Hard Evidence and Mechanism Design*

Jesse Bull[†] and Joel Watson[‡]

July 2002, Revised November 2005

Abstract

This paper addresses how hard evidence can be incorporated into mechanism-design analysis. Two classes of models are compared: (a) ones in which evidentiary decisions are accounted for explicitly, and (b) ones in which the players make abstract declarations of their types. Conditions are provided under which versions of these models are equivalent. The paper also addresses whether dynamic mechanisms are required for Nash implementation in settings with hard evidence. The paper shows that static mechanisms suffice in the setting of "evidentiary normality" and that, in a more general environment, one can restrict attention to a class of three-stage dynamic mechanisms. *JEL* Classification: C70, D74, K10.

*The authors thank Jean-Jacques Laffont, Joel Sobel, Lars Stole, the anonymous associate editor, and the two anonymous referees for their generous comments. The paper originated as a section of Bull and Watson (2001).

[†]Department of Economics, Florida International University, Miami, Florida 33199. Email: bullj@fiu.edu.

[‡]Department of Economics, University of California, San Diego, La Jolla, California 92093-0508. Internet: http://weber.ucsd.edu/~jwatson/.

Many contractual and public-choice settings are analyzed using mechanism-design theory, which is a useful framework for studying the implications of informational asymmetries. The typical mechanism-design model comprises a set of states (representing the parties' information), a set of physical outcomes, and a specification of preferences over outcomes and states. The typical implementation exercise involves designing a game form, which induces a game to be played in each state. In essence, the game form specifies how the parties may send messages to each other and to an external enforcer, who will select the outcome on their behalf. The parties' rational behavior in each state implies an outcome function (the outcome as a function of the state).

In its standard "public choice" form, the mechanism-design framework abstracts from institutional and technological constraints beyond those that the modeler can represent in the definition of states, outcomes, and preferences. This abstraction can create a useful simplification, but it is not suitable for the analysis of some real constraints that we are interested in studying. For example, in many settings, parties can present *hard evidence*. The key features of hard evidence are that (a) whether to present evidence is an inalienable action, and (b) the existence of evidence depends on the state.

In this paper, we discuss how hard evidence can be incorporated into the mechanism-design framework. We model the parties' inalienable actions (evidence production) as distinct from an arbitrarily-designed message form. A mechanism specifies the message form and how the public action is a function of messages and evidence. Our work is thus along the lines of Myerson (1982,1991), whose mechanism-design analysis nicely distinguishes between inalienable individual and public actions.

Our main objectives are to (i) address the meaning and validity of the revelation principle, and (ii) compare the efficacy of static and dynamic mechanisms in the context of hard evidence. We focus on weak (Bayesian) Nash implementation, which is the appropriate concept for analysis of the revelation principle and is nontrivial in many important economic environments (in particular, contractual settings with renegotiation, as in Maskin and Moore 1999 and Segal and Whinston 2002).[1] On (ii), we find that dynamic mechanisms are essential for implementing some outcome functions; this result stands in stark contrast to the standard case without hard evidence (where the static/dynamic distinction is irrelevant for Nash implementation).[2] We examine both situations of incomplete and complete information.

Our analysis proceeds in four stages. First, we prove a "weak revelation result," which justifies focusing on mechanisms in which, in addition to their inalienable

[1]With renegotiation and quasilinear utility, weak Nash implementation is unique in payoffs. Weak Nash implementation is also nontrivial in settings with one player or where there is no "bad" public action that can be used effectively to punish the players; it is trivial when there are two or more players and there is a very inefficient public action that can be used to punish the players off the equilibrium path.

[2]We think it is worthwhile to understand these issues in the context of weak Nash implementation before examining strong/unique implementation.

actions about what evidence to present, the parties are instructed to simultaneously report their own types (and they report truthfully in equilibrium).

Second, we address the degree to which each party's behavior in a mechanism can be interpreted as an abstract "declaration of his type." In standard mechanism-design environments, this issue is resolved by the standard revelation principle, which justifies analyzing truthful reporting in direct-revelation game forms.[3] In settings with hard evidence, however, this issue is more complicated because evidence disclosure is inalienable and therefore cannot be freely translated into a direct-revelation form. Our "strong revelation result" provides an intuitive condition—called *evidentiary normality*—under which, for each party, there is a one-to-one mapping between types and report/evidence pairs that holds across all implementation exercises. For each type θ_i of party i, there is evidence $\overline{e}_i^{\theta_i}$ such that, when party i presents $\overline{e}_i^{\theta_i}$ and sends the message "θ_i," it means "party i declares his type to be θ_i" and this meaning is fixed over all implementation exercises.

Third, under evidentiary normality, we show how a setting with hard evidence can be translated into an "abstract-declaration model," in which each party simply declares his type but is limited in what he can say. We provide an equivalence result for the two models.

Fourth, we address whether there is a difference between what dynamic mechanisms and static mechanisms can Bayesian-Nash implement. In static mechanisms, all messages and evidence disclosure occur simultaneously, whereas in dynamic mechanisms these actions occur sequentially. We present an example of an outcome function that can only be implemented using a dynamic mechanism; this is an interesting result because, since Nash implementation is a static concept, dynamic mechanisms are not needed for Nash implementation in conventional models. However, we prove that static mechanisms suffice under evidentiary normality. We also show that when evidentiary normality does not hold, one can constrain attention to simple three-stage dynamic mechanisms in which the players first send private messages to the external enforcer, then they receive private messages, and then the players disclose evidence.

Our modeling exercise amounts to a reformulation of Green and Laffont's (1986) "limited verifiability" analysis. These authors were the first to study the abstract-declaration model.[4] The contribution of our approach beyond that of Green and Laffont (1986) is threefold. First, we provide a link between real evidence and abstract declarations, which clarifies when the abstract-declaration environment has an intuitive interpretation. Part of this link involves justifying a fixed action space for mechanism design. Second, we lend support to Green and Laffont's "nested range condition" by arguing that evidentiary normality (the equivalent in our context) is

[3] See Dasgupta et al. (1979), for example.

[4] Others that examine settings with partial verifiability include Milgrom and Roberts (1986), Hart and Moore (1988), Okuno-Fujiwara et al. (1990), Shin (1994), Lipman and Seppi (1995), and Seidman and Winter (1997). Bull (2001a) and Deneckere and Severinov (2001) examine moderate costs of evidence production.

commonly satisfied in reality. Third, we extend the analysis to settings with more than one player. We find that evidentiary normality justifies studying the abstract-declaration model, but, without evidentiary normality, it is not appropriate to focus on static mechanisms.

Our analysis is related to that of Lipman and Seppi (1995). In Lipman and Seppi's model, several agents jointly observe a state and individually send messages to a receiver, who then takes an action. The senders are constrained to the same state-contingent set of messages. Lipman and Seppi consider the design of dynamic game forms in which the players send messages sequentially. In our analysis of dynamic mechanisms, we perform a similar exercise to determine "standard form" mechanisms that are sufficient for implementation in various settings.[5]

Also closely related is the concurrent work of Deneckere and Severinov (2001) and Forges and Koessler (2003). These papers study aspects of mechanism design with limits on communication. They are motivated differently than is our study (they take different perspectives and focus on different settings), but their analysis and results significantly overlap with ours. In the Conclusion, we discuss in detail the relation between these papers and the modeling exercise presented herein.

The next Section describes the standard public-action mechanism-design model, to which hard evidence is added in Section 2. The characterization results for static implementation are presented in Section 3 and the relation to models with abstract declarations is explored in Section 4. Dynamic mechanisms are the focus of Section 5. Section 6 offers concluding remarks and a discussion of related papers. The Appendix contains proofs that do not appear in the body of the paper, with the exception of Theorems 1 and 2 whose simple proofs are omitted (they follow the standard method of establishing the basic revelation principle).

