the set of weakly undominated Nash equilibria in $G(c)$:

$$UEA(c) = \{(s_1, s_2) \in S \mid (s_1, s_2) \text{ is a N.E. in } G(c) \text{ and} \\ s_i \in S_i \ (i = 1, 2) \text{ is not weakly dominated strategy in } G(c)\} \quad (6)$$

where s_i is a weakly dominated strategy in $G(c)$ if there is a strategy $\tilde{s}_i \in S_i$ such that $c_i(u_i(\tilde{s}_i, s_j)) \geq c_i(u_i(s_i, s_j))$ for all $s_j \in S_j$ and $c_i(u_i(\tilde{s}_i, s_j)) > c_i(u_i(s_i, s_j))$ for some $s_j \in S_j$. We employ (6) as the first condition of our solution concept, instead of (5). Noting that $UEA(c)$ can take empty set for some c , the first condition of U-SPNE is defined as the following:

$$\begin{aligned} r^*(c) \in UEA(c) \quad \forall c \in C \text{ with } UEA(c) \neq \phi \\ r^*(c) \in EA(c) \quad \forall c \in C \text{ with } UEA(c) = \phi \end{aligned} \quad (7)$$

Next, given the action profile played at the last stage, we consider agents' simultaneous participation choices at the second stage. Recall that the last stage game will be reached only when both agents decide to participate. Therefore, for a given strategy profile of (c_1, c_2, l_1, l_2) , the payoff matrix at the second stage can be described as follows:

G	P	NP
P	$c_1(u_1(r_1(c), r_2(c))), c_2(u_2(r_1(c), r_2(c)))$	ϵ, ϵ
NP	ϵ, ϵ	ϵ, ϵ

Figure 3: Agents' payoff matrix at the second stage by backward induction

Depending on the magnitudes of $c_i(u_1(r_1(c), r_2(c)))$ and ϵ , three cases could happen: (A) P is weakly dominant strategy, (B) NP is weakly dominant strategy, and (C) P and NP are completely indifferent. For case (A) and (B), it is totally natural to assume that the weakly dominant strategy would be played. For simplicity, we make the standard assumption that if the agent is indifferent between participating and not participating, he chooses to participate. Thus, in the case of (C), agent i is assumed to choose P . Putting it all together, we introduce the following as the second condition of U-SPNE:

$$\text{For every } c \in C, d_i^*(c) = P \text{ iff } c_i(u_i(r_1^*(c), r_2^*(c))) \geq \epsilon. \quad (8)$$

The last step is to consider the first stage in which principals propose compensation schemes. The last condition on U-SPNE is the standard one:

$$U_1^p(c_1^*, c_2^*, l_1^*, l_2^*) \geq U_1^p(c_1, c_2^*, l_1^*, l_2^*) \quad \forall c_1 \in C_1 \text{ and same for principal 2.} \quad (9)$$

As a result, our solution concept applied to the delegation game is expressed as the following definition:

Definition 2.1

$(c_1^*, c_2^*, l_1^*, l_2^*)$ is U-SPNE if it satisfies the conditions of (7), (8), and (9)

Then we define implementation and full implementation.

Definition 2.2

c^* implements $\omega \in \mathbf{R}^2$ in U-SPNE via $l^* \in L$ iff (c^*, l^*) is a U-SPNE of $D(G)$ with (1) $u(r^*(c^*)) = \omega$ and (2) $d_1^*(c^*) = d_2^*(c^*) = P$.

Definition 2.3

c^* fully implements $\omega \in \mathbf{R}^2$ in U-SPNE via $l^* \in L$ iff c^* implements ω via l^* in U-SPNE and the following holds:

$$\text{if } c^* \text{ also implements } \tilde{\omega} \in \mathbf{R}^2 \text{ in U-SPNE via some } \tilde{l}, \text{ then } \tilde{\omega} = \omega. \tag{10}$$

A fully implementations compensation scheme is attractive for principals in the sense that payoff vector ω is realized for sure, without depending on a specific choice of equilibrium by their agents. Thus, once we require full 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.

3 Results in 2×2 Games

In this section, we consider the case where an underlying game G is 2×2 game, described in Fig.4.

G	s_{21}	s_{22}
s_{11}	α_1, α_2	β_1, γ_2
s_{12}	γ_1, β_2	δ_1, δ_2

Figure 4: 2×2 Underlying Game G

All payoffs are assumed to be non-negative and the following three cases will be examined.

$$\text{Case 1} \quad \alpha_i > \gamma_i, \delta_i \geq \beta_i, \alpha_i > \delta_i \quad (i = 1, 2) \quad (11)$$

$$\text{Case 2} \quad \alpha_i > \gamma_i, \delta_i \geq \beta_i, [\delta_i < \alpha_i \text{ and } \delta_j \geq \alpha_j] \quad (i = 1, 2, i \neq j) \quad (12)$$

$$\text{Case 3} \quad \alpha_i < \gamma_i, \beta_i < \delta_i, \delta_i < \alpha_i \quad (i = 1, 2) \quad (13)$$

In Case 1 and Case 2, first two conditions are identical: (s_{11}, s_{21}) is a strict Nash equilibrium and (s_{12}, s_{22}) is a (not necessarily strict) Nash equilibrium. The difference between these cases is the third condition. In Case 1, (α_1, α_2) is strictly preferred to both parties, while Case 2 exhibits conflicts between (α_1, α_2) and (δ_1, δ_2) . Therefore, Case 1 includes *coordination game*, and Case 2 includes *battle of sex* as a typical case, respectively. In the last case, s_{i2} is the dominant strategy, thus socially optimal outcome (α_1, α_2) is not a Nash equilibrium in G . *Prisoners' dilemma* is included in this case.

