Full Implementation in Backwards Induction

Yi-Chun Chen† Yifei Sun‡

May 22, 2014

Abstract

In a complete-information environment with two or more players and a finite type space, we show that any social choice function can be fully implemented in backwards induction via a finite perfect-information stochastic mechanism with arbitrarily small transfers.

1 Introduction

In a complete-information environment with two or more players and a finite type space, we show that any social choice function can be fully implemented in backwards induction using a finite perfect-information stochastic mechanism. Our result is achieved by invoking (1) a dynamic stochastic mechanism; (2) arbitrarily small transfers; (3) the domain restriction which rules out identical preferences and preference orderings with complete indifference over all outcomes.

Subgame-perfect implementation is known to be more permissive than Nash implementation (Moore and Repullo (1988)). Our result can be contrasted with two existing perfect-information mechanisms which implement an arbitrary social choice function in subgame-perfect equilibrium. The mechanism in (Moore and Repullo, 1988, Section 5.1) imposes large off-equilibrium transfers, while the mechanism in Glazer and Perry (1996) requires at least three

---

*We thank Eddie Dekel, Peter Hammond, Takashi Kunitomo, Ehud Lehrer, Xiao Luo, Rene Saran, Satoru Takahashi, Olivier Terceux, Cheng-Zhong Qin, Shmuel Zamir, and conference/seminar participants at AMES 2013 and the RCGEB Workshop on Game Theory at Shandong University for very helpful comments. We also gratefully acknowledge financial support from the Singapore Ministry of Education Academic Research Fund Tier 1. All remaining errors are our own.

†Department of Economics, National University of Singapore, Singapore 117570, ecsycc@nus.edu.sg

‡Department of Economics, National University of Singapore, Singapore 117570, sunyifei@nus.edu.sg

1See (Glazer and Perry, 1996, p. 28) for a discussion of practical and theoretical reasons to favor sequential/perfect-information mechanisms. In particular, they argue that “sequential mechanisms, with backwards induction as their solution concept, seem to be more intuitive and simpler to understand than their simultaneous counterparts.”
players and it is virtual implementation, i.e., the desirable social outcome is obtained only with large probability.\footnote{See also (Osborne and Rubinstein, 1994, pp. 193-195) for an exposition of the result in Glazer and Perry (1996).} Both mechanisms have thus been criticized for their susceptibility to renegotiation (Jackson, 2001, pp. 689-690). In contrast, our mechanism obtains full implementation with small and only off-equilibrium transfers.\footnote{Admittedly, these mechanisms share the common problem of having lengthy game trees. Specifically, in (Moore and Repullo, 1988, Section 5.1) the length is proportional the number of players; in Glazer and Perry (1996) the length is proportional to the number of players and the probability of obtaining desirable outcomes; in our mechanism, the length is proportional to the number of players and inversely proportional to the amount of small transfers being allowed for.}

In a generic perfect-information game, the backwards induction outcome is induced by several notions of extensive-form rationalizability.\footnote{These solution concepts include, for example, the subgame rationalizability in Bernheim (1984) and the extensive-form rationalizability in Pearce (1984). See also Battigalli and Siniscalchi (2002) for an epistemic characterization of extensive-form rationalizability.} As we allow for small transfers, our mechanism can be made generic to implement any social choice function in these notions of extensive-form rationalizability. In contrast, Bergemann, Morris, and Tercieux (2011) show that a stronger version of the monotonicity condition due to Maskin (1999) is necessary for implementation in normal-form rationalizability.

Our result can also be contrasted with the static mechanism in Abreu and Matsushima (1994) which fully implements any social choice function in one round deletion of weakly dominated strategies followed by iterated deletion of strictly dominated strategies. Glazer and Rubinstein (1996) argue that an extensive-form game provides a “guide” for solving a normal-form game and thereby reduces the computational burden on the players. Glazer and Rubinstein (1996) define a solution concept called \textit{guided iteratively undominated strategies} and prove that a social choice function is implementable in guided iteratively undominated strategies if and only if it is implementable in subgame-perfect equilibrium in a perfect-information mechanism. It follows that our mechanism also implements any social choice function in guided iteratively undominated strategies.

The paper is organized as follows. Section 2 describes the environment and the main result. Section 3 presents the mechanism. Section 4 provides the proof. Section 5 concludes.

## 2 Environment

Let $N = \{1, 2, \ldots, n\}$ denote the set of players. The set of pure social alternatives is denoted by $A$, and $\Delta (A)$ denotes the set of all probability distributions over $A$ with countable supports. In this context, $a \in A$ denotes a pure social alternative and $l \in \Delta (A)$ denotes a lottery on $A$. 