1 The Standard Mechanism-Design Model

We consider a public-choice setting with n players, where a *public action* p is taken by an external enforcer. Let P be the set of feasible public actions, which we also call *outcomes*. The players' preferences over public actions depend on the *state* θ, which is a vector of types for all of the players; we write $\theta = (\theta_1, \theta_2, \ldots \theta_n)$, where θ_i is player i's type, for $i = 1, 2, \ldots, n$. We let Θ denote the set of feasible states and we assume this is finite. The set of feasible types for player i is denoted Θ_i, so $\Theta \subset \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$. In state θ, player i observes his own type θ_i and has a belief about the types of the other "$-i$" players that is described by the probability distribution $\mu_{-i}(\cdot \mid \theta_i)$, where $\mu_{-i}(\theta_{-i} \mid \theta_i)$ is the probability that player i believes the other players have type profile θ_{-i}. The external enforcer does not observe the state.

[5] Lipman and Seppi (1995) focus on provability assumptions that imply "robust inference" (meaning that, with a particular mechanism, the receiver can be sure to learn about the state regardless of the agents' preferences). Their objectives are thus tangential to ours; they also use a stronger notion of implementation across a wider range of preferences than are appropriate for our purposes.

Preferences are represented by the utility function $u : P \times \Theta \to \mathbf{R}^n$. We write $u(p, \theta)$ as the payoff vector of public action p in state θ. Player i's payoff is $u_i(p, \theta)$.

This treatment of types and beliefs allows for players having a common prior, but does not require it. A special case that we will specifically consider is that of complete information, where the state is commonly observed by the players. For settings of complete information, we will conveniently redefine (or, as is said, abuse) notation and write $\Theta_i = \Theta$ for each i; in this case, we also have $\mu_{-i}(\theta, \theta, \ldots, \theta \mid \theta) = 1$. Also, when analyzing settings with one player, we write $\Theta_1 = \Theta$.

An *outcome function* $g : \Theta \to P$ associates an outcome with each state. The set of implementable outcome functions are those that can be induced via equilibrium play of fixed mechanisms. A mechanism is a game form, specifying a game tree with actions for the players and outcomes (from P) at terminal nodes. For a given state θ, the game form and $u(\cdot, \theta)$ define the game played in state θ.

In a typical contractual application, the players are contracting parties and the external enforcer is a court or other authority. The public action represents transfers and other actions that the court compels. The state signifies an event that occurs during the contractual relationship, such as specific investments or some random draw. The contract that the parties form specifies how they are to interact with the court and how the court should respond, following the realization of the state. Thus, the contract specifies the mechanism. The parties design the mechanism to implement an outcome function of their choice.

2 Incorporating Hard Evidence

The standard mechanism-design framework imposes no structure other than what is given by the set of states, the set of outcomes, and the players' preferences. To model hard evidence, however, some additional structure must be assumed. To see why this is so, consider an example.

Suppose there are two states, Y and N. State Y signifies that player 1 has found a particular rare coin while on an archaeological expedition, whereas state N signifies that he has not found the coin. In other words, in state Y player 1 possesses the coin, whereas in state N he does not have it. Suppose further that, once the state is realized, player 1 can present any of his possessions to the external enforcer for inspection. However, there is no way for the external enforcer to search every possible place where player 1 may have the coin; player 1 can easily hide or dispose of it. The coin can be disclosed by player 1, if he has it, or not disclosed.

There are two important aspects of this hard evidence (the coin) that must be incorporated into the mechanism-design analysis. First, we have the fact that the coin can be disclosed in state Y but not in state N. Second, whether to disclose the coin is player 1's *inalienable* action.[6]

[6] Other recent examples of models in which there are inalienable decisions within the mechanism

To incorporate the hard evidence, we view player 1's inalienable action as something that cannot be tampered with in the design of a mechanism. In other words, the mechanism should be constrained to explicitly include player 1's choice of whether to disclose the coin. There are then state-contingent limits on the actions that player 1 can take. In particular, in state N, player 1 is not allowed to pick the "disclose the coin" action. The game form may include other actions, such as messages sent to the court. The public action is a function of these messages and whether player 1 discloses the coin.[7]

Returning to the general framework, we suppose that each player takes an inalienable *evidentiary action* (which we also call *evidence production*). Specifically, player i chooses e_i from a set E_i. We assume that evidence production is costless; that is, the player's evidentiary actions are not directly payoff-relevant. Thus, we do not need to include evidence in the definition of the "outcome." We capture evidentiary limits as type-contingent constraints on each player's ability to produce evidence. Specifically, player i of type θ_i is constrained to select e_i from some nonempty set $E_i^{\theta_i} \subset E_i$. These subsets define the *evidentiary structure*. Define

$$E^{\theta} \equiv \{(e_1, e_2, \ldots, e_n) \mid e_i \in E_i^{\theta_i} \text{ for each } i = 1, 2, \ldots, n.\}$$

and let $E \equiv \bigcup\{E^{\theta} \mid \theta \in \Theta\}$ be the set of possible evidence profiles.

In practice, evidence may take the form of *documents* that the players can disclose. We use the word "document" broadly to include papers, such as canceled checks or bills of sale, as well as any other physical objects or tasks that players can disclose or perform. This interpretation implies extra analytical structure whereby e_i represents a collection of documents that player i presents. Formally, then, e_i is a subset of some grand set of documents D_i. The state determines the available sets of documents for each player. That is, for every type θ_i of player i, $E_i^{\theta_i}$ is some subset of D_i, and E_i is a subset of the power set of D_i. Bull and Watson (2004) models evidence in this way and studies a complete information setting in which public actions are transfers.

We are particularly interested in evidentiary structures with the following property.

Definition 1: *The evidentiary structure is called* **normal** *if, for each player i and every type θ_i, there is an evidentiary action $\overline{e}_i^{\theta_i} \in E_i^{\theta_i}$ with the following property. For every $\theta_i' \in \Theta_i$, if $\overline{e}_i^{\theta_i} \in E_i^{\theta_i'}$ then $E_i^{\theta_i} \subset E_i^{\theta_i'}$.*

include Brusco (2000) and Evans (2002).

[7]An alternative way of incorporating the hard evidence is to include *in the definition of the outcome* whether the coin is disclosed. To capture the idea that the coin cannot be disclosed in state N, we could assume that disclosure of the coin in state N gives player 1 an arbitrarily large negative payoff. To capture that whether to disclose is player 1's inalienable action, we then must restrict the class of allowable game forms to those in which disclosure of the coin (a component of the outcome) only depends on player 1's actions. Further, player 1 must have a "nondisclosure" action. We prefer to simply model evidentiary actions as individual, inalienable actions that the players take. For more on the difference between modeling an action as public and individual, see Watson (2003).

Normality means that there is "maximal evidence" available to each type, which is sufficient to use in possibly distinguishing this type from others. For a given type θ_i of player i, we call $\overline{e}_i^{\theta_i}$ the *maximal evidentiary action* (though it need not be unique). Maximal evidence disclosure by all players in a given state θ implies that $\overline{e}^{\theta} \equiv (\overline{e}_1^{\theta_1}, \overline{e}_2^{\theta_2}, \ldots, \overline{e}_n^{\theta_n})$ is disclosed. Though our focus is different than that of Lipman and Seppi (1995), normality is essentially equivalent to their full reports condition.[8]

If one considers evidence in terms of documents, as briefly discussed above, a sufficient (but not necessary) condition for normality is that each player is freely able to present any combination of documents in his possession. When player i can present any combination of his existing documents, we can define $\overline{e}_i^{\theta_i}$ to be $D_i^{\theta_i}$; in words, player i's maximal evidence entails disclosing all of the documents he has.[9]

We first study static mechanisms, in which all players' evidence and messages are submitted simultaneously. In Section 5, we discuss conditions under which implementation relies on the use of sequential mechanisms. In settings with hard evidence, a static mechanism has two components. First, it prescribes an arbitrary message-game form to be played by the players simultaneously with their inalienable actions as to what evidence to produce. Specifically, the mechanism specifies message spaces $M_1, M_2, \ldots, M_n$. Simultaneously and independently, the players choose messages and evidence, with player i selecting a message from M_i and an evidentiary action from E_i. We use the term *action* (with no qualifier) to describe a player's choice of a message and evidentiary action, and we write

$$a_i = (m_i, e_i) \in A_i \equiv M_i \times E_i.$$

We write the players' action profile as an element of $A \equiv M \times E$, where $M \equiv M_1 \times M_2 \times \cdots \times M_n$. The second component of a mechanism is a mapping $f : A \to P$, which prescribes the public action for each profile of messages and evidence. We represent a mechanism as (M, f).

