

Implementation with Evidence Disclosure

Elchanan Ben-Porath¹ and Bart Lipman²

November 2007

Preliminary and Incomplete !!

Abstract

The standard theory of implementation and mechanism design is limited in two ways: First it is assumed that the social choice function depends only on the profile of preferences of the agents over the set of social alternatives. Second, it is assumed that agents cannot support their statements with hard evidence. The general goal of this research is to extend the theory by studying a more general model which is not limited in these two ways. More specifically we study the conditions on the proof technology, the preferences of the players, and the social choice function which make the implementation of the social choice function possible.

Our main result establishes that a simple condition on the evidence structure that is necessary for the implementation of a SCF (social choice function) f when the preferences of the agents are state-independent is also a sufficient condition for the implementation of f for every profile of preferences (including state-dependent preferences) in environments where the social planner

¹The department of Economics and the center for rationality, the Hebrew University of Jerusalem. [e-mail: benporat@math.huji.ac.il](mailto:benporat@math.huji.ac.il)

²Department of Economics, Boston University, Boston MA 02215. [e-mail: blipman@bu.edu](mailto:blipman@bu.edu)

can perform monetary transfers among the players. Furthermore, there exists a mechanism that implements f in a way that is robust to the preferences of the players.

1 Introduction

The standard theory of implementation and mechanism design is limited in two ways: First it is assumed that the social choice function depends only on the profile of preferences of the agents over the set of social alternatives. Second, it is assumed that agents cannot support their statements with hard evidence. The general goal of this research is to extend the theory by studying a more general model which is not limited in these two ways.

Let $\mathcal{I} = \{1, \dots, n\}$ be a set of agents, A a set of social alternatives, and S a finite set of states. A state $s \in S$ specifies all the parameters that are relevant for the determination of the alternative $f(s)$ that a social planner (henceforth, SP) would like to implement. The function $f, f : S \rightarrow A$, is called the social choice function. We assume that SP does not know the true state but that every agent $i \in \mathcal{I}$ does know it. (Thus, in this paper we discuss environments with complete information.) Each agent $i \in \mathcal{I}$ is characterized by a preference, $\succsim_{i,s}$, for every $s \in S$ over the set of alternatives and by a proof technology, $M_i = \{M_i(s)\}_{s \in S}$, where $M_i(s)$ is the set of events (partial sets of S) that player i can prove in the state s . To see how this formulation generalizes the standard model we observe that in a standard model S is the set of profile of preferences of the agents over A and $M_i(s)$ is the empty set.

The general goal of our research is to study the conditions on the proof technology, the preferences of the players, and the social choice function which make the implementation of the social choice function possible. We are especially interested in identifying conditions on the proof technology that allow for “robust” implementation. That is, we are looking for conditions which require minimal knowledge on the part of SP on the preferences of the agents and on their proof technologies.

We show that a simple condition that is necessary for the implementation of a SCF f when the preferences are state-independent is also a sufficient condition for the implementation of f for every preference structure in en-

vironments where SP can perform monetary transfers among the players. Furthermore, there exists a mechanism Γ^f that implements f for every preference structure. Thus, Γ^f performs an implementation of f that is robust to the preferences of the players. Theorem 1 establishes the existence of robust implementation with an unbounded mechanism that involves only “small” (epsilon) monetary transfers while theorem 2 establishes robust implementation with a bounded and simple mechanism that involve “large” monetary transfers.

To further demonstrate and motivate the issues and questions that our research addresses we present three examples. The first two examples refer to situations where the preferences of the players are independent of the state (each player prefers more money upon less). In such cases the standard theory of implementation is irrelevant because if players cannot present any evidence then clearly for any mechanism the set of equilibrium outcomes is the same set in every state. The third example motivates our interest in robust implementation, it refers to a situation where the preferences of the players are irrelevant for the determination of the alternative that SP wants to implement. Furthermore, SP does not know the preferences of the players. Clearly in such a situation SP would like to use a mechanism which will implement for every profile of preferences of the players.

