Evidence Reading Mechanisms*

Frédéric Koessler† Eduardo Perez-Richet‡

March 13, 2017

Abstract

We study implementation with privately informed agents who can produce evidence. We characterize social choice functions that are implementable by mechanisms that are robust to the designer's commitment power, and simply apply the social choice function to a reading of the evidence. In this class of mechanisms, our results provide conditions on the evidence structure such that (i) a function that is implementable with transfers is also implementable with evidence but no transfer, (ii) under private value, the efficient allocation is implementable with budget balanced and individually rational transfers, and (iii) in single-object auction and bilateral trade environments with interdependent values, the efficient allocation is implementable with budget balanced and individually rational transfers.

Keywords: Implementation, Mechanism Design, Evidence, Hard Information.

JEL classification: C72; D82.

1 Introduction

The usual mechanism design approach assumes cheap talk communication: messages are equally accessible to all types of agents, regardless of what they know. Under this assumption, a principal who wants to implement a contingent plan of action must be able to deter all possible lies. In this paper, following a thin but grounded tradition in the mechanism design literature, we assume that privately informed agents have access to evidence. We provide sufficient conditions for implementability by simple mechanisms and show that they apply to familiar environments like auctions and bilateral trade. These conditions also allow us to characterize evidence structures such that a social choice function that can be implemented with transfers can also be implemented with evidence but no transfer.

*We thank Elhanan Ben Porath, Jeanne Hagenbach, Johannes Hörner, Matt Jackson, Emir Kamenica, Navin Kartik, David Levine, Ludovic Renou, Phil Reny and Olivier Tercieux for useful discussions, comments and suggestions. We also thank seminar and workshop participants at Ecole Polytechnique, PSE, Concordia University, the University of Warwick, the workshop on Mathematical Aspects of Game Theory and Applications in Roscoff, and the Transatlantic Theory Workshop. Eduardo Perez-Richet thanks Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), and iCODE (Institute for Control and Decision), research project of the Idex Paris-Saclay, for financial support. Frédéric Koessler thanks the French National Research Agency for financial support.

†Paris School of Economics – CNRS, e-mail: frederic.koessler@psemail.eu

‡Economics Department, Sciences Po Paris, e-mail: eduardo.perez@sciencespo.fr

We focus on *reading mechanisms*. Such mechanisms simply apply the contingent plan of action of the principal (the social choice function) to a reading of the evidence, that is, a consistent interpretation of each message profile. To justify this approach think of the social choice function as the first-best plan of action for the principal, which she would implement if she knew the agents' types. The principal may know her first-best action plan but be unsure about the remaining details of her preferences. The use of a *reading mechanism* ensures that, for all preferences of the principal that are compatible with her first-best, the actions mandated by the mechanism are optimal given beliefs that are consistent with the evidence provided by the agents. In particular, if her first-best is implementable by a reading mechanism with *accurate readings* on the equilibrium path, then, regardless of her particular preferences, it can be implemented as a perfect Bayesian equilibrium of the disclosure game in which the principal would choose her action as mandated by the mechanism *after* having observed the agents' reports.[1] Hence reading mechanisms are robustly immune against a lack of commitment by the principal, a property that makes them more credible. Reading mechanisms also have additional properties that make them desirable in practice and may be perceived as more legitimate. For example, an agent may sue an institution for treating him in a way that is not compatible with its mission (the social choice function) given the evidence. Finally, they are, by construction, deterministic.

Implementation by an accurate reading mechanism requires two intuitive conditions. First, the evidence structure must be sufficiently rich to provide each informational type of each agent with a message that conveys her information. This message must be such that no other informational type of the same agent would be both willing to and capable of using the same message. We call this the *evidence base condition*. Second, the principal must be able to deter participants from using other messages than those in the evidence base. Her only lever to do so is her reading of evidence. Thus, implementation is possible whenever the principal can have a skeptical and consistent interpretation of each participant's message. This is the case if the set of types that could have sent a given message admits a *worst case type*, that is a type that no other type for whom the message was available would have been willing to masquerade as.

The literature on evidence, or hard information, starts with the sender-receiver models of Grossman and Hart (1980), Grossman (1981) and Milgrom (1981). One branch of the ensuing literature has tried to identify

[1] Glazer and Rubinstein (2006) show that this perfect Bayesian equilbrium property holds in their binary action persuasion framework with given preferences for the principal, Sher (2011) and Hart, Kremer, and Perry (2017) extend it to non binary environments. Ben-Porath, Dekel, and Lipman (2017) show in a different environment that there is no value to commitment, in the sense that the principal could get the same payoff in the optimal mechanism as in the disclosure game.

conditions that preclude the full disclosure result of the seminal papers (Milgrom and Roberts, 1986, Shin, 2003, Wolinsky, 2003, Dziuda, 2011). More importantly for this paper, another branch has sought to extend the full revelation result to more general settings: Okuno-Fujiwara, Postlewaite, and Suzumura (1990) consider information disclosure preceding a game, while Seidmann and Winter (1997) and Giovannoni and Seidmann (2007) consider sender-receiver games à la Crawford and Sobel (1982) and introduce the notion of worst case type which we use in this paper. Hagenbach, Koessler, and Perez-Richet (2014) work with this notion to build a theory that includes and extends former full revelation results, and develop ideas that play an important role in this paper. Lipman and Seppi (1995) give full revelation results in a different environment, with symmetrically informed senders and sequential communication.

The literature on mechanism design with evidence starts with Green and Laffont (1986). In a principal agent model, they provide a necessary and sufficient condition on the evidence structure (the nested range condition) for a revelation principle to hold. Singh and Wittman (2001) show how to implement social choice functions without this condition but with monotonic preferences in the allocation. Bull and Watson (2007), Deneckere and Severinov (2008) and Strausz (2016) extend the revelation principle to more general mechanism design setups and message structures, while Forges and Koessler (2005) develop similar results for communication games.

