Implementation with Partial Provability[1]

Elchanan Ben-Porath[2] Barton L. Lipman[3]

First Draft
September 2008

Current Draft
October 2011

[1]We thank Eddie Dekel, Jacob Glazer, Sean Horan, Navin Kartik, Phil Reny, Olivier Tercieux, various seminar audiences, and an anonymous associate editor and referee for helpful comments and the US–Israel Binational Science Foundation for supporting this research. Lipman also thanks the National Science Foundation for support.

[2]Department of Economics and Center for Rationality, Hebrew University. Email: benporat@math.huji.ac.il.

[3]Department of Economics, Boston University. Email: blipman@bu.edu.

Abstract

We extend implementation theory by allowing the social choice function to depend on more than just the profile of preferences of the agents and by allowing agents to support their statements with hard evidence. We show that a simple condition on the evidence structure which is necessary for the implementation of a social choice function f when the preferences of the agents are state independent is also sufficient for implementation for any preferences (including state dependent) if the social planner can perform small monetary transfers beyond those called for by f and there are at least three players. If transfers can be large, f can be implemented in a game with perfect information when there are at least two players under an additional boundedness assumption. In both cases, transfers only occur off the equilibrium path. In the special but important case of allocation problems, under weak conditions, f can be implemented in a perfect information game with at least two players and *no* transfers. In all cases, the use of evidence enables implementation which is robust in the sense that the social planner needs very little information about the preferences, beliefs, and evidence of the agents and the agents need little information about each others' preferences. Furthermore, our results still hold if evidence can be forged at an arbitrarily small but strictly positive cost. Finally, we relate our results to the classical work of Maskin (1977) and Moore and Repullo (1988) on implementation without evidence.

1 Introduction

This paper addresses a pair of related limitations present in the theory of implementation since the seminal work of Hurwicz (1972) and Maskin (1977, 1999). By implementation, we refer to what is sometimes called full implementation — that is, the requirement that *every* outcome selected by the solution concept under consideration in the game induced by the mechanism leads to the social alternative specified by the social choice function. The first limitation is the assumption that the social alternative to be implemented depends *only* on the profile of preferences of the agents.[1] As we argue below, this excludes many important situations of economic interest. The second limitation is the assumption that the agents cannot provide hard evidence. Obviously, hard evidence can establish the veracity of statements which would not be incentive compatible. Furthermore, use of hard evidence is very common in real world institutions, so the omission of this possibility from the theory of implementation is clearly an important gap to fill.[2] As will become clear shortly these two issues are related in that it is impossible to address the first limitation without introducing evidence. In addition, we will show that implementation with evidence has robustness properties which are not possible in standard models.

The general goal of our research is to extend the standard theory by considering a more general model that is not restricted by these two assumptions. In this paper we study complete information implementation, primarily focusing on subgame perfect equilibrium.

Let $\mathcal{I} = \{1, \ldots, I\}$ be a set of agents, A a set of social alternatives, and S a 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. As we explain shortly, a state also determines all relevant characteristics of the agents. The function $f : S \to A$ is called the social choice function (SCF). We assume that SP does not know the true state but that every agent $i \in \mathcal{I}$ does know it. (This is the assumption of complete information.) As in the usual model, each agent $i \in \mathcal{I}$ in each state $s \in S$ has a preference $\succeq_{i,s}$ over the set of alternatives. (This is phrased differently in the usual model, but this difference is one of notation, not substance.) The more significant change is that we also assume each agent i has a proof technology, $M_i = \{M_i(s)\}_{s \in S}$, where $M_i(s)$ is the set of events (subsets of S) that player i can prove in state s. So $E \in M_i(s)$ if in state s, player i has some evidence he can present which proves that the true state lies in the event E. Thus a state also specifies the hard evidence available to the agents.

[1] In a sense, this is a result, not an assumption. As we explain below, without evidence, it is impossible to implement an outcome that depends on more than the preferences of the agents.

[2] It would be fair to say that the extensive literature on mechanism design (where we only require *an* equilibrium with the outcome specified by the social choice function) has also, by and large, the same limitations. We cite some notable exceptions in the sequel.

To see that the standard model is a special case of this model, note that the assumption that players cannot prove anything corresponds to $M_i(s) = S$ for every $i \in \mathcal{I}$ and $s \in S$. Similarly, the assumption that the social choice depends only on the preference profile translates to the requirement that if $\succ_{i,s}=\succ_{i,s'}$ for every $i \in \mathcal{I}$, then $f(s) = f(s')$. In our model, this condition need not hold, in which case none of the standard implementation results (*e.g.*, Maskin (1977), Moore and Repullo (1988), Abreu and Sen (1990), Palfrey and Srivastava (1991), or Abreu and Matsushima (1992)) apply. To see this, simply note that if the preferences of the agents do not vary across states and if agents cannot present evidence (as the usual model assumes), then the game induced by any mechanism in the usual model will not vary across states, so the set of equilibria cannot vary across states. Hence the social choice cannot depend on more than the preference profile if there is no evidence available.

Our goal is to study the conditions on the proof technology, the preferences of the players, and the social choice function which make implementation of the social choice function possible. We identify a simple condition on the relationship between the proof technologies of the agents and the social choice function, f, which we call *measurability*, which is necessary for implementation of f when the preferences are state independent.

We then show that measurability is sufficient for implementation of f for *every* preference structure if SP can perform monetary transfers among the players beyond those called for by f. More specifically, our first sufficiency result, Theorem 1, shows that even if we require use of the very natural equilibrium notion of backward induction in a perfect information game, measurability implies implementation is possible if there are at least two players under a boundedness assumption. Our second sufficiency result, Theorem 2, shows that even if we restrict monetary transfers to be arbitrarily small, measurability implies implementation is possible if there are at least three players in a mechanism which involves an integer game. This mechanism has only one stage and hence it implements in Nash equilibrium as well. For both theorems, transfers only occur off the equilibrium path. We also show in Theorem 3 that for the problem of allocating a set of goods among a set of agents, we can achieve implementation in a perfect information mechanism without monetary transfers under weak conditions.

We emphasize that the measurability condition puts no restrictions on how the preferences of the agents relate to the evidence available and the social alternative to be implemented. Thus, if measurability is satisfied, SP can implement without knowing anything about the preference structure and its relation to the proof technologies and social choice function. We also show that the agents do not need more than minimal information about the preferences of other agents to play equilibrium strategies. As a result, the planner does not need to know whether the agents know each other's preferences or what they believe about the preferences of others. A different and surprising form of robustness is that our results also hold if evidence can be forged at an arbitrary

strictly positive cost.

Finally, we relate our results to the classical work of Maskin (1977) and Moore and Repullo (1988).[3] In particular, we explain how our condition of measurability relates to Maskin's monotonicity and Moore and Repullo's preference reversals in a modified model with an extended outcome space which includes evidence.[4]

To further motivate the issues, we present three examples. In the first two examples, the preferences of the players are independent of the state. In such cases, the standard theory of implementation is irrelevant: if players cannot present any evidence, then for any mechanism the set of equilibrium outcomes is independent of the state. The third example illustrates a situation where SP has limited information about 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.

1. Consider a personal injury trial where the problem of SP is to determine the level of compensation that the defendant should pay the plaintiff. Suppose that SP (the judge) wants to set the compensation equal to 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 state specifies the level of damages. Clearly, 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.

2. Consider the problem of an organization allocating 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 regardless of the state. On the other hand, SP wants to implement an allocation which depends on which department can most efficiently use the organization's resources to further its goals, an objective which depends on the state of the world.

3. Consider a situation where the set of players is a public commission that has to decide how to allocate money between different projects. Each member of the commission may have a personal preference among the projects. For instance, some may be close to his home or 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 experts who may have personal biases and from whom information should be extracted. Theorems 4 and 5 show that if the social choice function satisfies measurability, then SP can implement when he has no information about the preferences of the players and even when they have no information about each others' preferences.

[3] See also Abreu and Sen (1990).

[4] We thank an anonymous referee for drawing our attention to this relationship.

The literature on communication with verifiable information is quite extensive, so our account must be brief and very partial. Starting with Grossman (1981), Grossman and Hart (1980), and Milgrom (1981), there is a growing literature which examines persuasion games. In these games, a set of agents (often only one agent) make verifiable statements to a "receiver" who takes an action. See, for example, Fishman and Hagerty (1990), Glazer and Rubinstein (2001, 2004, 2006), Lipman and Seppi (1995), Milgrom and Roberts (1986), Seidman and Winter (1997), Sher (2008), and Shin (1994). See also Forges and Koessler (forthcoming) and Okuno-Fujiwara, Postlewaite, and Suzumura (1990) for other interesting models of games where players make verifiable statements.

There has been some work that examines communication of verifiable information in the context of mechanism design (Alger and Ma (2003), Bull and Watson (2004, 2007), Deneckere and Severinov (2008), Green and Laffont (1986)) and also more recently communication of verifiable information with a mediator (Forges and Koessler (2005)). These papers analyze environments with incomplete information and primarily focus on the question of identifying an appropriate form of the Revelation Principle — that is, when some form of a direct revelation mechanism suffices for achieving an equilibrium with a particular outcome. By contrast, we focus on (full) implementation, where we seek games for which *every* equilibrium has the desired outcome.

There are a few papers that discuss implementation with evidence in specific contexts. Hurwicz, Maskin, and Postlewaite (1995) and Postlewaite and Wettstein (1989) consider the implementation of the Walrasian social choice function in a model where a player may misrepresent his endowment by destroying or hiding some of it. In this setting, showing (part of) one's endowment proves that the player has at least that much. Bull and Watson (2004) study Nash implementation of a "zero–sum" social choice function that specifies monetary transfers between a set of agents.

Perhaps the closest predecessor 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. While their results are phrased in terms of equilibrium outcomes of a fixed game, many of the results can be directly translated into statements of implementation of a social choice function. 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, for the most part, 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 Section 3 cannot be obtained by such mechanisms.

We discuss another closely related paper, Kartik and Tercieux (2011), in Section 3 after a presentation of our results.

In Section 2, we define the model. We turn to the results in Section 3. After showing

that measurability is necessary for implementation with state independent preferences, we turn to environments with monetary transfers. In Section 3.1, we show that measurability is sufficient for implementation in a perfect information game in bounded environments. In Section 3.2, we show that measurability is sufficient for implementation in a one–stage simultaneous move game with ε transfers. In Section 4, we discuss the sense in which the implementation is robust with respect to the planner's information about the agents and to the possibility of forging evidence. In Section 5.1, we explain how our result relates to standard implementation results on an enlarged outcome space. While this connection provides one means of proving our main theorems, our proofs are more direct as we explain. In Section 5.2, we provide a series of examples demonstrating that our results are tight. Section 6 concludes. Proofs are contained in the Appendix.

2 Model

A social environment Ψ is a tuple $\langle \mathcal{I}, A, S, (\succeq_{i,s}, M_i(s))_{i\in I, s\in S}\rangle$. $\mathcal{I} = \{1, .., I\}$ is the set of agents and A is a set of social alternatives. S is the set of states of the world. $\succeq_{i,s}$ is the preference relation of agent i over A at state s.

The more novel part of the model is $M_i(s)$. This is the set of "hard evidence" messages i has available in state s. When agent i presents a message $m_i \in \cup_{s\in S} M_i(s)$, he presents proof of the event $E(m_i) = \{s \,|m_i \in M_i(s)\}$. 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 or hard evidence m_i with the event $E(m_i)$ and thus we define $M_i(s)$ as the set of events that i can prove at state s.

This definition implies two important properties of $M_i(s)$. First, for all $E \in M_i(s)$, we must have $s \in E$. That is, agent i cannot prove something that is false. Second, $E \in M_i(s)$ implies $E \in M_i(s')$ for every $s' \in E$. In words, if the agent can prove event E in some state, he must be able to prove it in every state in E. In subsequent discussions, we refer to this property as *consistency*. To understand this property intuitively, note that it simply embodies the idea that if a particular piece of evidence proves the event E, no more and no less, then this piece of evidence must exist exactly when E is true.

For one of our results, we also assume that there are no restrictions on the amount of his evidence any agent can present. Formally, we say that the evidence structure is *normal* if for every i and every s,

$$\bigcap_{E\in M_i(s)} E \in M_i(s).$$

That is, for a normal evidence structure, it is always feasible for i to prove the event that corresponds to presenting all the evidence in his possession. Thus normality implies that

there are no effective constraints on the time and effort that are required to establish a claim.[5]

It would be very natural to also assume that $S \in M_i(s)$ for all s. That is, any player i can always present a trivial message that proves nothing. However, such an assumption is not needed for our results.

Let $M_i = \cup_{s \in S} M_i(s)$. We say that a message $m_i \in M_i$ *refutes* state $s' \in S$ if $m_i \notin M_i(s')$. That is, m_i could not have been sent at state s'. Equivalently, m_i refutes s' if $s' \notin m_i$. Note that player i can refute state s' at state s iff $M_i(s) \not\subseteq M_i(s')$.

For a simple example of what this framework allows, suppose there is one agent and two states of the world, p (where the agent can play the piano) and np (where he cannot). Then a natural specification of the evidence available to player 1 is $M_1(p) = \{S, \{p\}\}$ and $M_1(np) = \{S\}$. To see this, note that if player 1 cannot play the piano, all he can do is make random noises with it. However, this does not prove he cannot play the piano, since he could make the same noises if he does know how to play. Hence in state np, the only event player 1 can prove is the trivial event S. On the other hand, in state p, the agent could provide the trivial proof of S by making random noises or could play a piece which demonstrates his ability, hence proving $\{p\}$.[6] The example represents a general phenomenon — in many situations, proving a negative proposition ("agent i cannot do x" or "agent i does not have y") is difficult or impossible.[7]

We refer to the set $\{\succeq_{i,s} | i \in I, s \in S\}$ as the *preference structure*. We say that a preference structure has *state independent preferences* if the preference of each player over A is independent of s. That is, for every $i \in \mathcal{I}$ and $s, s' \in S$, we have $\succeq_{i,s} = \succeq_{i,s'}$.

A *social choice function* (SCF) for Ψ is a function $f : S \to A$. The social choice function represents the outcome a social planner (SP) wishes to achieve as a function of the state. The social planner does not know the true state but every player in $\mathcal{I}$ does know it, as in the standard complete information implementation problem.

[5]Lipman and Seppi (1995) first identified this condition, calling it the full reports condition. Most papers refer to this assumption as normality following Bull and Watson (2007). While most of the literature assumes normality, see Lipman and Seppi (1995), Glazer and Rubinstein (2001, 2004, 2006), Bull and Watson (2007), and Kartik and Tercieux (2010) for models which relax the assumption. We thank Kartik and Tercieux for pointing out that a previous draft of this paper stated an unnecessarily strong version of this property.

[6]The piano player example appears in Lipman and Seppi (1995) who attribute it to Mike Peters.

[7]For example, Hurwicz, Maskin, and Postlewaite (1995) and Postlewaite and Wettstein (1989) assume that a player who has x units of a good can prove that he has *at least* y units for every $y \leq x$ by simply showing y units. However, the player cannot prove that he has *exactly* x units because he might be hiding some units. See Okuno-Fujiwara, Postlewaite, and Suzumura (1990) for a similar model. See also Shin (1994) for a model where an agent can prove that he has evidence by providing it, but can never prove that he does not have evidence.

