Bayesian implementation with verifiable information

Esteban Peralta

Dept. of Economics, University of Michigan, 611 Tappan Ave. (48109), Ann Arbor, MI, USA
eperalta@umich.edu

April 15, 2019

Summary

This note studies the class of allocations that are fully Bayesian implementable in the presence of verifiable information. I identify a condition, termed evidence distinguishability, that together with incentive compatibility is both necessary and sufficient for full Bayesian implementation in direct mechanisms that elicit verifiable information. I show, however, that every evidence distinguishable allocation that is incentive compatible when verifiable information is not elicited is Bayesian monotonic at every verifiable deception; namely, at every verifiable joint report of information at which the direct mechanism delivers an undesirable outcome. To the extent that requiring Bayesian monotonicity to hold at every verifiable deception is a demanding condition, only indirect mechanisms can exploit the presence of verifiable information to fully implement incentive compatible allocations.

Keywords: Mechanism design; Bayesian implementation; verifiable information.
JEL codes: C72, D71, D82.

1 Introduction

The standard analysis of full implementation problems—where the goal is to find a mechanism that produces no undesirable outcome as an equilibrium—assumes that agents can manipulate their information in any way. In many situations, however, agents' information is verifiable. For example, taxpayers must present evidence about their income, and employers often monitor their employees' effort to limit their ability to falsely report their performance. In these situations, the designer might then be able to achieve goals that run against the agents' incentives.[1]

Ben-Porath and Lipman (2012) and Kartik and Tercieux (2012) have recently shown that in complete information environments, the presence of verifiable information allows the designer to fully Nash implement allocations that fail to be monotonic (Maskin, 1999).

[1]The consequences of having agents whose information can be verified have been widely studied since the publication of seminal papers by Green and Laffont (1986), Grossman (1981), Milgrom (1981), and Milgrom and Roberts (1986).

These results suggest that the designer's ability to exploit the presence of verifiable information might extend to environments of incomplete information.[2] This would be important as Bayesian monotonicity—the extension of monotonicity to environments of incomplete information—is a demanding condition (see, e.g., Serrano and Vohra (2005)). The above results, however, hinge heavily on the use of "indirect" mechanisms; namely, mechanisms that both elicit more than the agents' information and involve complicated "integer" games. Since the use of these mechanisms has been widely criticized (see, e.g., Kartik *et al.* (2014), Jackson (1992), and Ortner (2015))), whether the presence of verifiable information enlarges the class of fully Bayesian implementable allocations by means of "simple" mechanisms or not becomes unclear. This note tackles this question by investigating when, in environments of incomplete information, the presence of verifiable information can be exploited by direct mechanisms.

I incorporate verifiable information into an otherwise standard model by assuming that agents' types describe not only a payoff-relevant parameter, but also a set of messages. These messages are private but verifiable, in the sense that different types might possess different messages. As a consequence, the disclosure of some messages could prove that certain types have not occurred—but disclosing messages that are not available is unfeasible. Verifiable messages are payoff-irrelevant, but I restrict attention to allocation rules that depend on both parameters and messages. For this class of models, Deneckere and Severinov (2008) show that when agents can send every combination of their available messages and the direct mechanism elicits not only the agents' parameters but also their verifiable messages, a generalization of incentive compatibility is necessary and sufficient for the truthful revelation of information to be a Bayesian Nash equilibrium.[3] This generalization, which I term "evidence incentive compatibility," requires that no agent–type finds strictly beneficial to report either a false parameter or a proper subset of its verifiable messages when every other agent reports her true type.

While truthful implementation guarantees that the direct mechanism possesses an equilibrium that delivers the desired outcome at every state, full implementation also demands that agents do not coordinate on a feasible deception; that is, on a false but feasible joint report of types for which the direct mechanism does not deliver the outcome prescribed by the desired allocation rule. I identify a condition, termed "evidence distinguishability," that, together with evidence incentive compatibility, is necessary and sufficient for full Bayesian implementation in the direct mechanism (Proposition 1). Evidence distinguishability requires that for every feasible deception, there is some agent–type who finds the outcome obtained by some feasible deviation strictly better than the deceptive outcome.