The mechanism implies what we call the *message and disclosure game*, which is a Bayesian game with type-contingent restrictions on actions. In state θ, player i's

[8] To see the equivalence of normality to Lipman and Seppi's full report condition, here is their definition extended to the setting in which agents may possess different hard evidence. For every θ, there is evidence $\overline{e}_i^{\theta_i} \in E_i^{\theta_i}$ such that $T_i(\overline{e}_i^{\theta_i}) = \cap_{e_i \in E_i^{\theta_i}} T_i(e_i)$, where $T_i(e_i) \equiv \{\theta_i \in \Theta_i \mid e_i \in E_i^{\theta_i}\}$. The interpretation of T_i is that it is the set of types of player i for which evidence e_i is feasible. In this sense, it is the information content of evidence e_i. The interpretation of the condition is that for every type, there is a single evidentiary action that summarizes what all available evidentiary actions for this type would imply.

[9] As noted above, it is not necessarily the case that each type has a unique maximal evidentiary action. For example, let $\Theta_1 = \{X, Y\}$ and suppose player 1's set of documents is $D_1 = \{Q, R\}$. Suppose neither of these documents exists in state X, so that $E_1^X = \{\emptyset\}$; suppose that both exist and can be presented in state Y, so that $E_1^Y = \{\emptyset, \{Q\}, \{R\}, \{Q, R\}\}$. Note that, in state Y, player 1 can disclose Q, R, neither, or both. This is a normal evidentiary environment, with $\overline{e}_1^X = \emptyset$ and $\overline{e}_1^Y$ chosen as *any* element of $E_1^Y \setminus \emptyset$. To see that the ability to disclose any combination of documents is not necessary for normality, observe that, in this example, we could have specified $E_1^Y = \{\emptyset, \{Q\}, \{R\}\}$ and still have normality.

action space is restricted to $A_i^{\theta_i} \equiv M_i \times E_i^{\theta_i}$; that is, although he can send any message from M_i, he can only produce feasible evidence. We denote player i's strategy by $\alpha_i : \Theta_i \to A_i$, where $\alpha_i(\theta_i) \in M_i \times E_i^{\theta_i}$ is assumed for each θ_i. A strategy profile is denoted by α and can be written as a function $\alpha : \Theta \to A$. We also write α_{-i} as the profile of strategies for the players other than player i. Payoffs in state θ, as a function of the action profile, are given by u. Thus, if the action profile is $a \in A$, then the payoff vector is $u(f(a), \theta)$.

We study (Bayesian) Nash implementation.[10] A strategy profile α is a Nash equilibrium of the message and disclosure game if, for every player i and each type $\theta_i \in \Theta_i$, $\alpha_i(\theta_i)$ solves

$$\max_{a_i \in A_i^{\theta_i}} \sum_{\theta_{-i} \in \Theta_{-i}(\theta_i)} \mu_{-i}(\theta_{-i} \mid \theta_i) u_i(f(a_i, \alpha_{-i}(\theta_{-i})), (\theta_i, \theta_{-i})),$$

where $\Theta_{-i}(\theta_i) \equiv \{\theta_{-i} \mid (\theta_i, \theta_{-i}) \in \Theta\}$.

Definition 2: *A mechanism (M, f) is said to* **implement** *outcome function g if the implied message and disclosure game has a Bayesian Nash equilibrium α such that $f(\alpha(\theta)) = g(\theta)$ for every state $\theta \in \Theta$. An outcome function is said to be* **implementable** *if there is a mechanism that implements it.*[11]

3 Characterization Results

This section addresses two key issues: (1) the extent to which mechanism-design analysis can be simplified by restricting attention to a particular class of mechanisms and equilibrium behavior, and (2) whether there is a mapping between states and actions that allows us to interpret—in a way that is uniform across implementation exercises—each player's message and evidence production as an abstract "declaration of his type."

Our first result is a weak revelation result; it justifies constraining attention to the following type of implementation.

[10] We focus on pure strategy equilibrium. It is straightforward to include mixed strategy equilibrium in the analysis, but it would require extra notation and the consideration of public randomization. Our analysis leans heavily on the assumption that evidence is costless to disclose when it exists. In many real cases, evidence production entails a moderate (monetary or psychic) cost. If the cost is relatively small, our model is probably a good approximation. In general, we think it is important to distinguish between (a) existence of evidence, (b) costs of producing evidence, and (c) any exogenous constraints on mechanisms. In our modeling exercise, (b) and (c) are assumed away.

[11] Recall our comments in the Introduction regarding the justification for studying the "weak" version of implementability. Mookherjee and Reichelstein (1990) study a variant of the revelation principle for strong implementation in a standard mechanism-design setting.

Definition 3: *We say that outcome function g is* **implementable with direct and truthful messages** *if there is a mechanism (M, f) with the following properties. First, $M_i = \Theta_i$ for each i. Second, the implied message and disclosure game has a Bayesian Nash equilibrium α such that $f(\alpha(\theta)) = g(\theta)$ for every state $\theta \in \Theta$ and, for each player i and each type θ_i, it is the case that $\alpha_i(\theta_i) = (\theta_i, e_i^{\theta_i})$ for some $e_i^{\theta_i} \in E_i^{\theta_i}$.*

In other words, the mechanism has "direct-revelation" message spaces—where $M_i = \Theta_i$ for every player i—and implementation is achieved with an equilibrium in which each player's message is a truthful report of his realized type.

Theorem 1 (Weak Revelation Result): *If an outcome function is implementable then it is implementable with direct and truthful messages.*

This theorem is a version of Myerson's (1982) Proposition 2, in the context of state-contingent constraints on individual actions.[12] The theorem shows that, in the implementation of a given outcome function, there is a one-to-one relation between states and each player's equilibrium action. Thus, player i's action $(\theta_i, e_i^{\theta_i})$ can be interpreted as "player i declares that his type is θ_i." However, Theorem 1 does not guarantee that the meanings of specific actions are uniform over the implementation of various outcome functions. In other words, when implementing g the action that means "player i declares that his type is θ_i" may be a different action than the one with this meaning when implementing g'. Hence, we call Theorem 1 the Weak Revelation Result.

We are then led to ask whether, in general, there is a way of limiting ourselves to a fixed meaning of evidence. The answer is "no," as the following example demonstrates. Consider the setting in which $n = 1$, $\Theta = \{X, Y, Z\}$, $P = \{\underline{p}, \overline{p}\}$, and $E_1 = \{Q, R\}$. Suppose payoffs are given by $u(\underline{p}, \theta) = 0$ and $u(\overline{p}, \theta) = 1$, for each $\theta \in \Theta$. Feasible evidence is given by $E_1^X = \{Q\}$, $E_1^Y = \{Q, R\}$, and $E_1^Z = \{R\}$. In words, player 1's evidentiary action must be Q in state X, he can present evidence Q or R in state Y, and he must choose R in state Z. Note that this evidentiary structure is not normal.

In this example, one can implement the outcome function g defined by $g(X) = g(Y) = \overline{p}$ and $g(Z) = \underline{p}$. Further, it is not difficult to confirm that this can only be accomplished if player 1 produces evidence Q in state Y; otherwise, in state Z he could pretend to be in state Y and would get the preferred outcome $\overline{p}$. Likewise, we can implement the function g' defined by $g'(X) = \underline{p}$ and $g'(Y) = g'(Z) = \overline{p}$, but doing so requires giving player 1 the incentive to produce evidence R in state Y.