Glazer and Rubinstein (2004) and Glazer and Rubinstein (2006) consider the role of evidence in a simple persuasion problem in which an informed sender seeks to persuade an uninformed receiver to take an action that she only wishes to take in particular circumstances. Glazer and Rubinstein (2004) considers a mechanism in which the sender first declares her type, and then may be asked to produce evidence. They develop a linear programming approach that allows them to characterize an optimal mechanism which is of the claim and verify form, and show how the receiver can benefit from randomization at the verification stage. Closer to our paper, Glazer and Rubinstein (2006) assume that the sender directly sends evidence, so a mechanism is a *persuasion rule* that states which messages are deemed convincing by the receiver. They show that randomization is not needed, and provide a procedure to find an optimal persuasion rule. In both papers, the optimal mechanism is shown to be credible in the sense that the behavior of the sender and the receiver constitutes a perfect Bayesian equilibrium of the associated disclosure game (see also Sher, 2011, 2014, Hart et al., 2017 and Ben-Porath et al., 2017).

Sher and Vohra (2014) considers the possibility of using evidence in a monopolistic price discrimi-

nation framework and solve a modified version of the program that gives the optimal direct mechanism. This modified program is obtained by erasing the incentive constraints from a type t to a type s if the evidence structure precludes t from showing all the evidence available to s.

This work is also connected to the literature on full implementation à la Maskin (1999) with evidence. Kartik and Tercieux (2012) consider Nash implementation, whereas Ben-Porath and Lipman (2012) consider subgame perfect implementation. Ben-Porath, Dekel, and Lipman (2014) consider costly information verification.

For simplicity, the main text only contains sufficient conditions for ex-post implementation. In the appendix, we present the analogue of our general results for interim implementation, and we provide necessary and sufficient conditions for interim and ex-post implementation under additional constraints.

2 The Model

There is a set N of n agents indexed by i, and a set of alternatives denoted by $\mathcal{A}$. Each agent has a type t_i which encodes her privately observed information. The set of possible realizations of this random variable is a finite set $\mathcal{T}_i$, and $\mathcal{T} = \mathcal{T}_1 \times \cdots \times \mathcal{T}_n$ is the set of type profiles. The utility of agent i when alternative a is implemented is $u_i(a;t)$, where $t = (t_1, \cdots, t_n)$. We say that i has private values if $u_i(a;t)$ is independent of t_{-i}.

The evidence structure is defined by a finite message space $\mathcal{M}_i$, and a correspondence $M_i : \mathcal{T}_i \rightrightarrows \mathcal{M}_i$ for each agent, where $M_i(t_i)$ is the set of messages available to agent i of type t_i. A subset $\mathcal{S}_i \subseteq \mathcal{T}_i$ is certified by a message m_i if $M_i^{-1}(m_i) = \mathcal{S}_i$, where $M_i^{-1}(m_i) \equiv \left\{t_i \in \mathcal{T}_i \,|\, m_i \in M_i(t_i)\right\}$. $\mathcal{S}_i$ is certifiable if there exists a message m_i that certifies $\mathcal{S}_i$, that is, a message which is available to all the types in $\mathcal{S}_i$, and to none other. We say that the evidence structure $\left(\mathcal{M}_i, M_i\right)_{i=1}^n$ satisfies own type certifiability if, for every agent i, and every $t_i \in \mathcal{T}_i$, the set $\{t_i\}$ is certifiable.

The set of consistent interpretations of a message profile m, which we call the *set of readings* of m, is given by $R(m) = M_1^{-1}(m_1) \times \cdots \times M_n^{-1}(m_n) \subseteq \mathcal{T}$. A *reading of the evidence* is a function $\rho : \mathcal{M} \to \mathcal{T}$ such that, for every m, $\rho(m) \in R(m)$. It is an interpretation of each possible message profile as a type profile that is consistent with the evidence.

A *social choice function* is a mapping $f : \mathcal{T} \to \mathcal{A}$. We consider only deterministic and static mechanisms. Since we take the evidence structure as given, a mechanism is then simply given by a

deterministic *outcome function* $g : \mathcal{M} \to \mathcal{A}$, which determines the alternative chosen by the designer following every possible message profile. In the game defined by the mechanism $g(\cdot)$, each agent chooses a messaging strategy $\mu_i : \mathcal{T}_i \to \mathcal{M}_i$ such that $\mu_i(t_i) \in M_i(t_i)$.

A messaging strategy profile $\mu(\cdot)$ is an ex post equilibrium of the game generated by the mechanism $g(\cdot)$ if, for every type profile t, every agent i, and every message $m_i \in M_i(t_i)$, $u_i\big(g(\mu(t); t\big) \geq u_i\big(g(m_i, \mu_{-i}(t_{-i})); t\big)$. A mechanism $g(\cdot)$ ex post implements the social choice function $f(\cdot)$ if there exists an ex post equilibrium $\mu(\cdot)$ of the game generated by $g(\cdot)$, such that $g\big(\mu(t)\big) = f(t)$ for every $t \in \mathcal{T}$. In the remainder of the paper, we will refer to ex post implementation simply as implementation, and to ex post equilibrium as equilibrium.[2]

A mechanism $g(\cdot)$ is a *reading mechanism* if the alternative chosen by the mechanism designer is always consistent with the messages she receives and the social choice function she wants to implement; that is, $g(m) \in f\big(R(m)\big)$ for every message profile $m \in \mathcal{M}$.

The outcome function of a reading mechanism is completely pinned down by a reading of the evidence. Indeed, the outcome function $g(\cdot)$ of a reading mechanism can always be defined as the action $f\big(\rho(m)\big)$ for some reading $\rho(\cdot)$. In such mechanisms, the designer only decides how to read the evidence, and that determines the outcome. To each reading corresponds a unique reading mechanism, but different readings may generate the same mechanism if the social function is not one to one.

We say that a reading mechanism $\rho(\cdot)$ *accurately implements* $f(\cdot)$ if it does so with *accurate readings* on the equilibrium path, that is, if there exists a strategy profile $\mu(\cdot)$ that forms an equilibrium of the game induced by $\rho(\cdot)$ and satisfies $\rho(\mu(t)) = t$, for every $t \in T$.