For each case, we will prove that every socially efficient outcome (say $\omega \in \mathbf{R}^2$) can be fully implemented in U-SPNE. The way to prove all the theorems below – Theorem 3.1, 3.2 and 3.3 – can be summarized under the same structure;

$$\left\{ \begin{array}{l} \text{Step 1} \quad \text{Construct } (c^*, l^*) \equiv (c_1^*, c_2^*, (d_1^*, r_1^*), (d_2^*, l_2^*)) \\ \text{Step 2} \quad \text{Check } u(r^*(c^*)) = \omega \text{ and } d_1^*(c^*) = d_2^*(c^*) = P \\ \text{Step 3} \quad \text{Check the condition of (7) to (9).} \\ \text{Step 4} \quad \text{Prove the fullness of } (c^*, l^*) \end{array} \right.$$

With this setup, we obtain the following results.

Theorem 3.1

Under the condition of (11), there exists a strategy profile (c^, l^*) such that c^* fully implements (α_1, α_2) in U-SPNE via l^* if the following is satisfied:*

$$0 < \epsilon \leq \min\{\alpha_1 - \gamma_1, \alpha_2 - \gamma_2\}.$$

Proof. Step 1. Construct the following $(c^*, l^*) = (c_1^*, c_2^*, (d_1^*, r_1^*), (d_2^*, r_2^*))$ such that:

$$c_i^*(\alpha_i) = \epsilon, \quad c_i^*(\beta_i) = c_i^*(\gamma_i) = c_i^*(\delta_i) = 0^3 \quad (i = 1, 2) \quad (14)$$

$$d_i^*(c) \text{ satisfying (8)} \quad (i = 1, 2). \quad (15)$$

$$r_1^*(c) = \begin{cases} s_1^d & \text{such that } s_1^d \text{ is weakly dominant strategy (w.d.s)} & \text{if there exists such } s_1^d \\ \tilde{s}_1 & \text{such that } \arg \max_{s_1} c_1(u_1(s_1, s_2^d)) & \text{if } \nexists \text{ w.d.s } s_1^d \text{ and } \exists \text{ w.d.s } s_2^d \\ s_{12} & \text{otherwise} \end{cases} \quad (16)$$

and same for agent 2.

Step 2. Note that $G(c^*)$ is described as Fig.5.

$G(c^*)$	s_{21}	s_{22}
s_{11}	ϵ, ϵ	$0, 0$
s_{12}	$0, 0$	$0, 0$

Figure 5: Induced Game $G(c^*)$

Since s_{11} is weakly dominant strategy in $G(c^*)$ and from the construction of $r_i^*(c)$, $u(r^*(c^*)) = u(s_{11}, s_{21}) = (\alpha_1, \alpha_2)$. Furthermore, $d_i(c^*) = P$ since $c_i^*(u_i(r_1^*(c^*), r_2^*(c^*))) = c_i^*(u_i(s_{11}, s_{21})) = \epsilon$.

Step 3. We check (7) First. For each $c \in C$, its induced game $G(c)$ can be divided into the following three cases of (A) to (C). (A) If both agents have weakly dominant strategies in $G(c)$, s_1^d and s_2^d , then each of them uses this strategy and $(s_1^d, s_2^d) \in UEA(c)$. (B) W.l.o.g. suppose that a_1 has no weakly dominant strategy and that a_2 has a weakly dominant strategy s_2^d . Then a_2 chooses s_2^d and, from the construction of r_1^* , a_1 chooses his best response to s_2^d , denoted by \tilde{s}_1 . Note that \tilde{s}_1 is not weakly dominated since a_1 has only two strategies and none of them is assumed to be a weakly dominant strategy. Therefore, $(\tilde{s}_1, s_2^d) \in UEA(c)$. (C) If none of the players has weakly dominant strategy, then they uses s_{12} . Since c is non-decreasing and (s_{12}, s_{22}) is N.E in the underlying game, $(s_{12}, s_{22}) \in UEA(c)$.

Next we check (8) but it is obvious from (15).

³In FJK's paper, c^* is named a *target compensation function*, characterized by "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."