In summary, implementation of g requires associating evidentiary action Q with state Y, whereas implementation of g' requires associating R with state Y. In the first case, production of Q is integral to the abstract statement "player 1 declares the

[12] The model can be translated into Myerson's formulation by modeling evidentiary constraints in utility terms (as described in footnote 7) and by defining a new public action that maps E to P.

state to be Y." In the second case, production of R has this interpretation. Thus, in general, evidence can have no fixed meaning.

However, we next show that a fixed meaning always exists if the evidentiary structure is normal. Recall that, under a normal evidentiary structure, there is maximal evidence in every state: $\overline{e}^{\theta}$. We define $\overline{a}_i^{\theta_i} \equiv (\theta_i, \overline{e}_i^{\theta_i})$ and we write $\overline{a}^{\theta} = (\overline{a}_1^{\theta_1}, \overline{a}_2^{\theta_2}, \ldots, \overline{a}_n^{\theta_n})$, for every θ. Similarly, $\overline{\alpha}(\theta) \equiv (\theta, \overline{e}^{\theta})$ for every θ.

Definition 4: *Assume that the evidentiary structure is normal. We say that outcome function g is* **implementable with direct and truthful messages and with maximal evidence production** *if there is a mechanism (M, f) with the following properties. First, $M_i = \Theta_i$ for each i. Second, $\overline{\alpha}$ is a Bayesian Nash equilibrium of the implied message and evidence game. Third, $f(\overline{\alpha}(\theta)) = g(\theta)$ for every state θ.*

Theorem 2 (Strong Revelation Result): *Assume that the evidentiary structure is normal. If an outcome function is implementable then it is implementable with direct and truthful messages and with maximal evidence production.*

The proof of this theorem is a constructive argument, whereby a given mechanism and equilibrium behavior is mapped into another mechanism and maximal evidence production. Normality comes into play when showing that any deviation in the latter setting is associated with a feasible deviation in the former setting, which enables the definition of the new mechanism so that $\overline{\alpha}$ is an equilibrium.

The Strong Revelation Result ensures that there is a meaningful and uniform notion of "truthful declaration." Player i declares his type to be θ_i by choosing action $\overline{a}_i^{\theta_i}$, that is, by sending message "θ_i" and providing evidence $\overline{e}_i^{\theta_i}$. Because we can restrict attention to equilibria in which type θ_i of player i selects $(\theta_i, \overline{e}_i^{\theta_i})$ in the message and disclosure game, this action can always be interpreted to mean "player i declares that his type is θ_i." This meaning can be *fixed* across all implementation exercises.

4 Models with Abstract Declarations

In this section, we investigate whether one can translate a setting of mechanism-design with hard evidence into a mechanism-design setting in which each player simply declares his type but faces type-contingent constraints on the declaration he can make. We study a model in which the abstract message "θ_i" takes the place of the real message and evidence $(\theta_i, \overline{e}_i^{\theta_i})$.

We start by defining an *abstract-action model* as follows. For each player i, there is a *fixed* set of actions denoted A_i; this set is fundamental and cannot be designed as in the model of the previous sections. The players are constrained in that, given his type θ_i, player i's action is limited to be from some nonempty set $A_i^{\theta_i} \subset A_i$. Letting

$$A \equiv \{(a_1, a_2, \ldots, a_n) \mid \text{there is a state } \theta \text{ such that } a_i \in A_i^{\theta_i} \text{ for each } i\},$$

a mechanism thus specifies a function of the form $f : A \rightarrow P$. The players play a game in which they simultaneously select feasible actions and payoffs are given by $u(f(\cdot), \theta)$. An outcome function g is implementable if there is a function $f : A \rightarrow P$ such that the implied game has an equilibrium α with $f(\alpha(\theta)) = g(\theta)$ for every $\theta \in \Theta$.

The abstract-action model is precisely Green and Laffont's (1986) general model, with two exceptions. First, and mainly rhetorical, we suggest no interpretation of the abstract actions.[13] Green and Laffont named some of the player's actions as direct reports of his type and they assumed that a player can always report his realized type ($\theta_i \in A_i^{\theta_i}$ for all $\theta_i \in \Theta_i$). Second, our model is defined for an arbitrary number of players, whereas Green and Laffont focused on the case of one player.

The only difference between the abstract-action model and our model of hard evidence is that, in the abstract-action model, the set A_i is fixed, whereas, in the model developed in Section 2, A_i is endogenous to a particular implementation exercise (because M is arbitrary). Recognize that, unless some version of the revelation principle holds, there is no justification for constraining attention to a *fixed* set of actions for mechanism design. Thus, as Green and Laffont recognized at the outset of their modeling exercise, adopting an abstract-action model amounts to an assumption that is not necessarily innocuous. However, our Weak Revelation Result implies that any setting with hard evidence can be translated into an equivalent abstract-action model, where one simply defines $A_i \equiv \Theta_i \times E_i$ and $A_i^{\theta_i} \equiv \Theta_i \times E_i^{\theta_i}$, for each i and every θ_i. That is, one can focus on these fixed action spaces. Thus, we conclude that Green and Laffont's (1986) modeling approach is general and always valid, *within the class of static mechanisms and subject to the interpretive issues just noted.*

A special case of the abstract-action model, which we call the *abstract-declaration model*, has $A_i \equiv \Theta_i$ for all i. In words, this model has each player simply reporting his type to the external enforcer. The abstract-declaration model is the most basic one studied by Green and Laffont (1986). Any setting of mechanism design with hard evidence that has a normal evidentiary structure can be translated into an abstract-declaration model as follows. Let the sets $E_i^{\theta_i}$ (for each player i and every type θ_i) be given by any setting with hard evidence. By normality of the evidentiary structure, for each player i and every state θ, we have an element $\overline{e}_i^{\theta_i} \in E_i^{\theta_i}$ such that $\overline{e}_i^{\theta_i} \in E_i^{\theta'_i}$ if and only if $E_i^{\theta_i} \subset E_i^{\theta'_i}$. We define

$$A_i^{\theta_i} \equiv \{\theta'_i \mid \overline{e}_i^{\theta'_i} \in E_i^{\theta_i}\},$$

for each i and every $\theta_i \in \Theta_i$. In words, given that he is type θ_i player i can declare his type to be θ'_i if and only if all of the evidence available to type θ'_i is also available to type θ_i. Note that this construction satisfies Green and Laffont's (1986) assumption that $\theta_i \in A_i^{\theta_i}$ for all θ_i. Also note that this translation is not generally possible

[13]For this reason, we refer to this as an "abstract-declaration model" instead of a revelation mechanism. As noted above, this differs from the more standard revelation mechanism in that each player's feasible actions depend on his type.

without a normal evidentiary structure.[14]

Our next result establishes the connection between a complete-information setting with hard evidence and its corresponding abstract-declaration model. Let α^* denote the strategy profile in which the players report truthfully ($\alpha^*(\theta) = \theta$ for each θ).

Definition 5: *With reference to a given abstract-declaration model, we say that outcome function g is* **implementable with truthful reporting** *if there is a function $f : \Theta \to P$ such that α^* is an equilibrium of the revelation game and $f(\theta) = g(\theta)$ for every state θ.*

We have the following special result for settings with complete information and more than one player.

Theorem 3: *Take as given any complete-information setting with hard evidence that has a normal evidentiary structure and $n \geq 2$. Consider its translation into an abstract-declaration model. An outcome function is implementable in the setting with hard evidence if and only if it is implementable in the abstract-declaration model. Furthermore, if an outcome function is implementable in the abstract-declaration model then it is implementable with truthful reporting.*

This theorem states that, as long as the underlying setting with hard evidence has a normal evidentiary structure and complete information, we are justified in analyzing truthful reporting in the corresponding abstract-declaration model. The abstract-declaration model is generally simpler in the sense that direct reports of each player's type take the place of possibly complicated descriptions of evidence.

Note that the meaning of abstract actions is closely tied to the evidentiary structure. If the evidentiary structure is normal, and assuming $n \geq 2$, then one can analyze an abstract-declaration model with a fixed interpretation of a player "declaring his type."[15] On the other hand, if the evidentiary structure is not normal, then abstract actions have no fixed meaning and there is no reason to name them after the types. Thus, although the abstract-action model is always justified, it is not always meaningful or insightful to call a particular action "θ_i" and to assume that $\theta_i \in A_i^{\theta_i}$, as Green and Laffont (1986) do.