We turn now to a description of these examples:

1. Consider a tort trial where the problem of SP is to determine the level of compensation that the defendant should pay the plaintiff. Suppose that SP wants to set the compensation to be the damage that has been caused to the plaintiff but does not know what that damage is. Thus, the set of alternatives is the set of possible compensations and the set of states is the different levels of damages that may have been caused. Clearly, in this case the preferences of the two players over the set of alternatives are independent of the state- the plaintiff always prefers a higher compensation and the defendant a lower one. The set of social choice functions that can

be implemented is determined by the kind of evidence that each player can present.

2. Allocation of a fixed budget among a set of individuals or departments.

Here again it is reasonable to assume that each player wants to get as much as possible independent of the state while SP wants to implement an allocation which depends on factors such as, efficiency in the use of the money, needs, importance of the tasks that the units are required to perform and so forth.

3. Consider a situation where the set of players is a committee in a public office that has to choose a project from a set of alternatives (or, alternatively, has to decide how to allocate money between different projects). Each member in the committee may have a personal preference among the projects (some maybe close to home, some may benefit family members or friends). Clearly, personal benefits that members stand to gain from the projects are irrelevant for the selection rule which SP would like to implement. Thus, from the point of view of SP the members of the committee are a group of experts who may have personal biases and from which information should be extracted. If SP does not know the personal biases of the members then clearly SP would like to use a mechanism which would implement for every profile of preferences of the members.

The literature on communication of verifiable information is by now fairly extensive and given the space limitation our account must be brief and very partial. Many interesting papers are not mentioned here. Starting with Grossmann (1981), Grossmann and Hart (1980), and Milgrom (1981) there has been work on games where players make verifiable statements regarding their private information before taking actions (Fishman and Hagerty, 1990, Forges and Koessler, 2005b, Glazer and Rubinstein, 2006, Lipman and Seppi, 1995, Milgrom Roberts 1986, Okuno-Fujiwara et al, 1990, Seidman and Winter 1997, Shin, 1994.) The papers that are closer to our agenda are those that focus on persuasion games where each player in a set of “senders”

(some of the work considers just one sender) tries to persuade a receiver to take an action which is beneficial to the given sender. In most of these papers the main goal is to identify conditions that would lead to full revelation of the true state. Among these papers the closest to the current research is Lipman and Seppi (1995). They consider a game of persuasion and obtain necessary and sufficient conditions for robust inference of the true state. However, their analysis is limited by several assumptions. First, they assume that all the players have the same proof technology. Second, they assume that players have conflicting preferences. Finally, they restrict attention to mechanisms where each player sends a message only once and players move in a sequence. The results which we present in sections 3 cannot be obtained by such mechanisms.

There has been some work that examines communication of verifiable information in the context of mechanism design (Bull and Watson, 2004, Deneckere and Severinov, 2003, Green and Laffont, 1986) and also more recently communication of verifiable information with a mediator (Forges and Koessler, 2005a). These papers examine the validity of the revelation principle. Thus, this literature identifies conditions under which direct revelation mechanisms would suffice but it does not characterize the set of social choice functions that can be implemented which is the subject of our research.

This draft is organized as follows:

In section 2 we define the model. Theorems 1 and 2 are presented in section 3. Section 4 contains the proofs.

2 The Model

A social environment Ψ is a tuple $\langle \mathcal{I}, A, S, (\succsim_{i,s}, M_i(s))_{i \in \mathcal{I}, s \in S} \rangle$ where

$\mathcal{I} = \{1, \dots, n\}$ – The set of agents.

A – A set of social alternatives.

S – A finite set of states.

$\succsim_{i,s}$ – The preference relation of agent i on A at the state s .

$M_i(s)$ – The set of messages that are “hard evidence”. When agent i presents a message $m_i \in \cup_{s \in S} M_i(s)$ he presents a proof of the event $E(m_i) = \{s \mid m_i \in M_i(s)\}$ (that is, he proves the event that consists of the states in which m_i can be presented). It is convenient to identify the message/hard evidence m_i with the event $E(m_i)$ and thus define $M_i(s)$ as the set of events that i can prove at the state s . This definition implies:

1. For all $E \in M_i(s)$ $s \in E$, (which means that agent i cannot prove something that is false.)
2. $E \in M_i(s)$ implies $E \in M_i(s')$ for every $s' \in E$.