We think of a state $s \in S$ as a specification of all facts about the world which are relevant for the implementation problem. In particular, a state specifies every fact that is relevant for the determination of the alternative that SP wishes to implement, whether these facts affect the preferences of the agents or not. As in the examples in the introduction, SP's preferred outcome may vary across states even when the preferences of the agents do not. In addition, as is clear from the definition, a state specifies the preference of each player over the social alternatives and the evidence that he can present.

There is one conceptual question we must address before defining mechanisms in this environment, namely, whether the mechanism can require or forbid an agent to present certain evidence. In the usual definitions, SP is completely free to determine the set of messages an agent may use in any given situation. However, when the messages include hard evidence, can the mechanism require use of some messages (when the agent has them available)? Alternatively, can it forbid use of some messages?

Turning to the possibility of compulsion first, one can imagine situations where the social planner knows that the evidence an agent has is kept in a specific location (say, documents in a bank safe) and has legal authority to seize this evidence. In a scenario like this, the social planner can force the agent to reveal the evidence he has. Clearly, in such a case, there are no interesting incentive questions surrounding such evidence — the social planner can always learn it whether the agent has an incentive to reveal it or not. In reality, there are many situations where the planner does not know whether the agent has evidence and, if so, where it could be found. In such a case, even the most draconian legal requirements do not *force* presentation of evidence, but merely provide very strong incentives for agents to *choose* to provide it. To see the point concretely, consider the piano player example above. Can the mechanism say that in state p, player 1 has no option other than to present the evidence $\{p\}$? We assume that this is not possible. The mechanism can give incentives for the agent to provide this evidence by, for example, punishing the agent severely if he does not provide it. However, this is quite different from constructing a mechanism in which in state p, it is not feasible for player 1 to avoid providing this proof. For one thing, punishing the agent for not providing evidence implies that he is punished in state np as well.

As for forbidding some messages, again, in principle, this can be valuable for SP.[8] However, it is difficult to draw a line between requiring presentation of evidence (which we prohibit) and forbidding the use of relatively uninformative evidence. In any case, given our other assumptions, our results do not depend on whether we allow SP to forbid some messages or not.[9] Hence for simplicity, we assume he cannot.

[8] We thank Kartik and Tercieux for an illuminating exchange on this subject.

[9] Our necessary condition is necessary whether or not SP is allowed to forbid some messages and our sufficiency results do not use such prohibitions.

In short, we assume that whenever an agent has an opportunity to present evidence in a mechanism, the mechanism must allow him to present whatever of his available evidence as he chooses. Of course, the way outcomes depend on his choices is unconstrained.

More specifically, a mechanism is a game form with finitely many stages which specifies at each history h of the game a subset of players who act simultaneously at h after observing h.[10] That is, the players who act at the history h are fully informed about all players' choices in prior stages. An action for a player i is a pair that consists of a cheap talk message c_i (that is, c_i can be played in every state s) and a hard evidence message m_i. The assumption that player i can present any evidence he has means that at state s, player i can play any message m_i in the set $M_i(s)$. In addition, the mechanism defines an outcome function which specifies the social alternative $a \in A$ that SP chooses as a function of the complete path of play as summarized by the terminal history.

For notational simplicity, our definition of a mechanism imposes more restrictions than this. In particular, we consider only multistage games where the set of "nonevidence" messages available to a player is constant across his information sets. However, we emphasize that this is only to avoid excessive notation. As will be obvious from the proof, our necessity result would hold for any class of games and any equilibrium notion. Since we can prove sufficiency by constructing any game, the restriction to multistage games is without loss of generality for sufficiency results.

Formally,[11] a *mechanism* consists of the following objects. For each player i, there is a set C_i, interpreted as the "cheap talk" messages available for i. We also have a set of histories, H, where a history $h \in H$ is a finite sequence $(h_1, \ldots, h_k)$ for some k. We assume that the empty sequence, denoted $\emptyset$, is an element of H and that if $(h_1, \ldots, h_k, h_{k+1}) \in H$, then $(h_1, \ldots, h_k) \in H$. A history $h = (h_1, \ldots, h_k) \in H$ is *terminal* if there is no h_{k+1} such that $(h_1, \ldots, h_k, h_{k+1}) \in H$. Let H_T denote the set of terminal histories. We also have a function $P : H \setminus H_T \to 2^{\mathcal{I}} \setminus \{\emptyset\}$ where $P(h)$ is the set of players who choose actions at history h where the interpretation is that these players move simultaneously after having observed history h. We require that if $(h_1, \ldots, h_k, h_{k+1}) \in H$, then $h_{k+1} \in \prod_{i \in P(h_1,\ldots,h_k)} C_i \times M_i$. (Recall that $M_i = \cup_{s \in S} M_i(s)$). That is, each component of a history is tuple of cheap talk messages and evidence statements, one for each player in the subset of players who move at that stage and these choices are made simultaneously. In line with this, we write $h_{k+1} = \{(c_i, m_i)\}_{i \in P(h_1,\ldots,h_k)}$

Since, at state s, player i can only provide evidence in the set $M_i(s)$, the set of actions available to each player i at each history $h \in H$, $i \in P(h)$, is restricted to the set $C_i \times M_i(s)$. Say that a history $(h_1, \ldots, h_k)$ is *feasible* in state s if for $n = 1, \ldots, k$, letting $h_n =$

[10] It is easy to show that our necessity result does not rely on the finite bound on the number of stages of the mechanism and, as we show, sufficiency can be proved with such mechanisms. Thus our results would not change if we allowed mechanisms with an unbounded number of stages.

[11] Our definition is a slight variation on that of Osborne and Rubinstein (1994), Section 6.3.2.

$\{(c_{in}, m_{in})\}_{i \in P(h_1,\ldots,h_{n-1})}$, we have $m_{in} \in M_i(s)$ for all $i \in P(h_1, \ldots, h_{n-1})$. Formally, our assumption that players cannot be compelled or forbidden to present evidence takes the following form: If $(h_1, \ldots, h_k, h_{k+1}) \in H$ is feasible in state s and $h'_{k+1} \in \prod_{i \in P(h_1,\ldots,h_k)} C_i \times M_i(s)$, then $(h_1, \ldots, h_k, h'_{k+1}) \in H$.

Finally, we have a function $g : H_T \to A$ specifying the outcome selected by the mechanism for each terminal node. A mechanism Γ is the tuple $\langle H, P, g, (C_i)_{i \in \mathcal{I}} \rangle$.

We say that the mechanism Γ *uses limited evidence* if for every history $(h_1, \ldots, h_k) \in H$, for every player i, there is at most one $\ell < k$ such that $i \in P(h_1, \ldots, h_\ell)$. That is, the mechanism uses limited evidence if no player can present evidence more than once. To understand the reason this property may be of interest, recall that an evidence structure is normal if it is possible to present all of one's evidence in a single message. If we consider nonnormal evidence structures but allow an agent enough opportunities to present evidence, we effectively restore normality. Thus in considering results that do not rely on normality, we focus on mechanisms which use limited evidence.

The mechanism Γ is a *one–stage mechanism* if every nonempty history has length one. The mechanism has *perfect information* if for every nonterminal $h \in H$, $P(h)$ is a singleton.

For every $s \in S$, a mechanism Γ induces a game, denoted by $\Gamma(s)$, where the set of histories is the set of histories that are feasible in state s and the preference of player i over the set of terminal histories is the preference that is induced by $\succeq_{i,s}$.

We say that a mechanism Γ *implements* a social choice function f if for every $s \in S$ and every profile of pure strategies $\sigma^s = (\sigma_1^s, ..., \sigma_I^s)$ that is a subgame perfect equilibrium in the game $\Gamma(s)$, we have $g(\sigma^s) = f(s)$ (where $g(\sigma^s)$ is the alternative that is selected by g at the terminal history that is reached under σ^s). Obviously, if Γ is a one–stage mechanism, then the set of subgame perfect equilibria of the induced game is the same as the set of Nash equilibria.

The definitions for our robustness results are given in Section 4.

Remark 1 In the text, we focus on implementation in pure strategy equilibria. Most of the proofs cover mixed strategy equilibria also and Appendix F shows that all results carry over to implementation in mixed equilibria.

3 Main Results

This section contains our main implementation results, while the robustness results are presented in Section 4.

We first present a simple condition, *measurability*, which is necessary for the implementation of a social choice function when the preferences are state independent. We then turn to environments where the social planner can perform monetary transfers among the players and show that this condition is also sufficient for implementation of f for *every* preference structure, state independent or otherwise. Theorem 1 establishes that it is possible to implement any SCF f satisfying measurability with a perfect information mechanism when there are at least two players under a boundedness assumption. This result does not require that the evidence structure be normal. Theorem 2 establishes that, with three or more players, any measurable f can be implemented with only "small" (epsilon) monetary transfers though at the cost of using a mechanism which involves an integer game and assuming normal evidence. Furthermore, in this case, f can be implemented in a one–stage mechanism, establishing that it can be implemented in Nash equilibrium as well. Finally, in the special but important case of allocation problems, Theorem 3 establishes implementation with no transfers at all under weak conditions, again regardless of whether we assume normality.

For both Theorems 1 and 2, we present examples which relate the mechanisms we use to the classical mechanisms used by Maskin (1977) and Moore and Repullo (1988) and demonstrate the robustness of our mechanisms. These issues are discussed in more detail in Section 5.1 and Section 4 respectively.

We turn now to a formal statement of our results.

Definition 1 *Given a social environment Ψ, we say that the SCF f satisfies* measurability *if for every pair of states s and s' with $M_i(s) = M_i(s')$ for all i, we have $f(s) = f(s')$.*

In other words, define two states to be equivalent if every agent has the same set of available evidence in the two states. Then f satisfies measurability if f is measurable with respect to the partition of S induced by this equivalence relation.

Put differently, measurability 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 state s' when the true state is s (*i.e.*, $M_i(s) \not\subseteq M_i(s')$) or can refute state s at s' ($M_i(s') \not\subseteq M_i(s)$).

Proposition 1 *Let Ψ be a social environment with state independent preferences and let f be an SCF for Ψ. If f can be implemented, then f satisfies measurability.*

Proof: Suppose f is implemented by the mechanism Γ. Fix any pair of states $s, s' \in S$ such that $M_i(s) = M_i(s')$ for all i. It is easy to see, then, that a history h is feasible in s iff it is feasible in s'. Since the preferences in states s and s' are also the same, we have $\Gamma(s) = \Gamma(s')$. Hence the set of subgame perfect equilibrium outcomes in $\Gamma(s)$ and $\Gamma(s')$ are the same. Since Γ implements f, then, we must have $f(s) = f(s')$, so f is measurable. ▌

Clearly, this result is not driven by our restriction to multistage games or our focus on subgame perfect equilibrium and does not depend on whether we assume normality or not. The result holds for any class of mechanisms and any equilibrium concept simply because a mechanism cannot induce different games in states s and s' unless some agent's preference or evidence differs in the two states. The result highlights the obvious fact that for the outcome to vary across states, either preference variation or evidence variation is necessary.

Definition 2 *We say that Ψ is a social environment with* monetary transfers *if the following conditions hold. First, there is a set $\hat{A}$ such that the set of alternatives A can be written as*

$$A = \{(\hat{a}, t_1, \ldots, t_I) \in \hat{A} \times \mathbf{R}^I \mid \sum_{i \in \mathcal{I}} t_i \leq 0\}.$$

Second, for each i, the preference relations $\succeq_{i,s}$ for $s \in S$ can be represented by a utility function $u_i : A \times S \to \mathbf{R}$ of the form

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

for some function $v_i : \hat{A} \times S \to \mathbf{R}$. Finally, player i's preferences over lotteries over A are represented by the expectation of u_i.

As we explain below, the linearity of u_i in t_i is not critical for our results. The assumption on preferences over lotteries is made in order to establish that our results are valid not only for *pure* subgame perfect equilibrium but also for mixed equilibria.

The following terminology will be convenient. We say that an SCF $f : S \to A$ is *essential* if for every $s \in S$, there exists an alternative $\hat{a}(s) \in \hat{A}$ such that $f(s) = (\hat{a}(s), 0, \ldots, 0)$. We have the following interpretation in mind. The social planner is interested in implementing a function which maps S to $\hat{A}$, not A. To obtain this goal he uses monetary transfers as incentives to induce revelation of the true state. We emphasize that the social choice function f may itself call for transfers and that the t_i's are "above

and beyond" what f calls for. That is, $\hat{A}$ itself may have a product structure which allows for transfers. In this sense, the restriction to essential f's is without loss of generality.

Given any $\varepsilon > 0$, we say that a mechanism uses ε *monetary transfers* if for every terminal history h, if $g(h) = (\hat{a}, t_1, \ldots, t_I)$, we have $|t_i| \leq \varepsilon$ for all i. We say that a mechanism is *budget–balanced* if for every terminal history h, $\sum_{i \in \mathcal{I}} t_i = 0$ where $g(h) = (\hat{a}, t_1, \ldots, t_I)$.

3.1 Implementation in a Perfect Information Game

Let

$$V(\Psi) = \sup_{i \in \mathcal{I},\, s \in S,\, \hat{a}, \hat{a}' \in \hat{A}} v_i(\hat{a}, s) - v_i(\hat{a}', s).$$

Thus $V(\Psi)$ is an upper bound on the monetary value of a change in the selection of an alternative in $\hat{A}$ for every player i and every state s. We say that an environment is *bounded* if $V(\Psi)$ is finite.

Our first sufficiency result yields implementation in a class of games where subgame perfection seems particularly natural, namely games of perfect information. We note that this result does not require normality of the evidence structure.

Theorem 1 *Let Ψ be a bounded social environment with monetary transfers and at least two agents. Let f be an essential SCF for Ψ that satisfies measurability. Then there exists a perfect information mechanism Γ_f using limited evidence that implements f. If there are at least three agents, there is such a mechanism satisfying budget–balance.*

Remark 2 Theorem 1 has the following simple consequence. Consider a bounded social environment Ψ with monetary transfers and at least two agents such that for every s and s', with $s \neq s'$, there is some i with $M_i(s) \neq M_i(s')$. Then *every* essential social choice function f can be implemented with a perfect information mechanism which uses limited evidence. If there are at least three agents, the same is true restricting to budget–balanced mechanisms.

To see the intuition for Theorem 1, consider the following example.

Example 1.

There are two players, 1 and 2, and two states, s_1 and s_2. Let $f(s_i) = a_i = (\hat{a}_i, 0, 0)$, $i = 1, 2$, where $\hat{a}_1 \neq \hat{a}_2$. Let $M_1(s_1) = \{S, \{s_1\}\}$ and $M_1(s_2) = M_2(s_1) = M_2(s_2) = \{S\}$.

That is, in state s_1, 1 can prove that s_1 is the true state, but 1 cannot prove anything in s_2 and 2 can never prove anything. Assume $V(\Psi)$ is finite.

We can implement f with no assumptions on the preferences aside from the possibility of monetary transfers and boundedness. The mechanism we use is an adaptation of the mechanism used by Moore and Repullo (1988)[12] as we explain in more detail below.