[14] Further, no alternative translation is guaranteed sufficient for determining all implementable outcome functions. As the example in Section 3 demonstrates, without normality a single type of player may need to use different evidence to separate himself from other types in the implementation of different outcome functions. Mapping these different evidentiary actions to the same abstract action yields a model that over-represents what can be implemented.

[15] Furthermore, our analysis shows that Green and Laffont's (1986) generalized revelation result (for the abstract-action model, extended to the case of $n \geq 2$) is in some sense redundant because, whenever their general nested range condition holds, one can translate the abstract-action model into an equivalent abstract-declaration model. Specifically, one can think of arbitrary abstract actions as evidentiary actions. The generalized nested range condition then implies normality, allowing the translation.

The second claim of Theorem 3 (regarding truthful reporting) is the straightforward n-player extension of Green and Laffont's (1986) main revelation result. It is not difficult to see that normality of the evidentiary structure implies Green and Laffont's "nested range condition" in the abstract-declaration model. The nested range condition holds when, for each player i and any two types $\theta_i, \theta_i' \in \Theta_i$, $\theta_i' \in A_i^{\theta_i}$ implies $A_i^{\theta_i'} \subset A_i^{\theta_i}$. Green and Laffont show that the claim holds if and only if the nested range condition is satisfied.

Regarding the first claim of Theorem 3, the "sufficiency" ("if") direction is easily proved. The "necessity" ("only if") direction requires a more elaborate argument, because there are, in a sense, more ways to deviate in the setting with hard evidence than there are in the abstract-declaration model. The key is to separate the possible actions in the hard-evidence setting into two groups, one comprising actions that translate into bona fide type-declarations (which may be dishonest) and another comprising gibberish. When a player deviates with gibberish, his deviation is ignored in favor of the other players' declarations. This construction relies on there being at least two players and complete information. In particular, complete information is required so that, in the event that one player makes a "gibberish" declaration, all of the necessary information about the desired outcome is contained in the other players' declarations.

The conclusion of Theorem 3 is not valid in settings of incomplete information or with one player, as the following example indicates. Suppose that $\Theta_1 = \{X, Y\}$, $E_1 = \{Q, R, S\}$, $E_1^X = \{Q, R\}$, $E_1^Y = \{R, S\}$, and $P = \{\underline{p}, \overline{p}\}$; further, suppose that type X of player 1 strictly prefers outcome $\underline{p}$ and that type Y strictly prefers $\overline{p}$. Note that the evidentiary setting is normal, with $\overline{e}_1^X = Q$ and $\overline{e}_1^Y = S$. The associated abstract-declaration model has $A_1^X = \{X\}$ and $A_1^Y = \{Y\}$. Suppose we want to implement $\overline{p}$ for type X and $\underline{p}$ for type Y. This can obviously be implemented in the abstract-declaration model, where player 1 cannot lie. However, in the underlying setting of hard evidence, it cannot be implemented because both types can deviate by disclosing R (and sending some arbitrary message) and, whatever outcome is specified for this action, one of the types will have the strong incentive to deviate so. A similar argument applies with incomplete information when $n \geq 2$.[16]

The key issue here is that there are more ways of deviating in the setting with hard evidence than there are in the abstract-declaration model. A version of Theorem 3 can be recovered, however, if one considers an abstract-action model with more actions than there are states. In fact, under a strengthened normality condition, a single extra action suffices.

Definition 6: *The evidentiary structure is called* **strongly normal** *if it is normal*

[16]To see this, consider a two-player setting with Θ_1, E_1, E_1^X, E_1^Y, P, and player 1's preferences as above, but with $\Theta_2 = \{a, b\}$, $E_2 = \{q, r\}$, $E_2^a = \{q\}$, $E_2^b = \{r\}$. Suppose we want to implement $\overline{p}$ in states Xa and Xb, and want to implement $\underline{p}$ in states Ya and Yb. This breaks down as above. With two or more players and complete information, we can always prevent a player from unilaterally deviating with gibberish, but it is not the case with incomplete information.

and, for each player i, there is an evidentiary action $\underline{e}_i$ such that $\underline{e}_i \in E_i^{\theta_i}$ for every $\theta_i \in \Theta_i$.

Strong normality requires the existence of "minimal evidence" that can be presented in every state, in addition to the maximal evidence mandated by normality. For motivation, consider again the interpretation of evidence as disclosure of documents. Assuming that it is always feasible for player i to disclose nothing (that is, using the notation for documents introduced in Section 2, $\emptyset \in \mathcal{D}_i^{\theta_i}$ for every θ_i), the evidentiary action $\underline{e}_i$ can represent $\emptyset$, satisfying the condition of strong normality.

Theorem 4: *Take as given any setting with hard evidence that has a strongly normal evidentiary structure. There is an abstract-action model with $A_i \equiv \Theta_i \cup \{\emptyset\}$, for each i, that has the following property. An outcome function is implementable in the setting with hard evidence if and only if it is implementable in the abstract-action model. Furthermore, if an outcome function is implementable in the abstract-action model then it is implementable with truthful reporting (meaning that type θ_i of player i reports "θ_i").*

The extra declaration "$\emptyset$" is not necessary if, for each player i, there is a "minimal type" whose evidence can be presented by every other type. This is because, in the setting with hard evidence, any deviation that does not mimic another type can be treated as though the player has identified himself as his minimal type.

Corollary: *Take as given any setting with hard evidence that has a strongly normal evidentiary structure and suppose that, for each player i, there is a type θ_i such that $\overline{e}_i^{\theta_i} \in E_i^{\theta_i'}$ for every $\theta_i' \in \Theta_i$. An outcome function is implementable in the setting with hard evidence if and only if it is implementable in the corresponding abstract-declaration model. Furthermore, if an outcome function is implementable in the abstract-declaration model then it is implementable with truthful reporting.*

5 On Dynamic Mechanisms

Our analysis to this point, as well as that of Green and Laffont (1986) and some others in the related literature, has been limited to static mechanisms, meaning that all interaction (messages and evidence disclosure) occurs at one time.[17] The usual defense of static mechanisms is that they are sufficient for the analysis of Nash implementation. However, this logic is not necessarily valid for settings with hard evidence, as the following example shows.

Consider the complete-information setting in which $n = 2$, $\Theta_1 = \Theta_2 = \Theta = \{X, Y, Z\}$, $P = \{\underline{p}, \overline{p}\}$, $E_1 = \{\emptyset\}$, and $E_2 = \{Q, R\}$. Suppose payoffs are given by

[17] As briefly discussed in the Introduction, Lipman and Seppi (1995) study dynamic mechanisms in a setting where a player's message space may contain evidence. They are concerned with the minimum that evidence has to prove in order for the decision maker's inference, which is based on the messages sent by players, to be robust. They do not consider the use of static mechanisms.

$u(\underline{p},\theta) = (0,0)$ and $u(\overline{p},\theta) = (1,-1)$, for each $\theta \in \Theta$. Feasible evidence is given by $E_1^X = E_1^Y = E_1^Z = \{\emptyset\}$, $E_2^X = \{Q\}$, $E_2^Y = \{Q,R\}$, and $E_2^Z = \{R\}$. In this example, player 2's evidentiary structure is the same as that in the example of Section 3. We shall analyze whether it is possible to implement outcome function g defined by $g(X) = g(Z) = \overline{p}$ and $g(Y) = \underline{p}$.

We first demonstrate that g cannot be implemented by a static mechanism. Invoking the Weak Revelation Result, we focus on implementation with direct and truthful messages, so that $A_1 = \{X,Y,Z\}$ and $A_2 = \{X,Y,Z\} \times \{Q,R\}$. We shall represent player 2's individual actions as XQ, XR, and so forth. Note that, if g is implementable, then we can assume f is defined so that $f(X,XQ) = f(Z,ZR) = \overline{p}$, because (X,XQ) and (Z,ZR) will be the equilibrium action profiles in the message and disclosure game in states X and Z. Furthermore, in state Y the equilibrium action profile is either (Y,YQ) or (Y,YR); in the former case we have $f(Y,YQ) = \underline{p}$, whereas in the latter case we have $f(Y,YR) = \underline{p}$.