In addition we assume:

3. $S \in M_i(s)$ so it is always possible not to prove anything.
4. $M_i(s)$ is closed under intersection, that is, $E, E' \in M_i(s)$ implies $E \cap E' \in M_i(s)$. This simply means that if player i can prove E and can also prove E' then he can prove E and E' together.³

We say that Ψ is a social environment with state-independent preferences if the preference of each player over A is independent of s . That is, for every $i \in I$, $s, s' \in S$, $\succsim_{i,s} = \succsim_{i,s'}$.

It will be convenient to refer to the set $\{\succsim_{i,s} \mid i \in I, s \in S\}$ as the preference structure.

A social choice function (SCF) for Ψ is a function $f : S \rightarrow A$.

We are considering a situation where a social planner (SP) is interested in implementing a social choice function f . The social planner does not know the true state but every player in I does know it. (Thus, we are interested in implementation in an environment with complete information.)

A mechanism Γ is a tuple $\langle H, Z, P, g \rangle$ ⁴ where H is a set of histories, $Z \subset H$ is the set of terminal histories, $P : H \setminus Z \rightarrow 2^I$ is a function that specifies

³The assumption is not as innocent as it might seem. It implies that there are no effective constraints on the time and effort that are required to establish a claim (event). See Glazer and Rubinstein (2001), Glazer and Rubinstein (2006), and Lipman and Seppi (1995) for models which relax this assumption.

⁴We adopt here the modelling of Osborne and Rubinstein (1994).

the set of players that act at each non-terminal history, and $g, g : Z \rightarrow A$, associates with each terminal history z an alternative $g(z)$ that SP selects at z . We require that for every $h \in H \setminus Z$ and $i \in P(h)$ the set of actions of player i at h is $C_i \times \cup_{s \in S} M_i(s)$ where C_i can be thought of as the set of cheap talk messages for player i . This requirement is a matter of convenience. It makes the formal description simpler because it ensures that the mechanism defines which players will be called to make statements and present evidence after each history h that is possible at some state $s \in S$. Obviously, at the state s player i can only provide evidence for events in the set $M_i(s)$. Thus, we define $\Gamma(s)$ to be the perfect information extensive-form game that is obtained from Γ by restricting the set of actions of each player i at each history $h \in H$, $i \in P(h)$, to be the set $C_i \times M_i(s)$. In $\Gamma(s)$ the preference of player i over the set of terminal histories is the preference that is induced by $\succsim_{i,s}$.

The mechanism Γ is unbounded if the sets $C_i, i \in I$, are infinite. The mechanism Γ is simple and bounded if the sets $C_i, i \in I$, are finite and in each non-terminal history h there is just one player who acts.

We say that a mechanism Γ implements a social choice function f if for every $s \in S$ and every profile of strategies $\sigma^s = (\sigma_1^s, \dots, \sigma_I^s)$ that is a subgame perfect equilibrium (henceforth, s.p.e) in the game $\Gamma(s)$ we have $g(\sigma^s) = f(s)$ (where $g(\sigma^s)$ is the alternative that is selected by g in the terminal history that is induced by σ^s .) Thus, we are focusing on subgame perfect implementation.

3 The Results

We consider social environments where the social planner can perform monetary transfers among the players. We show that a simple condition that is necessary for the implementation of a SCF when the preferences are state-independent is also a sufficient condition for the implementation of f for

every preference structure. Furthermore, there exists a mechanism Γ^f that implements f for every preference structure. Thus, Γ^f performs an implementation of f that is robust to the preferences of the players. Theorem 1 establishes the existence of robust implementation with an unbounded mechanism that involves only “small” (epsilon) monetary transfers while theorem 2 establishes robust implementation with bounded and simple mechanism that involve “large” monetary transfers.