Finally, we check (9). W.l.o.g, consider p_1 's deviation from $(c_1^*, c_2^*, l_1^*, l_2^*)$ to $(\tilde{c}_1, c_2^*, l_1^*, l_2^*)$. Note that $U_i^p(c_1^*, c_2^*, l_1^*, l_2^*) = \alpha_1 - \epsilon$, and for all $G(\tilde{c}_1, c_2^*)$, one of the following two cases, (A) and (B), must hold. (A) Suppose there exists no a_1 's weakly dominant strategy in $G(\tilde{c}_1, c_2^*)$. Since s_{21} is a_2 's weakly dominant strategy in $G(\tilde{c}_1, c_2^*)$, $r_2^*(\tilde{c}_1, c_2^*) = s_{21}$ and a_1 uses \tilde{s}_1 in (16), that is, a best response to s_{21} . Since $\alpha_1 - \epsilon \geq \gamma_1$, the only possibility for the deviation to \tilde{c}_1 to be strictly better for p_1 than c_1^* given (c_2^*, l_1^*, l_2^*) is $r_1^*(\tilde{c}_1, c_2^*) = s_{11}$ and $\tilde{c}_1(\alpha_1) = \delta$ with $\delta < \epsilon$. However, this implies $d_1(\tilde{c}_1, c_2^*) = NP$ by (15), which results in zero payoff to p_1 . Therefore, \tilde{c}_1 cannot be a strictly preferable deviation. (B) For (\tilde{c}_1, c_2^*) , suppose there exists a_1 's weakly dominant strategy, s_1^d . From the non-decreasing property of the compensation function, $s_1^d = s_{11}$, thus $r^*(\tilde{c}_1, c_2^*) = (s_{11}, s_{21})$. Similarly, for \tilde{c}_1 to be strictly better than c_1^* , the only possibility is that $\tilde{c}_1(\alpha_1) = \delta$ with $\delta < \epsilon$. However, this implies $d_1(\tilde{c}_1, c_2^*) = NP$ by (15), which results in zero payoff to p_1 . Therefore, \tilde{c}_1 cannot be a strictly preferable deviation. Note that the assumption of $0 < \epsilon \leq \min\{\alpha_1 - \gamma_1, \alpha_2 - \gamma_2\}$ guarantees that the same argument holds for principal 2's deviation from $(c_1^*, c_2^*, r_1^*, r_2^*)$.

Step 4. Suppose there exists \tilde{l} such that c^* implements $\tilde{\omega} \in \mathbf{R}^2$ in U-SPNE via \tilde{l} . Since s_{12} is weakly dominated strategy in $G(c^*)$, s_{11} has to be played in $G(c^*)$. Therefore, $\tilde{\omega} = (\alpha_1, \alpha_2)$. ■

Theorem 3.1 implies that the socially optimal outcome can be fully implemented in U-SPNE. To compare our result with FJK, consider the following example.

Example 3.1

Suppose that the underlying game G is the following coordination game:

G	s_{21}	s_{22}
s_{11}	2, 2	0, 0
s_{12}	0, 0	1, 1

Figure 6: Underlying Game G : Coordination Game

For this underlying game, our equilibrium compensation scheme, c^* in (14), and FJK's compensation function constructed in their proof (See Fershtman, Judd, and Kalai [3]) generate the same induced game which is shown in Fig.7.

$G(c^*)$	s_{21}	s_{22}
s_{11}	ϵ, ϵ	$0, 0$
s_{12}	$0, 0$	$0, 0$

Figure 7: Induced Game $G(c^*)$

In Fig.7, Nash equilibria in the induced game among agents are (s_{11}, s_{21}) and (s_{12}, s_{22}) . If we employ SPNE as the solution concept to this game, (s_{12}, s_{22}) cannot be excluded thus the problem of multiple equilibria would remain. Both of our solution concept of U-SPNE and FJK's implementability with MRA, on the other hand, succeeds to exclude (s_{12}, s_{22}) . However, the underlying logic is different: In U-SPNE, (s_{12}, s_{22}) is excluded because they are weakly dominated strategy. On the other hand, in FJK's arguments, SPNE, which they employed as the equilibrium concept, imposes no restriction on each agent's behavior about which equilibrium is chosen. Instead of employing refined equilibrium concept, FJK used refined implementability. Specifically, MRF requires agent i to play s_{i1} because (s_{11}, s_{21}) gives ϵ to BOTH agents. They argued that;

"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."

Two remarks should be made. First, if we take the basic approach of mechanism design into account, such refinement should be imposed not on implementability side but on the equilibrium concept. Second, within the notion of SPNE, there is no reason why agent 1 is sure that agent 2 does attend and takes s_{21} . In our notions, on the other hand, agent i thinks that j takes s_{j1} for sure because s_{j1} is the only undominated strategy. ■

Next, we turn to the second case of (12).

Theorem 3.2

Under the condition of (12), there exists a strategy profile (c^*, l^*) such that c^* fully implements (α_1, α_2) in U-SPNE via r^* if the following is satisfied:

$$\epsilon \leq \min\{\alpha_i - \gamma_i, \alpha_j - \gamma_j\} \text{ and } \epsilon \leq \alpha_i - \delta_i.$$

Proof. Here we prove the case where $i = 1$ and $j = 2$.

Step 1. From the assumption, we can construct the following c_1^* :

$$c_1^*(\alpha_1) = \epsilon, \quad c_1^*(\beta_1) = c_1^*(\gamma_1) = c_1^*(\delta_1) = 0. \quad (17)$$

On the other hand, assumptions regarding p_2 's utility can be summarized;

$$\gamma_2 < \alpha_2 \leq \delta_2 \quad \text{and} \quad \beta_2 \leq \delta_2. \quad (18)$$

Now depending on magnitude between α_2 and β_2 , consider the following c_2^* .⁴

$$c_2^* : \begin{cases} c_2^*(\alpha_2) = c_2^*(\delta_2) = \epsilon, \quad c_2^*(\beta_2) = c_2^*(\gamma_2) = 0 & \text{if } \beta_2 < \alpha_2 \\ c_2^*(\alpha_2) = c_2^*(\delta_2) = c_2^*(\beta_2) = \epsilon, \quad c_2^*(\gamma_2) = 0 & \text{if } \beta_2 \geq \alpha_2 \end{cases} \quad (19)$$

and a_i 's strategy is same as (15) and (16).