The payoff for agent i to masquerade as another type s_i when she is really of type t_i, under a social choice function $f(\cdot)$, is given by the following masquerading payoff function:

$$v_i(s_i|t_i; t_{-i}) = u_i\big(f(s_i, t_{-i}); t_i, t_{-i}\big).$$

These payoff functions represent the incentives of agents of given types to masquerade as other types. These incentives are determined by the social choice function and the preferences of the agents. For each agent i, and each type profile t_{-i}, they can be summarized by an oriented graph on $\mathcal{T}_i$, such that a type t_i points to a type s_i if t_i has an incentive to masquerade as s_i. Following Hagenbach et al. (2014), we call the relation that defines this graph a *masquerade relation*: we say that t_i wants to

[2]Interim implementation is discussed in the appendix.

masquerade as s_i given t_{-i}, denoted by $t_i \xrightarrow{\mathfrak{M}[t_{-i}]} s_i$, if and only if $v_i(s_i|t_i;t_{-i}) > v_i(t_i|t_i;t_{-i})$. For a generic masquerade relation, we will use the notation $\to$.

We can use this relation to define a *worst-case type* (given t_{-i}) for $\mathcal{S}_i \subseteq \mathcal{T}_i$ as a type in $\mathcal{S}_i$ that no other type in $\mathcal{S}_i$ would like to masquerade as. We denote the set of such types as

$$\text{wct}(\mathcal{S}_i|t_{-i}) := \left\{ s_i \in \mathcal{S}_i \,|\, \nexists\, t_i \in \mathcal{S}_i, t_i \xrightarrow{\mathfrak{M}[t_{-i}]} s_i \right\}.$$

Graphically, a worst case type is a type in $\mathcal{S}_i$ with no incoming arc from any other type in $\mathcal{S}_i$. The set of worst case type may be empty, or have more than one element. A masquerade relation $\to$ on $\mathcal{T}_i$ admits a cycle $(t_i^1, \ldots, t_i^k)$ if $t_i^1 \to t_i^2 \to \cdots t_i^k \to t_i^1$.

Lemma 1 (Acyclicity and Worst Case Types)**.** *Let $\to$ be a masquerade relation on an individual type set $\mathcal{T}_i$. The following points are equivalent: (i) $\to$ is acyclic; (ii) Every nonempty subset $\mathcal{S}_i \subseteq \mathcal{T}_i$ admits a worst case type; (iii) There exists a function $w : \mathcal{T}_i \to \mathbb{R}$ such that $t_i \to s_i \;\Rightarrow\; w(s_i) > w(t_i)$.*

Proof. See Proposition 1 in Hagenbach et al. (2014). □

If condition (iii) holds, we say that the masquerade relation is *weakly represented* by the function $w(\cdot)$. The representation is weak because we have an implication rather than an equivalence. If it were an equivalence, the masquerade relation would be a linear ordering of types, as with the utility representation of rational preferences.

An *evidence base* for agent i is a base of messages $\mathcal{E}_i \subseteq \mathcal{M}_i$ such that there exists a one-to-one function $e_i : \mathcal{T}_i \to \mathcal{E}_i$ that satisfies $e_i(t_i) \in M_i(t_i)$ and $t_i \in \bigcap_{t_{-i} \in \mathcal{T}_{-i}} \text{wct}\big(M_i^{-1}\big(e_i(t_i)\big)\big|t_{-i}\big)$ for every t_i. Whenever own type certifiability is satisfied for an agent, her message correspondence admits an evidence base, regardless of preferences and the social choice function under consideration.

Example 1 (Evidence Base)**.** As an illustration, consider a single privately informed agent i with three possible types, $\mathcal{T}_i = \{t^1, t^2, t^3\}$, whose masquerade relation is given by $t^1 \to t^2 \to t^3$. The message correspondence $M_i(t^1) = \{m^1, m^3, m^4\}$, $M_i(t^2) = \{m^1, m^2, m^4\}$, $M_i(t^3) = \{m^1, m^2, m^3\}$ admits two evidence bases: $\mathcal{E}_i = \{m^1, m^2, m^3\}$ and $\mathcal{E}_i = \{m^4, m^2, m^3\}$. On the contrary, the message correspondence $M_i(t^1) = \{m^1, m^4\}$, $M_i(t^2) = \{m^1, m^2, m^3, m^4\}$, $M_i(t^3) = \{m^1, m^3\}$ does not admit any evidence base because type t^3 has no message certifying an event for which it is a worst case type. ◇

In the usual mechanism design formulation, the designer is allowed to design a space of messages that are accessible for free to all types. We have not done that because if such messages exist, they can be described in our framework. However, it is interesting to think about the introduction of such messages as a way to relax the evidence base condition in the characterization results of the next sections. To see that, suppose that the designer can create a set of additional messages for each agent that are accessible for free irrespective of types, and let $\hat{\mathcal{M}}_i$ be the set of such messages for agent i. We consider the new evidence structure $\tilde{\mathcal{M}}_i = \mathcal{M}_i \times \hat{\mathcal{M}}_i$ in which agent i of type t_i can send any message in $M_i(t_i) \times \hat{\mathcal{M}}_i$. We refer to structures that can be obtained in this way as *cheap talk completions* of $\big(\mathcal{M}_i, M_i(\cdot)\big)_{i=1}^n$.

Proposition 1. *There exists a cheap talk completion of* $\big(\mathcal{M}_i, M_i(\cdot)\big)_{i=1}^n$ *such that agent* i *has an evidence base if and only if, for every* t_i*, there exists a message* $m_i \in \mathcal{M}_i$ *such that* $t_i \in \bigcap_{t_{-i}} \text{wct}\big(M_i^{-1}(m_i)|t_{-i}\big)$.