We turn now to a formal description.

Let Ψ be a social environment and let f be a SCF for Ψ . We say that f satisfies condition (N) if:

(N) For every pair of states $s, s' \in S$. If $f(s) \neq f(s')$ then there exists a player $i \in I$ s.t. $M_i(s) \neq M_i(s')$.

Condition (N) says that if the alternatives that are selected by f at the states s and s' are different from each other then there exists some player i who can either refute the state s' at the state s ($M_i(s) \not\subseteq M_i(s')$) or can refute the state s at s' ($M_i(s') \not\subseteq M_i(s)$).

Proposition (0):

Let Ψ be a social environment with state-independent preferences and let f be a SCF for Ψ . If f can be implemented then f satisfies condition (N).

Proof:

Suppose there that there exists a pair of states $s, s' \in S$ such that $f(s) \neq f(s')$ and for every $i \in I$ $M_i(s) = M_i(s')$. Assume by contradiction that there exists a mechanism Γ that implements f and let σ^s be a s.p.e. in $\Gamma(s)$. Since Γ implements f $g(\sigma^s) = f(s)$. Now since $M_i(s) = M_i(s')$ for every $i \in I$ the game $\Gamma(s)$ is identical to the game $\Gamma(s')$. It follows that σ^s is a s.p.e. in $\Gamma(s')$. Thus, there exists a s.p.e. σ^s in $\Gamma(s')$ s.t. $g(\sigma^s) \neq f(s')$. This contradicts the assumption that Γ implements f .

Definition:

Let Ψ be a social environment and let f be a SCF for Ψ . We say that a mechanism Γ performs a robust implementation of f if Γ implements f in every environment Ψ' that may differ from Ψ only in the preference structure.

Definition:

Let $\epsilon > 0$. We say that Ψ is a social environment with monetary transfers (ϵ -monetary transfers) if the set of alternatives A can be written as $A = \widehat{A} \times R^n$ ($A = \widehat{A} \times [-\epsilon, \epsilon]$) for some set \widehat{A} and the preference relations of agent i on A at the different states can be represented by a utility function u_i , $u_i : A \times S \rightarrow R$, that has the following form:

$$u_i((\widehat{a}, t_1, \dots, t_n), s) = v_i(\widehat{a}, s) + t_i$$

for some function v_i , $v_i : \widehat{A} \times S \rightarrow R$.

We have the following interpretation in mind: The social planner is interested in implementing a function $f, f : S \rightarrow \widehat{A}$. To obtain this goal he uses monetary transfers as incentives to induce revelation of the true state. The following terminology will be convenient: A SCF $f, f : S \rightarrow A$, is essential if it does not involve monetary transfers. More precisely, f is essential if for every $s \in S$ there exists an alternative $a = a(s) \in A$ s.t. $f(s) = (a, 0, \dots, 0)$.

Theorem 1:

Let $\epsilon > 0$ and let Ψ be a social environment with ϵ -monetary transfers and where the number of players is at least three. Let f be an essential SCF for Ψ that satisfies the condition (N). Then there exists an unbounded mechanism Γ_f that performs a robust implementation of f .

Definition:

A social environment Ψ satisfies the condition (N*) if:

(N*) For every $s, s' \in S$ there exists a player $i \in I$ s.t. $M_i(s) \neq M_i(s')$.

Corollary:

Let Ψ be a social environment with ϵ -monetary transfers that satisfies the condition (N*) and where the number of players is at least three. Then for every SCF f there exists an unbounded mechanism Γ_f that performs a robust implementation of f .

Theorem 2:

Let Ψ be a social environment with monetary transfers where the number of players is at least two and let f be an essential SCF for Ψ that satisfies the condition (N). Then there exists a simple bounded mechanism Γ_f that performs a robust implementation of f .

Corollary:

Let Ψ be a social environment with monetary transfers that satisfies the condition (N*) and where the number of players is at least two. Then for every SCF f there exists a simple bounded mechanism Γ_f that performs a robust implementation of f .