Within the class of allocation rules that are incentive compatible when verifiable messages are *not* elicited, however, every evidence distinguishable allocation is Bayesian monotonic at every feasible deception (Proposition 2). This condition is weaker than Bayesian monotonicity, and I show by example that the result cannot be strengthened to require Bayesian monotonicity. To the extent that requiring Bayesian monotonicity to hold at every feasible deception is still demanding, it follows that the direct mechanism exploits the presence of verifiable information *only if* this presence relaxes some incentive constraint. Put another way, when truthful revelation does not require the elicitation of verifiable information, eliciting this information might nonetheless prevent "bad" coordinations, but only when it is done by an indirect mechanism. The results might therefore shed some light on how much

[2]The standard references on full implementation problems with incomplete but unverifiable information are Jackson (1991), Mookherjee and Reichelstein (1990), Palfrey and Srivastava (1989) and Postlewaite and Schmeidler (1986).

[3]A similar result was obtained by Bull and Watson (2007).

the presence of verifiable information relaxes the full implementation problem in practical applications, in which only direct revelation mechanisms can be used.

The rest of the paper is organized as follows. Section 2 introduces the basic model and Section 3 presents the notions of Bayesian full implementation, evidence incentive compatibility, and Bayesian monotonicity. Section 4 introduces the notion of evidence distinguishability, characterizes the class of allocations that are fully Bayesian implementable by their direct mechanism, and concludes that the presence of verifiable information can generally be exploited only by indirect mechanisms. Finally, Section 5 offers a brief discussion of the main assumptions of the paper.

2 Model

I assume the existence of $n \geq 2$ agents, $N = \{1, 2, ..., n\}$, and denote by i a generic agent. I let A denote a finite set of outcomes, Θ_i a finite set of payoff-relevant parameters for player i and $\mathcal{E}_i$ a finite collection of messages for player i. For any player i, I denote by $\theta_i \in \Theta_i$ a generic parameter and by $E_i \subseteq \mathcal{E}_i$ a generic subset of messages. The set of types of player i is defined as $T_i := \Theta_i \times 2^{\mathcal{E}_i}$.[4] Thus, the set of all type profiles is denoted $T := \times_{i \in N} T_i$. Messages in $\mathcal{E}_i$ are verifiable in the sense that they need not be part of the collection of messages of every type of player i. I will denote by $t_i = (\theta_i, E_i)$ a typical element of T_i, where $E_i \subseteq \mathcal{E}_i$ and $\theta_i \in \Theta_i$. I will write $E_i(t_i)$ and $v(t_i)$ for the collection of messages and parameter of type $t_i = (\theta_i, E_i)$.[5] For any $t \in T$, I will write $v(t_{-i}) = (v(t_1), ..., v(t_{i-1}), v(t_{i+1}), ..., v(t_n))$ for the profile of payoff-relevant parameters of every player but i. The utility function of player i is given by $U^i : A \times \Theta \to \mathbf{R}$, where $\Theta := \times_i \Theta_i$. Hence, I assume that verifiable messages are payoff-irrelevant.

I assume the existence of a common and diffuse prior on T; namely, a common prior μ that puts positive probability on every $t \in T$. After receiving signal t_i, player i forms a belief $\mu(t_{-i}|t_i)$ about the realization of types of the other players. A social choice function (SCF) or allocation rule is a map $x : T \to A$. Hence, allocation rules depend on both parameters and verifiable messages. Given an allocation x, I denote i's interim utility at t_i by:

$$V^i(x, t_i) := \sum_{t_{-i} \in T_{-i}} U^i(x(t_i, t_{-i}), v(t_i), v(t_{-i}))\mu(t_{-i}|t_i).$$

3 Implementation

A mechanism $\mathcal{M} = (\mathcal{C}, g)$ is a collection of messages $\mathcal{C}_i$ for each player i and a map $g : \mathcal{C} \times 2^{\mathcal{E}_1} \times ... \times 2^{\mathcal{E}_n} \to A$, where $\mathcal{C} := \times_{i \in N} \mathcal{C}_i$. Each message $c_i \in C_i$ represents a cheap talk message; namely, a message that every type can send. Notice that the outcome prescribed by the mechanism depends both on cheap talk and verifiable messages. The latter, however, are private information and so are not part of the design. Given a prior μ, a mechanism $\mathcal{M}$ induces a game of incomplete information that I denote by $\mathcal{M}(\mu)$. A strategy for player i in game $\mathcal{M}(\mu)$ is a map $\sigma_i : T_i \to \mathcal{C}_i \times 2^{\mathcal{E}_i}$.[6] A strategy σ_i is *feasible* if $\sigma_i(t_i) \in \mathcal{C}_i \times 2^{E_i(t_i)}$ for every t_i. A strategy profile is a tuple $\sigma = (\sigma_1, ..., \sigma_n)$. A profile σ is feasible if σ_i is feasible

[4] Every type is assumed to have a non-empty set of messages, so this notation entails a slight abuse.