Take the case in which the equilibrium action profile in state Y is (Y,YQ), requiring $f(Y,YQ) = \underline{p}$, and we will reveal a contradiction. To deter player 1 from sending message "X" in state Y, it must be that $f(X,YQ) = \underline{p}$. However, to deter player 2 from sending message "Y" in state X, we need $f(X,YQ) = \overline{p}$. A similar contradiction occurs in the case in which the equilibrium action profile in state Y is (Y,YR), where $f(Y,YR) = \underline{p}$; we need $f(Z,YR) = \underline{p}$ to deter player 1 from saying "Z" in state Y, but we also need $f(Z,YR) = \overline{p}$ to deter player 2 from saying "Y" in state Z. Thus, g is not implementable using static mechanisms.

Interestingly, there is a simple dynamic mechanism that implements g:

> First, player 1 must announce the state. After observing player 1's announcement, player 2 then must take his evidentiary action (Q or R). The outcome is prescribed to be $\underline{p}$ if player 1 said "Y" or if player 2's evidence shows player 1 to have lied about the state (that is, if player 1 said "X" and player 2 presented R, or if player 1 said "Z" and player 2 presented Q). Otherwise, the outcome is $\overline{p}$.

Clearly, player 2 can always reveal whether announcements "X" or "Z" by player 1 are lies, and it is in player 2's interest to reveal a lie. Player 1 therefore has the incentive to tell the truth.

In this example, the dynamic mechanism is effective because player 2 has the ability to tailor his evidence to the actual announcement made by player 1.[18] In particular, if player 1 lies in state Y, then player 2 chooses the appropriate evidence

[18] Lipman and Seppi's (1995) Example 1 (two lobbyist who in every state each have a "not message" for every other state, which proves that the state is not that other state, and are constrained to each send only one "not message") has a similar flavor in that their "not messages" are used to reveal a lie. Though they do not address the issue, the evidence environment in their example is not normal. Of further interest, Lipman and Seppi suggest that time constraints, such as in a debate, may limit the amount of claims that a party can prove. An argument along this line could suggest a reason for why, in practice, some evidence environments are not normal.

to reveal the lie. Player 2's disciplining function cannot be replicated in a static mechanism because there is no *single* evidentiary action that reveals both the "X" and "Z" lies. The failure of normality creates this problem.

Our next result shows that static mechanisms are sufficient for implementation in settings with normal evidentiary structures. We consider the class of dynamic mechanisms (extensive-game forms) with moves of Nature. Each decision node is either a "message" node, with an arbitrary number of branches, or an "evidentiary" node, where a player must take his evidentiary action. Extensive forms are restricted to those in which each player makes an evidentiary decision exactly once in every path through the tree. The information structure is arbitrary; a player's messages or evidence may or may not be observed by others. Outcomes are prescribed at terminal nodes.[19]

To study rationality for dynamic mechanisms, we use Bayesian Nash equilibrium (so the result also holds for any refinement). For a given extensive form, let S_i denote the set of pure strategies for player i, $i = 1, 2, \ldots, n$, and let S_0 denote the pure strategies of Nature. Each player-type θ_i must select a strategy for the extensive form. Type θ_i is limited to strategies in the set $S_i^{\theta_i}$ that choose elements in $E_i^{\theta_i}$ at evidentiary decision nodes. Without loss of generality, we can focus on the case in which each type of each player selects a pure strategy and where Nature has a finite number of strategies.[20]

Let $\beta_i : \Theta_i \to S_i$ give player i's strategy in the mechanism as a function of player i's type; that is, $\beta_i(\theta_i)$ is the extensive-form strategy in $S_i^{\theta_i}$ that is played by type θ_i of player i. A strategy profile for the mechanism can thus be written $\beta : \Theta \to S_1 \times S_2 \times \cdots S_n$. We write β_{-i} for the profile of type-contingent strategies for the players other than player i. Furthermore, let $F(s_0, s_i, s_{-i})$ denote the public action at the terminal node that is reached when strategy profile (s_i, s_{-i}) is played and Nature selects s_0 in the extensive form. This is an equilibrium if, for every player i and each type $\theta_i \in \Theta_i$, $\beta_i(\theta_i)$ solves

$$\max_{s_i \in S_i^{\theta_i}} \sum_{\theta_{-i} \in \Theta_{-i}(\theta_i)} \sum_{s_0 \in S_0} \mu_{-i}(\theta_{-i} \mid \theta_i)\sigma_0(s_0)u_i(F(s_0, s_i, \beta_{-i}(\theta_{-i})), (\theta_i, \theta_{-i})),$$

where σ_0 is the probability distribution over Nature's strategies. An outcome function g is implemented by the given dynamic mechanism if there is an equilibrium β such that $g(\theta) = F(s_0, \beta(\theta))$ for each $\theta \in \Theta$ and every $s_0 \in S_0$.

Theorem 5: *Assume that the evidentiary structure is normal. If an outcome function is implemented by a dynamic mechanism then there is also a static mechanism that implements it.*

[19] This exercise is related to Lipman and Seppi's (1995) analysis of game forms in which the players sequentially send messages.

[20] For each mechanism and a mixed-strategy equilibrium, we can find an equivalent mechanism and pure-strategy equilibrium where Nature provides private randomization devices for the players.

Proof: Suppose that g is implemented by some dynamic mechanism and take as given the equilibrium β played in the dynamic mechanism. Construct a static mechanism so that $M_i \equiv \Theta_i$, for each player i, and define the function f as follows. For each player i, we translate player i's action a_i into a strategy in the original extensive form. Specifically, if $a_i = \overline{a}_i^{\theta_i}$ for some $\theta_i \in \Theta_i$, then this is translated into the equilibrium strategy $\beta_i(\theta_i)$ of type θ_i for the original dynamic mechanism. If $a_i \equiv (\theta_i, e_i) \neq \overline{a}_i^{\theta_i}$ for all $\theta_i \in \Theta_i$, then a_i can be translated into any strategy that takes evidentiary action e_i at the information set(s) where player i must present evidence. Then $f(a)$ is defined as the outcome at the terminal node of the original extensive form that would be reached if the players used the strategies corresponding to action profile a.

For the static mechanism, let us prescribe that each player truthfully report his type and present maximal evidence. Clearly, when the players follow this prescription, they obtain $g(\theta)$ in state θ. One can also easily verify that a feasible deviation for type θ_i of player i in the static game translates into a feasible deviation in the extensive form for type θ_i. For example, suppose that player i can choose $\overline{a}_i^{\theta'_i}$ when his type is θ_i. This means that $\overline{e}_i^{\theta'_i} \in E_i^{\theta_i}$, which, by evidentiary normality, implies $E_i^{\theta'_i} \subset E_i^{\theta_i}$. Thus, in the original extensive form, it is feasible for player i of type θ_i to play the strategy that he would have played in equilibrium had his type been θ'_i. Note that without normality there is not maximal evidence $\overline{e}_i^{\theta_i}$ for every type and it is not possible, in general, to translate some a_i in the static game into an equilibrium strategy in the extensive form. (In state Y of the above example, there is no way to effectively translate a message and disclosure of Q or R into player 2's strategy in the extensive form. This is because in the static game any feasible a_2 in state Y is also feasible some other state.) Because we started with equilibrium strategies for the extensive form, the players have the incentive to follow our prescription in the static mechanism. Thus, the static mechanism implements g. *Q.E.D.*

The example and theorem demonstrate another way in which normality plays a role in determining the proper mechanism-design methodology. Even holding aside the issues discussed earlier in this paper, the example shows that extending Green and Laffont's (1986) analysis to settings with two or more players requires the consideration of dynamic mechanisms. When the evidentiary structure is not normal, dynamic mechanisms can be useful in allowing players to tailor their evidence disclosure to reveal deviations by others.