Proof. Suppose that there exists a cheap talk completion with an evidence base for agent i. Since all types of i have access to the messages in $\hat{\mathcal{M}}_i$, the fact that t_i has a message $(m_i, \hat{m}_i) \in \tilde{\mathcal{M}}_i$ for which it is a worst case type, implies that it is a worst case type of the set certified by $m_i \in M_i(t_i)$. Suppose now that every type has a message such that it is a worst case type of the set certified by this message. Then the only reason why an evidence base may not exist in the initial evidence structure would be that several types, say two types t_i and s_i are worst case types of the same message m_i. But then, allowing t_i and s_i to send the same piece of evidence m_i, and another cheap talk message, $\hat{m}_i$ for t_i, and $\hat{m}_i'$ for s_i, would mean that, in the new structure obtained by completing with $\hat{m}_i$ and $\hat{m}_i'$, they each have a different message for which they are a worst case type. □

Intuitively, when we allow for such completions, we will have an evidence base for agent i as long as any type of agent i has access to a piece of evidence that rules out any type s_i that would like to masquerade as t_i, under some profile t_{-i}. In a framework where masquerading incentives are monotonic, as in the seller-buyer model of Milgrom (1981), an agent of a given type (for example quality), must be able to rule out any lower type. This condition is related to the distinguishability condition[3] in Kartik and Tercieux (2012), and to the incentive compatibility conditions derived in Deneckere and Severinov (2008).

[3]In their work on Nash implementation, a type must be able to disprove any other type such that the pair would violate Maskin monotonicity.

3 Sufficient Conditions for Implementation

The existence of an evidence base is important for implementation, because it allows the agents to convey their type with a message that no other type would both want and be able to imitate. For implementation to be possible, the second important requirement is for the principal to be able to punish deviators. With reading mechanisms, the principal can only punish a deviator with a consistent but skeptical reading, that is by attributing, for every realization of t_{-i}, the piece of evidence m_i sent by the deviator to a type t_i which is a worst case type among the types that could have sent m_i. These intuitions are formalized in the following theorem. The proof is by construction, and is similar to the construction of a fully revealing equilibrium in Hagenbach et al. (2014). The evidence base of an agent provides a natural candidate for her equilibrium strategy, so we construct a mechanism that reads each message from this evidence base accurately. Then we can complete the reading by interpreting each message profile comprising a unilateral deviation from equilibrium messages as the correct type profile for the non deviators, and a worst case type for the deviator. It is easy to see that such readings make unilateral deviation non profitable.

Theorem 1. *There exists a reading mechanism that accurately implements $f(\cdot)$ if the following conditions hold for every agent i: (i) For every $t_{-i} \in \mathcal{T}_{-i}$, and every message $m_i \in \mathcal{M}_i$, the set $M_i^{-1}(m_i)$ admits a worst case type given t_{-i}; (ii) $M_i(\cdot)$ admits an evidence base.*

Proof. We construct an accurate reading as follows. For every i, let $e_i : \mathcal{T}_i \to \mathcal{M}_i$ be a one-to-one mapping associated with an evidence base of agent i. Consider a message profile m such that for every $i \neq j$, the message m_i is in the range of e_i. Then if m_j is also in the range of e_j, the reading of the message profile is $\rho_j(m_j, m_{-j}) = e_j^{-1}(m_j)$, and $\rho_i(m_j, m_{-j}) = e_i^{-1}(m_i)$ for every $i \neq j$. If on the other hand, m_j is not in the range of e_j, then $\rho_i(m_j, m_{-j}) = e_i^{-1}(m_i)$ for every $i \neq j$, whereas the message of agent j is interpreted as a type in $\text{wct}\big(M_j^{-1}(m_j)|\rho_{-j}(m_j, m_{-j})\big)$.

Then the strategy profile e is fully revealing. It is also an (ex-post) equilibrium. Indeed if all agents but i use this strategy profile, then a message m_i of agent i that does not belong to the range of e_i is interpreted as a type in $\text{wct}\big(M_i^{-1}(m_i)|t_{-i}\big)$ for every t_{-i}. Hence such a deviation does not benefit to agent i. Another possible deviation would be to send a message in the range of e_i that differs from $e_i(t_i)$, call it $e_i(t_i')$, when i's type is really t_i. But then this message is interpreted as t_i' regardless of t_{-i}, and because t_i' is a worst case type of $e_i(t_i')$ given any t_{-i}, agent i does not gain from the deviation if her true type is t_i. □

In order to do optimal mechanism design, that is to find the optimal contingent plan for a principal with given preferences, it is useful to be able to characterize the set of implementable contingent plans in a tractable manner. The conditions of Theorem 1 are not easy to deal with. Furthermore, it is easier to work with a condition that involves only the preferences of the agent, rather than the message structure. We can characterize the set of social choice functions that are implementable under any message structure that satisfies own type certifiability as being the set of social choice functions that generate acyclic masquerade relations.

Corollary 1. *There exists a reading mechanism that accurately implements $f(\cdot)$ under any message structure that satisfies own type certifiability if for every agent i and every $t_{-i} \in \mathcal{T}_{-i}$ the masquerade relation $\xrightarrow{\mathfrak{M}[t_{-i}]}$ is acyclic on $\mathcal{T}_i$.*

Proof. Own type certifiability ensures that the evidence base condition is satisfied. Then to implement $f(\cdot)$ it is sufficient to satisfy the worst case type condition of Theorem 1. By Lemma 1, we know that it is satisfied for every certifiable subset if and only if the masquerade relation is acyclic. □

4 Evidence and Transfers

In this section, we assume that the agents have quasilinear preferences. The preferences of agent i over alternatives are still represented by the function $u_i(a;t)$, which we now interpret as the valuation of the agent. If agent i is given a transfer τ_i, her utility is given by $u_i(a;t) + \tau_i$. Our goal is to compare transfers and evidence as tools for implementation, and to give a first assessment of what can be achieved by using them as complements. For that, we start by introducing a few notations.