Remark:

While our general proof uses “large” monetary transfers, there are cases of interest where “small” transfers will suffice. We mention two important examples:

1. Allocation of a set of indivisible objects.
2. Allocation of a fixed amount of money.

In both these cases it possible to implement with ϵ - monetary transfers if each player cares only about the bundle (amount) that he receives and weakly prefers more objects (strictly prefers more money) to less.

In section 3 we explain how a simple modification of the proof of theorem 2 establishes this result.

4 Proofs

Proof of Theorem 1:

The mechanism Γ_f is a stage game where the stage depends on the set of states that have not been refuted yet. Let E denote the set of possible states in the current stage. Say that a state $\hat{s} \in E$ is regular w.r.t E if there exists a player $i \in I$ and a state $s' \in E$ s.t. $M_i(\hat{s}) \not\subseteq M_i(s')$. Thus, in a regular state there is some player i who can provide evidence that the true state belongs to E' where E' is strictly contained in E . With each regular state \hat{s} associate some player $i(\hat{s})$ who can refute some other state in E at \hat{s} . Define E_{IR} to be the set of irregular states in E . Our first observation is that condition (N) implies that if E_{IR} is not empty then there exists some alternative $\tilde{a}(E)$ such that for every $s \in E_{IR}$ $f(s) = \tilde{a}(E)$. (To see that we note that if $s, s' \in E_{IR}$ then $M_i(s) = M_i(s')$ for every $i \in I$.)

Define now the mechanism Γ_f as follows:

First, if $E_{IR} = E$ then the alternative $\tilde{a}(E)$ is selected and the game terminates. So assume, now, that $E_{IR} \subset E$.

Each player chooses a piece of evidence and a message in $E \times A \times Z$ where Z is the set of integers. Let $c_i = (s_i, a_i, z_i)$ denote a message of player i and let $c = (c_1, \dots, c_n)$ denote a profile of messages. First, if some player presents evidence E' (where E' is strictly contained in E) we move to the stage that corresponds to E' . So assume that no player presents such evidence and distinguish now between several cases:

1. There exists a regular state \hat{s} so that for every player $j \neq i(\hat{s})$ $s_j = \hat{s}$. (s_i maybe any element in S .) In this case the alternative $f(\hat{s})$ is selected and the game terminates.

2. For every i $s_i = \tilde{s}$ where \tilde{s} is non-regular. In this case we let the players $i(s), s \in E \setminus E_{IR}$, move in a sequence so that at his turn the player $i(s)$ is given a (second) opportunity to present evidence. If no one presents evidence then the alternative $\tilde{a}(E)$ is selected and the game terminates. Otherwise, let i be

the first player that presents evidence. Then the alternative that is selected is $\tilde{a}(E)$ but in addition player i gets $\$ \epsilon$ and every other player pays $\$ \frac{1}{n-1} \cdot \epsilon$.

3. There is a set of $n - 1$ players, \tilde{I} , such for every $i \in \tilde{I}$, $s_i = \tilde{s}$, where \tilde{s} is non-regular. For the player j that is not in \tilde{I} we have $s_j \neq \tilde{s}$. In this case the alternative that is selected is $\tilde{a}(E)$.

4. For any profile of messages that does not fit the patterns in (1)-(3) the alternative that is selected is determined by the integer game. Specifically, let i be a player who chose the highest integer z_i (if there is a non-singleton set of players who chose the highest integer then take i to be the smallest number in this set.) The alternative that is selected is a_i and in addition player i gets $\$ \epsilon$ and every other player pays $\$ \frac{1}{n-1} \cdot \epsilon$.

We now prove that the mechanism Γ_f implements f .

The proof is by induction on the size of the set of states S . The proof for the case where $|S| = 1$ is immediate. Assume now that $|S| = m$ and that the theorem has been proved for social environments in which the number of states is smaller than m .