In the example, it is important that player 2's evidentiary action follows player 1's message. More generally, a game form may specify that the players' evidentiary actions occur sequentially. For instance, it could be specified that player 1 sends a message, then player 2 observes player 1's message and selects evidence, and then player 1 observes player 2's move and selects his own evidence. A key question is whether implementation sometimes requires sequential evidentiary actions. Our final result establishes that the answer is "no."

We use the term *special three-stage mechanism* for any mechanism of the following form. In the first stage, the players privately announce their types to the external enforcer. In the second stage, the players receive private messages from the enforcer. In the third stage, the players simultaneously take their evidentiary actions. A special three-stage mechanism can be viewed as a procedure in which each player makes a claim about his type and then is asked to provide evidence to reinforce his claim.[21]

Theorem 6: *If an outcome function is implemented by a dynamic mechanism then there is also a special three-stage mechanism that implements it with truthful reporting in equilibrium. Furthermore, in settings of complete information the second stage is not required and one can assume that first-stage messages are public.*

The class of special three-stage mechanisms is similar to Myerson's (1982,1991) framework of mechanism design with moral hazard, where a communication system precedes a prespecified game. Thus, our evidentiary actions are analogous to fixed moral hazard variables, although we limit attention to the technology in which, in each state, a given evidentiary action is either feasible and free or it is infeasible. The issue addressed by Theorem 6—whether the timing of evidentiary actions should be designed as sequential—has not been addressed in the prior literature.[22]

Proof of Theorem 6: Suppose that outcome function g is implemented by a given dynamic mechanism—call it the *original extensive form*—and let β denote the equilibrium that supports the implementation. Also, let F be the mapping from strategy profiles (including Nature's strategy) to public actions as defined above.

We define a special three-stage mechanism as follows. First, the players send private reports of their types. Let θ' denote the report profile. Second, a private message is sent to each player, identifying an information set for this player in the original extensive form. The information sets are selected at random and correspond to the players' reports and Nature's strategy. Specifically, a pure strategy for Nature, s_0, is secretly drawn (not known to the players) from Nature's probability distribution. Note that the profile $(s_0, \beta(\theta'))$ determines a unique path in the original extensive form. Along this path, each player i has exactly one information set where he takes his evidentiary action; call this information set I_i. Further, the players are assigned numbers according to the order in which they take evidentiary actions along the path. Let k_i be player i's number. Each player i is then privately informed only of I_i. That

[21] This type of mechanism roughly mimics the typical trial process, where the parties make opening statements in which they claim particular events have occured, and then they present evidence. A similar theme is present in Lipman and Seppi's (1995) "believe unless refuted" rule (Section 3 of their paper). Also related is the modeling exercise of Glazer and Rubinstein (2004), who study a game of persuasion between a listener (decision maker) and a speaker. The speaker sends a message and then, before making a decision with payoff consequences, the decision maker can engage in limited monitoring that amounts to the production of hard evidence. Glazer and Rubinstein examine how the listener should optimally verify an aspect of the state.

[22] It is easy to see that dynamic mechanisms are generally required for implementation in settings with payoff-relevant inalienable actions. A study along these lines would be instructive.

is, the player is told that he should consider what he would do in the original extensive form at information set I_i.

Third, the players simultaneously disclose evidence. Let e denote the profile of evidentiary actions. Also, let $s_i^{\theta_i'} \equiv \beta_i(\theta_i')$ for each player i. If $e_i = s_i^{\theta_i'}(I_i)$ for each player i (that is, player i's evidentiary action is exactly the one that would be chosen by strategy $\beta_i(\theta_i')$ at information set I_i in the original extensive form), then the external enforcer compels public action $F(s_0, \beta(\theta'))$. Otherwise, it must be that one or more of the players has taken an evidentiary action that is inconsistent with his reported type. In this case, let j indicate the player who deviates first along the path: j solves $\min_i k_i$ subject to $s_i^{\theta_i'}(I_i) \neq e_i$. Then the external enforcer compels public action $F(s_0, s_j', \beta_{-j}(\theta_{-j}'))$, where s_j' is defined by $s_j'(I_j) = e_j$ and $s_j'(h) \equiv s_j^{\theta_i'}(h)$ for every other information set h for player j. Note that the path generated by $(s_0, s_j', \beta_{-j}(\theta_{-j}))$ diverges at information set I_j from the path generated by $(s_0, \beta(\theta'))$. Further, these paths are consistent with any type of player j that can select e_j.

If every player reports honestly in the first stage and chooses evidence as prescribed (with player i selecting $s_i^{\theta_i}(I_i)$), then the public action is exactly that which would have resulted in the equilibrium of the original extensive form. Any deviation of type θ_i of player i (through his report and/or evidence disclosure) induces the distribution of outcomes that would have occurred if he had played some other strategy s_i'' in the original extensive form. Importantly, this strategy s_i'' is feasible for type θ_i, in that $s_i'' \in S_i^{\theta_i}$. Because type θ_i does not have the incentive to deviate to s_i'' in the original extensive form, he does not gain by deviating in the special three-stage mechanism. Thus, the prescribed strategies are an equilibrium. Furthermore, in settings of complete information, clearly we can dispense with the second stage and simply have the first-stage reports be public. *Q.E.D.*

6 Conclusion

We have provided a model of mechanism-design with hard evidence that is a straightforward extension of the standard mechanism-design framework. Our results clarify the revelation principle and identify conditions under which one can limit attention to certain classes of mechanisms. Our analysis supports and further develops Green and Laffont's (1986) modeling exercise, in particular with respect to the meaning of abstract actions and the extension to settings with more than one player.

Our findings demonstrate the importance of explicitly modeling hard evidence in terms of its real technology. Incorporating inalienable actions into the mechanism-design framework allows us to make precise the connection between an abstract theoretical model and actual technological and institutional constraints, which is important for understanding these constraints (as emphasized by Watson 2003). We encourage further research along this line. Our own future plans include analyzing dynamic mechanisms in settings of hard evidence and costly disclosure.

As noted in the Introduction, the analysis shown herein is closely related to the contemporaneous work of Deneckere and Severinov (2001) and the more recent work of Forges and Koessler (2003). Here are more details regarding the connection. Deneckere and Severinov (2001) examine principle-agent settings in which the agent has a type-contingent set of feasible messages but can communicate with the principle over multiple periods and can send any, and only one, of his messages in each period. Forges and Koessler (2003) examine the same kind of communication constraints in the context of a multiple-player, moral hazard (individual action) setting with mediated communication. The individual messages in these models are basically the same as the "documents" we discussed in Section 2, except that, for Deneckere-Severinov and Forges-Koessler, multiple documents can be disclosed only sequentially. In these models, mechanisms consist of multiple rounds of communication.

Deneckere and Severinov (2001) show that, in their model, any implementable outcome function can be implemented by a mechanism in which the agent first announces his type (cheap talk) and then, in subsequent periods, sends every individual message that he is able to send. That is, the agent announces his type and discloses all of his documents—what we call maximal evidence production. Thus, Deneckere and Severinov's theorem is a version of the Strong Revelation Result (Theorem 2); it holds because their setting exhibits evidentiary normality, which follows from the fact that all combinations of documents can be eventually submitted by the agent. Forges and Koessler (2003) have a similar result (their Theorem 1) that relates implementation with multiple periods of communication to static implementation in a setting in which players can, using our language, disclose any set of documents at one time. Thus, Forges and Koessler's result is an affirmation that sequential document disclosure is tantamount to disclosing sets of documents at once and, thus, that evidentiary normality holds.[23]

Forges and Koessler present a second result (their Theorem 2) that, accounting for differences in the settings analyzed, duplicates our Theorem 5. Under a "minimal closure condition," they find that one-period communication schemes suffice for implementation. The minimal closure condition is defined relative to the sets of types that an individual type can distinguish himself from; it requires that if a given type has ways of distinguishing itself from various other types then it has a way of simultaneously distinguishing itself from all of these other types. Translated into our modeling apparatus, this is evidentiary normality.