Note that the induced game for each case is the following.

$G(c^*)$	s_{21}	s_{22}	$G(c^*)$	s_{21}	s_{22}
s_{11}	ϵ, ϵ	$0, 0$	s_{11}	ϵ, ϵ	$0, 0$
s_{12}	$0, 0$	$0, \epsilon$	s_{12}	$0, \epsilon$	$0, \epsilon$

Figure 8: (Left) First case of (19) (Right) Second case of (19)

Step 2. In the left case of Fig.8, a_1 chooses s_{11} since this is the weakly dominant strategy, while a_2 chooses s_{21} since a_2 has no weakly dominant strategy and s_{21} is the best response to s_{11} . Therefore, (s_{11}, s_{21}) is played. In the right case, (s_{11}, s_{21}) is also played since both of these are the weakly dominant strategies. Thus $u(r^*(c^*)) = u(s_{11}, s_{21}) = (\alpha_1, \alpha_2)$ is true for both cases and $d_i(c^*) = P$ holds.

Step 3. The condition (7) holds by the exactly same reason as Theorem 3.1 (Recall that r_i^* is same as case 1 and the argument does not depend on the compensation scheme.). The condition (8) is also obviously satisfied.

⁴Such construction is necessary so as not to violate non-decreasing assumption of c_2^* . For example, $\beta_2 < \alpha_2$ together with (18) does not guarantee $\gamma_2 < \beta_2$, which implies under this condition, the second line of (19) violates non-decreasing assumption.

Then we check the condition (9) by considering each case of Fig.8 separately.

Left case of Fig.8. First, pick p_1 's deviation from $(c_1^*, c_2^*, r_1^*, r_2^*)$ to $(\tilde{c}_1, c_2^*, r_1^*, r_2^*)$, and note that $U_1^p(c_1^*, c_2^*, r_1^*, r_2^*) = \alpha_1 - \epsilon$. One of the following two cases of (A) or (B) must hold. (A) For (\tilde{c}_1, c_2^*) , suppose there exists no a_1 's weakly dominant strategy in $G(\tilde{c}_1, c_2^*)$. Since both players have no weakly dominant strategy and from (16), (s_{12}, s_{22}) is played. Therefore p_1 gets at most δ_1 . However, since $\epsilon \leq \alpha_1 - \delta_1$, such deviation cannot be strictly preferable to p_1 . (B) For (\tilde{c}_1, c_2^*) , suppose there exists a_1 's weakly dominant strategy, s_1^d . From the non-decreasing property of the compensation function, $s_1^d = s_{11}$. On the other hand, from \tilde{s}_2 in (16), a_2 takes s_{21} , which is the unique best response to s_{11} under c_2^* . Thus (s_{11}, s_{21}) is played. For \tilde{c}_1 to be strictly better than c_1^* , the only possibility is that $\tilde{c}_1(\alpha_1) = \delta$ with $\delta < \epsilon$. However, from (8), $d_1^*(\tilde{c}_1, c_2^*) = NP$ for such (\tilde{c}_1, c_2^*) since $\tilde{c}_1(u_1(r_1^*(\tilde{c}_1, c_2^*), r_2^*(\tilde{c}_1, c_2^*))) = \tilde{c}_1(u_1(s_{11}, s_{21})) = \delta < \epsilon$. This results in zero payoff to p_1 . Therefore, \tilde{c}_1 cannot be a strictly preferable deviation. Next, consider p_2 's deviation from $(c_1^*, c_2^*, r_1^*, r_2^*)$ to $(c_1^*, \tilde{c}_2, r_1^*, r_2^*)$. Note that $U_2^p(c_1^*, c_2^*, r_1^*, r_2^*) = \alpha_2 - \epsilon$ and for every $G(c_1^*, \tilde{c}_2)$, a_1 takes s_{11} because this is weakly dominant strategy under c_1^* . (A) For (c_1^*, \tilde{c}_2) , suppose there exists no a_2 's weakly dominant strategy in $G(c_1^*, \tilde{c}_2)$. (1) Suppose a_2 takes s_{21} , then (s_{11}, s_{21}) is played thus $\tilde{c}_2(\alpha_2) = \delta$ with $\delta < \epsilon$ has to be satisfied for \tilde{c}_2 to be a strictly preferable deviation for p_2 . However, this implies $d_2(c_1^*, \tilde{c}_2) = NP$ by (15), which results in zero payoff to p_2 . Therefore, \tilde{c}_2 cannot be a strictly preferable deviation. (2) If a_2 takes s_{22} , then (s_{11}, s_{22}) is played thus $\tilde{c}_2(\gamma_2) = \delta$ with $\alpha_2 - \epsilon < \gamma_2 - \delta$ has to be satisfied for \tilde{c}_2 to be strictly preferable. Again, however, this implies $\delta < \epsilon - (\alpha_2 - \gamma_2) < \epsilon$. (B) For (c_1^*, \tilde{c}_2) , suppose there exists a_2 's weakly dominant strategy, s_2^d . (1) If $s_2^d = s_{21}$, then (s_{11}, s_{21}) is played. Thus $\tilde{c}_2(\alpha_2) = \delta$ with $\delta < \epsilon$ is required for \tilde{c}_2 to be a strictly better deviation. However, from (8), $d_2^*(c_1^*, \tilde{c}_2) = NP$ by (15), which results in zero payoff to p_2 . Therefore, \tilde{c}_2 cannot be a strictly preferable deviation. (2) Suppose $s_2^d = s_{22}$, then (s_{11}, s_{22}) is played. Thus, $\tilde{c}_2(\gamma_2) = \delta$ with $\alpha_2 - \epsilon < \gamma_2 - \delta$ has to be satisfied for \tilde{c}_2 to be strictly preferable. Again, however, this implies $\delta < \epsilon - (\alpha_2 - \gamma_2) < \epsilon$.