Fix any $F > V(\Psi)$ and any $\varepsilon > 0$. Consider the following perfect information mechanism. First, in Stage 1, player 1 makes a cheap talk claim of either s_1 or s_2. Next, in Stage 2, player 2 can challenge or not. If 2 does not challenge, the game ends with outcome $f(s)$ where s is 1's claim. If 2 does challenge, we go to Stage 3.

In Stage 3, 1 has a chance to provide evidence. Whether he does or not, the outcome has $\hat{a} = \hat{a}_i$ where s_i is the state claimed by player 1 in Stage 1. Only the transfers depend on 1's evidence presentation in Stage 3. More specifically, the transfers to 1 and 2 are given in the table below:

	claimed s_1	claimed s_2
proves s_1	$(-F, -F)$	$(-F, F)$
no proof	$(-F-\varepsilon, F)$	$(-F-\varepsilon, -F)$

To see that this mechanism implements the social choice function in subgame perfect equilibrium, consider any subgame leading to Stage 3. It is easy to see that player 1 will present evidence $\{s_1\}$ at this point if he has it since presenting this evidence has no effect on $\hat{a}$ but lowers the amount 1 must pay by ε.

So consider Stage 2. Player 2 knows that 1 will present evidence if and only if he has it — that is, if and only if the state is s_1. It is easy to see that this implies 2 will challenge if and only if 1's claim was false. If 1's claim is false, then challenging earns F for 2, while if 1's claim is true, challenging costs 2 F. In either case, the alternative $\hat{a}$ which is selected is not affected by the challenge.

In light of this, it is easy to see that 1 will claim the true state in Stage 1. A false claim will be challenged in Stage 2, leading 1 to be fined at least F, an amount outweighing any gain from changing $\hat{a}$.

Note that none of the reasoning above changes if we assume that player 1 can forge evidence at some small cost. More specifically, suppose 1 could provide the evidence "proving" s_1 when the state is actually s_2 at a cost of $\delta > \varepsilon$. Since the gain to 1 in providing evidence when challenged is only ε, he would not forge evidence and hence the mechanism still implements f. We discuss the extension of our results to the case where

[12] See their Section 5.

evidence can be forged at a positive cost in the next section.

Finally, note that evidence is only presented once in this mechanism, in response to challenges. Hence the mechanism uses limited evidence.

To see the relationship to the Moore and Repullo mechanism, consider a different problem where there is no evidence but where there exists some $\hat{a}', \hat{a}'' \in \hat{A}$ such that player 1 strictly prefers $\hat{a}'$ to $\hat{a}''$ in state s_1 and has the opposite strict preference in state s_2. Consider exactly the mechanism above but where we change Stage 3 to the following. If there is a challenge, player 1 is offered a choice between $\hat{a}'$ and $\hat{a}''$. The outcome has whichever $\hat{a}$ he chooses along with the following transfers:

	claimed s_1	claimed s_2
chooses $\hat{a}'$	$(-F, -F)$	$(-F, F)$
chooses $\hat{a}''$	$(-F, F)$	$(-F, -F)$

It is easy to see that the analysis of this mechanism follows the same lines as the previous discussion. In Stage 3, player 1 will effectively reveal the true state by his choice, so in Stage 2, player 2 will challenge if and only if 1's claim is false. Hence 1 will claim the true state in Stage 1.

This mechanism is essentially the mechanism of Moore and Repullo (1988). What they require more generally is that for every pair of states, s_1 and s_2, there is some agent i who has a *preference reversal* between the two — that is, for some alternatives a and b, i strictly prefers a to b in state s_1 but has the opposite preference in state s_2. In this case, agent i can, in effect, "prove" the true state by choosing between a and b.

Our mechanism uses evidence to create the same effect as a preference reversal. In both states, player 1 would like to prove s_1. However, he can only do so in state s_1. Equivalently, this action costs him 0 in s_1 and costs a prohibitive amount in s_2. In the latter interpretation, we can say that he prefers to prove s_1, taking costs into account, in state s_1 but has the opposite preference in s_2.

This analogy suggests that we might be able to prove our results by appropriately reinterpreting standard theorems. While there is a linkage between our results and standard theorems, the connection is not strong enough to make our results corollaries.

To see the connection and why it is imperfect, consider how we might embed our model with evidence into the standard, evidence–less framework. Suppose that we expand the set of outcomes to include both the social alternative $a \in A$ and also the evidence presented by each player. To define preferences over such outcomes, it would be natural to say that a player who presents evidence in a state where that evidence is not feasible bears a very high, perhaps infinite, cost, but that presenting evidence in a state where

the evidence does exist is costless. Thus the state dependent evidence costs induce state dependent preferences. Hence we may be able to exploit standard approaches to constructing mechanisms to demonstrate sufficiency.

While this intuition is quite useful in understanding what we do, there are three issues that complicate the derivation of our results from those of Maskin and Moore–Repullo. First, evidence presentation is supposed to be under the control of the agent, not something that can be forced upon him by the social planner. Thus we must construct our mechanisms to ensure that the agent will choose to present the evidence the mechanism calls for. Second, evidence which is not available in a state is supposed to be completely infeasible, not just expensive, in such a state. This means that the mechanism in the extended space has more subgames than it is supposed to have. In principle, this could prevent theorems in the extended space from carrying over to our setting. Finally, evidence presentation is not part of the outcome we seek to implement. Thus we must ask whether there exists an extended outcome (including evidence presentation) which equals the outcome we want on the original space and which can be implemented. Thus even if the usual results do carry over, it remains to characterize which outcome functions on the original space can be implemented.

In Section 5.1, we return to these issues and discuss the connection of our results to standard theorems in more detail.

Turning to other issues, note that in this game, budget balance does not obtain if player 2 challenges 1 even when 1 claimed the true state. The fines charged to the two players cannot be given to either of them without interfering with their incentives. On the other hand, if there were a third player, he could receive the fines in this situation, allowing budget balance. This is why we obtain budget balance with at least three players but not, in general, with only two.[13]

Also, note that the only thing SP needs to know about players 1 and 2's preferences in order to set up this mechanism is how large the fines need to be. As long as he can bound the v_i differences, he does not need to know anything else about the players' preferences. Similarly, these are the only facts players 1 and 2 need to know about each other's preferences. In Section 4, we formalize this notion of robustness and generalize this observation.

Also, it is worth noting that, while the mechanism used in this example relies on information about which player has evidence, it is not difficult to give a more symmetric (but more complex) version of the mechanism which does not rely on SP knowing which player has evidence.[14] Thus SP does not need to know much about the available evidence

[13]The difficulty of achieving budget balance with two agents is well–known; for example, Moore and Repullo encounter a similar issue.

[14]See the mechanism used in the proof of this result in our working paper, Ben Porath and Lipman

either.

3.2 Implementation with ε Transfers

The mechanism of Theorem 1 does not require the planner to know much about the players or the players to know much about each other and has the appealing feature that it is a perfect information game. However, the mechanism relies on large monetary fines off the equilibrium path. The next result improves on this, though at the cost of moving to a mechanism which relies on an integer game and assuming the evidence structure is normal (as well as ruling out the case of two players). In this case, the planner needs even less information about the preferences of the players as the boundedness assumption is no longer needed. Also, the mechanism we use is a one–stage mechanism, so implementation is achieved in Nash equilibrium. Note that implementation in Nash equilibrium is more difficult to achieve in the sense that implementation in Nash equilibrium implies implementation in subgame perfect equilibrium but not conversely.[15]

Theorem 2 *Fix any $\varepsilon > 0$. Let Ψ be a social environment with monetary transfers, an evidence structure that is normal, and at least three players. Let f be an essential SCF for Ψ that satisfies measurability. Then there exists a budget–balanced one–stage mechanism with ε transfers Γ_f that implements f.*

Remark 3 As in the case of Theorem 1, we get a simple but striking corollary for social environments with monetary transfers, normal evidence, and at least three agents such that for every s and s', with $s \neq s'$, there is some i with $M_i(s) \neq M_i(s')$. For such environments, *every* essential social choice function can be implemented by a budget–balanced mechanism with ε monetary transfers.

As in the case of Theorem 1, our proof exploits the analogy to the standard implementation problem, this time the results of Maskin (1977), particularly the famous proof of Maskin's theorem due to Repullo (1987). In this regard, it is worth noting that the existence of monetary transfers ensures that Maskin's no veto power condition is satisfied.

To see the intuition, consider the following variation on Example 1.

Example 2.

(2009). The mechanism used there does require normality, though.

[15] To see this, simply note that if a mechanism implements in Nash equilibrium, then, viewing the mechanism as a one–stage game, we see that it also implements in subgame perfect equilibrium.

We now add a third state, s_3, and a third player. As before, we assume $M_1(s_1) = \{\{s_1\}, S\}$ and $M_1(s) = \{S\}$ for $s \neq s_1$. That is, player 1 can prove s_1 if it is true and nothing otherwise. We now assume $M_2(s_2) = \{\{s_2\}, S\}$ and $M_2(s) = \{S\}$ for $s \neq s_2$, so player 2 can prove s_2 if it is true and nothing otherwise. Finally, we assume $M_3(s) = \{S\}$ for all s, so player 3 never has any evidence.

Consider the following mechanism. The players move simultaneously, sending to the mechanism four things: evidence, a claim of a state, a requested outcome $\hat{a} \in \hat{A}$, and an integer. Let i's message be denoted $(E_i, c_i, \hat{a}_i, z_i) \in M_i \times S \times \hat{A} \times Z$ where $M_i = \cup_s M_i(s)$ and Z denotes the integers. (Thus $\hat{a}_i$ now refers to the $\hat{a}$ named by i, not the $\hat{a}$ in $f(s_i)$.)

The outcome is determined as a function of the claims as follows. If all three players claim the same state s and present all the evidence they would have in that state, the outcome is $f(s)$ with no transfers. If two players claim s and present all evidence they would have in that state, but the third either claims $s' \neq s$ or claims s but does not present all the evidence he has in s, then there are two cases. First, if the third player does not prove that the other two are lying — i.e., if his evidence is consistent with the state being s — then the outcome is still $f(s)$ with no transfers. Alternatively, if he proves the state is not s, the outcome is still $f(s)$ but with each of the other two players paying $\varepsilon/2$ to the third player. Finally, for any other strategy tuple, the outcome is determined by the player who chose the largest integer. Specifically, if i chose the largest integer (where we break ties in any fashion), then the outcome has $\hat{a} = \hat{a}_i$. In addition, each of the other two players pays $\varepsilon/2$ to i.

To see that this mechanism implements f, we first show that for every state s, there is a Nash equilibrium with outcome $f(s)$. In this equilibrium, each player i sets $c_i = s$ and provides all his evidence for state s. It is easy to see that no feasible unilateral deviation by any player can change the outcome since no player can disprove the true state. Hence this is an equilibrium with outcome $f(s)$.

To see that there is no other Nash equilibrium outcome, first note that there is no equilibrium in which the outcome is determined by the integers chosen. If so, any player who did not choose the largest integer would deviate to a larger integer since he could change the $\hat{a}$ part of the outcome (if desired) and replace paying $\varepsilon/2$ with receiving a payment of ε. Thus in any equilibrium, at least two of the three players make the same claim and present all the evidence they would have if that claim were true.

Suppose exactly two out of three players make the same claim and present all evidence for it. Then one state is going "unclaimed," so either of these two could deviate to claiming it instead, which makes the integers determine the outcome. This player i could choose an integer larger than that selected by any player and $\hat{a}_i$ equal to the $\hat{a}$ that would have obtained otherwise. The deviation leaves $\hat{a}$ unchanged but earns i a "reward" of ε, while his transfer would have been 0 or $-\varepsilon/2$ with no deviation. Obviously, then, this

would be a profitable deviation.

Hence in every equilibrium, all three players must make the same claim and present all the evidence they would have if that claim were true. Clearly, if this claim is s_1 or s_2, then the claim must be true since this is the only way player 1 or 2 would be able to provide all the evidence he would have in that state. Hence in either of these cases, the equilibrium achieves the right outcome.

So suppose the claim is s_3. Again, if the claim is true, the mechanism generates the desired outcome, so suppose the state is not s_3. In this case, one of players 1 and 2 must be able to prove that the claim is false. By deviating to a claim of the truth and presenting his evidence, this player doesn't change $\hat{a}$ but goes from no transfer to receiving a transfer of ε. Hence this is a profitable deviation.

It is not hard to show that this mechanism continues to implement f if any agent could forge evidence at a cost $\delta > \varepsilon$. To see this, consider first our argument above that it is an equilibrium in any state for each player to report the true state and present all his evidence in that state. We noted that no player could change the outcome since none could disprove the true state. Now a player could "disprove" the true state by providing false evidence. However, the reward he receives for this is ε and the cost of forgery is $\delta > \varepsilon$. Hence no player would have an incentive to deviate.

Above, we argued that if there is any other equilibrium in state s, it must be that all three players make the same claim and present all the evidence they would have if that claim were true. It is not hard to see that this argument is unaffected by the possibility of forgery as is our contradiction of the possibility that the false claim is s_3. So suppose the false claim is s_i, $i \in \{1, 2\}$. In this case, player i must be forging evidence that the state is s_i. But then he could deviate to not forging this evidence. This would not change the outcome but would save himself the forgery cost, making him strictly better off.

This mechanism is quite similar in spirit to the one originally proposed by Repullo (1987) to prove Maskin's (1977) characterization of Nash implementation. To see how Repullo's mechanism works, consider a variation on this environment where there is no proof but where preferences differ across states. The players simultaneously send to the mechanism a claim of a state, a requested alternative $\hat{a}$, and an integer. If all three players claim state s, the outcome is $f(s)$. If two out of three claim s while the third claims s' and requests $\hat{a}$, there are two cases. If the third player prefers $\hat{a}$ to $f(s)$ in state s, then we ignore his claim and the outcome is $f(s)$. On the other hand, if he prefers $f(s)$ to $\hat{a}$ in state s, the outcome is $\hat{a}$. Finally, if all three make different claims, the outcome is determined by the integers just as in our mechanism.

The key difference between Repullo's mechanism and ours is similar to the difference between the mechanism we used for Theorem 1 and the Moore–Repullo mechanism. In

both cases, the earlier work can be thought of as using revealed preference as a form of proof. In the case of Repullo's mechanism, the player who deviates proves that the others are lying when they claim s by asking for an outcome $\hat{a}$ that he would like *less* than $f(s)$ if they were telling the truth. By contrast, in our mechanism, the deviator changes the outcome only when he provides physical evidence that the others are lying.

We discuss the connections between Theorem 2 and Maskin's theorem in more detail in Section 5.1.

Turning to other issues, note that the analysis of equilibria did not use any properties of the agent's preferences other than the fact that more money is better than less. Thus this is all the planner needs to know about the preferences of the players to set up this mechanism and know it will implement. Similarly, this is all the players need to know about each other. Theorem 4 in the next section formalizes and generalizes this observation. Finally, as noted, the mechanism ensures implementation even when players can forge evidence at a small cost.