First, we verify that in every state s there exists a subgame perfect equilibrium (s.p.e) in $\Gamma_f(s)$ where $f(s)$ is selected. Assume, first, that s is regular w.r.t. S and consider the following profile of actions: Player $i(s)$ presents evidence and every player $j \neq i(s)$ sends a message c_j such $s_j = s$. We claim that given the induction hypothesis and given part (1) in the definition of the mechanism this behavior is part of a subgame-perfect equilibrium that leads to the selection of $f(s)$. To see that we note that the induction hypothesis implies that the presentation of evidence by $i(\hat{s})$ will lead to the revelation of $f(\hat{s})$ in a s.p.e of the continuation game. (The induction hypothesis is applied by observing that the continuation of the game after the presentation of evidence S' is equivalent to a game that starts with S' as the set of possible states.) It follows that every player $j, j \neq i(\hat{s})$, cannot benefit from a deviation. Part (1) in the definition of Γ_f implies that the player $i(\hat{s})$ cannot

benefit from a deviation.

Assume now that $s = \tilde{s}$ where \tilde{s} is a non-regular state w.r.t. S . Consider a profile of actions where each player i sends a message c_i such that $s_i = \tilde{s}$. Since no player can present any evidence at \tilde{s} it is easy to see that parts (2) and (3) in the definition of the mechanism imply that this behavior is part of an equilibrium that leads to the selection of $f(\tilde{s}) = \tilde{a}(S)$.

We now show that for every state s it is the case that every s.p.e in $\Gamma_f(s)$ leads to the selection of $f(s)$.

First, we note that if any evidence $S' \subset S$ is presented then the claim follows from the induction hypothesis. So assume that no evidence is presented. Now, we observe that except for the case where the profile of messages is such that for every i $s_i = \tilde{s}$ (where \tilde{s} is a non-regular state) we have that at every other profile of messages c , either (1) An integer game is played at c , or (2) There is at least one player j who could by changing his message induce a profile of messages $c' = (c'_j, c_{-j})$ in which the integer game is played. Since by playing the highest integer a player i can get an alternative that is strictly better than any other outcome he can get in a s.p.e of the continuation of the game (player i can get his best alternative in \hat{A} plus a monetary payment of $\$ \epsilon$) it follows that c cannot be part of a s.p.e. Thus, the only case that is left for us to consider is the possibility that the true state is some regular state \hat{s} and yet every player i announces $s_i = \tilde{s}$ where \tilde{s} is a non-regular state. However, in this case each player $i(s)$, $s \in S \setminus S_{IR}$, is offered a second opportunity to submit evidence and the first player who can submit evidence, call him \tilde{i} , (there is at least one such player, i.e., $i(\hat{s})$) will do that because he would get the alternative $\tilde{a}(S)$ plus $\$ \epsilon$ instead of $\tilde{a}(S)$ (if no one presents evidence) or $\tilde{a}(S)$ minus $\$ \frac{\epsilon}{n-1}$ (if some other player presents evidence.) It follows that a profile of messages where each player j announces \tilde{s} cannot be part of an equilibrium in \hat{s} because every player j who is different than the player \tilde{i} has to pay player \tilde{i} $\$ \frac{\epsilon}{n-1}$. Each such player j would be better off deviating to some message c'_j where $s_j \neq \tilde{s}$ and in this way (by part (3)

in the definition of the mechanism) ensuring the selection of the alternative $\tilde{a}(S)$ without paying anything.

Proof of Theorem 2

Define $V \equiv \max_{i \in I} \max_{s \in S} \max_{a, a' \in \hat{A}} v_i(a, s) - v_i(a', s)$ and let F_1 and F_2 be two numbers s.t. $F_2 > F_1 > V$. Thus, for any player i a monetary fine or reward of $\$F_1$ or $\$F_2$ outweighs any change in the alternative that is selected from \hat{A} .