Right case of Fig.8. Again, first pick p_1 's deviation from $(c_1^*, c_2^*, r_1^*, r_2^*)$ to $(\tilde{c}_1, c_2^*, r_1^*, r_2^*)$, and note that $U_1^p(c_1^*, c_2^*, r_1^*, r_2^*) = \alpha_1 - \epsilon$. Furthermore, for every $G(\tilde{c}_1, c_2^*)$, a_2 takes s_{21} because this is weakly dominant strategy. (A) For (\tilde{c}_1, c_2^*) , suppose there exists no a_1 's weakly dominant strategy in $G(\tilde{c}_1, c_2^*)$. (1) Suppose a_1 takes s_{11} , then (s_{11}, s_{21}) is played thus $\tilde{c}_1(\alpha_1) = \delta$ with $\delta < \epsilon$ has to be satisfied for \tilde{c}_1 to be a strictly preferable deviation for p_1 . However, this implies $d_1(\tilde{c}_1, c_2^*) = NP$

by (15), which results in zero payoff to p_1 . Therefore, \tilde{c}_2 cannot be a strictly preferable deviation.

(2) If a_1 takes s_{12} , then (s_{12}, s_{21}) is played thus $\tilde{c}_1(\gamma_1) = \delta$ with $\alpha_1 - \epsilon < \gamma_1 - \delta$ has to hold. Again, however, this implies $\delta < \epsilon - (\alpha_1 - \gamma_1) < \epsilon$. (B) For (\tilde{c}_1, c_2^*) , suppose there exists a_1 's weakly dominant strategy, s_1^d . In both cases of $s_1^d = s_{11}$ and $s_1^d = s_{12}$, exactly same arguments as (1) and (2) above hold, respectively. Next, consider p_2 's deviation from $(c_1^*, c_2^*, r_1^*, r_2^*)$ to $(c_1^*, \tilde{c}_2, r_1^*, r_2^*)$ and note that $U_2^p(c_1^*, c_2^*, r_1^*, r_2^*) = \alpha_2 - \epsilon$. Furthermore, for every $G(c_1^*, \tilde{c}_2)$, a_1 takes s_{11} because this is weakly dominant strategy. (A) For (c_1^*, \tilde{c}_2) , suppose there exists no a_2 's weakly dominant strategy in $G(c_1^*, \tilde{c}_2)$. (1) Suppose a_2 takes s_{21} , then (s_{11}, s_{21}) is played thus $\tilde{c}_2(\alpha_2) = \delta$ with $\delta < \epsilon$ has to be satisfied for \tilde{c}_2 to be a strictly preferable deviation for p_2 . However, this implies $d_2(c_1^*, \tilde{c}_2) = NP$ by (15), which results in zero payoff to p_2 . Therefore, \tilde{c}_2 cannot be a strictly preferable deviation. (2) If a_2 takes s_{22} , then (s_{11}, s_{22}) is played thus $\tilde{c}_2(\gamma_2) = \delta$ with $\alpha_2 - \epsilon < \gamma_2 - \delta$ has to be satisfied for \tilde{c}_2 to be strictly preferable. Again, however, this implies $\delta < \epsilon - (\alpha_2 - \gamma_2) < \epsilon$. (B) For (c_1^*, \tilde{c}_2) , suppose there exists a_2 's weakly dominant strategy, s_2^d . (1) If $s_2^d = s_{21}$, then (s_{11}, s_{21}) is played. Thus $\tilde{c}_2(\alpha_2) = \delta$ with $\delta < \epsilon$ is required for \tilde{c}_2 to be a strictly better deviation. However, from (8), $d_2^*(c_1^*, \tilde{c}_2) = NP$ by (15), which results in zero payoff to p_2 . Therefore, \tilde{c}_2 cannot be a strictly preferable deviation. (2) Suppose $s_2^d = s_{22}$, then (s_{11}, s_{22}) is played. Thus, $\tilde{c}_2(\gamma_2) = \delta$ with $\alpha_2 - \epsilon < \gamma_2 - \delta$ has to be satisfied for \tilde{c}_2 to be strictly preferable. Again, however, this implies $\delta < \epsilon - (\alpha_2 - \gamma_2) < \epsilon$.

Step 4. The fullness of implementability is obvious because, in both cases, the *only* element of $UEA(c^*)$ is (s_{11}, s_{21}) . ■

Example 3.2

Suppose that the underlying game G is the following "battle of sexes":

G	s_{21}	s_{22}
s_{11}	2, 1	0, 0
s_{12}	0, 0	1, 2

Figure 9: Underlying Game G : Battle of sexes

For this underlying game, our equilibrium compensation scheme and FJK's compensation scheme generate the same induced game which is shown in Fig.10.