[5] I model verifiable messages as abstract messages to capture a variety of situations in which the agents' ability to manipulate their information is limited. Thus, for example, a message may represent a document, a physical item (e.g., evidence), or some (in)action (e.g., inability to lie). See Deneckere and Severinov (2008) for a discussion.

[6] Notice that I am assuming that for every i, every $t_i \in T_i$ can send any combination (i.e., subset) of messages in $E_i(t_i)$. See section 5.3 for a discussion.

for every i. I will restrict attention to pure strategies.[7] Thus, I will focus on pure Bayesian Nash equilibria:

Definition 1. A feasible profile σ is a **Bayesian Nash Equilibrium** (BNE) of $\mathcal{M}(\mu)$ if for every i, every feasible $\tilde{\sigma}_i : T_i \to \mathcal{C}_i \times 2^{\mathcal{E}_i}$ and every t_i:

$$\sum_{t_{-i}\in T_{-i}} U^i(g(\sigma), v(t_i), v(t_{-i}))\mu(t_{-i}|t_i) \geq \sum_{t_{-i}\in T_{-i}} U^i(g(\tilde{\sigma}_i(t_i), \sigma^{-i}(t_{-i})), v(t_i), v(t_{-i}))\mu(t_{-i}|t_i).$$

Accordingly, I will consider implementation in Bayesian Nash equilibrium. Formally:

Definition 2. x is ***fully implementable*** if there is a mechanism $\mathcal{M} = (\mathcal{C}, g)$ such that $\mathcal{M}(\mu)$ has a BNE and for *every* BNE σ of $\mathcal{M}(\mu)$, $g(\sigma(t)) = x(t)$ for every $t \in T$.

This is the standard notion of full implementation in settings of incomplete information. Given an allocation x, the direct mechanism is denoted by $\Gamma_d = (\Theta, x)$. A strategy for player i in Γ_d is therefore a map $\alpha^i : T_i \to T_i$. A strategy α^i is **feasible** for player i if $E_i(\alpha^i(t_i)) \subseteq E_i(t_i)$ for every t_i. A profile α is feasible if α^i is feasible for every i. A profile α is a **deception** if $x(\alpha(t)) \neq x(t)$ for some $t \in T$.

Definition 3. x is ***truthfully implementable*** if there is a mechanism $\mathcal{M} = (\mathcal{C}, g)$ such that $\mathcal{M}(\mu)$ has *some* BNE σ such that $g(\alpha(t)) = x(t)$ for every $t \in T$.

Deneckere and Severinov (2008) prove that when players can send any combination of their available messages, any allocation that is truthfully implementable by some mechanism can be truthfully implemented by Γ_d in a truth-telling equilibrium; namely, the profile α where α^i is the identity map for every i is a BNE of Γ_d. Moreover, they offer the following characterization of the class of allocations that are truthfully implementable in terms of the incentive constraints that must hold:

Lemma 1. (Deneckere and Severinov (2008), Corollary 1)
x is truthfully implementable if and only if the following incentive constraints hold for all i, and all t_i, t_i' such that $E_i(t_i') \subseteq E_i(t_i)$:

$$\sum_{t_{-i}} U^i(x(t_i, t_{-i}), v(t_i), v(t_{-i}))\mu(t_{-i}|t_i) \geq \sum_{t_{-i}} U^i(x(t_i', t_{-i}), v(t_i), v(t_{-i}))\mu(t_{-i}|t_i). \tag{1}$$

The lemma captures the key difference, for truthful implementation, between the standard approach in which no verifiable information exists and models in which the agents' ability to manipulate their information is limited; namely, the designer does not have to give incentives to types that want but cannot mimic other types. In what follows I will say that an allocation is **evidence incentive compatible** (EIC) whenever (1) holds for every i and every t_i, t_i' such that $E_i(t_i') \subseteq E_i(t_i)$. Accordingly, I will say that an allocation is **incentive compatible** (IC) whenever (1) holds for every i and *every* t_i, t_i'.