In an evidence-free message structure, every mechanism is a reading mechanism. In this case, the following incentive compatibility conditions are necessary and sufficient conditions for implementability

Definition 1 (Evidence-Free Incentive Compatibility)**.** *A social choice function satisfies ex post incentive compatibility if, for every $t \in \mathcal{T}$, every agents i and every $s_i \in \mathcal{T}_i$*

$$v_i(s_i|t_i, t_{-i}) \leq v_i(t_i|t_i, t_{-i}). \tag{EPIC}$$

When using transfers, the mechanism designer can modify the incentives of the agents. We will therefore consider ex post transfer functions $\tau_i : \mathcal{T} \to \mathbb{R}$, and the corresponding modified masquerading

payoff $V_i(s_i|t_i;t_{-i}) = v_i(s_i|t_i;t_{-i}) + \tau_i(s_i;t_{-i})$. We start with a simple example showing that reading mechanisms can sometimes achieve implementation in situations where transfers cannot.

Example 2 (Evidence 1 – Transfers 0)**.** Consider a setup with one agent of two possible types t and t', and an evidence structure that satisfies own type certifiability. The social choice function selects action a when the type is t, and action a' when the type is t'. The preferences of the agent are given by $u(a,t) = u(a',t') = 0$, $u(a',t) = 2$, $u(a,t') = -1$. Therefore t wants to masquerade as t', but t' does not want to masquerade as t, hence the masquerade relation is acyclic, and the social choice function is implementable with evidence. It is not implementable with transfers, because any transfer that is sufficient to discourage t from claiming t' makes t' claim to be t. ◇

In fact, every social choice function that is implementable with transfers can be accurately implemented by a reading mechanism as long as the evidence base condition is satisfied. The intuition is simple: the worst case type associated with any subset of types is the type that would have received the highest transfer. To see that, note that if some other type, with a lower transfer, could obtain a better outcome by pretending to be this highest transfer type, then this type would be even more incentivized to claim being the highest transfer type under the transfer scheme because she would also get a higher transfer. But then, that would contradict the fact that the transfer scheme implements the social choice function. More precisely, the proof shows that the negative of the transfer function provides a weak representation of the masquerade relation, and then concludes by Lemma 1.

Theorem 2. *If $f(\cdot)$ is implementable with transfers and no evidence, it is also accurately implementable by a reading mechanism under any evidence structure such that each $M_i(\cdot)$ admits an evidence base. Furthermore, there exist social choice functions that are implementable by a reading mechanism under any evidence structure such that each $M_i(\cdot)$ admits an evidence base, but not with transfers.*

Proof. The social choice function is implementable with transfers if and only if $V_i(s_i|t_i;t_{-i})$ satisfies (EPIC). This implies that $-\tau_i(\cdot;t_{-i})$ is a weak representation of the masquerade relation of agent i given t_{-i}, allowing us to conclude by Lemma 1 and Theorem 1. The second part of the theorem is proved by Example 2. □

5 Efficient Mechanism Design with Private Values

In this section, we consider the problem of implementing an efficient social choice function $f(t) \in \arg\max_{a\in\mathcal{A}} \sum_i u_i\big(a;t\big)$ and allow the mechanism designer to use both evidence and transfers. Given a social choice function $f(\cdot)$, an ex-post transfer scheme τ_i for $i = 1,\dots,n$ is individually rational if, for every agent i, and every type profile t, $V_i(t_i|t_i;t_{-i}) = v_i(t_i|t_i;t_{-i}) + \tau_i(t) \geq 0$. It is budget balanced if, for every type profile t, $\sum_i \tau_i(t) \leq 0$. It fully extracts surplus if, for every type profile t, $\sum_i \tau_i(t) = -\sum_i v_i(t_i|t_i;t_{-i})$. To make it possible to satisfy both individual rationality and budget balance, we assume that, for every type profile t, $\sum_i v_i(t_i|t) \geq 0$.

Theorem 3. *Under private values, any efficient social choice function with any transfer scheme such that an evidence base is available for each agent can be accurately implemented by a reading mechanism. In particular, under own-type certifiability, the transfer scheme can be chosen to satisfy individual rationality and budget balance, and even extract full surplus.*

Proof. The proof is related to the classical Vickrey-Clarkes-Groves mechanism. Let $h_i(t) = \tau_i(t) - \sum_{j\neq i} u_j\big(f(t);t_j\big)$ denote what remains of agent i's transfer after subtracting her externality on other participants. Then i's incentive to masquerade as s_i when her type is t_i is given by

$$\begin{aligned} V_i(s_i|t_i,t_{-i}) - V_i(t_i|t_i,t_{-i}) &= u_i\big(f(s_i,t_{-i});t_i\big) + \sum_{j\neq i} u_j\big(f(s_i,t_{-i});t_j\big) + h_i(s_i,t_{-i}) \\ &\quad - u_i\big(f(t_i,t_{-i});t_i\big) - \sum_{j\neq i} u_j\big(f(t_i,t_{-i});t_j\big) - h_i(t_i,t_{-i}) \leq h_i(s_i,t_{-i}) - h_i(t_i,t_{-i}), \end{aligned}$$

where the inequality is a consequence of the fact that $f(t_i,t_{-i})$ maximizes the sum $\sum_i u_i\big(a;t_i\big)$. But then $h_i(\cdot,t_{-i})$ is a weak representation of i's masquerade relation given t_{-i} which is therefore acyclic by Lemma 1. We can conclude with Theorem 1. □

In a way, this result is almost a corollary of Theorem 2. Since, under private values, an efficient social choice function can be implemented with transfers by a VCG mechanism, then it can also be accurately implemented with evidence and no transfers. The value added of Theorem 3 is to show that, with evidence, transfers can be chosen to satisfy individual rationality and budget balance, which is not possible in general with VCG mechanisms.

Since full surplus extraction can be achieved under private values, one might wonder whether, with evidence, it is ever necessary to pay an information rent in order to achieve efficiency. In the next

sections, we show that full surplus extraction can be achieved in single-object auctions and bilateral trade with interdependent valuations.

6 Auctions

In this section, we explore the consequence of relaxing the private value assumption in auction environments. We also provide examples of situations where reading mechanisms fail.