Remark 4 While we assume that utility is linear in transfers, this is not necessary for Theorems 1 and 2. As the examples above suggest, the only properties we use in the proofs are the following. First, both theorems use the assumption utility is strictly increasing in one's own transfer and independent of any other agent's transfer. Second, Theorem 1 uses the assumption that there exists an i such that for any $\hat{a}$ and $\hat{a}'$, there are transfers t_i and $\hat{t}_i$ such that i strictly prefers $(\hat{a}, t_i)$ to $(\hat{a}', 0)$ to $(\hat{a}, \hat{t}_i)$ at every state s.

Next, we discuss a special but important case where implementation is achieved in a perfect information game with no transfers at all. Consider the problem of the allocation of a set of goods among a set of agents. It turns out that a simple modification of the proof of Theorem 1 establishes that in such a problem, every SCF that satisfies measurability and assigns each agent a positive amount of at least one divisible good can be implemented by a perfect information mechanism without monetary transfers. The basic intuition is simple: the role played by large fines in the mechanism of Theorem 1 can be played by giving the goods an agent would have received according to f to some other agent instead.

More formally, we say that a social environment Ψ is an *allocation environment*[16] if the following three statements hold. First, there exists an integer $K \geq 0$, nonempty sets

[16]We use this term rather than the more commonly used "economic environments" for two reasons. First, the phrase "economic environments" has been used by many authors for many different but conceptually similar notions — see, for example, Moore and Repullo (1988), Jackson (1991), Baliga (1999), Kartik and Tercieux (2011), and Healy and Mathevet (2011). Second, while some papers use this term to mean environments where the allocation of goods is determined as in our case (e.g., Healy

$\bar{A}_k \subseteq \mathbf{R}$, $k = 1, \ldots, K$, and positive numbers $\bar{x}_k$, $k = 1, \ldots, K+1$, such that

$$A = \left\{ (x^1, \ldots, x^I) \mid x^i \in \left[\prod_{k=1}^{K} \bar{A}_k \right] \times \mathbf{R}_+ \text{ and } \sum_{i=1}^{I} x_k^i \leq \bar{x}_k, \ k = 1, \ldots, K+1 \right\}.$$

That is, A is a set of allocations of $K+1$ goods, at least one of which is divisible. Second, for every $i \in \mathcal{I}$ and $s \in S$, $(x^i, x^{-i}) \sim_{i,s} (x^i, \bar{x}^{-i})$. In other words, in any state, agent i is indifferent between allocations which give him the same goods. Finally, for every $i \in \mathcal{I}$ and every $s \in S$, $(x^i, x^{-i}) \succ_{i,s} (\bar{x}^i, x^{-i})$ if $x^i > \bar{x}^i$ (where $>$ denotes the vector ordering of weakly larger in every component, strictly larger in at least one component).

Given $\varepsilon > 0$, define

$$A^\varepsilon = \left\{ x \in A \mid x_{K+1}^i \geq \varepsilon, \ \forall i \in \mathcal{I} \right\}.$$

Theorem 3 *Fix any allocation environment with at least two players and any $\varepsilon > 0$. Let $f : S \to A^\varepsilon$ be a SCF that satisfies measurability. Then there exists a perfect information mechanism Γ_f using limited evidence that implements f.*

Kartik and Tercieux (2011) study Nash implementation in a model where agents can send messages with state–dependent costs in addition to cheap talk messages.[17] Their framework includes the case of hard evidence because a hard evidence message can be viewed as a message with zero cost in some states and a prohibitively high cost in the rest. They give a condition called evidence–monotonicity which they show is necessary and, given some additional conditions, sufficient for Nash implementation. In the case of hard evidence, roughly speaking, evidence–monotonicity says that a pair of states should either satisfy our measurability condition or be related in the way described by Maskin monotonicity.[18]

The fact that their evidence–monotonicity condition refers both to evidence and to preferences allows them to obtain strengthened analogs of Proposition 1 and Theorem 2. On the other hand, the use of preference variation also implies that they do not obtain the robustness results mentioned earlier and discussed in detail in the next section.

Finally, they also relate their results to standard implementation on an extended outcome space, as we comment further on in Section 5.1.

---

and Mathevet (2011)), other papers use this term to mean something more similar to what we call environments with monetary transfers (e.g., Moore and Repullo (1988)).

[17] The initial version of their paper was written after the circulation of preliminary drafts of this paper which considered only subgame perfect implementation. Their results on Nash implementation were developed concurrently with ours.

[18] Intuitively, if a pair of states satisfy Maskin's monotonicity condition, then the variation in preferences across the two states can be exploited to implement different outcomes without evidence just as in Maskin's mechanism. To be precise, our description of their condition applies under the assumptions of monetary environments and normal evidence.

4 Robustness

In this section, we define the notion of robustness discussed informally so far and demonstrate that our results are robust in this sense. In particular, we show that when the social choice function satisfies measurability, the planner can implement even if he does not know the agent's preferences in any state $s \in S$ (beyond that they satisfy the requirements of a monetary environment and, for one result, are bounded in a certain sense) or the agent's beliefs about other agents. Furthermore, we do not need the agents to know anything more than the planner about other agents' preferences or even to have a common prior. In addition, since measurability imposes very little on the structure of evidence, the planner requires little information about the evidence available to the agents. We obtain such robustness because we exploit variation in evidence across states rather than variation in preferences. Finally, we show that our results are robust to allowing evidence to be forged at an arbitrary strictly positive cost.

To demonstrate robustness with respect to the planner's knowledge of preferences and beliefs, we make some changes in the model which apply only in this section. As in the rest of the paper, we let S denote the set of states, A the set of social alternatives, $f : S \to A$ the social choice function, $\mathcal{I}$ the set of players, and $M_i(s)$ the set of subsets of S that i can prove in state s. Our goal is to show robustness with respect to (1) the planner's knowledge about the preferences and beliefs of the agents and (2) the knowledge of each agent about the preferences and beliefs of other agents. Therefore, we now think of a state s as a specification of all the parameters that are relevant for the determination of the social alternative that SP wishes to implement and a specification of the evidence of each agent. However, s may not include enough information to identify the preferences of each agent or the beliefs of agents about the preferences of others.

Formally, we extend the model by adding *types* for each agent. For each i, let Θ_i denote i's set of types. A type $\theta_i \in \Theta_i$ determines the preferences of i and his beliefs over Θ_{-i}, both as a function of the state s which, as before, is assumed to be common knowledge among the agents. Of course, i's type is private information. For simplicity, we assume that each Θ_i is finite or countable. The set of all *full states* is Ω where $\Omega \subseteq S \times \prod_{i \in \mathcal{I}} \Theta_i$. Note that we do not require the set of full states to have a product structure. Thus we are not imposing any restrictions on whether types are correlated across players or are correlated with s. In particular, this formulation allows any relationship between the preferences of the agents and the planner's desired outcome. Given any s, we refer to a tuple $(s, \theta) \in \Omega$ as a full state consistent with s. We also say that θ_i is *consistent with* s if there exists θ_{-i} with $(s, \theta_i, \theta_{-i}) \in \Omega$. Let $\Theta_i(s)$ denote the set of θ_i consistent with s.

Agent i's utility function in full state $\omega = (s, \theta)$ is assumed to depend only on s and θ_i. That is, while we demonstrate robustness with respect to uncertainty about the preferences of other agents, we retain the assumption that each agent knows his own

payoffs. Thus we write

$$u_i(\hat{a}, t_1, \ldots, t_I, s, \theta_1, \ldots, \theta_I) = v_i(\hat{a}, s, \theta_i) + t_i.$$

The belief of θ_i at s over the types of the other players is denoted by $\mu_i(s, \theta_i) \in \Delta(\Theta_{-i})$. (As usual, Θ_{-i} is the set of profiles of types of the agents other than i and $\Delta(\Theta_{-i})$ is the set of probability distributions over Θ_{-i}.) We do not require the agents to have a common prior. However, we do impose a common support on the agent's beliefs in the sense that

$$\text{supp}(\mu_i(s, \theta_i)) = \{\theta_{-i} \in \Theta_{-i} \mid (s, \theta_i, \theta_{-i}) \in \Omega\}$$

for all i, $s \in S$, and all $\theta_i \in \Theta_i$ consistent with s.

In this section only, we refer to a tuple

$$\Psi^* = \left\langle \mathcal{I}, A, S, (M_i(s))_{i \in I, s \in S}, (\Theta_i)_{i \in \mathcal{I}}, \Omega, (\mu_i(s, \theta_i), v_i(\cdot, s, \theta_i))_{i \in \mathcal{I}, s \in S, \theta_i \in \Theta_i(s)} \right\rangle$$

as a *monetary environment with partial information*. We use the term partial information instead of the more common term "incomplete information" to emphasize that we are considering an environment where the state s is common knowledge among the agents. An SCF for an environment with partial information is a function $f : S \to A$. The definition of a mechanism is unchanged. Given an environment with partial information Ψ^* and a state s, a mechanism Γ now induces a game of incomplete information among the agents. We use $\Gamma(s \mid \Psi^*)$ to denote this game. Note that a strategy for player i in this game is a function of θ_i, so an equilibrium outcome is a function of θ.

Since we use different equilibrium notions for our two results, we state the definition of implementation for an arbitrary equilibrium concept. We say that Γ implements f in environment Ψ^* if for every $s \in S$, for every equilibrium of $\Gamma(s \mid \Psi^*)$, the equilibrium outcome given θ is $f(s)$ for every θ such that $(s, \theta) \in \Omega$.

The robustness properties of our mechanisms differ slightly across results for two reasons. First, Theorem 1 requires that the environment be bounded, while Theorem 2 does not. Second, Theorem 2 is based on a one–stage mechanism, so when we introduce incomplete information among the players, Bayes–Nash equilibrium is the appropriate solution concept. On the other hand, Theorem 1 uses a game of perfect information, so the robust version will require use of perfect Bayesian equilibrium. Since Theorem 2 is simpler on both criteria, we begin with it.

Theorem 4 (Robust Version of Theorem 2) *Fix any $\varepsilon > 0$. Fix any monetary environment with partial information Ψ^* with at least three players. Let f be an essential SCF for Ψ^* that satisfies measurability and assume the evidence structure is normal. Then there exists a budget–balanced one–stage mechanism with ε transfers Γ_f that implements f in Bayes–Nash equilibrium.*

For any monetary environment with partial information Ψ^*, let

$$V(\Psi^*) = \sup_{i\in\mathcal{I},\, s\in S,\, \theta_i\in\Theta_i(s),\hat{a},\hat{a}'\in\hat{A}} v_i(\hat{a}, s, \theta_i) - v_i(\hat{a}', s, \theta_i).$$

We say that Ψ^* is bounded if $V(\Psi^*) < \infty$.

Since we analyze environments with partial information in this section, the game induced by the mechanism we used in Theorem 1 is no longer a game of perfect information. Hence we use the term *sequential mechanism* to refer to a mechanism for which each information set is either a singleton or contains nodes that differ only in Nature's moves. That is, all past actions are observable and there are no simultaneous moves. We emphasize that this is just a change of name — the mechanism is the same one we used earlier.

Theorem 5 (Robust Version of Theorem 1) *Let Ψ^* be a bounded monetary environment with partial information and at least two agents. Let f be an essential SCF for Ψ^* that satisfies measurability. Then there exists a sequential mechanism Γ_f using limited evidence that implements f in perfect Bayesian equilibrium. If there are at least three agents, there is such a mechanism satisfying budget–balance.*

Remark 5 Definitions of perfect Bayesian equilibrium in the literature impose a variety of conditions on beliefs off the equilibrium path. We put no restrictions on such beliefs.

Since Theorem 3 is essentially a generalization of Theorem 1, we omit the straightforward generalization of Theorem 5 which gives a robust version of it.

Summarizing, Theorems 4 and 5 formalize the robustness results we discussed in the previous sections. More specifically, subject to the boundedness assumption in the case of Theorem 5,

1. The social planner does not need to know anything about the preferences of the players — nothing about how their preferences relate to the evidence available, the social choice to be implemented, or each other.

2. No agent needs to know anything about the preferences of other agents.

In addition, we note that the only information the planner needs to have about evidence is what is summarized by the measurability condition. That is, he needs to know that if $f(s) \neq f(s')$, then *some* agent has different evidence in the two states, but

he does not need to know anything about *which* agent does. In this sense, the planner requires very little information about the evidence. We do not provide a formal proof of this claim as it is a notationally complicated but straightforward extension of our other results.[19]

We conclude this section by discussing a different kind of robustness; namely, we show that our results still hold if there is a strictly positive cost of forging evidence. More specifically, change our model by assuming that there is a $\delta > 0$ such that if $E \notin M_i(s)$, agent i can still send evidence E in state s but at a cost of δ. It is easy to show that all results stated in this paper still go through with no changes. In particular, we can use the same mechanisms as the ones in the proofs of Theorems 1 and 2 so long as we specify that the ε of those mechanisms is smaller than δ. For Theorem 1, this conclusion is immediate; the slightly more involved argument that is required to demonstrate this claim for Theorem 2 is included with the proof of the theorem in Appendix B. For our Nash implementation result, this can be thought of as a generalization of Theorem 1 in Dutta and Sen (2011) and Corollary 1 of Kartik and Tercieux (2011), both of which show Nash implementation with agents who have arbitrarily small costs of lying.[20]

The next section will clarify the connection of our results to the classic implementation results of Maskin and Moore–Repullo, along the way demonstrating why small costs are sufficient for implementation.

5 Discussion

5.1 Comparison to Maskin and Moore–Repullo

In Section 3, we discussed the relationship between our mechanisms and those used in Maskin and Moore–Repullo in the context of two simple examples. In this section, we generalize and formalize this relationship, explaining in more detail how our results relate to theirs.

As discussed earlier, the key idea is to relate our theorems on implementation with

[19]More precisely, it is easy to obtain versions of Theorems 1, 2, 4, and 5 where the evidence available for each agent is common knowledge among the agents but is not determined by the state s. Hence the social planner does not know what evidence each agent has at each state s. He just knows that if $f(s)$ is different from $f(s')$, then some agent has different evidence across the two states. This result requires normal evidence.

[20]The Dutta–Sen and Kartik–Tercieux results can be thought of as modeling the situation where in every state, there is an agent who can prove what the true state is, but this evidence can be forged at a small cost. Of course, this assumption on evidence is much stronger than measurability.

evidence with outcome space A to versions of the classical implementation theorems without evidence but on an enlarged space which includes both the outcome in A and the evidence presented by the players. That is, we treat the problem as one where the social planner can specify the evidence presented by each player as a function of the state.[21]

We noted earlier that there are three reasons why our results do not follow directly as corollaries to standard results applied to the extended space. First, given a social choice function prescribing an outcome in A, moving to the larger space requires specifying how the evidence presentation will vary with s as well. Second, the implementation problem on the enlarged space treats "infeasible" evidence presentation as if it were costly but feasible. Finally, the implementation problem on the enlarged space treats evidence decisions as made by the planner, not the players.

In this section, we show how to solve the first problem, explain why the second causes no difficulties, and describe how one can address the third. Because all three issues can be addressed, we could prove our theorems as implications of the standard implementation theorems. However, as the discussion below will clarify, this is not straightforward and our direct proofs are simpler and more transparent.

Before showing how to extend the social choice function, we first recall some standard definitions from Maskin and Moore–Repullo. We state these for an arbitrary outcome space, denoted $\mathcal{A}$, but will soon apply them for our specific setting.