I now define the well–known Bayesian monotonicity notion introduced by Jackson (1991).[8] To do this, the following notation will prove useful. First, for any t_i, any i, and any allocation y, I let the allocation y_{t_i} be defined as $y_{t_i}(s) := y(t_i, s_{-i})$ for every $s \in T$. Second, for any allocation y and profile α, I denote by y_α the allocation that results when every i follows the strategy α^i. Third, for any profile α and any allocation y, I define $y_{\alpha^{-i}}$ as $y_{\alpha^{-i}}(t) := y(t_i, \alpha^{-i}(t_{-i}))$ for every $t \in T$.

[7] See section 5.4 for a discussion.

[8] A weaker version of Bayesian monotonicity as defined by Jackson (1991) was analyzed by Palfrey and Srivastava (1989). Within the class of full support priors, however, the two notions coincide.

Definition 4. x is ***Bayesian monotonic*** (BM) if, for any deception α, there exist i, t_i, and an allocation rule y constant on T_i such that:

$$V^i(y_\alpha, t_i) > V^i(x_\alpha, t_i) \text{ while } V^i(x, t_i') \geq V^i(y_{\alpha^i(t_i)}, t_i') \ \forall t_i'. \tag{2}$$

I will say that an allocation x is **Bayesian monotonic at every feasible deception** if, for any feasible deception α, there exist i, t_i, and y constant on T_i such that (2) holds.

4 Implementation in direct mechanisms

This section introduces a condition that will be useful to identify the class of allocations that can be fully implemented, in the presence of verifiable information, by the direct mechanism.

Definition 5. x is ***evidence distinguishable*** (ED) if, for any deception α, either α is unfeasible or there exist i and t_i such that:

$$V^i(x(\hat{t}_i, \alpha^{-i}(t_{-i})), t_i) > V^i(x_\alpha, t_i) \text{ for some } \hat{t}_i : E_i(\hat{t}_i) \subseteq E_i(t_i). \tag{3}$$

In words: An allocation is evidence distinguishable whenever for any feasible deception there is some agent–type who finds the outcome of some feasible deviation strictly better than the deceptive allocation. The following proposition is the first result of the paper:

Proposition 1. *x is fully implementable by its direct mechanism if and only if x is evidence distinguishable and evidence incentive compatible.*

The proof is in the appendix, but two comments are in order. First, this result relies on neither some restriction on the environment nor the existence of more than two players. Second, the result sheds some light on the role of verifiable information, vis–a–vis that of indirect mechanisms, in enlarging the class of fully implementable allocations. The next proposition shows, however, that in some situations evidence distinguishability can be a demanding condition:

Proposition 2. *If x is evidence distinguishable and incentive compatible, then x is Bayesian monotonic at every feasible deception.*

The proof is in the appendix. Notice that Propositions 1 and 2 jointly entail the following corollary:

Corollary 1. *If x is fully implementable by its direct mechanism but is not Bayesian monotonic at some feasible deception, then x is not incentive compatible.*

Corollary 1 says that when Bayesian monotonicity fails at some feasible deception, fully implementing an incentive compatible allocation by eliciting verifiable information can *only* be done by some indirect mechanism. Since environments in which IC holds but BM does not are the canonical examples that motivate the need for indirect mechanisms, Corollary 1 is then saying that this need cannot be relaxed by the presence of verifiable information. Thus, the value of verifiable information in enlarging the class of fully implementable allocations might be limited in some applications.

The following example illustrates that Proposition 2 cannot be strengthened; namely, that EIC and ED do not necessarily entail Bayesian monotonicity. Thus, the example highlights the difference between Bayesian monotonicity and Bayesian monotonicity at every feasible deception.