We end with comments on an extension. In practice, hard evidence may arise stochastically as a function of productive actions that the players take.[24] Note that

[23]We favor thinking in terms of evidentiary actions, rather than sequences of type-constrained messages, and we emphasize the benefit of being very explicit about the distinction between real evidentiary actions and abstract messages in the context of mechanism design.

[24]Bull (2001b) studies a setting where the probability evidence is revealed depends on a player's effort to suppress the evidence. See Bull and Watson (2004) for a detailed treatment of productive actions and contract with hard evidence.

our framework can incorporate random evidence by defining the state to be a vector comprising an important payoff-relevant event (such as the player's prior investments or other productive actions) and the outcome of a random draw that determines the availability of hard evidence. For example, player 1's type might be L, H1, or H0, where L denotes that he failed to invest earlier, H1 denotes that he invested earlier and has hard evidence to prove it (a document, say), and H0 denotes that he invested earlier but has no proof. The analysis of implementable outcome functions would proceed just as characterized herein. However, note that a key question will be, for example, whether player 1 can be given the incentive to invest, which may depend on implementing an outcome function that makes the weighted average of states H1 and H0 attractive for player 1, relative to state L. Thus, in applications, it will be important to examine convex combinations of implementable state-contingent payoffs.

Appendix: Other Proofs

Proof of Theorem 3: Because this is a setting of complete information, we write $\Theta_i = \Theta$ for each i. As noted in the text, the second claim of the theorem is a simple extension of Green and Laffont's (1986) result. Using this fact, and invoking Theorem 2, we can focus on (i) direct and truthful messages and maximal evidence production in the setting of mechanism design with hard evidence, and on (ii) truthful-declaration equilibria in the abstract-declaration model. It is straightforward to verify that any outcome function that is implementable in the setting with hard evidence is also implementable in the abstract-declaration model. This follows from the observation that any declaration "θ" in the abstract-declaration model can be translated into the message and evidence that the player would send in the other model.

Regarding the other direction, take an outcome function g that is implementable in the abstract-declaration model and let f' be the function that implements it. We will first define a function f for the setting with hard evidence. To this end, for each player i and each $a_i \in \Theta \times E_i$, define $\hat{\theta}_i(a_i)$ as follows. If $a_i = \overline{a}_i^\theta$ for some $\theta \in \Theta$ then specify $\hat{\theta}_i(a_i) \equiv \theta$; otherwise, specify $\hat{\theta}_i(a_i) \equiv \phi$, where ϕ is an indicator value that is not a member of Θ. The idea is that, in the message and evidence game, player i will obtain the "ϕ" label if his action does not correspond to the prescribed equilibrium action in some state.

For every action profile $a \in \Theta^n \times E$ satisfying $\hat{\theta}_i(a_i) \neq \phi$ for every i, let $\hat{\theta}(a) \equiv (\hat{\theta}_1(a_1), \hat{\theta}_2(a_2), \ldots, \hat{\theta}_n(a_n))$. For these action profiles, prescribe $f(a) \equiv f'(\hat{\theta}(a))$. Group all other action profiles into two sets, which are distinguished by whether more than one player has the "ϕ" label. If there is a player i such that $\hat{\theta}_i(a_i) = \phi$ and $\hat{\theta}_j(a_i) = \hat{\theta}_k(a_i) = \theta \neq \phi$ for every $j, k \neq i$ (that is, everyone else agrees on state θ, while player i send gibberish), then prescribe $f(a) = g(\theta)$. For every other action profile, $f(a)$ may be arbitrarily defined.

With f thus specified, action profile $\overline{a}^\theta$ is a Nash equilibrium in the message and disclosure game in state θ. If player i deviates by successfully pretending to be in another state, he would obtain the payoff of deviating in this way in the abstract-declaration game, which deters the deviation. What is important here is that, by the definition of the abstract-declaration model, a feasible deviation in the setting of hard evidence (say, $(\theta', \overline{e}_i^{\theta'})$ in state θ) translates into a feasible deviation in the abstract-declaration model (that is, θ'). If player i deviates with gibberish—which gives him the "ϕ" label, then his payoff is the same as if he did not deviate. By construction of f, this implements g. *Q.E.D.*

Proof of Theorem 4: The second claim of the theorem is easily proved using the standard argument. To prove the first claim, we use the translation in the proof of Theorem 3 with $\emptyset$ in place of ϕ, whereby the report "θ_i" in the abstract-action model corresponds to $\overline{a}_i^{\theta}$ in the setting with hard evidence. To study how implementation in the abstract-action model implies implementation in the setting with hard evidence, we translate report "$\emptyset$" into $(\theta_i, \underline{e}_i)$ for any arbitrarily chosen θ_i. To study how implementation in the setting with hard evidence implies implementation in the abstract-declaration model, we translate every action a_i that satisfies $a_i \neq \overline{a}_i^{\theta_i}$ for all θ_i into the declaration "$\emptyset$." It is not difficult to confirm that feasible deviations in one model translate into feasible deviations in the other. *Q.E.D.*

References

Brusco, S., 2000. Unique implementation of action profiles: necessary and sufficient conditions. manuscript, Universidad Carlos III de Madrid.

Bull, J., 2001a. Costly evidence production and the limits of verifiability. manuscript, The Wharton School, University of Pennsylvania.

Bull, J., 2001b. Costly evidence and systems of fact-finding. manuscript, Florida International University.

Bull, J., Watson, J., 2001. Evidence disclosure and verifiability. UC San Diego Working Paper 2000-16.

Bull, J., Watson, J., 2004. Evidence disclosure and verifiability. J. Econ. Theory 118, 1–31.

Dasgupta, P., Hammond, P., Maskin, E., 1979. The implementation of social choice rules: some general results on incentive compatibility. Rev. Econ. Stud. 46, 185–216.

Deneckere, R., Severinov, S., 2001. Mechanism design and communication costs. manuscript, University of Wisconsin.

Evans, R., 2002. Efficient contracts in complex environments. manuscript, St. John's College, Cambridge.

Forges, F., Koessler, F., 2003. Communication equilibria with partially verifiable types. manuscript, Université de Cergy-Pontoise.

Glazer, J., Rubinstein, A., 2004. On optimal rules of persuasion. Econometrica 72, 1715–1736.

Green, J., Laffont, J., 1986. Partially verifiable information and mechanism design. Rev. Econ. Stud. 53, 447–456.

Hart, O., Moore, J., 1988. Incomplete contracts and renegotiation. Econometrica 56, 755–785.

Lipman, B. L., Seppi, D. J., 1995. Robust inference in communication games with partial provability. J. Econ. Theory 66, 370–405.

E. Maskin, Moore J., 1999. Implementation and renegotiation. Rev. Econ. Stud. 66, 39–56.

Milgrom, P., Roberts, J., 1986. Relying on the information of interested parties. Rand J. Econ. 17, 18–32.

Mookherjee, D., Reichelstein, S., 1990. Implementation via augmented revelation mechanisms. Rev. Econ. Stud. 57, 453–75.

Myerson, R. B., 1982. Optimal coordination mechanisms in generalized principal-agent problems. J. Math. Econ. 10, 67–81.

Myerson, R. B., 1991. Game theory: analysis of conflict. Harvard University Press, Cambridge.

Okuno-Fujiwara, M., Postlewaite, A., Suzumura, K., 1990. Strategic information revelation. Rev. Econ. Stud. 57, 25–47.

Segal, I., Whinston, M., 2002. The Mirrlees approach to mechanism design with renegotiation (with applications to hold-up and risk-sharing. Econometrica 70, 1-45.

Seidmann, D., Winter, E., 1997. Strategic information transmission with verifiable messages. Econometrica 65, 163–169.

Shin, H. S.. 1994. The burden of proof in a game of persuasion. J. Econ. Theory 64, 253–264.

Watson, J., 2003. Contract, mechanism design, and technological detail. UC San Diego Economics Working Paper 2002-04.