$G(c^*)$	s_{21}	s_{22}
s_{11}	ϵ, ϵ	$0, 0$
s_{12}	$0, 0$	$0, \epsilon$

Figure 10: Induced Game $G(c^*)$

Again, Nash equilibria in $G(c^*)$ are (s_{11}, s_{21}) and (s_{12}, s_{22}) . Agent 1 has weakly dominant strategy s_{11} , while no weakly dominant strategy for Agent 2. U-SPNE advocate that Agent 2 thinks Agent 1 plays s_{11} for sure. Given this, Agent 2 chooses the best response to s_{11} .

Note that, exactly same argument holds for the full implementability of $(1, 2)$. In order to investigate which outcome is more likely to be implemented, we need to examine the uniqueness of U-SPNE. In this paper, as well as FJK's original argument, the issue of uniqueness of equilibrium in the whole game is out of focus and put for further research. ■

Next, we turn to the final case of (13).

Theorem 3.3

Under the condition of (13), there exists a strategy profile (c^*, l^*) such that c^* fully implements (α_1, α_2) in U-SPNE via r^* if the following is satisfied:

$$0 < \epsilon \leq \min\{\alpha_1 - \delta_1, \alpha_2 - \delta_2\}. \quad (20)$$

Proof. Step 1. Construct the following $(c^*, l^*) = (c_1^*, c_2^*, (d_1^*, r_1^*), (d_2^*, r_2^*))$ such that:

$$c_i^*(\alpha_i) = c_i^*(\gamma_i) = \epsilon, \quad c_i^*(\beta_i) = c_i^*(\delta_i) = 0 \quad (i = 1, 2) \quad (21)$$

$$d_i^*(c) \text{ satisfying (8)} \quad (i = 1, 2). \quad (22)$$

$$r_1^*(c) = \begin{cases} s_{11} & \text{if } c = c^* \\ s_{12} & \text{if otherwise.} \end{cases} \quad (23)$$

and same for agent 2.

Step 2. Note that $G(c^*)$ is described as Fig.11.

$G(c^*)$	s_{21}	s_{22}
s_{11}	ϵ, ϵ	$0, \epsilon$
s_{12}	$\epsilon, 0$	$0, 0$

Figure 11: Induced Game $G(c^*)$

From the construction of $r_i^*(c)$, $u(r^*(c^*)) = u(s_{11}, s_{21}) = (\alpha_1, \alpha_2)$. Furthermore, $d_i(c^*) = P$ since $c_i^*(u_i(r_1^*(c^*), r_2^*(c^*))) = c_i^*(u_i(s_{11}, s_{21})) = \epsilon$.

Step 3. In $G(c^*)$, $(s_{11}, s_{21}) \in UEA(c^*)$. Furthermore, for every c , $(s_{12}, s_{22}) \in UEA(c)$ is true since s_{12} is the dominant strategy in the underlying game. Thus (7) holds. (8) also holds from (22). Finally we check (9). Note that $U_i^p(c_1^*, c_2^*, l_1^*, l_2^*) = \alpha_i - \epsilon$. For p_i 's any deviation, (s_{12}, s_{22}) is played, thus p_i 's payoff is at most δ_i . However, $\delta_i \leq \alpha_i - \epsilon$ holds from (20) thus there exists no strictly preferable deviation from $(c_1^*, c_2^*, l_1^*, l_2^*)$ for p_i .

Step 4. Suppose, in negation, that c^* also implements $\tilde{\omega} = (\tilde{\omega}_1, \tilde{\omega}_2) \in \{(\beta_1, \gamma_2), (\gamma_1, \beta_2), (\delta_1, \delta_2)\}$ in U-SPNE via some $\tilde{l} = ((\tilde{d}_1, \tilde{r}_1), (\tilde{d}_2, \tilde{r}_2))$. Then, from the first condition of Definition 2.2, $u(\tilde{r}(c^*)) = \tilde{\omega}$. However, for any $\tilde{\omega}$, there exists $i \in \{1, 2\}$ such that $c_i(u_i(\tilde{r}_1(c^*), \tilde{r}_2(c^*))) = c_i(\tilde{\omega}_i) = 0 < \epsilon$. From (8), this implies $\tilde{d}_i(c^*) = NP$, which contradicts to the second condition of Definition 2.2. ■

Example 3.3

Suppose that the underlying game G is the following "prisoners' dilemma ":

G	s_{21}	s_{22}
s_{11}	$4, 4$	$0, 5$
s_{12}	$5, 0$	$1, 1$

Figure 12: Underlying Game G : Prisoners' Dilemma

For this underlying game, our equilibrium compensation scheme and FJK's compensation scheme generate the same induced game which is shown in Fig.13.

$G(c^*)$	s_{21}	s_{22}
s_{11}	ϵ, ϵ	$0, \epsilon$
s_{12}	$\epsilon, 0$	$0, 0$

Figure 13: Induced Game $G(c^*)$

In $G(c^*)$, $UEA(c^*) = S_1 \times S_2$, that is, all outcomes are in weakly undominated Nash equilibrium. Compensation scheme c^* succeed to make (s_{11}, s_{21}) be a undominated Nash equilibrium.