Example 1. Let there be two agents, i, and j and let $A = \{a, b, c\}$ be the set of outcomes. Let $\mathcal{E}_i = \{\beta_0, \beta_1\}$ and $\mathcal{E}_j = \{\beta_0\}$ and suppose that $\Theta_i = \{\theta_i\}$ and $\Theta_j = \{\theta_j, \tilde{\theta}_j\}$. Consider a uniform distribution over $T_i \times T_j$ and the following utilities and social choice function x:

(θ_i, θ_j)	$(\theta_i, \tilde{\theta}_j)$	θ_j	$\tilde{\theta}_j$
$c(3)$	$b(4)$	$a(3)$	$c(2.5)$
$b(3)$	$c(4)$	$b(2)$	$a(2.1)$
$a(2)$	$a(3)$	$c(1)$	$b(1)$

	$t_j^1 = (\theta_j, \{\beta_0\})$	$t_j^2 = (\tilde{\theta}_j, \{\beta_0\})$
$t_i^3 = (\theta_i, \{\beta_0\})$	a	b
$t_i^2 = (\theta_i, \{\beta_1\})$	a	c
$t_i^1 = (\theta_i, \{\beta_0, \beta_1\})$	b	a

It is not hard to see that x is both incentive compatible and evidence distinguishable. By Proposition 2, x is then Bayesian monotonic at every feasible deception. However, x fails to be Bayesian monotonic at the unfeasible deception α where $\alpha^i(t_i^2) = t_i^3$, $\alpha^i(t_i^3) = t_i^2$, $\alpha^i(t_i^1) = t_i^1$, and α^j is the identity map. To see this, first notice that the utility of i is independent of her types. Thus, the only possibility for BM to hold at α lies in agent j's incentives. While there are 27 social choice functions constant on T_j that one can check, it is not hard to see that none of these allocations gives the preference reversal required by BM. First, notice that $V^j(x_\alpha, t_j^1) = V^j(x, t_j^1) = 8/3$ and $V^j(x_\alpha, t_j^2) = V^j(x, t_j^2) = (5.6)/3$. Since t_i^2 and t_i^3 mimic each other, the set of outcomes that both t_j^1 and t_j^2 receive with positive probability is the same as what they receive when agent i reports truthfully. As a consequence, every allocation y (constant on T_j) that gives either t_j^1 or t_j^2 an expected payoff larger than $8/3$ or $(5.6)/3$, respectively, would also be larger than what they get with y when agent i reports truthfully. Thus, no allocation constant on T_j can grant the preference reversal required by BM. $\square$

5 Discussion

5.1 Indirect mechanisms

Since Corollary 1 states that direct mechanisms cannot fully implement incentive compatible allocations when BM fails at some feasible deception, one might then be interested in understanding whether, indeed, indirect mechanisms can help. This subsection discusses this question briefly. For any type t_i, let $[t_i] := \{t_i' : E_i(t_i) \subseteq E_i(t_i')\}$ denote the collection of types that could "masquerade" as t_i. The following notion generalizes both BM and ED:

Definition 6. x is ***Bayesian evidence monotonic*** (BEM) if, for any deception α, either α is unfeasible or there exist i, t_i, and an allocation rule y constant on $[t_i]$ such that:

$$V^i(y_{\alpha^{-i}}, t_i) > V^i(x_\alpha, t_i) \text{ while } V^i(x, t_i') \geq V^i(y_{t_i}, t_i') \ \ \forall t_i' : E_i(t_i) \subseteq E_i(t_i'). \tag{4}$$

Two comments are in order. First, notice that BEM generalizes BM because a deception at which Bayesian monotonicity fails need not be feasible. Second, BEM generalizes both BM at every feasible deception and ED because BEM might allow the designer to obtain a preference reversal at some feasible deception where both BM and ED fail. This is easy to see with respect to ED, as an indirect mechanism allows the use of cheap talk messages–other than those made in relation to the preference parameter. To see why BEM also generalizes BM at every feasible deception, notice that a failure of Bayesian monotonicity at some feasible deception implies that enlarging the space of messages—from that in Γ_d—does not help the designer eliminate the deception without compromising the truthful equilibrium. In the presence of verifiable information, however, the same enlargement might help because the state-dependent part of the message some player might use to flag the deceptive behavior might not be feasible to any other type that could disrupt the truthful revelation. This is why the allocation rule requested by BEM need not be independent of i's types, but only constant on every type that could masquerade as the "deviating" type. Thus, BEM is also weaker than requiring BM to hold at every feasible deception.[9] It is not hard to see, moreover, that EIC and ED jointly entail BEM.