The agents have quasilinear utilities, and agent i's valuation of the single object for sale is given by a function $\mathfrak{u}_i(t) \geq 0$ that depends on the full type profile t. An auction (a social choice function) is a rule for allocating the object to one of the agents $\alpha : \mathcal{T} \to N$, and a positive[4] price function $\pi : \mathcal{T} \to \mathbb{R}_+$ for the winner of the auction.

An auction is individually rational if it never requires the winner to pay a price higher than her valuation, that is $\pi(t) \leq \mathfrak{u}_{\alpha(t)}(t)$. It is efficient if it allocates the good to one of the agents with the highest valuation, that is $\alpha(t) \in \arg\max_i \mathfrak{u}_i(t)$. It is fully extractive if it is efficient and $\pi(t) = \mathfrak{u}_{\alpha(t)}(t)$.

Theorem 4 (Single-Object Auctions)**.** *Any individually rational auction such that an evidence base is available for each agent is accurately implementable by a reading mechanism. In particular, under own-type certifiability, the fully extractive auction is accurately implementable.*

Proof. Pick an agent i, and fix t_{-i}. We can split the type set of agent i into two regions, the set of types for which she does not get the good, $\mathcal{T}_i^0$, and the set of types for which she obtains the good, $\mathcal{T}_i^+$. First, note that any type masquerading as a type in $\mathcal{T}_i^0$ forgoes the good and gets a payoff of 0. Second, any type in $\mathcal{T}_i^+$ obtains a nonnegative payoff by masquerading as her true type, because the auction is individually rational. These two observations imply that no type wants to masquerade as a type in $\mathcal{T}_i^0$, and therefore, if the masquerade relation $\xrightarrow{\mathfrak{M}[t_{-i}]}$ admits a cycle on $\mathcal{T}_i$, then all the types involved in the cycle must lie in $\mathcal{T}_i^+$. Because all types in $\mathcal{T}_i^+$ obtain the good, the gain in payoff obtained by a type $t_i \in \mathcal{T}_i^+$ by masquerading as another type $s_i \in \mathcal{T}_i^+$ is given by the difference of prices $\pi(t_i, t_{-i}) - \pi(s_i, t_{-i})$. This implies that the masquerading relation $\xrightarrow{\mathfrak{M}[t_{-i}]}$ restricted to $\mathcal{T}_i^+$ is weakly represented by the function $-\pi(\cdot, t_{-i})$. Therefore it is acyclic by Lemma 1. Then we can conclude by Theorem 1. □

Note that Dasgupta and Maskin (2000) exhibit an ex post incentive compatible efficient auction in a framework with interdependence and no evidence. However, they must impose a one dimensionality

[4] An auction is therefore budget balanced by definition.

Environment 1

	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$u_1(s_1)$	0	7	3	10
$u_2(s_1)$	0	5	4	9
$u_1(t_1)$	0	10	2	12
$u_2(t_1)$	0	15	1	16

Environment 2

	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$u_1(s_1)$	0	2	8	10
$u_2(s_1)$	0	0	9	9
$u_1(t_1)$	0	8	5	13
$u_2(t_1)$	0	9	1	10

Table 1: Two Multi-Object Auction Environments

assumption on the type set. The equivalent of a one dimensionality assumption in our framework would be an assumption that the type set of each agent can be linearly ordered so that $t_i > t_i'$ if and only if $v_i(t_i, t_{-i}) > v_i(t_i', t_{-i})$, for every t_{-i}. Clearly, we do not need such an assumption with evidence.[5]

Jehiel, Meyer-Ter-Vehn, Moldovanu, and Zame (2006) pointed out that, in environments with multidimensional types, interdependent valuations, transfers and no evidence, the only ex post implementable social choice functions are the constant ones. Our result implies that this limitation of ex post implementation does not apply when evidence is available.

The next example shows that with multiple objects, individually rational and efficient auctions may generate cycles in the masquerade relations of the agents. This is important for several reasons. First, it shows that, even under own-type certifiability, reading mechanisms do have limitations. Second, when full extraction is not possible it may be possible to achieve efficiency and individual rationality by leaving an information rent to the agents.

Example 3 (Two Multiple-Object Auctions)**.** Consider auction environments with two agents, and two goods. The set of possible bundles that can be allocated to an agent is $\{\emptyset, \{1\}, \{2\}, \{1,2\}\}$. Agent 1 has information, encoded in the type set $\mathcal{T}_1 = \{s_1, t_1\}$, while agent 2 has no information. We consider two payoff environments, for which the valuations of the different bundles are given in Table 1, where the squares indicate the efficient allocation. In the environment, full extraction cannot be achieved, but efficiency and individual rationality can be achieved by foregoing an information rent. In the second one, individual rationality and efficiency cannot be achieved together.

[5]Dasgupta and Maskin (2000) consider a continuum of types. It would not affect our result to work with a continuum of types provided that all certifiable subsets are compact and the auction would have to use a pricing scheme that is upper semi continuous in the type of the agent that is getting the good.

First, consider environment 1. In the fully extractive auction, each agent pays her value for the bundle she receives. Therefore, all agents get a payoff of 0 if the auction proceeds according to the true type of agent 1. Suppose that agent 1 convinces the auctioneer that her type is t_1 instead of s_1. In this case, agent 1 obtains good 2 instead of good 1, at a price of 2. Since her true type is s_1, her payoff is $\mathfrak{u}_1\big(\{2\}|s_1\big) - 2 = 1 > 0$. Therefore, $s_1 \xrightarrow{\mathfrak{M}} t_1$. Now suppose that agent 1 convinces the auctioneer that her type is s_1 instead of t_1. Then she obtains good 1 instead of good 2, at a price of 7, so her masquerading payoff is $\mathfrak{u}_1\big(\{1\}|t_1\big) - 7 = 3 > 0$. Therefore, $t_1 \xrightarrow{\mathfrak{M}} s_1$.

It is, however, possible to find an individually rational and efficient auction that leads to an acyclic masquerade. Consider, for example, changing the price of object 1 from 7 to 6 when the type is s_1. Then agent 1 of type s_1 has a payoff of 1 under truthful revelation, and does no longer profit by masquerading as t_1. This change of price makes the incentive of t_1 to masquerade as s_1 stronger, but this is not a concern since the cycle is broken. The information rent that has to be paid in this auction is 1 if the type is s_1, and 0 otherwise. It is easy to check that this is in fact the revenue maximizing auction among individually rational efficient auctions. The expected information rent is therefore equal to the probability of type s_1.