On the other hand, if we would want to derive unique equilibrium in an induced game, “stronger” equilibrium concept may be needed. However, it should be noted that if principals directly plays the underlying game G , there always an incentive for agent i to deviate from s_{i1} to s_{i2} . By introducing agents, such deviation is deterred because it triggers agents’ further deviation to play (s_{12}, s_{22}) . ■

4 An Example of 3×3 Games - The Problem of Multiple Equilibria

In this section, we show that, by using an example of 3×3 coordination game, not only does our standard procedure attain full implementability, our solution concepts can also solve problem of multiple equilibria, an issue where FJK’s solution concepts fall short. Consider the following example:

Example 4.1

Suppose the underlying game is the following:

G	s_{21}	s_{22}	s_{23}
s_{11}	k_1, k_2	$k_1, 0$	$0, 0$
s_{12}	$0, k_2$	k_1, k_2	$0, 0$
s_{13}	$0, 0$	$0, 0$	$0, 0$

Figure 14: Underlying Game $G : 3 \times 3$ Coordination Game

For this underlying game, we establish the following result.

Remark 4.1

For the underlying game of Fig.14, there exists a strategy profile (c^*, l^*) such that c^* fully implements (k_1, k_2) in U-SPNE via l^* if the following is satisfied:

$$0 < \epsilon \leq \min\{k_1, k_2\} \tag{24}$$

Before proving the theorem, we first show the strategy profile (c^*, l^*) in order to explain why our approach works in this game compared to FJK’s solution concepts.

Construct the following $(c^*, l^*) = (c_1^*, c_2^*, (d_1^*, r_1^*), (d_2^*, r_2^*))$ such that:

$$c_i^*(k_i) = \epsilon, \quad c_i^*(0) = 0 \quad (i = 1, 2) \quad (25)$$

$$d_i^*(c) \text{ satisfying (8)} \quad (i = 1, 2). \quad (26)$$

$$r_1^*(c) = \begin{cases} s_1^d & \text{such that } s_1^d \text{ is weakly dominant strategy (w.d.s)} & \text{if there exists such } s_1^d \\ \tilde{s}_1 & \text{such that } \arg \max_{s_1} c_1(u_1(s_1, s_2^d)) & \text{if } \nexists \text{ w.d.s } s_1^d \text{ and } \exists \text{ w.d.s } s_2^d \\ s_{13} & \text{otherwise} \end{cases} \quad (27)$$

and same for agent 2.

Note that $G(c^*)$ is described as Fig.15.

	s_{21}	s_{22}	s_{23}
s_{11}	ϵ, ϵ	$\epsilon, 0$	$0, 0$
s_{12}	$0, \epsilon$	ϵ, ϵ	$0, 0$
s_{13}	$0, 0$	$0, 0$	$0, 0$

Figure 15: Induced Game $G(c^*)$

In FJK's argument, either (s_{11}, s_{21}) or (s_{12}, s_{22}) is assumed to be played, but there's no justification which outcome is more likely to be played. In other words, "coordination failure" could happen and thus the difficulty of multiple equilibrium remains under the assumption of mutually rational agents. On the other hand, since s_{i2} is a weakly dominated strategy, these are never played under U-SPNE. Thus our solution concepts can justify to play (s_{11}, s_{21}) and exclude the possibility of playing (s_{12}, s_{22}) , since $UEA(c^*) = \{(s_{11}, s_{21})\}$ is singleton.

Such difference comes from the characteristics of our solution concepts. In FJK, the assumption of mutually rational players is imposed on payoff profiles or on agents' collective action. On the other hand, U-SPNE is a refinement based on individual behaviors: weakly dominated strategy.

Now we complete the proof.

Proof of Remark4.1 Step 1. See (25) to (27).

Step 2. Since s_{i1} is weakly dominant strategy in $G(c^*)$ and from the construction of $r_i^*(c)$, $u(r^*(c^*)) = u(s_{11}, s_{21}) = (k_1, k_2)$. Furthermore, $d_i(c^*) = P$ since $c_i^*(u_i(r_1^*(c^*), r_2^*(c^*))) = c_i^*(k_i) = \epsilon$.

Step 3. We check (7) First. For each $c \in C$, its induced game $G(c)$ can be divided into the following three cases of (A) to (C). (A) If both agents have weakly dominant strategies in $G(c)$, s_1^d and s_2^d , then each of them uses this strategy and $(s_1^d, s_2^d) \in UEA(c)$. (B) W.l.o.g. suppose that a_1 has no weakly dominant strategy and that a_2 has a weakly dominant strategy s_2^d . Then a_2 chooses s_2^d and, from the construction of r_1^* , a_1 chooses his best response to s_2^d , denoted by \tilde{s}_1 . Note that \tilde{s}_1 is not weakly dominated since a_1 has only two strategies and none of them is assumed to be a weakly dominant strategy. Therefore, $(\tilde{s}_1, s_2^d) \in UEA(c)$. (C) If none of the players has weakly dominant strategy, then they uses s_{i2} . Since c is non-decreasing and (s_{12}, s_{22}) is N.E in the underlying game, $(s_{12}, s_{22}) \in UEA(c)$. Next we check (8) but it is obvious from (26).