On the other hand, one can show that EIC and BEM are both necessary for full implementation by *any* mechanism. In fact, one can find environments in which every allocation that is both EIC and BEM can be fully implemented by an indirect mechanism.[10] The following figure illustrates the relationship among the main notions considered in this paper:

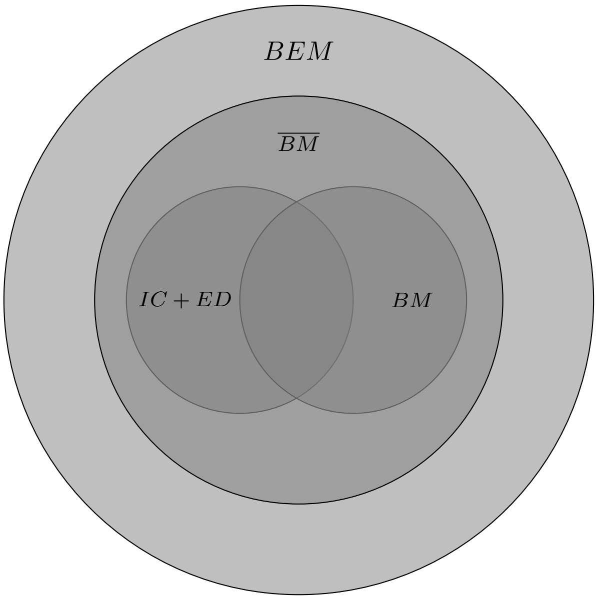

Figure 1: $\overline{BM}$ denotes the class of allocations that are BM at every feasible deception.

5.2 Diffuse priors

The paper assumes the existence of a (common) prior that has full support. Relaxing the assumption does not affect Proposition 2, but affects the full implementation problem (with and without verifiable information) because the mechanism used by the designer is only restricted to mimic the desired allocation on the set of type profiles that receive positive probability. As Jackson (1991) demonstrates in environments without verifiable information, it follows that the designer can fully implement an allocation that is neither incentive compatible nor Bayesian monotonic; all that is needed is the existence of some equivalent allocation that satisfies these properties.[11] The same observation can be used to show that Proposition 1 extends to cases in which the prior does not have a full support; indeed, an

[9]BEM also extends the notion of evidence monotonicity of Kartik and Tercieux (2012) when their analysis is generalized to incomplete information environments but restricted to satisfy the assumptions made in this paper. Indeed, they do not restrict attention to social choice functions or models with hard evidence, and impose no restriction on agents' communication abilities.

[10]An example of such an environment is one in which there are more than two players and each has a type-independent best outcome that, in addition, is preferred over the best outcome of any other agent. For this environment, a straightforward extension of the indirect, integer mechanism in Palfrey and Srivastava (1993) (pp. 27) can be used to fully implement any allocation that is both EIC and BEM. Whether indirect, but "simple," mechanisms can exploit the presence of verifiable information to fully implement allocations that fail to be BM is an interesting open question.

[11]Two allocations x and x' are equivalent if $x(t) = x'(t)$ for every t that receives positive probability under the common prior.

allocation x is fully implementable by *a* direct mechanism if and only if there is an equivalent allocation x' that is both evidence incentive compatible and evidence distinguishable.[12] However, this implies that Corollary 1 does not necessarily hold when the full support assumption is relaxed. Nonetheless, a weaker version of Corollary 1 holds: namely, If an allocation is both evidence incentive compatible and evidence distinguishable but fails to be Bayesian monotonic at some feasible deception, then the allocation is not incentive compatible.

5.3 Normal evidence structures

Assuming that agents can send every combination of their available messages means that their sets of verifiable messages satisfy a condition the literature often calls "normality" (see, for example, Ben-Porath and Lipman (2012)).[13] Roughly, normality requires agents to be able to send, at once, all of their available messages. Whenever not every combination of messages can be sent but normality holds, every result presented above continues to hold. To see this, notice that the inability of some type to send some combination of messages simply means that the type cannot mimic the type that possesses exactly those messages in such a combination. As a consequence, both implementation problems, truthful and full, are "relaxed."[14]

5.4 Mixed strategies

The paper restricts attention to the class of pure strategies. While a large part of the literature has followed a similar restriction, ignoring the possibility that the mechanism being used has some mixed deceptive equilibrium seems particularly troubling, given the spirit of the full implementation exercise.