By contrast, for environment 2, no individually rational and efficient auction can prevent a masquerading cycle. Indeed, individual rationality implies that the price of good 1 is at most 2 under s_1, and the price of good 2 is at most 5 under t_1. Therefore, the gain of s_1 from masquerading as t_1 is at least $(8-5)-2=1$, and the gain of t_1 from masquerading as s_1 is at least $(8-2)-5=1$. If we relax the constraint of positive prices, however, efficiency and individual rationality can be obtained by setting the price of good 1 to -1 in state s_1. Then, budget balance is also satisfied because the auctioneer can price good 2 at 9 in state s_1. ◇

7 Bilateral Trade

In this section, we consider the bilateral trade problem of Myerson and Satterthwaite (1983). We enlarge the traditional environment by considering interdependent valuations, so that the private information of the seller may enter in the valuation of the buyer, and vice versa. Bilateral trade with evidence has been considered in Singh and Wittman (2001) and Deneckere and Severinov (2008). Both papers consider Bayesian implementation, and assume private values. Furthermore, the mechanisms they build are not reading mechanisms. We show that implementation by a reading mechanism is

possible.

There are two agents with quasilinear preferences and one object. Agent 1 owns the object and is a potential seller, and agent 2 is a potential buyer. The seller's value for the item is given by $\varsigma(t_1, t_2) \geq 0$, and the buyer's value for the item is $\beta(t_1, t_2) \geq 0$, where $t_1 \in \mathcal{T}_1$ is the type of the seller and $t_2 \in \mathcal{T}_2$ is the type of the buyer.

A social choice function for this problem is called a trading rule. It determines whether trade takes place, and the transfers to each agent. Hence, it is characterized by three functions $\lambda : \mathcal{T}_1 \times \mathcal{T}_2 \to \{0, 1\}$, $\tau_1 : \mathcal{T}_1 \times \mathcal{T}_2 \to \mathbb{R}$ and $\tau_2 : \mathcal{T}_1 \times \mathcal{T}_2 \to \mathbb{R}$, where $\lambda(t_1, t_2)$ takes value 1 if trade takes place, and 0 otherwise, and $\tau_1(t_1, t_2)$ and $\tau_2(t_1, t_2)$ are respectively the transfers to the seller and the buyer. Then, the masquerading payoffs of the seller and the buyer are given by $v_1(s_1|t_1, t_2) = \tau_1(s_1, t_2) + \big(1 - \lambda(s_1, t_2)\big)\varsigma(t_1, t_2)$, and $v_2(s_2|t_2, t_1) = \tau_2(t_1, s_2) + \lambda(t_1, s_2)\beta(t_1, t_2)$.

A trading rule is *efficient* if trade occurs whenever $\beta(t_1, t_2) > \varsigma(t_1, t_2)$, and trade does not occur whenever $\beta(t_1, t_2) < \varsigma(t_1, t_2)$. It is *budget balanced* if, for every t_1, t_2, we have $\tau_1(t_1, t_2) + \tau_2(t_1, t_2) \leq 0$. It is *individually rational* if the following implications hold

$$\lambda(t_1, t_2) = 1 \;\Rightarrow\; \tau_1(t_1, t_2) \geq \varsigma(t_1, t_2) \;\text{ and }\; \tau_2(t_1, t_2) \geq -\beta(t_1, t_2)$$
$$\lambda(t_1, t_2) = 0 \;\Rightarrow\; \tau_1(t_1, t_2) = \tau_2(t_1, t_2) = 0.$$

Let $\mathcal{G}(t_1, t_2) = \beta(t_1, t_2) - \varsigma(t_1, t_2)$ denote the gains from trade. We will consider efficient trading rules that split the gains from trade between the seller, the buyer and the designer. Therefore the transfer functions are given by $\tau_1(t_1, t_2) = \lambda(t_1, t_2)\big\{\varsigma(t_1, t_2) + \alpha^s(t_1, t_2)\mathcal{G}(t_1, t_2)\big\}$, and $\tau_2(t_1, t_2) = -\lambda(t_1, t_2)\big\{\beta(t_1, t_2) - \alpha^b(t_1, t_2)\mathcal{G}(t_1, t_2)\big\}$, where $\lambda(t_1, t_2)$ is an efficient trading rule, $\alpha^b(t_1, t_2) \geq 0$ and $\alpha^s(t_1, t_2) \geq 0$ are such that $\alpha^b(t_1, t_2) + \alpha^s(t_1, t_2) \leq 1$, and represent the respective shares of the gains from trade obtained by the buyer and the seller. These trading rules thus give a share $\alpha^d(t_1, t_2) = \big(1 - \alpha^b(t_1, t_2) - \alpha^s(t_1, t_2)\big)$ of the gains from trade to the designer. They are efficient, budget balanced and individually rational by construction. In fact, they span all the set of efficient, budget balanced and individual rational trading rules.

Theorem 5 (Bilateral Trade)**.** *Any efficient, budget balanced and individually rational trading rule such that an evidence base is available for each agent is accurately implementable by a reading mechanism.*

Proof. We show that these trading rules lead to acyclic masquerade relations for the seller and the

buyer. We start with the seller, so we fix the information of the buyer to some type t_2. First, note that a trading type never wants to masquerade as a non-trading type. Indeed, a trading type t_1 gets more than her value for the good since $\tau_1(t_1, t_2) = \varsigma(t_1, t_2) + \alpha^s(t_1, t_2)\mathcal{G}(t_1, t_2) \geq \varsigma(t_1, t_2)$, whereas, if she masqueraded as a non-trading type, she would have to keep the good.