We turn now to the definition of the mechanism Γ_f . Two preceding comments are required. First, when there are two players there is an out-of-equilibrium history where both players are punished and in such a case the budget is not balanced. When the number of players is greater or equal than three budget balance can be maintained by simply transferring the fines of the players that are punished to other players, say, by dividing them equally between these players. Thus, when we say that at a given history of the game each player in a set \mathcal{I}' of players pays some fine it is understood that the fines that are collected are distributed equally between the players in $\mathcal{I} \setminus \mathcal{I}'$. Second, when SP punishes one, or some, of the players (which is an out-of-equilibrium event) the element in \hat{A} that is selected by SP is immaterial as long as it is the same element for each history that involves punishments. Thus, we will omit reference to this selection from the description of the mechanism.

The mechanism Γ_f consists of two phases, the *claiming* phase and the *refutation* phase. Each phase consists of n stages, a stage for each player. In both phases the players move in a sequence starting with player 1.

In the *claiming* phase player 1 starts by claiming a state s_1 and presenting evidence M_1 . If $M_1 \neq M_1(s_1)$ player 1 pays the fine F_1 and the game terminates. If $M_1 = M_1(s_1)$ we proceed to the next stage, player 2 claims a state s_2 and presents evidence M_2 . If $s_2 \neq s_1$ or if $M_2 \neq M_2(s_1)$ both players

1 and 2 pay a fine of F_1 and the game terminates. Otherwise, we proceed to the next stage. If the game reaches stage i and player i makes the statement s_i and provides the evidence M_i , then if $s_i \neq s_1$ or $M_i \neq M_i(s_i)$ then both player 1 and player i pay a fine of F_1 and the game terminates. Otherwise, if $i < n$, we proceed to the next stage, if $i = n$ the *claiming* phase is completed and the game moves to the *refutation* phase. Here again the players move in a sequence starting with player 1. If the game reaches stage i then player i is invited to refute s_1 , that is, to present evidence that will prove that the true state is different than s_1 . (If the true state is s then player i can refute s_1 iff $M_i(s) \not\subseteq M_i(s_1)$). If player i refutes s_1 then all the players different from i pay a fine of F_2 , player i gets all these fines as a reward, and the game terminates. If player i does not refute s_1 the game moves to the next stage. If no player refutes then there are no fines (and no rewards), s_1 is selected and the game terminates.

We now claim that if σ^s is a s.p.e in $\Gamma_f(s)$ then $g(\sigma^s) = f(s)$.

The proof consists of two steps. First we show that if $\sigma_{s, M_1(s)}^s$ is a s.p.e in the subgame that follows the truthful statement, $(s, M_1(s))$, by player 1 then $g(\sigma^s) = f(s)$. Thus, by stating the truth player 1 can obtain the alternative $f(s)$. Then we show that if s_1 is the statement of player 1 in the first stage then $f(s_1) = f(s)$ and $f(s)$ is selected. Start with the first step. If player 1 claims the true state s then (obviously) no player can refute s because s is the truth. It follows that in the *claiming* phase it is better for player $i \neq 1$ to say the truth $(s, M_i(s))$ rather than contradict player 1, or not report all the evidence in $M_i(s)$, because contradicting player 1 or failing to report all the evidence will result in a punishment while saying the truth avoids it. Thus, we obtain that in a s.p.e of the subgame that follows a truthful statement by player 1 there are no punishments and rewards and at the end of the game SP selects $f(s)$. We turn now to the second step of the proof. Assume that there exists a s.p.e, $\bar{\sigma}^s$, where at the first stage player 1 makes a statement $(s_1, M_1) \neq (s, M_i(s))$. We will show that $f(s) = f(s_1)$ and $f(s)$

is selected. First, we observe that $M_1 = M_1(s_1)$ because otherwise the game terminates with a punishment for player 1 while as we have just seen saying the truth avoids it. Since s is the true state it follows that $M_1(s_1) \subseteq M_1(s)$. Now since $\bar{\sigma}^s$ is a s.p.e no player i contradicts player 1 in the *claiming* phase (otherwise, player 1 is punished and would have been better of saying the truth at the beginning). Since the truth is s it follows that $M_i(s_1) \subseteq M_i(s)$ for every $i \in \mathcal{I}$. (Otherwise, player i could not have presented all the evidence in $M_i(s_1)$.) If $\bar{\sigma}^s$ is an equilibrium then it must be the case that there is no player who refutes s_1 at the refutation phase. (If some player j refutes s_1 then each one of the other players would have been better of stating the truth at the claiming phase.) However, the condition stated in the theorem implies that if $f(s) \neq f(s_1)$ then there exists at least one player j for which $M_j(s_1) \subset M_j(s)$. In this case there would be some player j who would refute s_1 (i.e., presents a message $m_j \notin M_j(s_1)$) and get the big prize of $\$F_2$. It follows that $f(s_1) = f(s)$ and the alternative $f(s)$ is selected.