A basic difficulty that arises in the presence of mixed strategies is that the very notion of full implementation becomes unclear. While it seems clear that the mechanism should deliver the desired outcome at (every profile of pure strategies in the support of) every mixed Bayesian Nash equilibrium, it is not so clear whether one should require, as Serrano and Vohra (2010) do, the existence of some pure Bayesian Nash equilibrium for which the mechanism delivers the desired outcome.[15] Regardless of the notion of mixed full implementation one adopts, it is not hard to see that ED is still necessary if one restricts attention to direct mechanisms. In fact, the natural strengthening of ED that accounts for mixed deceptions—which requires a feasible (and pure) profitable deviation by some agent-type from any mixed deception—is also necessary for mixed full implementation. Thus, if one adopts the notion of mixed full implementation of Serrano and Vohra (2010), Proposition 1 can be extended to the case of mixed strategies; namely, IC and the strengthening of ED jointly characterize the class of allocations that are mixed fully implementable by the direct mechanism.

[12] Notice that when the prior does not have full support, the direct revelation mechanism is not uniquely defined. On the other hand, notice that the necessity of the notion of Bayesian evidence monotonicity discussed above can also be weakened; namely, an allocation x is fully implementable only if there is some equivalent allocation x' that is Bayesian evidence monotonic.

[13] I am grateful to Navin Kartik for pointing this out.

[14] On the other hand, relaxing normality would not be innocuous; truthful implementation might require the use of dynamic mechanisms (see, e.g., Deneckere and Severinov (2008)).

[15] See Mezzetti and Renou (2012) for a discussion of this restriction within complete information environments. In incomplete information environments, the difference is important because, for example, relaxing the requirement by allowing for the existence of some *mixed* Bayesian Nash equilibrium for which the mechanism delivers the desired outcome entails that incentive compatibility is no longer necessary for full implementation. The notion of full implementation adopted by Serrano and Vohra (2010) extends, naturally, the notion in Maskin (1999).

On the other hand, Proposition 2 is still valid regardless of the notion of mixed full implementation one adopts. Moreover, a stronger version would hold: If an allocation satisfies both IC and the strengthening of ED discussed above, then the allocation must satisfy, at every mixed feasible deception, the notion of mixed Bayesian monotonicity proposed by Serrano and Vohra (2010).[16] Accordingly, the natural strengthening of Corollary 1 would also hold.

Acknowledgments

This note is a substantially revised version of one of the chapters of my dissertation submitted to Yale University. I am indebted to Navin Kartik, Takashi Kunimoto and Larry Samuelson for helpful discussions and really grateful to two anonymous referees and the editor in charge of the paper for their comments and suggestions. I thank Yale University, the Cowles Foundation, and the University of Michigan for financial support. Any errors are, of course, my own.

Appendix

Proof of Proposition 1

That x is evidence incentive compatible whenever it is fully implementable follows from Lemma 1. To see that evidence distinguishability is also necessary, suppose that x is fully implementable by Γ_d but not evidence distinguishable. Thus, there is some deception α such that inequality (3) in the paper does not hold. Then, for every i, every t_i and every $\hat{t}_i$ such that $t_i \in [\hat{t}_i]$:

$$V^i(x_\alpha, t_i) \geq V^i(x(\hat{t}_i, \alpha^{-i}(t_{-i})), t_i).$$

Since α is feasible, it follows that α is a BNE of Γ_d; contradicting that x is fully implementable by Γ_d. To see the other direction, notice that since x is evidence incentive compatible, Lemma 1 entails that x is truthfully implementable. Hence, it is sufficient to show that no undesirable equilibrium arises in Γ_d. Suppose that there is some deception α that is a BNE of Γ_d. Thus:

$$V^i(x_\alpha, t_i) \geq V^i(x(\hat{t}_i, \alpha^{-i}(t_{-i})), t_i),$$

for every i, every t_i, and every $\hat{t}_i$ such that $t_i \in [\hat{t}_i]$. Since α is feasible, it follows that x is not evidence distinguishable. $\square$

Proof of Proposition 2

Take any feasible deception α. Since x is evidence distinguishable, there is an agent i and a type t_i such that, for some $t_i \in [\hat{t}_i]$:

[16] Mixed Bayesian monotonicity extends BM by requiring a preference reversal at every mixed deception. Interestingly, Serrano and Vohra (2010) show that when one works with *random* social choice functions, mixed Bayesian monotonicity is equivalent to Bayesian monotonicity. It is not hard to see that the same equivalence would hold between BEM and mixed BEM, the natural extension of BEM that accounts for mixed deceptions. It is unclear to me whether these equivalences hold when one restricts attention to deterministic social choice functions (as the present paper does). However, no equivalence between ED and its strengthening exists when one focuses attention on deterministic social choice functions. Indeed, in example 1 above, the mixed deception where t_i^1 reports t_i^1 and t_i^2 with probability 1/2 and every other type of both agents reports truthfully, is a mixed Bayesian Nash equilibrium of the direct mechanism.