We say that a pair of states s and s' exhibit a *preference reversal* if there exists an agent i and two alternatives $a, a' \in \mathcal{A}$ such that $u_i(a, s) > u_i(a', s)$ and $u_i(a', s') > u_i(a, s')$. We say that f is *preference measurable* if every pair of states s and s' with $f(s) \neq f(s')$ exhibits a preference reversal. Preference measurability is the key assumption in Moore and Repullo.

We say that a social choice function $f : S \to \mathcal{A}$ satisfies *monotonicity* if for all $s, s' \in S$, if $u_i(f(s), s) \geq u_i(a, s)$ implies $u_i(f(s), s') \geq u_i(a, s')$ for all $a \in \mathcal{A}$ and all i, then $f(s') = f(s)$. Equivalently, if $f(s) \neq f(s')$, then there exists i and a such that $u_i(f(s), s) \geq u_i(a, s)$ but $u_i(f(s), s') < u_i(a, s')$. Monotonicity is the property Maskin shows is necessary and almost sufficient for implementation in Nash equilibrium in environments without evidence.

[21] Kartik and Tercieux (2011) also discuss the relationship between Nash implementation with evidence and implementation on an enlarged outcome space. Specifically, their Theorem 4 shows that their main condition, evidence–monotonicity, is equivalent to the existence of an extension of the social choice function to the larger space which satisfies monotonicity. This is the analog of our result below which shows that measurability implies monotonicity on the extended outcome space. Their result was developed concurrently with our analysis in this section. Our analysis also includes an analogous result for Moore–Repullo and subgame perfect implementation, an issue they do not discuss.

We now show that measurability in our implementation problem implies that we can extend the social choice function to the enlarged outcome space in such a way that we can satisfy either preference measurability or monotonicity on the enlarged space.

First, we show how to "translate" our model to the enlarged outcome space. Recall that

$$A = \{(\hat{a}, t_1, \ldots, t_I) \in \hat{A} \times \mathbf{R}^I \mid \sum_i t_i \leq 0\}$$

and that $M_i = \cup_{s \in S} M_i(s)$. Let $M_i^* = M_i \cup \{\emptyset\}$ where $\emptyset$ is interpreted as the "evidence presentation outcome" for a player who does not get an opportunity to present evidence. Let

$$\mathcal{A} = A \times M_1^* \times \ldots \times M_I^*.$$

Recall that $u_i((\hat{a}, t_1, \ldots, t_I), s) = v_i(\hat{a}, s) + t_i$. We extend the utility functions to $\mathcal{A} \times S$ by

$$\bar{u}_i((\hat{a}, t_1, \ldots, t_I, m_1, \ldots, m_I), s) = \begin{cases} u_i((\hat{a}, t_1, \ldots, t_I), s), & \text{if } m_i \in M_i(s) \text{ or } m_i = \emptyset \\ u_i((\hat{a}, t_1, \ldots, t_I), s) - K, & \text{otherwise} \end{cases}$$

where $K > 0$. Intuitively, this simply says that we switch from assuming that in state s, the agent cannot present any evidence not in $M_i(s)$ to assuming that he can do so, but only at a cost.

Given a social choice function $f : S \to A$ on the original outcome space, we consider two different extensions of it to the larger outcome space. First, we define $f_{MR} : S \to \mathcal{A}$ by $f_{MR}(s) = (f(s), \emptyset, \ldots, \emptyset)$. In other words, for every $s \in S$, f_{MR} specifies the same $a \in A$ as f and no evidence presentation. Second, we define $f_M : S \to \mathcal{A}$ by $f_M(s) = (f(s), \bar{E}_1(s), \ldots, \bar{E}_I(s))$ where $\bar{E}_i(s)$ is the event i proves if he presents all his evidence in s. That is, $\bar{E}_i(s) = \cap_{E \in M_i(s)} E$. Recall that for Theorem 2, we assume that the evidence structure is normal which means that $\bar{E}_i(s) \in M_i(s)$ for all i and s.

We now show the following claims. First, if Ψ is bounded and f is measurable, then for all $K > 0$, f_{MR} is preference measurable. Second, if f is measurable and the evidence structure satisfies normality, then for all $K > 0$, f_M is monotonic.

For the first claim, fix any $K > 0$. Suppose $f_{MR}(s) \neq f_{MR}(s')$. We construct a preference reversal for s and s'. By definition of f_{MR}, $f_{MR}(s) \neq f_{MR}(s')$ implies $f(s) \neq f(s')$. By measurability, $f(s) \neq f(s')$ implies there exists some i with $M_i(s) \neq M_i(s')$. Without loss of generality, assume $M_i(s) \not\subseteq M_i(s')$. (Otherwise, reverse the roles of s and s'.) Let m_i^* be any element of $M_i(s) \setminus M_i(s')$. Fix any $\hat{a} \in \hat{A}$ and any $\varepsilon \in (0, K)$. Let t' denote a vector of 0's and m' a vector of $\emptyset$'s. Define t by setting $t_j = t'_j = 0$ for all $j \neq i$ and $t_i = \varepsilon$. Define m by setting $m_j = m'_j = \emptyset$ for all $j \neq i$ and $m_i = m_i^*$. Finally, let $a = (\hat{a}, t, m_1, \ldots, m_I)$ and $a' = (\hat{a}, t', m'_1, \ldots, m'_I)$. Thus in outcome a, player i presents

evidence m_i^* and gets a reward of ε, while in outcome a', i does not present any evidence and gets no reward.[22] Clearly,

$$\bar{u}_i(a, s) = v_i(\hat{a}, s) + \varepsilon > v_i(\hat{a}, s) = \bar{u}_i(a', s)$$

while $K > \varepsilon$ implies

$$\bar{u}_i(a', s') = v_i(\hat{a}, s') > v_i(\hat{a}, s') + \varepsilon - K = \bar{u}_i(a, s').$$

Hence we have a preference reversal, so preference measurability holds.

For the second claim, suppose normality holds so that f_M is well–defined. Again, fix any $K > 0$. Suppose we have states s and s' with $f_M(s) \neq f_M(s')$. Again, by measurability, $M_i(s) \neq M_i(s')$ for some i. First, suppose that $M_i(s) \not\subseteq M_i(s')$, so $\bar{E}_i(s) \notin M_i(s')$. Let a denote the outcome which differs from $f_M(s)$ only in that i presents evidence $\bar{E}_i(s')$ instead of $\bar{E}_i(s)$. Then

$$\bar{u}_i(f_M(s), s) = u_i(f(s), s) \geq \bar{u}_i(a, s)$$

while

$$\bar{u}_i(f_M(s), s') = u_i(f(s), s') - K < \bar{u}_i(a, s').$$

Hence f_M satisfies monotonicity.

Second, suppose $M_i(s) \subseteq M_i(s')$ but $M_i(s) \neq M_i(s')$. That is, $M_i(s) \subset M_i(s')$. Now let a denote the outcome equal to $f_M(s)$ but with a transfer of $\varepsilon \in (0, K)$ to i and i presenting evidence $\bar{E}_i(s')$ instead of $\bar{E}_i(s)$. Then we have

$$\bar{u}_i(f_M(s), s) = u_i(f(s), s) > u_i(f(s), s) + \varepsilon - K = \bar{u}_i(a, s),$$

and

$$\bar{u}_i(f_M(s), s') = u_i(f(s), s') < u_i(f(s), s') + \varepsilon = \bar{u}_i(a, s').$$

Hence f_M satisfies monotonicity in this case as well.

It is also easy to show that Maskin's no veto power condition is trivially satisfied since no two players can have the same most favorable outcome.[23]

These observations imply that if f is measurable, then the extended versions of f can be implemented in a model without evidence where we treat the planner as able to "assign" evidence to players and where we replace infeasibility with costs. This observation, by itself, does not prove our results since it still remains to show that (a) replacing

[22] To be precise, at a', i does not get an opportunity to present evidence. In this sense, if i chooses a', he is choosing to not have a chance to present evidence.

[23] The most favorable outcome for player i must involve him receiving the largest possible transfer from the other players and hence cannot be the most favorable outcome for any other player.

infeasibility with costs is innocuous and (b) that the mechanisms still work when players choose their evidence presentation instead of the mechanism. While we omit the details, it is true that one can complete this line of argument and prove our results that way. Instead, our proofs construct mechanisms which are natural analogs to the mechanisms of Maskin and Moore–Repullo and then proving directly that the analog mechanisms do implement. This direct proof is more straightforward and transparent than verifying that a mechanism for a different model can be appropriately reinterpreted.

While we omit the proofs, it is not hard to see why the two difficulties mentioned above do not cause problems. Regarding the first problem, clearly, if we make the costs of presenting "infeasible" evidence high enough, no equilibrium will involve a player choosing to presenting such evidence. Hence the equilibrium set will not be affected.[24]

In the case of Theorem 1 and Moore–Repullo, the second problem is also easily addressed. As seen in the example in Section 3.1, the only time evidence is relevant is in the response to a challenge when the agent being challenged has a choice between two outcomes. In the reinterpretation of our model where evidence is part of the outcome, we would give the agent being challenged a choice between two extended outcomes where the only evidence presented in either case is presented by the agent himself. Obviously, there is no real distinction between having the agent choose between two outcomes which differ in the evidence he presents versus a choice between which of the two pieces of evidence to present with the rest of the outcome depending on his choice.[25]

On the other hand, this point is much more subtle in addressing the relationship between Theorem 2 and Maskin's theorem. To see why, recall that Maskin's result is about one–stage mechanisms, so exploiting this result requires us to follow suit. In particular, then, we cannot have a separate phase of evidence presentation as we did for Theorem 1 — all evidence that is to be presented must be presented at once. In other words, part of the outcome must be irrevocably determined by the presentation choices of the agents.

Because of this, we cannot simply translate Repullo's (1987) mechanism for proving Maskin's theorem directly to this setting to prove our theorem. To see the point, recall that in Repullo's mechanism, if $I-1$ players name a particular state s and outcome $a = f(s)$ but the remaining player i deviates to s' and a', then the outcome is $f(s)$ if i prefers a' to $f(s)$ at state s and a' otherwise. But once i deviates, all evidence has been presented. Thus the mechanism is "locked in" to the evidence provision part of the outcome. Clearly, depending on the nature of i's deviation, it may be impossible for the mechanism to prescribe $f(s)$ or a'.

[24]This is not necessarily true in considering Nash equilibria of an extensive form game since off equilibrium strategies could involve use of such strategies.

[25]We can use large fines to ensure that he won't choose some evidence presentation which is not one of the options we want him to consider.

To address this difficulty, we can modify the Repullo mechanism so that players present both cheap talk claims and evidence. The evidence part of the outcome is the evidence that is presented by the players, while the $(\hat{a}, t)$ part of the outcome is determined by both the cheap talk claims and evidence. This mechanism can be interpreted either as a "standard" mechanism for the implementation problem on the extended outcome space or as a mechanism with evidence. To see why we are able to follow Repullo's proof, consider again our proof that measurability implies that our extended outcome function f_M satisfies monotonicity. We showed this by constructing an alternative allocation a where the only change in the evidence component of the outcome was for exactly one player. This enables us to essentially follow Repullo's proof of Maskin's theorem since it means that we can specify a mechanism where the response to a deviation by a single player to an "unexpected" presentation of evidence respects that presentation of evidence and changes at most only the (a, t) part of the outcome.[26] Again, our proof is similar to this, but much more direct.

The connection between our results and those of Maskin and Moore–Repullo give another way to understand our robustness results of Section 4. Those results showed when measurability holds, we can implement even when the planner knows very little about the preferences of the players and the players know little about each other's preferences. Above, we showed that very minimal assumptions on preferences are sufficient for measurability to imply monotonicity and preference reversal. Hence one interpretation of our robustness results is that measurability eliminates the need to use preferences over social outcomes (that is, over A) to obtain monotonicity or preference reversals over the extended outcome space.

The connection also clarifies why our results hold even if there is a small cost of forging evidence. Our results showing that measurability implies monotonicity and preference reversal on the extended outcome space allow for arbitrarily small costs. Since the potential difficulties in using this to obtain implementation are solved without having to increase these costs, we implement even when the cost of "lying" is small.

[26] In particular, the monotonicity of f_M ensures that if in state s', each player j announces state $s \neq s'$ and presents $\bar{E}_j(s)$, then there is some player i who could deviate to $\bar{E}_i(s')$ and a pair $(\hat{a}, t)$ such that he prefers $((\hat{a}, t), (\bar{E}_i(s'), \bar{E}_{-i}(s)))$ to $((f(s), 0), \bar{E}(s))$ at s' but not at s. Since the evidence part of the outcome of the deviation is the evidence presented by the players, SP can select this outcome without "dictating" to the players the evidence to present.

5.2 Tightness

In this subsection, we use a series of examples to show that our results are tight in the sense that stronger results cannot be obtained without additional hypotheses.[27] In particular, we demonstrate the following:

1. With one player, it may not be possible to implement a measurable f even with unbounded transfers. Thus Theorems 1 and 2 do not extend to the one agent case.

2. With two players, it may not be possible to implement a measurable f with ε transfers, whether or not we require budget–balance or restrict attention to perfect information games. Thus Theorem 1 does not extend to ε transfers and Theorem 2 does not extend to the case of two players.

3. Even with three players, it may not be possible to implement a measurable f with ε transfers in a perfect information game, even if we do not require budget–balance. Thus Theorem 1 does not extend to ε transfers even with more than two players and Theorem 2 does not extend to perfect information games even with more than two players.

4. For any number of players, it may not be possible to implement a measurable f without monetary transfers even when we allow for a general multistage mechanism. Thus Theorem 2 does not extend to the case where there are no transfers.

All the examples we present are variations on Example 1 in Section 3. Throughout this section, we assume normality to make clear that we cannot extend these results even if we assume normal evidence.

We begin with the first point above. So consider Example 1 but now with only agent 1. Suppose his preference is $f(s_2) \succ_{1,s} f(s_1)$ for all s. Then it is impossible to implement f, robustly or otherwise, with large fines or small. To see this, suppose that f can be implemented. Hence there is a mechanism with outcome $f(s_2)$ in state s_2. But then player 1 can use the same strategy in state s_1 that he uses in s_2 to obtain $f(s_2)$ in state s_1. Since he prefers this to $f(s_1)$, it cannot be true that we implement f.[28]

Turning to the second point, recall that in our discussion of Example 1 in the previous section, we made no assumptions about preferences other than boundedness and showed that we could implement with a perfect information mechanism and large transfers. Now

[27] As we show in Section F of the Appendix, these claims hold whether or not we restrict attention to pure strategy equilibria.

[28] See Glazer and Rubinstein (2004, 2006) for an interesting treatment of the one agent case.

we demonstrate the second point by giving specific preferences for which we cannot implement with any perfect information mechanism restricted to "small" transfers. So consider the state independent preferences where $v_1(\hat{a}_2) = v_2(\hat{a}_1) = 1$ and $v_1(\hat{a}_1) = v_2(\hat{a}_2) = 0$. As in Example 1, the SCF is $f(s_k) = (\hat{a}_k, 0, 0)$. Also, player 1 can prove the state is s_1 in state s_1 and can prove nothing in s_2, while 2 cannot prove anything in either state. Clearly, if there are other feasible outcomes in $\hat{A}$ which serve the same role as transfers, then a bound on transfers is irrelevant. So we assume that $\hat{A} = \{\hat{a}_1, \hat{a}_2\}$. We assume transfers are small in the sense that we consider only mechanisms such that for every terminal history h, if $g(h) = (\hat{a}, t_1, t_2)$, then $|t_i| < 1/2$ for $i = 1, 2$.