Finally, we check (9). W.l.o.g, consider p_1 's deviation from $(c_1^*, c_2^*, l_1^*, l_2^*)$ to $(\tilde{c}_1, c_2^*, l_1^*, l_2^*)$. Note that $U_1^p(c_1^*, c_2^*, l_1^*, l_2^*) = k_1 - \epsilon$, and for all $G(\tilde{c}_1, c_2^*)$, one of the following two cases, (A) and (B), must hold. (A) Suppose there exists no a_1 's weakly dominant strategy in $G(\tilde{c}_1, c_2^*)$. Since s_{21} is a_2 's weakly dominant strategy in $G(\tilde{c}_1, c_2^*)$, $r_2^*(\tilde{c}_1, c_2^*) = s_{21}$ and a_1 uses \tilde{s}_1 in (27), that is, a best response to s_{21} . Since $k_1 - \epsilon \geq 0$, the only possibility for the deviation to \tilde{c}_1 to be strictly better for p_1 than c_1^* given (c_2^*, l_1^*, l_2^*) is $r_1^*(\tilde{c}_1, c_2^*) = s_{11}$ and $\tilde{c}_1(k_1) = \delta$ with $\delta < \epsilon$. However, this implies $d_1(\tilde{c}_1, c_2^*) = NP$ by (15), which results in zero payoff to p_1 . Therefore, \tilde{c}_1 cannot be a strictly preferable deviation. (B) For (\tilde{c}_1, c_2^*) , suppose there exists a_1 's weakly dominant strategy, s_1^d . From the non-decreasing property of the compensation function, $s_1^d = s_{11}$, thus $r^*(\tilde{c}_1, c_2^*) = (s_{11}, s_{21})$. Similarly, for \tilde{c}_1 to be strictly better than c_1^* , the only possibility is that $\tilde{c}_1(k_1) = \delta$ with $\delta < \epsilon$. However, this implies $d_1(\tilde{c}_1, c_2^*) = NP$ by (15), which results in zero payoff to p_1 . Therefore, \tilde{c}_1 cannot be a strictly preferable deviation. Note that the assumption of $0 < \epsilon \leq \min\{k_1, k_2\}$ guarantees that the same argument holds for principal 2's deviation from $(c_1^*, c_2^*, r_1^*, r_2^*)$.

Step 4. Suppose there exists \tilde{l} such that c^* implements $\tilde{\omega} \in \mathbb{R}^2$ in U-SPNE via \tilde{l} . Since s_{i2} and s_{i3} is weakly dominated strategy in $G(c^*)$, s_{i1} has to be played in $G(c^*)$. Therefore, $\tilde{\omega} = (k_1, k_2)$. ■

5 Conclusion

In this paper, we studied delegation games. First, we reviewed the model and solution concepts employed in FJK's analysis, then modified these in line with mechanism design methodology.

That is, we first constructed the delegation game where agents' participation decision is explicitly built in, then employed the weakly undominated subgame perfect Nash equilibrium (U-SPNE) as a refined equilibrium, while making the basic notion of the implementability unchanged. Given this, we showed that in certain classes of 2×2 game - including prisoners' dilemma, general coordination games, and battle of sexes - every socially optimal outcome is fully implementable in U-SPNE. We also argued that, by using an example of 3×3 coordination game, not only does our standard procedure attain full implementability, our solution concepts could also solve problem of multiple equilibria, an issue where FJK's solution concepts fall short.

We conclude this paper by pointing out several issues for further research. First, our delegation framework can obviously be extended to any strategic form games. A delegation game where its underlying game is Bertrand price competition or Cournot duopoly could be intriguing. Next, in this paper, we assume that, if either of agents chooses not to participate the underlying game, principals get a constant payoff by exerting an outside option. This setup is natural to describe a situation where considerable expertise – any knowledge or experiences – is needed to enter the underlying game and only agents have such expertise. On the other hand, an alternative story, for example, a principal enters the game instead of its agent would be intriguing for further research. Finally, as discussed in Introduction and Example 3.2, we could alternatively take uniqueness of (U-) SPNE into consideration, instead of uniqueness of an outcome. Whether such uniqueness of the equilibrium can be obtained in our setting could be an interesting question.

References

- [1] Austen-Smith.D and Banks.J (2005). *Positive Political Theory II* Ann Arbor, University of Michigan Press.
- [2] Fershtman.C and Judd.K.L (1987). Equilibrium Incentives in Oligopoly, *American Economic Review* 77, 927-940.
- [3] Fershtman.C, Judd.K.L and Kalai.E. (1991). Observable Contracts: Strategic Delegation and Cooperation, *International Economic Review* 32, 551-559.

- [4] Jackson.M.O, Palfrey.T.R and Srivastava.S (1994). Undominated Nash Implementation in Bounded Mechanisms, *Games and Economic Behavior* **6**, 474-501.
- [5] Palfrey.T.R and Srivastava.S (1991). Nash Implementation Using Undominated Strategies, *Econometrica* **59**, 479-501.
- [6] Persson.T, Roland.G and Tabellini.G (1998). Towards Micropolitical Foundations of Public Finance, *European Economic Review* **42**, 685-694.
- [7] Persson.T and Tabellini.G (1995). Double-edged incentives: Institutions and Policy coordination, in G.Grossman and K.Rogoff (Eds.) *Handbook of International Economics III*, North-Holland, Amsterdam.
- [8] Schelling.T.C (1960). *The Strategy of Conflict*. Cambridge: Harvard University Press.
- [9] Sjostrom.T (1994). Implementation in Undominated Nash Equilibria without Integer Games, *Games and Economic Behavior* **6**, 502-511.
- [10] Yu.Y (2005). Public Goods Provision, Unit-by-Unit Cost-Sharing Rules and the Core, *Review of Economic Design* **9**, 363-376.