Remark:

While our general proof uses large fines, there are many cases of interest where small fines will suffice. For example, suppose that we must allocate a set of indivisible objects among the agents in a manner which depends on s . Assume each agent cares only about the objects he receives and that more objects are at least weakly preferred to fewer. Then we can make the fines F_1 and F_2 arbitrarily small as long as we maintain the order above and choose the alternative \hat{a} so that the agent or agents being fined receive no objects and an agent who receives a reward at the refutation gets all the objects. It is easy to see that the proof above works with obvious minor changes. Similarly, suppose that we have a fixed amount of money to be divided among the agents in a manner which depends on s , each agent cares only about how much money he receives, and more money is better than less. Suppose each agent's share according to f is bounded away from zero.

Then we can replace the fines with choices of outcomes where the agent being punished receives an amount of money strictly below the minimum share he should get according to f . As before, if we choose the right ordering of these punishments, the proof above works with minor modifications. In this case, we do not require the agents to have money which can be taken away from them — we simply give them less money in some situations.

References

1. Bull, J. and Watson, J., 2004, Hard Evidence and Mechanism Design, *Journal of Economic Theory*, 118, 1-31.
2. Deneckere, R. and Severinov, S., 2003, Mechanism Design and Communication Costs, mimeo.
3. Fishman, M. and Hagerty, C., 1990, The Optimal Amount of Discretion to Allow in Disclosures, *Quarterly Journal of Economics*, 105, 427-444.
4. Forges, F. and Koessler, F., 2005a, Communication Equilibria with Partially Verifiable Types, *Journal of Mathematical Economics*, 42, 793-811.
5. Forges, F. and Koessler, F., 2005b, Long Persuasion in Sender Receiver Games, mimeo.
6. Glazer, K. and Rubinstein, A., 2001, Debates and Decisions, On a Rationale of Argumentation Rules, *Games and Economic Behavior*, 36, 158-173.
8. Glazer, K. and Rubinstein, A., 2004, On Optimal rules of Persuasion, *Econometrica*, 72, 1715-1736.
9. Glazer, K. and Rubinstein, A., 2006, On The Pragmatics of Persuasion: A Game Theoretic Approach, *Theoretical Economics*, forthcoming.
10. Grossman, S., 1981, The Informational Role of Warranties and Private Disclosure about Product Quality, *Journal of Law and Economics*, 24, 461-483.
11. Grossman, S. and Hart, O., Disclosure Laws and Takeover Bids, *Journal of Finance*, 35.
12. Lipman, B. and Seppi, D., 1995, Robust Inference in Communication Games with Partial Provability, *Journal of Economic Theory* 66, 370-405.
13. Milgrom, P., 1981, Good News and Bad News, *Bell Journal of Economics*, 12, 380-391.
14. Milgrom P., and Roberts J., 1986, Relying on Information of Interested Parties, *Rand Journal of Economics*, 17, 18-32.

15. Okuno-Fujiwara, M., Postlewaite, A. and Sujumura, K., 1990, Strategic Information Revelation, *Review of Economic Studies*, 57, 25-47.
16. Osborne, M., Rubinstein, A., 1994, *A Course in Game Theory*, The MIT Press.
17. Seidman, D. and Winter, E., 1997, Strategic Information Transmission with Verifiable Messages, *Econometrica*, 65, 163-169.
18. Shin, H.S., 1994, The Burden of Proof in a Game of Persuasion, *Journal of Economic Theory*, 64, 253-264.