In Section E.1 of the Appendix, we prove the following.

Claim 1 *In this example, there is no mechanism with transfers bounded below 1/2 which implements f.*

For the third point, we show that we cannot combine the results of Theorems 1 and 2 to obtain implementation in a perfect information game with ε transfers, even with three or more agents. We show this by adding a third player to the previous example. Assume player 3 is exactly like player 2 in the sense that he has no evidence in any state and has the same utility function as player 2. By Theorem 1, we can implement f in a perfect information game without a bound on transfers. By Theorem 2, we can implement f in a game without perfect information but with only ε transfers. However, in Section E.2 of the Appendix, we prove the following.

Claim 2 *In this example, there is no perfect information mechanism with transfers bounded below 1/2 which implements f.*

In both of these arguments, budget–balance plays no role, so dropping such a restriction does not overcome the negative results in the examples.

For the fourth point, we show that for any integer n, there exists an environment with n players and a measurable SCF that cannot be implemented in a multistage mechanism without transfers. We show this by generalizing the previous example so that there are $n-1$ players in the position of player 2 above. That is, players $2, \ldots, n$ have no evidence in either state and have the same state–independent utility function $v_i(\hat{a}_1) = 1$ and $v_i(\hat{a}_2) = 0$. Player 1, as above, can prove s_1 in state s_1 and nothing in s_2. He has the state–independent utility function $v_1(\hat{a}_1) = 0$ and $v_1(\hat{a}_2) = 1$. By Theorem 2, f can be implemented with a mechanism which involves only ε monetary transfers. However, in Section E.3 of the Appendix, we prove the following.

Claim 3 *In this example, there is no multistage mechanism without transfers that implements f.*

6 Conclusion

We have extended implementation theory in two ways. First, we allow the social choice function to depend on more than just the preferences of the agents. Second, we allow agents to support their statements with hard evidence.

We have shown that the measurability condition which is necessary for the implementation of a social choice function f when preferences are state independent is also a sufficient condition, with or without state independence, for the subgame perfect implementation of f when the social planner can perform monetary transfers. Furthermore, f can be implemented even if the planner knows very little about the agents' evidence, preferences, or their beliefs about each other's preferences. Theorem 1 establishes implementation with a perfect information mechanism when there are at least two players and the social planner can perform "large" monetary transfers. In this case, the implementation is robust in the sense that the planner only needs to know an upper bound on the players' willingness to pay to change the outcome. Theorem 2 establishes that when there are at least three players and the evidence structure is normal, f can be implemented with a one–stage mechanism using only ε monetary transfers but which relies on an integer game. In this case, the planner does not need any information about the preferences, nor do the players need any information about each others' preferences. Finally, in the special but important case of allocation problems, Theorem 3 shows that we can implement under weak conditions using a perfect information game with no transfers. Again, this mechanism implements regardless of the preferences of the agents. In all cases, the only information the planner requires about evidence is that measurability holds. In addition, the results are unchanged if we allow the possibility of forging evidence at an arbitrary strictly positive cost. Finally, we have discussed the relationship between our results and the classical work of Maskin (1977) and Moore and Repullo (1988) on implementation without evidence. In particular, we showed that our condition of measurability of the evidence structure implies monotonicity and preference reversal in a modified model with an extended outcome space which includes evidence where infeasible evidence is replaced by feasible but costly evidence.

There are many interesting directions for future research. First, the mechanisms we use in our results appear to have a variety of additional robustness properties which may be worth formalizing and exploring further. It may be of interest to characterize the mechanisms which require the least information on the part of the planner and/or the agents.

Second, clearly, implementation with a perfect information mechanism is more appealing than implementation by a mechanism which relies on integer games. It may be interesting to determine what can be implemented using perfect information mechanisms that allow the social planner to randomize but which do not rely on large monetary transfers.

Other general directions of interest are results without monetary transfers (or other "structural" assumptions which serve the same role), models with incomplete information among the agents, restrictions on or costs of evidence provision, and models where the social planner has less commitment power.

A Proof of Theorem 1

Fix a bounded environment Ψ and an essential social choice function f satisfying measurability. We write f as $f(s) = (\hat{a}(s), 0, \ldots, 0)$. Let F satisfy $F > V(\Psi)$ and fix any $\varepsilon > 0$. (Recall that $V(\Psi) = \sup_{i \in \mathcal{I},\ s \in S, \hat{a}, \hat{a}' \in \hat{A}} v_i(\hat{a}, s) - v_i(\hat{a}', s)$. By boundedness, this is finite.) Thus, for any player i, a monetary fine or reward of F outweighs any utility gain from changing the alternative that is selected.

The mechanism works as follows. We have (at most) I stages, each divided into (at most) three parts. In Stage i.A, $i = 1, \ldots, I$, player i sends a cheap talk message $c_i^A \in S$, interpreted as a claim of a state. In Stage i.B, player $i + 1 \pmod I$ sends a cheap talk message $c_i^B \in S$ where $c_i^B = c_i^A$ is interpreted as agreeing with i's claim and $c_i^B \neq c_i^A$ is interpreted as a challenge. If $i + 1$ challenges i, we move to Stage i.C which is explained below. Otherwise, we move directly to Stage $(i + 1)$.A for $i \leq I - 1$. Finally, if we get all the way through Stage I with no challenges, the outcome is determined as follows. There are no transfers. If there is an $\hat{a}$ which is the unique value of $\hat{a}(s)$ for some s such that $M_i(s) = M_i(c_i^A)$ for all i, then this is the $\hat{a}$ chosen. Otherwise, it is $\hat{a}(c_1^A)$.

If we move to Stage i.C for some i, player i presents evidence E_i. The outcome has $\hat{a} = \hat{a}(c_i^A)$ with transfers which depend on E_i, c_i^A, and c_i^B. If $M_i(c_i^A) = M_i(c_i^B)$, then i and $i+1 \pmod I$ both get transfers of $-F$. If $M_i(c_i^A) \neq M_i(c_i^B)$, then it must be possible for i either to refute c_i^A if c_i^B is true (i.e., $M_i(c_i^B) \not\subseteq M_i(c_i^A)$) or to refute c_i^B if c_i^A is true (i.e., $M_i(c_i^A) \not\subseteq M_i(c_i^B)$) or both. If it is possible to refute c_i^A when c_i^B is true, then the transfers are

$$\begin{cases} (-F - \varepsilon, -F) & \text{if } c_i^A \in E_1 \\ (-F, F) & \text{otherwise} \end{cases}$$

where the first number is the transfer to i and the second is the transfer to the challenger $i + 1 \pmod I$. If it is not possible to refute c_i^A when c_i^B is true (and therefore is possible to refute c_i^B when c_i^A is true), the transfers are

$$\begin{cases} (-F - \varepsilon, F) & \text{if } c_i^B \in E_1 \\ (-F, -F) & \text{otherwise} \end{cases}$$

In all cases, if $I \geq 3$, the other $I - 2$ agents get the same transfers as one another, chosen to make the total transfer equal to 0.

Note that evidence is only presented by a player in the response to the challenge stage and the mechanism ends after this. Hence this mechanism uses limited evidence.

To see that this mechanism implements f, let s^* denote the true state. Consider any Stage i.A. First, suppose $M_i(s^*) \not\subseteq M_i(c_i^A)$. Then $i + 1$ could challenge i and claim s^*. It is clearly optimal for i to present evidence refuting c_i^A, so $i + 1$ will receive F, yielding a

higher payoff than could be earned in the equilibrium of the following subgame otherwise. Second, suppose $M_i(s^*) \subset M_i(c_i^A)$ (where this is strict inclusion). Now it is possible at c_i^A to refute s^* but not conversely. So if $i+1$ challenges claiming s^*, i will be unable to refute this since it is true and the challenger $i+1$ will receive F. Hence unless $M_i(s^*) = M_i(c_i^A)$, it is optimal for $i+1$ to challenge. Since this leads to a fine of at least F for i, i's optimal strategy in Stage i.A is to make a claim c_i^A satisfying $M_i(c_i^A) = M_i(s^*)$. Obviously, i could claim the true state, so such claims exist. Hence there are no challenges and thus no transfers in equilibrium.

Finally, note that there must be at least one s for which $M_i(s) = M_i(c_i^A)$ for all i, namely the true state, s^*. Furthermore, measurability implies that if another state s' also satisfies this property, then $\hat{a}(s') = \hat{a}(s^*)$. Hence the mechanism has the outcome $\hat{a}(s^*)$, so we implement. ▌

B Proof of Theorem 2

The proof is by construction of a one–stage mechanism Γ_f which implements f. In the mechanism we construct, every agent chooses a piece of evidence and a cheap talk message in $S \times \hat{A} \times Z$ where Z denotes the positive integers. We write a typical choice of cheap talk message for i as $(s_i, \hat{a}_i, z_i)$.

To define the outcome as a function of the profile of evidence and cheap talk reports, we distinguish between several cases. Let $\bar{E}_i(s)$ denote what i proves in state s if he presents all his evidence. That is, $\bar{E}_i(s) = \cap_{E \in M_i(s)} E$. (Recall that we assume the evidence structure is normal, so $\bar{E}_i(s) \in M_i(s)$ for all i and s.)

First, suppose there is a state s such that $(s_i, E_i) = (s, \bar{E}_i(s))$ for all i. In this case, the outcome is $f(s)$.

Second, suppose there is a state s and an agent j such that $(s_i, E_i) = (s, \bar{E}_i(s))$ for all $i \neq j$ but $(s_j, E_j) \neq (s, \bar{E}_j(s))$. There are two subcases. First, suppose $s \notin E_j$. In this case, the outcome is $f(s)$ with a transfer of $\varepsilon/2$ to j, $\varepsilon/2$ to that $i \neq j$ who chooses the largest z_i (breaking ties by choosing the largest i who names the largest z_i), and transfers of $-\varepsilon/(I-2)$ to the other agents. Second, suppose $s \in E_j$. In this case, the outcome is $f(s)$ with a transfer of ε to that $i \neq j$ who chooses the largest z_i (with ties broken as above) and $-\varepsilon/(I-1)$ to the other agents.

Finally, for any other profile of messages, the alternative that is selected is determined by the integers chosen. Specifically, let i be the player who chose the highest integer z_i (breaking ties as above). The alternative that is selected is $\hat{a}_i$ with a transfer of ε to

player i and $-\varepsilon/(I-1)$ to every other player.

It is easy to see that for each s, there is an equilibrium with outcome $f(s)$. Specifically, the strategies $((s, f(s), 0), \bar{E}_1(s)), \ldots, ((s, f(s), 0), \bar{E}_I(s))$ form a Nash equilibrium with outcome $f(s)$ as no feasible unilateral deviation by any player can improve the outcome for him.

We now show that there is no (pure or mixed) Nash equilibrium in state s with an outcome different from $f(s)$. Fix any mixed strategies σ that form an equilibrium of $\Gamma(s)$. Let H^* denote the set of pure strategy profiles that fall into any case where the outcome is determined by the integers. It is easy to see that if $h \in H^*$, then $\sigma(h) = 0$. Otherwise, there must be some player who could increase his integer and improve his expected payoff conditional on $h \in H^*$ and hence improve his unconditional expected payoff.

Hence for any h with positive probability, we have $(s_i, E_i) = (s', \bar{E}_i(s'))$ for all i. Clearly, this implies that every i has $(s_i, E_i) = (s', \bar{E}_i(s'))$ with probability 1. Otherwise, there is a positive probability of a realization $h \in H^*$.

The outcome under these strategies is $f(s')$. Suppose that the outcome is not $f(s)$, so $f(s) \neq f(s')$. By measurability, there is some i with $M_i(s) \neq M_i(s')$. Since every i is presenting evidence $\bar{E}_i(s')$, it must be true that $\bar{E}_i(s') \in M_i(s)$ for all i. This implies $M_i(s') \subseteq M_i(s)$. To see this, suppose to the contrary that there is $E \in M_i(s')$ with $E \notin M_i(s)$. By consistency, $s \notin E$. But then $s \notin \bar{E}_i(s')$ so we cannot have $\bar{E}_i(s') \in M_i(s)$, a contradiction.

So $M_i(s') \subseteq M_i(s)$ for all i. Hence measurability implies that there must be some i for whom this inclusion is strict. Obviously, then, this agent could deviate to the same c_i but to evidence $E_i \in M_i(s) \setminus M_i(s')$. Since $E_i \notin M_i(s')$, we have $s' \notin E_i$. This would not change $\hat{a}$, but would yield i a transfer of $\varepsilon/2$ and hence make him better off, a contradiction.

Finally, to extend the result to allow costly forging of evidence, recall that the cost of providing false evidence is $\delta > 0$. Fix any $\varepsilon \in (0, \delta)$ and consider the same mechanism as above. Just as before, the strategies $((s, f(s), 0), \bar{E}_1(s)), \ldots, ((s, f(s), 0), \bar{E}_I(s))$ form a Nash equilibrium in state s with outcome $f(s)$. To see this, suppose agent j deviates from this strategy. If his evidence presentation is still consistent with s, then he changes the outcome so that some other agent receives ε and he must pay $\varepsilon/(I-1)$, obviously making him worse off. If his evidence presentation is inconsistent with s, he must have forged the evidence at a cost δ. The $\hat{a}$ does not change, but j receives a transfer of $\varepsilon/2 < \varepsilon < \delta$. Hence the transfer is less than the cost of forgery, so he is worse off. Hence these strategies form a Nash equilibrium in state s.

Just as before, there can be no equilibrium, pure or mixed, where the integers affect the outcome. Hence every equilibrium in state s has $(s_i, E_i) = (s', \bar{E}_i(s'))$ for some s' for all i. The same argument as above shows that there can be no equilibrium in state s where every agent i presents evidence which is feasible (without forgery) in state s but the outcome is different from $f(s)$. Hence if there is an equilibrium in state s with an outcome different from $f(s)$, it must be true that some agent is forging evidence in the equilibrium. Let j denote such an agent. Suppose j deviates to $(s, \bar{E}_j(s))$ and $\hat{a}_j = f(s')$. Depending on whether this evidence is consistent with s' or not, j either earns a transfer of $\varepsilon/2$ or has to pay $\varepsilon/(I-1)$. However, he saves the forgery cost of δ. Hence, since $\delta > \varepsilon > \varepsilon/(I-1)$, j gains from the deviation in either case, so this cannot be an equilibrium. ▌

C Proof of Theorem 3

Fix an allocation environment Ψ and SCF f satisfying the conditions of the theorem. Let Γ_f denote the mechanism that was defined in the proof of Theorem 1. We will show that a simple modification of Γ_f implements f.