\textbf{2 See also (Osborne and Rubinstein, 1994, pp. 193-195) for an exposition of the result in Glazer and Perry (1996).}

\textbf{3 Admittedly, these mechanisms share the common problem of having lengthy game trees. Specifically, in (Moore and Repullo, 1988, Section 5.1) the length is proportional the number of players; in Glazer and Perry (1996) the length is proportional to the number of players and the probability of obtaining desirable outcomes; in our mechanism, the length is proportional to the number of players and inversely proportional to the amount of small transfers being allowed for.}

\textbf{4 These solution concepts include, for example, the subgame rationalizability in Bernheim (1984) and the extensive-form rationalizability in Pearce (1984). See also Battigalli and Siniscalchi (2002) for an epistemic characterization of extensive-form rationalizability.}
For each player $i \in N$, denote by $\Theta_i$ a finite set of types of player $i$. The utility index of player $i$ over the set $A$ is denoted by $v_i : A \times \Theta_i \rightarrow \mathbb{R}$, where $v_i(a, \theta_i)$ specifies the bounded utility of player $i$ from the social alternative $a$, when he is of type $\theta_i$. Player $i$’s expected utility from a lottery $l \in \Delta (A)$ under type $\theta_i$ is $u_i(l, \theta_i) = \sum_{a \in A} l(a) v_i(a, \theta_i)$, which is well defined since $v_i(a, \theta_i)$ is bounded.

Following Abreu and Matsushima (1992) and Glazer and Perry (1996), we assume that (i) for each $\theta_i \in \Theta_i$, $v_i(\cdot, \theta_i)$ is not a constant function on $A$; (ii) for any two distinct types $\theta_i$ and $\theta'_i$, $v_i(\cdot, \theta_i)$ is not a positive affine transformation of $v_i(\cdot, \theta'_i)$. This restriction guarantees the reversal property which is used to elicit players’ true type. We also assume that player $i$’s utility is quasilinear in transfers, denoted by $u_i(l, \theta_i) + t_i$.

A planner aims to implement a social choice function that is a mapping $f : \Theta \rightarrow \Delta (A)$, where $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$. We assume that the true type profile $\psi \in \Theta$ is commonly known to the players but unknown to the planner.

We assume that the planner can fine or reward a player $i \in N$ and denote by $t_i$ the transfer from player $i$ to the planner. A finite sequential stochastic mechanism is a finite perfect-information game tree $\Gamma$ together with an outcome function $\zeta$ which specifies for each terminal history a lottery $l \in \Delta (A)$ with a transfer profile $(t_1, t_2, ..., t_n)$. A sequential mechanism $(\Gamma, \zeta)$ has fines and rewards bounded by $\bar{t}$ if $|t_i| \leq \bar{t}$ for every $i \in N$ and every terminal history.

We now state our main result as follows.

**Theorem 1** For any $n \geq 2$, social choice function $f$, and $\bar{t} > 0$, there exists a finite sequential stochastic mechanism with fines and rewards bounded by $\bar{t}$ such that for each type profile $\psi$, $f(\psi)$ with no transfer is the unique subgame perfect equilibrium outcome of the mechanism.

### 3 The Mechanism

#### 3.1 Preliminaries

Given a social choice function $f$, since $\Theta_i$ is finite for any $i$, we let

$$\xi = \max_{\theta_i \in \Theta_i, \theta_i' \in \Theta_i, i \in N} |u_i(f(\theta), \theta_i) - u_i(f(\theta'), \theta_i)|. \quad (1)$$

\[5\] Here we follow Abreu and Matsushima (1994) and Glazer and Perry (1996) in assuming that the space of type profiles is a product space.
Choose an integer $K$ and $\varepsilon > 0$ such that
\[
\xi/K < \varepsilon < \bar{t}/6.
\] (2)

Hence, $K$ is large when $\bar{t}$ is small. For any distinct types $\theta_i$ and $\theta'_i$, let $x_{\theta_i,\theta'_i}$ and $x_{\theta'_i,\theta_i}$ be two lotteries such that
\[
u_i(x_{\theta_i,\theta'_i}, \theta_i) > \nu_i(x_{\theta'_i,\theta_i}, \theta_i);
\]
\[
u_i(x_{\theta_i,\theta'_i}, \theta'_i) < \nu_i(x_{\theta'_i,\theta_i}, \theta'_i).
\]
The existence of $x_{\theta_i,\theta'_i}$ and $x_{\theta'_i,\theta_i}$ is guaranteed by the assumption on the preference. Let $L \equiv \{x_{\theta_i,\theta'_i}, x_{\theta'_i,\theta_i}\}_{\theta_i \neq \theta'_i, i \in N}$. Observe that $L$ is a finite set since $\Theta_i$ and $N$ are both finite.