$$V^i(x(\hat{t}_i, \alpha^{-i}(t_{-i})), t_i) > V^i(x_\alpha, t_i).$$

Construct an allocation y by setting $y(t_i, t_{-i}) := x(\hat{t}_i, t_{-i})$ for every t_i and every t_{-i}. Hence, y is constant on T_i. Moreover, y_α is defined as

$$y_\alpha(t_i, t_{-i}) = y(\alpha^i(t_i), \alpha^{-i}(t_{-i})) = x(\hat{t}_i, \alpha^{-i}(t_{-i})).$$

Hence, it follows that

$$V^i(y_\alpha, t_i) > V^i(x_\alpha, t_i).$$

However, $y_{\alpha^i(t_i)}(t_i, t_{-i}) = x(\hat{t}_i, t_{-i})$ and x is incentive compatible. Hence

$$V^i(x, t_i') \geq V^i(y_{\alpha^i(t_i)}, t_i')$$

for every t_i'. Since this argument holds for any feasible deception α, it follows that x is Bayesian monotonic at every feasible deception. □

6 References

Ben-Porath, Elchanan, and Lipman, Barton L. 2012. Implementation with partial provability. *Journal of Economic Theory*, **147**(5), 1689–1724.

Bull, Jesse, and Watson, Joel. 2007. Hard evidence and mechanism design. *Games and Economic Behavior*, **58**(1), 75–93.

Deneckere, Raymond, and Severinov, Sergei. 2008. Mechanism design with partial state verifiability. *Games and Economic Behavior*, **64**(2), 487–513.

Green, Jerry R, and Laffont, Jean-Jacques. 1986. Partially verifiable information and mechanism design. *The Review of Economic Studies*, **53**(3), 447–456.

Grossman, Sanford J. 1981. The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, **24**(3), 461–483.

Jackson, Mathew O. 1991. Bayesian implementation. *Econometrica: Journal of the Econometric Society*, 461–477.

Jackson, Matthew O. 1992. Implementation in undominated strategies: A look at bounded mechanisms. *The Review of Economic Studies*, **59**(4), 757–775.

Kartik, Navin, and Tercieux, Olivier. 2012. Implementation with evidence. *Theoretical Economics*, **7**(2), 323–355.

Kartik, Navin, Tercieux, Olivier, and Holden, Richard. 2014. Simple mechanisms and preferences for honesty. *Games and Economic Behavior*, **83**, 284–290.

Maskin, Eric. 1999. Nash equilibrium and welfare optimality. *The Review of Economic Studies*, **66**(1), 23–38.

Mezzetti, Claudio, and Renou, Ludovic. 2012. Implementation in mixed Nash equilibrium. *Journal of Economic Theory*, **147**(6), 2357–2375.

Milgrom, Paul, and Roberts, John. 1986. Relying on the information of interested parties. *The RAND Journal of Economics*, 18–32.

Milgrom, Paul R. 1981. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380–391.

Mookherjee, Dilip, and Reichelstein, Stefan. 1990. Implementation via augmented revelation mechanisms. *The Review of Economic Studies*, **57**(3), 453–475.

Ortner, Juan. 2015. Direct implementation with minimally honest individuals. *Games and Economic Behavior*, **90**, 1–16.

Palfrey, Thomas R, and Srivastava, Sanjay. 1989. Implementation with incomplete information in exchange economies. *Econometrica: Journal of the Econometric Society*, 115–134.

Palfrey, Thomas R, and Srivastava, Sanjay. 1993. Bayesian implementation. **53**.

Postlewaite, Andrew, and Schmeidler, David. 1986. Implementation in differential information economies. *Journal of Economic Theory*, **39**(1), 14–33.

Serrano, Roberto, and Vohra, Rajiv. 2005. A characterization of virtual Bayesian implementation. *Games and Economic Behavior*, **50**(2), 312–331.

Serrano, Roberto, and Vohra, Rajiv. 2010. Multiplicity of mixed equilibria in mechanisms: A unified approach to exact and approximate implementation. *Journal of Mathematical Economics*, **46**(5), 775–785.