For every player i, we define the following four allocations:

- a_i^1. i receives $\varepsilon/2$ of good $K+1$ and zero units of every other good. Player $i+1 \pmod I$ receives everything else.
- a_i^2. i receives $\varepsilon/2$ of good $K+1$ and zero units of every other good. Player $i+1 \pmod I$ receives zero of every good. The remaining goods are allocated among the remaining players in any fashion which gives strictly positive amounts of every good to every other player.
- a_i^3. Player $i+1 \pmod I$ receives all goods.
- a_i^4. Players i and $i+1 \pmod I$ receive nothing. The goods are allocated among the remaining players in any fashion which gives strictly positive amounts of every good to every other player.

Define the mechanism $\hat{\Gamma}_f$ as follows. Only the allocations are changed from Γ_f, not the structure of moves, etc. When there is a challenge in Stage i, if the transfers in Γ_f to i and $i+1$ were $(-F, F)$, the outcome in $\hat{\Gamma}_f$ is a_i^1. If the transfers were $(-F, -F)$, the outcome is a_i^2. If the transfers were $(-F-\varepsilon, F)$, the outcome is a_i^3. Finally, if the transfers were $(-F-\varepsilon, -F)$, the outcome is a_i^4.

It is easy to see that the proof of Theorem 1 relies only on the following assumptions about preferences for any player i and states s and s'. First, outcome $\hat{a}(s')$ and receiving a payment of F is strictly preferred by i at s to any point in the range of f. Second, any point in the range of f is strictly preferred by i at s to outcome $\hat{a}(s')$ and paying a fine of F. Finally, outcome $\hat{a}(s')$ and paying a fine of F is strictly preferred by i at s to outcome $\hat{a}(s')$ and paying a fine of $F + \varepsilon$.

It is easy to see that all three properties hold for the replacements of the fines used by the mechanism $\hat{\Gamma}_f$. More specifically, since $f(s)$ gives every agent at least ε of the divisible good and a_i^3 gives $i+1$ everything and a_i^1 all but $\varepsilon/2$ of the divisible good, $i+1$ must strictly prefer a_i^1 and a_i^3 to every point in the range of f at every state. On the other hand, since a_i^1 and a_i^2 leaves i with only $\varepsilon/2$ of the divisible good and nothing else, this must be strictly worse for i than anything in the range of f at any state. Similarly, a_i^2 and a_i^4 must be strictly worse for $i+1$ than anything in the range of f at any state. Finally, since a_i^1 and a_i^2 give i $\varepsilon/2$ of the divisible good while a_i^3 and a_i^4 give him nothing, he prefers the former in every state. ▌

D Proofs of Robustness Results

D.1 Proof of Theorem 4

The proof parallels that of Theorem 2. The mechanism is exactly the same as the one used there.

Fix an environment with partial information Ψ^*. Fix any $s \in S$. As in the proof of Theorem 2, it is easy to see that there is a Bayes–Nash equilibrium with outcome $f(s)$ for all θ such that $(s, \theta) \in \Omega$. Specifically, take the strategy for every agent i to be $\sigma_i(\theta_i) = ((s, f(s), 0, 0), \bar{E}_i(s))$ for every $\theta_i \in \Theta_i$. Since no feasible unilateral deviation can change the outcome, these strategies form a Bayes–Nash equilibrium with outcome $f(s)$ in every full state $(s, \theta) \in \Omega$.

The proof of Theorem 2 showed that the original mechanism had no pure or mixed Nash equilibria with an outcome different from $f(s)$. Since no agent's utility is affected by any other agent's type, the effect of uncertainty about other agents' types is the same as allowing mixing. Because of this, it is easy to adapt that proof to show that there is no pure or mixed Bayes–Nash equilibrium whose outcome differs from $f(s)$ in any full state $(s, \theta) \in \Omega$. In particular, we replace H^* in that proof with the set of pure strategy profiles played with positive probability given some $(s, \theta) \in \Omega$ such that the outcome is determined by the integers. Then the proof only has to be modified to clarify which

type of a given player deviates to prevent alternative strategy profiles from forming an equilibrium. In all cases, any type who gives positive probability to the (s, θ) in question suffices.

For example, consider the part of the proof of Theorem 2 which establishes that H^* has zero probability in equilibrium. To extend this, fix some $(s^*, \theta^*) \in \Omega$ for which there is positive probability of a profile of pure strategies for which the integers determine the outcome. Let i be any agent who does not say the highest integer given this realization of the pure strategy profile and given (s^*, θ^*). Then i must give probability strictly less than 1 to the event that $z_i > z_j$ for all $j \neq i$ given s^* and θ_i^* and given that the integers determine the outcome.[29] Clearly, then, i would gain by changing his strategy when he is type θ_i^* to a larger integer (possibly also changing $\hat{a}$). Hence there is no such equilibrium. The other parts are extended analogously. ▌

D.2 Proof of Theorem 5

Let Ψ^* be a monetary environment with partial information and fix any $s \in S$. Let Γ denote the mechanism defined in the proof of Theorem 1 and let $\Gamma^*(s)$ denote the sequential game of incomplete information induced by Γ and Ψ^* at state s.

Fix any profile of types $\bar{\theta}$ such that $(s, \bar{\theta}) \in \Omega$ and let $\bar{\Gamma}(s)$ denote the complete information game defined by Γ in the environment where it is common knowledge that the utility functions are given by $v_i(\cdot, s, \bar{\theta}_i)$ for each i. We now show that for every θ with $(s, \theta) \in \Omega$, the equilibrium outcome of $\Gamma^*(s)$ is the same as the equilibrium outcome of $\bar{\Gamma}(s)$.

To be more precise, note that any history of actions in the mechanism corresponds to a collection of information sets in $\Gamma^*(s)$, one information set for each type of the player whose turn it is to move at this history. Thus, in general, there is a set of outcomes following a given history, one for each profile of types.

We establish that in any equilibrium, for every state s, for every player i and every type θ_i of i, i makes a claim c_i^A in Stage i.A satisfying $M_i(c_i^A) = M_i(s)$ and is not challenged by player $i+1$ in Stage i.B. (The specific claim made by i may depend on θ_i.) As we explain below, this claim together with the measurability of f imply that the outcome in any equilibrium must be $f(s)$.

To establish this claim, first, consider any history of actions which puts us in Stage i.C for some i where it is i's turn to present evidence. Clearly, i's action only affects his

[29] This statement uses both our common support assumption and the assumption that each Θ_i is at most countable. Together, this ensures that $\mu_i(s^*, \theta_i^*)(\theta_{-i}^*) > 0$.

transfer and nothing else, so his optimal strategy at this point depends only on s, not his type or his beliefs about his opponent's types.

The rest of the proof is by induction on the stage. So suppose we are at Stage I and for all $i < I$, the claim made at Stage i.A satisfied $M_i(c_i^A) = M_i(s)$ and was not challenged. It is easy to see that in Stage I.B, 1 will challenge if and only if a challenge will be successful in the sense that $M_I(c_I^A) \neq M_I(s)$. To see this, simply note that at 1's turn, he has the choice between the $\hat{a}$ which would result from not challenging and the fine or reward that would result from challenging. Regardless of his type, the reward he would earn from a successful challenge must be preferred strictly to $\hat{a}$ and the fine he would incur from an unsuccessful one would be strictly worse than $\hat{a}$. Hence he challenges if and only if the challenge would be successful — that is, if and only if $M_I(c_I^A) \neq M_I(s)$. Given this, player I certainly makes a claim satisfying $M_I(c_I^A) = M_I(s)$ since for every θ_I, his payoff if challenged is strictly lower than his payoff to any such claim.

Given this, consider Stage i for any $i < I$ and that we have shown the claim for all larger i. If i makes a claim such that $M_i(c_i^A) = M_i(s)$, then by the induction hypothesis, the outcome if $i+1$ does not challenge has no transfers. Hence the outcome from a failed challenge is strictly worse, so $i+1$ will not challenge. On the other hand, if $M_i(c_i^A) \neq M_i(s)$, $i+1$ can make a successful challenge and will necessarily be better off than from not challenging. Hence for every type, $i+1$ challenges if and only if $M_i(c_i^A) \neq M_i(s)$. Given this, it is easy to see that i will make some claim satisfying $M_i(c_i^A) = M_i(s)$, as claimed.

Since this goes back to the beginning of the game, we see that in every equilibrium, the equilibrium path must have claims satisfying $M_i(c_i^A) = M_i(s)$ for all i. By measurability, there is only one outcome $\hat{a}$ such that $\hat{a} = \hat{a}(s')$ for some s' satisfying $M_i(c_i^A) = M_i(s')$ for all i. Since $f(s)$ clearly is such an outcome, the unique equilibrium outcome is $f(s)$, so the mechanism implements. ▌

E Proofs of Tightness Claims

E.1 Proof of Claim 1

Suppose by contradiction that there exists a mechanism Γ that implements f. Let $\Gamma(s_k)$ and $\Sigma_i(s_k)$, $k = 1, 2$, $i = 1, 2$ denote respectively the game that is induced by Γ at s_k and the set of strategies of player i in $\Gamma(s_k)$. Let $U_i : \Sigma_1(s) \times \Sigma_2(s) \to \mathbf{R}$ denote the payoff functions for the normal form of $\Gamma(s)$. Since Γ implements f, there is a subgame perfect equilibrium and hence a Nash equilibrium, $(\hat{\sigma}_1, \hat{\sigma}_2)$, in $\Gamma(s_2)$ with outcome $f(s_2)$.

Thus $U_1(\hat{\sigma}_1, \hat{\sigma}_2) = 1$ and $U_2(\hat{\sigma}_1, \hat{\sigma}_2) = 0$. Since $(\hat{\sigma}_1, \hat{\sigma}_2)$ is a Nash equilibrium and since monetary transfers must be strictly less than $1/2$, we have that for *every* $\sigma_2 \in \Sigma_2(s_2)$, the probability that $\hat{a}_1$ is selected when the profile $(\hat{\sigma}_1, \sigma_2)$ is played is strictly less than $1/2$. To see this, suppose that it is not true. Then there is a strategy for player 2, say σ_2', such that when $(\hat{\sigma}_1, \sigma_2')$ is played, the probability of $\hat{a}_1$ is greater than or equal to $1/2$. If player 2 deviates to this strategy, his payoff is strictly positive since he pays a fine of strictly less than $1/2$. Hence $(\hat{\sigma}_1, \hat{\sigma}_2)$ is not a Nash equilibrium in s_2, a contradiction. Given this, we must have $U_1(\hat{\sigma}_1, \sigma_2) > 0$ for every $\sigma_2 \in \Sigma_2(s_2)$ since the alternative $\hat{a}_2$ is selected with probability at least $1/2$ and player 1 pays a fine strictly less than $1/2$.

Consider now the game $\Gamma(s_1)$. Since $\hat{\sigma}_1 \in \Sigma_1(s_1)$ and since $\Sigma_2(s_1) = \Sigma_2(s_2)$, player 1 can guarantee a strictly positive payoff by playing $\hat{\sigma}_1$. It follows that in every Nash equilibrium in $\Gamma(s_1)$, player 1 obtains a strictly positive payoff. Hence there is no Nash equilibrium (let alone a subgame perfect equilibrium) in $\Gamma(s_1)$ with outcome $f(s_1)$. Hence Γ cannot implement f, a contradiction. ▌

E.2 Proof of Claim 2

Fix any perfect information mechanism Γ such that if $g(h) = (\hat{a}, t_1, t_2, t_3)$ for some terminal history h, then $|t_i| < 1/2$, $i = 1, 2, 3$. We show that this game cannot implement f. The proof works by showing that if there is a subgame perfect equilibrium of $\Gamma(s_2)$ with outcome $\hat{a}_2$ and some vector of transfers t, then every subgame perfect equilibrium of $\Gamma(s_1)$ has outcome $\hat{a}_2$ plus some transfers $\hat{t}$. The proof is by induction on the depth of $\Gamma(s_2)$.

So suppose $\Gamma(s_2)$ has a subgame perfect equilibrium with outcome $(\hat{a}_2, t_1, t_2, t_3)$.

First, suppose the depth of $\Gamma(s_2)$ is 1. Suppose player 2 is the only one who moves. Then it must be true that given any choice he makes, the outcome is $\hat{a}_2$ and some vector of transfers $\hat{t}$. To see this, suppose that there is some choice which leads to outcome $\hat{a}_1$ and some vector of transfers $\bar{t}$. Player 2's payoff from this choice is $1 + \bar{t}_2$. Since the equilibrium outcome is $(\hat{a}_2, t_1, t_2, t_3)$, we must have

$$1 + \bar{t}_2 \leq t_2$$

or $t_2 - \bar{t}_2 \geq 1$. Hence

$$|t_2| + |\bar{t}_2| \geq |t_2 - \bar{t}_2| \geq 1,$$

so

$$\max\{|t_2|, |\bar{t}_2|\} \geq \frac{1}{2},$$

contradicting our bound on transfers. Hence, as asserted, every choice available to 2 in $\Gamma(s_2)$ leads to $\hat{a}_2$ and some vector of transfers.

Given this, consider state s_1. Since player 2 has no proof available, his strategy set in $\Gamma(s_1)$ is the same as his strategy set in $\Gamma(s_2)$. So in state s_1, the equilibrium outcome must be $\hat{a}_2$ with some vector of transfers. Obviously, a symmetric argument applies to player 3.

So now suppose it is player 1 who moves in this game. Then there must be some cheap talk message available to player 1 which yields outcome $\hat{a}_2$ and some vector of transfers t. So consider the game $\Gamma(s_1)$. Since player 1 has proof, his set of strategies is larger in this game. Also, it is possible that the game now has larger depth, since presentation of proof might lead to a subgame. However, it must still be true that player 1 has available a message which leads to outcome $\hat{a}_2$ and transfers t. Suppose that in $\Gamma(s_1)$, there is an equilibrium with outcome $\hat{a}_1$ and some vector of transfers $\bar{t}$. Then player 1 must prefer this outcome to $\hat{a}_2$ with transfers t. Similar reasoning to the above shows that this implies

$$\max\{|t_1|, |\bar{t}_1|\} \geq \frac{1}{2},$$

again violating our bound on transfers. Hence any subgame perfect equilibrium in state s_1 must have outcome $\hat{a}_2$ with some vector of transfers, as asserted.

Now consider the induction step. So suppose we have proved the claim for all mechanisms such that the depth of $\Gamma(s_2)$ is less than or equal to $k-1$ and consider a mechanism such that $\Gamma(s_2)$ has depth k. Suppose player 2 moves first. An argument similar to the one above for the depth 1 case shows that if a subgame perfect equilibrium has outcome $\hat{a}_2$ and some vector of transfers t, then it must be true that each choice available to player 2 leads to a subgame in which every subgame perfect equilibrium has outcome $\hat{a}_2$ and some vector of transfers. By the induction hypothesis, this remains true in state s_1, so every initial choice available to player 2 in $\Gamma(s_1)$ leads to a subgame in which every subgame perfect equilibrium has an outcome of $\hat{a}_2$ and some vector of transfers. It readily follows that every subgame perfect equilibrium of $\Gamma(s_1)$ has such an outcome. Thus we have proved the claim for this mechanism if player 2 moves first. A similar argument applies if player 3 moves first.