3.2 The Mechanism and Implementation

The mechanism has $K + 2$ rounds. In each round $k \leq K + 1$, the players move sequentially. Player 1 moves first, player 2 moves second, and so on. In round $k \leq K$, each player $i$ announces a type profile $m_i^k \in \Theta$.

Let
\[
l = \sum_{k=1}^{K} \frac{1}{K} f (m_i^k).
\]

Note that $l$ is fixed after the first $K$ rounds. Then, by finiteness of $L$ and $\Theta_i$, choose $p_l \in (0, 1)$ such that for any $l' \in L$ and any $i \in N$, any $\theta_i \in \Theta_i$,
\[
|\nu_i(l, \theta_i) - \nu_i((1 - p_l)l + p_l l', \theta_i)| < \varepsilon/2. \tag{3}
\]

Let
\[
x_{l, \theta_i, \theta'_i} = (1 - p_l)l + p_l x_{\theta_i, \theta'_i};
\]
\[
x_{l, \theta'_i, \theta_i} = (1 - p_l)l + p_l x_{\theta'_i, \theta_i}.
\]

Consequently, we have
\[
u_i(x_{l, \theta_i, \theta'_i}, \theta_i) > \nu_i(x_{l, \theta'_i, \theta_i}, \theta_i); \tag{4}
\]
\[
u_i(x_{l, \theta_i, \theta'_i}, \theta'_i) < \nu_i(x_{l, \theta'_i, \theta_i}, \theta'_i).
\]
In round $K + 1$, each player $i$ announces his own type, $m_i^{K+1} \in \Theta_i$. Let $m^{K+1} = (m_1^{K+1}, \ldots, m_n^{K+1})$.

In round $K + 2$, in the order of player $n + 1(\equiv 1)$, $n, \ldots, 2$, player $i$ has an opportunity to announce his predecessor’s preference $m_i^{K+2} \in \Theta_{i-1}$ if and only if $m_{j}^{K+2} = m_{j-1}^{K+1}$ for every $j > i$,

- if $m_i^{K+2} \neq m_{i-1}^{K+1}$, then player $i - 1$ chooses $x_{l,m_i^{K+1},m_i^{K+2}}$ or $x_{l,m_i^{K+2},m_i^{K+1}}$ and the game ends;
- if $m_i^{K+2} = m_{i-1}^{K+1}$, then the game continues and player $i - 1$ gets the opportunity to announce his predecessor’s preference $m_{i-1}^{K+2} \in \Theta_{i-2}$.

If $m_i^{K+2} = m_{i-1}^{K+1}$ for all $i$, then the social alternative is determined by the lottery $l$ and the game ends.

The transfers are specified as follows:

\[ t_i = \eta_i + \tau_i + \delta_i. \]

\[ \eta_i = \begin{cases} 
-3\varepsilon, & \text{if } m_i^{K+2} \neq m_{i-1}^{K+1}, \text{ and } i - 1 \text{ chooses } x_{l,m_i^{K+1},m_i^{K+2}}; \\
\varepsilon, & \text{if } m_i^{K+2} \neq m_{i-1}^{K+1}, \text{ and } i - 1 \text{ chooses } x_{l,m_i^{K+2},m_i^{K+1}}; \\
0, & \text{otherwise.}
\end{cases} \]

\[ \tau_i = \begin{cases} 
-2\varepsilon, & \text{if } m_i^{K+2} \neq m_i^{K+1}; \\
0, & \text{otherwise.}
\end{cases} \]

\[ \delta_i = \begin{cases} 
-\varepsilon, & \text{if } i \text{ is the last person who chooses } m_i^K \neq m_i^{K+1} \text{ for some } k \leq K; \\
0, & \text{otherwise.}
\end{cases} \]

Note that first, along any history, a player is fined at most $6\varepsilon$ and rewarded at most $\varepsilon$, which is bounded by $\tilde{l}$ (by (2)). Second, when $m_i^{K+2} \neq m_{i-1}^{K+1}$, $i - 1$ will be fined $2\varepsilon$ regardless of her choice between $x_{l,m_i^{K+1},m_i^{K+2}}$ and $x_{l,m_i^{K+2},m_i^{K+1}}$, whereas whether $i$ will get $\varepsilon$ or $-3\varepsilon$ depends on player $i - 1$’s choice. We draw the game tree for rounds $K + 1$ and $K + 2$ in Figure 1 on p. 10 and highlights the equilibrium path in boldface.

There are two differences between our mechanism and the GP mechanism. First, unlike the GP mechanism, we adopt a modified MR mechanism to elicit player’s true type in round $K + 1$ and round $K + 2$. The modified MR mechanism further differs from the MR mechanism in an essential way: by using randomization, we can make (by (3)) the lottery assigned to each terminal history arbitrarily close to lottery $l$ which is determined by the announcements from
round 1 to round $K$. Consequently, relative to the transfers, the announcement made in either round $K + 1$ or round $K + 2$ has a negligible effect on the lotteries associated to terminal histories. Therefore, we can elicit each player’s true type in round $K + 1$ without large transfers which are required in the MR mechanism. Second, the lottery $l$ is determined in the first $K$ rounds by a single player (i.e., player 1) rather than by the majority rule as in the GP mechanism. The feature enables us to accommodate the case of two players.6

## 4 Implementation

Denote the true type profile by $\psi$.

**Claim 1** In any subgame perfect equilibrium where player $i$ moves in round $K + 2$, player $i$ will announce $m_i^{K+2} = \psi_{i-1}$ if $m_{i-1}^{K+1} = \psi_{i-1}$ and announce $m_i^{K+2} \neq m_i^{K+1}$ otherwise.

**Proof.** First, consider player 2’s choice in round $K + 2$. This is the last move in the game tree. There are two cases:

Case 1. $m_1^{K+1} = \psi_1$: If player 2 announces $m_2^{K+2} = \psi_1$, then $l$ is implemented and $\eta_2 = 0$. If, instead, player 2 announces $m_2^{K+2} \neq \psi_1$, then by (4) player 1 will choose $x_{l,m_1^{K-1},m_2^{K+2}}$, while player 2 will be fined $\eta_2 = -3\varepsilon$. By (3), player 2 will announce $\psi_1$.

Case 2. $m_1^{K+1} \neq \psi_1$: If player 2 announces $m_2^{K+2} = m_1^{K+1}$, then $l$ is implemented and $\eta_2 = 0$. If, instead, player 2 announces $m_2^{K+2} = \psi_1$, then by (4) player 1 will choose $x_{l,m_1^{K+2},m_1^{K-1}}$, while player 2 will be rewarded $\eta_2 = \varepsilon$. By (3), player 2 will announce some $m_2^{K+2} \neq m_1^{K+1}$.

Inductively, each player $i$ with $2 \leq i \leq n$ will confirm his predecessor’s announcement in $K + 1$ (i.e., $m_i^{K+2} = m_i^{K+1}$) if $m_{i-1}^{K+1} = \psi_{i-1}$; while player $i$ will challenge his predecessor’s announcement in $K + 1$ (i.e., $m_i^{K+2} \neq m_i^{K+1}$) if $m_{i-1}^{K+1} \neq \psi_{i-1}$.

Now consider player 1 (i.e., player $n + 1$)’s choice in round $K + 2$. Again, there are two cases:

Case 1. $m_n^{K+1} = \psi_n$: If player 1 announces $m_1^{K+2} = \psi_n$, then $l$ is implemented, $\eta_1 = 0$, and player 1 will be fined $\tau_1 = -2\varepsilon$ in case of being challenged by player 2 later. If, instead, player 1 announces $m_1^{K+2} \neq \psi_n$, then by (4) player $n$ will choose $x_{l,m_n^{K+1},m_1^{K+2}}$, while player 1 will be fined $\eta_1 = -3\varepsilon$. By (3), player 1 will announce $\psi_n$.

Case 2. $m_n^{K+1} \neq \psi_n$: If player 1 announces $m_1^{K+2} = m_n^{K+1}$, then $l$ is implemented and $\eta_1 = 0$. If, instead, player 1 announces $m_1^{K+2} = \psi_n$, then by (4) player $n$ will choose $x_{l,m_1^{K+2},m_n^{K+1}}$, while player 1 will be rewarded $\eta_1 = \varepsilon$. By (3), player 1 will announce some $m_2^{K+2} \neq m_1^{K+1}$.

---

6The idea can also be used to show that the result of Glazer and Perry (1996) holds even when there are only two players.
Claim 2 In any subgame perfect equilibrium, every player truthfully announces his own type in round $K + 1$, i.e., $m_i^{K+1} = \psi_i$ for all $i \in N$.