So suppose player 1 moves first. Since there is a subgame perfect equilibrium with outcome $\hat{a}_2$ and transfers t in $\Gamma(s_2)$, player 1 has a cheap talk message available that leads to a subgame with a subgame perfect equilibrium that has this outcome in state s_2. By the induction hypothesis, every subgame perfect equilibrium of this subgame will have an outcome of $\hat{a}_2$ and some vector of transfers $\hat{t}$ in state s_1. Using the same reasoning as in the depth 1 case, the bound on transfers implies that this is better for player 1 than any feasible outcome with $\hat{a} = \hat{a}_1$. Hence in state s_1, 1's optimal choice must lead to an outcome with $\hat{a} = \hat{a}_2$. ▌

E.3 Proof of Claim 3

Fix any multistage mechanism Γ such that for *every* history h, $g(h) = (\hat{a}, 0, \ldots, 0)$ for some $\hat{a} \in \hat{A}$. That is, there are no transfers on any history. Let $\Sigma_i(s_k)$ denote the set of strategies for player i in $\Gamma(s_k)$. We show that Γ cannot implement f. The proof works by showing that if $\Gamma(s_2)$ has a subgame perfect equilibrium with outcome $\hat{a}_2$, then $\Gamma(s_1)$ has a subgame perfect equilibrium with outcome $\hat{a}_2$ as well. The proof is by induction on the depth of the game $\Gamma(s_2)$.

First, suppose that the depth of $\Gamma(s_2)$ is 1. Let $\sigma = (\sigma_1, \ldots, \sigma_n)$ be a subgame perfect equilibrium of $\Gamma(s_2)$ with outcome $\hat{a}_2$. We claim that $\hat{a}_2$ is a subgame perfect equilibrium outcome of $\Gamma(s_1)$ as well. To see this, define a profile of strategies $\hat{\sigma}$ in $\Gamma(s_1)$ as follows. At the first stage of the game, each player i plays the action σ_i. (Note that $\Sigma_i(s_2) \subseteq \Sigma_i(s_1)$ for all i, so this is feasible.) Let $b = (b_1, .., b_n)$ be some profile of actions that is played in the first stage. If b is a profile of actions which is feasible in $\Gamma(s_2)$, the game must terminate when b is played since the depth of $\Gamma(s_2)$ is 1. So suppose $\Gamma(s_1)$ does not terminate when b is played. Then b must not be feasible in s_2, implying that player 1's action, b_1, involved presentation of a proof that the state is s_1. Let $\Gamma^b(s_1)$ denote the subgame that is the continuation of $\Gamma(s_1)$ after the play of b and let τ_b be any subgame perfect equilibrium of $\Gamma^b(s_1)$. For any history h within the subgame $\Gamma^b(s_1)$, let $\hat{\sigma}(b, h) = \tau(h)$.

We now show that $\hat{\sigma}$ is a subgame perfect equilibrium of $\Gamma(s_1)$. Obviously, the outcome when $\hat{\sigma}$ is played is $\hat{a}_2$, so this will establish our claim for depth 1. Since the restriction of $\hat{\sigma}$ to every subgame of $\Gamma(s_1)$ that starts at the second stage of the game is a subgame perfect equilibrium of the subgame, the only thing we need to show is that no player i can gain from a unilateral deviation in the first stage of the game.

To see that this holds, note that player 1 gets his best possible outcome, $\hat{a}_2$, when $\hat{\sigma}$ is played so clearly he cannot gain from deviating. For any player $i \neq 1$, any action $b_i \neq \sigma_i$ that he can play at the first stage of $\Gamma(s_1)$ is an action that he can play in the game $\Gamma(s_2)$ as well. Hence the fact that $\Gamma(s_2)$ has depth 1 implies that any such action leads to the termination of the game $\Gamma(s_1)$ with an outcome that is identical to the outcome in $\Gamma(s_2)$ when i deviates to b_i. Since σ is an equilibrium in $\Gamma(s_2)$, we conclude that player i cannot gain from deviating. Hence $\hat{\sigma}$ is a subgame perfect equilibrium of $\Gamma(s_1)$ with outcome $\hat{a}_2$.

So suppose that the claim has been proved for every mechanism Γ' such that the depth of the game $\Gamma'(s_2)$ is smaller than K. Let Γ be a mechanism such that the depth of $\Gamma(s_2)$ is K and let σ be a subgame perfect equilibrium of $\Gamma(s_2)$ with outcome $\hat{a}_2$. We will show that $\Gamma(s_1)$ has a subgame perfect equilibrium $\hat{\sigma}$ with outcome $\hat{a}_2$.

For each profile of actions b which can be played in the first stage of $\Gamma(s_2)$ and does not terminate the game, let Γ^b denote the game form that is the continuation of Γ following a play of b in the first stage. Γ^b is a mechanism such that $\Gamma^b(s_2)$ has depth less than K. Let σ^b denote the profile of strategies in $\Gamma^b(s_2)$ that is induced by σ. Since σ is a subgame perfect equilibrium of $\Gamma(s_2)$, σ^b is a subgame perfect equilibrium of $\Gamma^b(s_2)$. If the outcome under σ^b is $\hat{a}_2$, then the induction hypothesis implies that $\Gamma^b(s_1)$ has a subgame perfect equilibrium, say $\tilde{\sigma}^b$, with outcome $\hat{a}_2$.

With this in mind, we construct a subgame perfect equilibrium $\hat{\sigma}$ of $\Gamma(s_1)$ with outcome $\hat{a}_2$ as follows. In the first stage of the game, each player i chooses his action according to σ_i. Let b be a profile of first stage actions which does not terminate the game. If b is feasible in state s_2 and if the outcome of $\Gamma^b(s_2)$ under σ^b is $\hat{a}_2$, then let $\hat{\sigma}$ on $\Gamma^b(s_1)$ be the subgame perfect equilibrium $\tilde{\sigma}^b$. For any other b, choose any subgame perfect continuation strategies.

To see that the outcome under $\hat{\sigma}$ is $\hat{a}_2$, note that the first stage strategies under $\hat{\sigma}$ are the same as those under σ. Hence the resulting profile of actions b must either terminate the game with outcome $\hat{a}_2$ or lead to a subgame where the continuation equilibrium has outcome $\hat{a}_2$. To see that $\hat{\sigma}$ is a subgame perfect equilibrium of $\Gamma(s_1)$, note that the restriction of $\hat{\sigma}$ to every proper subgame in $\Gamma(s_1)$ is subgame perfect so we just need to show that no player i can gain from a deviation at the first stage of the game. The outcome $\hat{a}_2$ is the best outcome for player 1 so, clearly, he will not gain by deviating.

So consider a deviation in the first stage by any player $i \neq 1$ to an action b_i which has zero probability under $\hat{\sigma}_i$. Suppose this deviation leads to outcome $\hat{a}_1$ with positive probability. Note that i's deviation is also feasible in s_2 and that the players other than i are playing the same actions as they played in s_2 under σ. Hence i's deviation would lead to a positive probability of outcome $\hat{a}_1$ under σ in $\Gamma(s_2)$ as well. Since i prefers $\hat{a}_1$ to $\hat{a}_2$, this contradicts σ being a subgame perfect equilibrium of $\Gamma(s_2)$. ▌

F Mixed Strategies

Our discussion has focused on implementation in pure strategy subgame perfect equilibrium. In this section, we show that all our results carry over for the case where the solution concept is mixed strategy subgame perfect equilibrium.[30] First, it is easy to see

[30]The fact that the mixed equilibria include the pure equilibria can make implementation with mixed strategies easier or more difficult. More specifically, let f be a SCF and Γ a mechanism. It is possible that Γ implements f in mixed strategies but not in pure strategies (because there may exist only a non–pure equilibrium in $\Gamma(s)$ with outcome $f(s)$) and it is possible that Γ implements f in pure strategies but not in mixed strategies (because there may exist a non–pure equilibrium which induces an outcome

that Proposition 1 holds for mixed equilibria as well as pure. It is also easy to see that the proofs given earlier for Theorems 2 and 4 and Claims 1 and 3 apply to implementation both in pure and in mixed strategies.

Finally, we turn to the extension for the results that refer to implementation in perfect information games, namely, Theorems 1, 3, and 5 and Claim 2. For Theorems 1, 3, and 5, we claim that the mechanisms which were defined in Parts A, C, and D.2 of the Appendix implement the SCF f in mixed strategies as well. To see this, consider the mechanism Γ_f which is defined in the proof of Theorem 1 (the arguments for Theorems 3 and 5 are identical). We first note that the proof of Theorem 1 establishes that for every state s, there exists a subgame perfect equilibrium in the game $\Gamma_f(s)$ with outcome $f(s)$. It is easy to see that the argument which establishes that there is no pure subgame perfect equilibrium with a different outcome than $f(s)$ also implies that there is no mixed subgame perfect equilibrium which gives positive probability to an outcome different from $f(s)$. Specifically, if every player i claims a state c_i^A such that $M_i(s) = M_i(c_i^A)$, then, in equilibrium, it must be the case that no players challenge any other player and the outcome is $f(s)$. On the other hand, if player i chooses a claim c_i^A such that $M_i(s) \neq M_i(c_i^A)$, then the argument given in the proof shows that he will end up with an outcome that is inferior to $f(s)$. Hence in equilibrium, player i will not make such a claim with positive probability.

For Claim 2, it is easy to see that the proof of the claim in Part E.2 of the Appendix establishes that if Γ is a perfect information mechanism such that there exists a *pure* subgame perfect equilibrium of $\Gamma(s_2)$ with outcome $\hat{a}_2$ plus some transfers t, then every *pure* subgame perfect equilibrium of $\Gamma(s_1)$ has outcome $\hat{a}_2$ plus some transfers $\hat{t}$. Since $\Gamma(s_2)$ is a perfect information game, it has a pure strategy equilibrium. Therefore, if Γ implements f in a subgame perfect equilibrium in mixed strategies, then there is a pure subgame perfect equilibrium in $\Gamma(s_2)$ with an outcome $\hat{a}_2$ and some vector of transfers t. Hence there is a (pure) subgame perfect equilibrium with outcome $\hat{a}_2$ and a vector of transfers $\hat{t}$ in $\Gamma(s_1)$, so f cannot be implemented.

different than $f(s)$ at state s while all the pure equilibria have outcome $f(s)$).

References

[1] Abreu, D., and A. Sen, "Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition," *Journal of Economic Theory*, **50**, April 1990, 285–299.

[2] Abreu, D., and H. Matsushima, "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, **60**, September 1992, 993–1008.

[3] Alger, I., and A. Ma, "Moral Hazard, Insurance, and Some Collusion," *Journal of Economic Behavior and Organization*, **50**, February 2003, 225–247.

[4] Baliga, S., "Implementation in Economic Environments with Incomplete Information: The Use of Multi–Stage Games," *Games and Economic Behavior*, **27**, 1999, 173–183.

[5] Ben Porath, E., and B. Lipman, "Implementation and Partial Provability," Boston University working paper, April 2009.

[6] Bull, J., and J. Watson, "Evidence Disclosure and Verifiability," *Journal of Economic Theory*, **118**, September 2004, 1–31.

[7] Bull, J., and J. Watson, "Hard Evidence and Mechanism Design," *Games and Economic Behavior*, **58**, January 2007, 75–93.

[8] Deneckere, R. and S. Severinov, "Mechanism Design with Partial State Verifiability," *Games and Economic Behavior*, **64**, November 2008, 487–513.

[9] Dutta, B., and A. Sen, "Nash Implementation with Partially Honest Individuals," working paper, 2011.

[10] Fishman, M. and K. Hagerty, "The Optimal Amount of Discretion to Allow in Disclosure," *Quarterly Journal of Economics*, **105**, May 1990, 427–444.

[11] Forges, F. and F. Koessler, "Communication Equilibria with Partially Verifiable Types," *Journal of Mathematical Economics*, **41**, November 2005, 793–811.

[12] Forges, F. and F. Koessler, "Long Persuasion Games," *Journal of Economic Theory*, forthcoming.

[13] Glazer, J., and A. Rubinstein, "Debates and Decisions: On a Rationale of Argumentation Rules," *Games and Economic Behavior*, **36**, August 2001, 158–173.

[14] Glazer, J., and A. Rubinstein, "On Optimal Rules of Persuasion," *Econometrica*, **72**, November 2004, 1715–1736.

[15] Glazer, J., and A. Rubinstein, "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach," *Theoretical Economics*, **1**, December 2006, 395–410.

[16] Green, J., and J.-J. Laffont, "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, **53**, 1986, 447–456.

[17] Grossman, S., "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, **24**, 1981, 461–483.

[18] Grossman, S., and O. Hart, "Disclosure Laws and Takeover Bids," *Journal of Finance*, **35**, May 1980, 323–334.

[19] Healy, P., and L. Mathevet, "Designing Stable Mechanisms for Economic Environments," *Theoretical Economics*, forthcoming.

[20] Hurwicz, L., "On Infomationally Decentralized Systems," in Radner, R., and C. B. McGuire, eds., *Decision and Organization*, Amsterdam: North Holland, 1972.

[21] Hurwicz, L., E. Maskin, and A. Postlewaite, "Feasible Implementation of Social Choice Rules when the Designer does not Know Endowments or Production Sets," in Ledyard, J., ed., *The Economics of Informational Decentralization: Complexity, Efficiency, and Stability*, Boston: Kluwer Academic Publishers, 1995, 367–433.

[22] Jackson, M., "Bayesian Implementation," *Econometrica*, **59**, March 1991, 461–477.

[23] Kartik, N., and O. Tercieux, "Implementation with Evidence," Columbia University working paper, 2011.

[24] Lipman, B., and D. Seppi, "Robust Inference in Communication Games with Partial Provability," *Journal of Economic Theory*, **66**, August 1995, 370–405.

[25] Maskin, E., "Nash Equilibrium and Welfare Optimality," MIT working paper, 1977.

[26] Maskin, E., "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, **66**, January 1999, 23–38.

[27] Milgrom, P., "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, **12**, Autumn 1981, 380–391.

[28] Milgrom, P., and J. Roberts, "Relying on the Information of Interested Parties," *Rand Journal of Economics*, **17**, Spring 1986, 18–32.

[29] Moore, J., and R. Repullo, "Subgame Perfect Implementation," *Econometrica*, **56**, September 1988, 1191–1220.

[30] Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura, "Strategic Information Revelation," *Review of Economic Studies*, **57**, January 1990, 25–47.

[31] Osborne, M., and A. Rubinstein, *A Course in Game Theory*, MIT Press, 1994.

[32] Palfrey, T., and S. Srivastava, "Nash Implementation Using Undominated Strategies," *Econometrica*, **59**, March 1991, 479–501.

[33] Postlewaite, A., and D. Wettstein, "Feasible and Continuous Implementation," *Review of Economic Studies*, **56**, October 1989, 603–611.

[34] Repullo, R., "A Simple Proof of Maskin's Theorem on Nash Implementation," *Social Choice and Welfare*, **4**, 1987, 39–41.

[35] Seidmann, D., and E. Winter, "Strategic Information Transmission with Verifiable Messages," *Econometrica*, **65**, January 1997, 163–169.

[36] Sher, I., "Persuasion and Limited Communication," University of Minnesota working paper, April 2008.

[37] Shin, H. S., "The Burden of Proof in a Game of Persuasion," *Journal of Economic Theory*, **64**, October 1994, 253–264.