Proof. Consider first player $n$. Suppose that player $n$ announces $m_n^{K+1} \neq \psi_n$. Since player 1 moves first in round $K + 2$, by Claim 1, this announcement will be challenged by player 1 and result in a penalty $\tau_n = -2\varepsilon$. In contrast with $m_n^{K+1} = \psi_n$, by (3), by announcing $m_n^{K+1} \neq \psi_n$, player $n$’s utility from the induced lottery is affected by an amount less than $\varepsilon$, and player $n$ potentially reduces the penalty $\delta_n = -\varepsilon$. Thus, player $n$ will announce $m_n^{K+1} = \psi_n$. Then, by Claim 1, player $n$ will have an opportunity move in round $K + 2$, and by a similar argument, $m_n^{K+1} = \psi_n$. We can inductively argue that $m_i^{K+1} = \psi_i$ for all $i \in N$.

Claim 3 In any subgame perfect equilibrium, if player $i$ is not the last one to announce a type profile that is different from $m^{K+1}$ along a history up to round $k \leq K$, then $m_i^k = \psi$.

Proof. Note that $m^{K+1} = \psi$ in any subgame perfect equilibrium by Claim 2. Consider player $n$’s decision in round $K$. Suppose that player $n$ is not the last one who lies along a given history. Then, player $n$ will be fined $\delta_n = -\varepsilon$ if he lies by announcing $m_n^K \neq \psi$ and will not be fined if he announces $m_n^K = \psi$. Since player $n$’s report will not affect the allocation, he strictly prefers to tell the truth.

Now suppose that the claim holds for any player $i$ with $1 \leq i \leq n - 1$ in round $K$ and we prove the claim for player $i$. Suppose that player $i' \neq i$ is the last one who lies along a given history. We divide two cases:

Case 1. $i > 1$. In this case, if player $i$ lies, then by the induction hypothesis, all the players will tell the truth in the following histories. Thus, player $i$ will be fined $\delta_i = -\varepsilon$. If player $i$ tells the truth, then by the induction hypothesis, all players $j \neq i'$ will tell the truth in the following histories. Thus, instead of player $i$, player $i'$ will be fined $\delta_{i'} = -\varepsilon$. Since $i > 1$, player $i$’s announcement will not affect the lottery chosen. Thus, $m_i^k = \psi$.

Case 2. $i = 1$. In this case, if player 1 lies, then by the induction hypothesis, all the players will tell the truth in the following histories. Thus, player 1 will be fined $\delta_1 = -\varepsilon$. If he tells the truth, instead of player 1, player $i'$ will be fined $\delta_{i'} = -\varepsilon$. Moreover, the maximal gain from the change in lottery chosen by lying is $\xi/K$. By (2), $\xi/K < \varepsilon$. It follows that truth-telling is strictly better for player 1 in round $K$.

We can conclude the claim by inductively arguing for all $i \in N$ in every round $k < K$ in a similar fashion.

Claim 4 In any subgame perfect equilibrium, $m_i^k = \psi$, for all $i \in N$, for all $1 \leq k \leq K$.

Proof. No player has lied in round $k = 1$. It then follows from Claim 3 that $m_i^1 = \psi$ for all $i$. Inductively, $m_i^k = \psi$ for all $i \in N$ and for all $1 \leq k \leq K$. ■
5 Concluding Remarks

We conclude with two remarks. First, if there are three or more players, our argument is essentially unaltered if the fines (resp. rewards) imposed on some player are to be paid to (resp. paid by) some other player instead of the planner. In other words, with three or more players, we can achieve budget balance (i.e., the transfers add up to zero) on and off the equilibrium path. Second, our result crucially relies on the assumption of complete information and is therefore subject to the criticism by Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012), namely, our mechanism still admits undesirable sequential equilibria when some information perturbation (as defined in Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012)) is introduced to the complete-information environment. An extension of our analysis to an incomplete-information environment is left for future research.

\footnote{When there are only two players, as in Moore and Repullo (1988), there may be an additional surplus generated off the equilibrium path.}
References


Figure 1: The game tree in round $K + 1$ and round $K + 2$ when there are two players, and each player has two types, $\psi_i$ and $\psi'_i$, where $\psi = (\psi_1, \psi_2)$ is the true type profile. For each payoff vector associated with the specific terminal node, the first coordinate is the lottery implemented and the second is the fine or reward imposed on player 1, while the third is the fine or reward imposed on player 2. The equilibrium path is indicated in boldface.