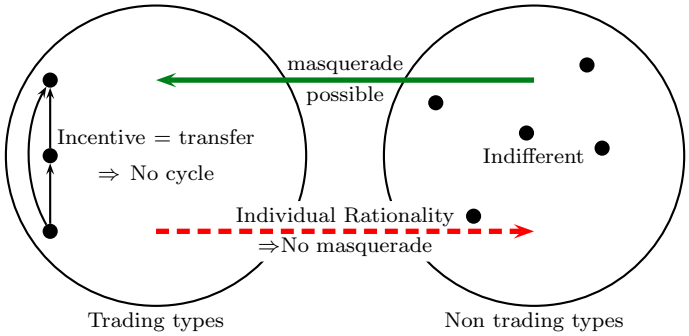
Second, a non-trading type never wants to masquerade as another non-trading type. Indeed, in both cases the seller gets to keep the good so she is indifferent. Therefore, a masquerading cycle can only occur among trading types. However, a trading type t_1 wants to masquerade as another trading type t_1' if and only if $\tau_1(t_1, t_2) < \tau_1(t_1', t_2)$. But this implies that the function $\tau_1(\cdot, t_2)$ is a weak representation of the masquerade relation restricted to trading types. Hence, by Lemma 1, there cannot exist a masquerading cycle among trading types.

For the buyer, the proof is symmetric. Start by fixing a seller type t_1. A trading type never wants to masquerade as a non-trading type, because when trading she pays less for the good than her valuation. A non-trading type never wants to masquerade as another non-trading type because it does not change anything. Finally, the masquerade over trading types can be weakly represented by the transfer function $\tau_2(t_1, \cdot)$, hence there can be no masquerading cycles among trading types. □

The intuition of the proof is similar as in the single-object auction case. Fixing the types of the other agent, the type set of an agent, say the seller, can be partitioned into trading types and non-trading types. Trading types do not want to masquerade as non-trading types because individual rationality implies that they are compensated for trading. Therefore a masquerading cycle can only consist of trading types. But a trading type wants to masquerade as another trading type only if she is getting a better transfer by doing so. Therefore the masquerade relation among trading types is governed by transfers, and has no cycle. Figure 7 illustrates this intuition.

8 Conclusion

In an environment where agents have access to evidence, we have defined a new class of mechanisms in which the designer must apply the social choice function to a profile of types that is consistent with the received profile of evidence, both along and off the equilibrium path. While restrictive, such mechanisms are robust to the principal's commitment ability and to his preferences (as long as the social choice function represents his first best), and they offer a solution to many classical mechanism design problems that are problematic under cheap talk. This is subject to having a rich



Partition of $\mathcal{T}_S$ for a given t_B

Figure 1: Bilateral Trade – intuition for the proof of Theorem 5.

evidence structure (say, own type certifiability). When this is the case, our results show that the set of implementable policies is considerably enlarged. While, in practice, evidence may not be as widely available as required for our results, they establish a benchmark, and contribute to drawing a picture of what can and cannot be achieved with evidence.

A Appendix

This appendix presents additional definitions to provide necessary and sufficient conditions for interim and ex-post implementation.

A.1 Additional Definitions

The interim belief of agent i about the types of the other agents is given by a distribution $p_i(\cdot|t_i) \in \Delta(\mathcal{T}_{-i})$. A messaging strategy profile $\mu : \mathcal{T} \to \mathcal{M}$ is an interim equilibrium[6] of this game if, for every i, every t_i, and every $m_i \in M_i(t_i)$,

$$E\Big(u_i\big(g(\mu(t));t\big)|t_i\Big) \geq E\Big(u_i\big(g(m_i,\mu_{-i}(t_{-i}));t\big)|t_i\Big).$$

A mechanism $g(\cdot)$ interim implements the social choice function $f(\cdot)$ if there exists an interim equilibrium $\mu(\cdot)$ of the game generated by $g(\cdot)$, such that $g\big(\mu(t)\big) = f(t)$ for every $t \in \mathcal{T}$.

The interim masquerading payoff of agent i is given by the function

$$v_i(s_i|t_i) = \sum_{t_{-i}\in\mathcal{T}_{-i}} u_i\big(f(s_i,t_{-i});t_i,t_{-i}\big)p_i(t_{-i}|t_i).$$

For the *interim masquerade relation*, we say that t_i wants to masquerade as s_i, denoted by $t_i \xrightarrow{\mathfrak{M}} s_i$, if and only if $v_i(s_i|t_i) > v_i(t_i|t_i)$. The set of interim worst-case types is denoted by

$$\text{wct}(\mathcal{S}_i) := \big\{s_i \in \mathcal{S}_i \,|\, \nexists\, t_i \in \mathcal{S}_i, t_i \xrightarrow{\mathfrak{M}} s_i\big\}$$

[6] As in Bergemann and Morris (2005, 2011), we use the term interim instead of Bayesian (equilibrium or implementation) to highlight the fact that we do not assume a common prior.

An *interim evidence base* for agent i is a set of messages $\mathcal{E}_i \subseteq \mathcal{M}_i$ such that there exists a one-to-one function $e_i : \mathcal{T}_i \to \mathcal{E}_i$ that satisfies $e_i(t_i) \in M_i(t_i)$, and $t_i \in \text{wct}\big(M_i^{-1}\big(e_i(t_i)\big)\big)$ for every t_i. If cheap talk completion of the evidence structure is allowed, the condition for an interim evidence base is that, for each t_i, there exists m_i such that $t_i \in \text{wct}\big(M_i^{-1}\big(m_i\big)\big)$.

A reading is *independent* if for every i the reading of the evidence satisfies $\rho_i(m_i, m_{-i}) = \rho_i(m_i, m'_{-i})$ for every m_i, m_{-i} and m'_{-i}. It means that agent i's evidence is interpreted independently of the evidence submitted by other agents.[7]

As in the main paper, we say that a reading mechanism *accurately implements* $f(\cdot)$ if it reads the evidence correctly on the equilibrium path of the corresponding equilibrium.

Definition 2 (Accurate Implementation)**.** *A reading mechanism with associated reading $\rho(\cdot)$ accurately (interim or ex post) implements $f(\cdot)$ if there exists an (interim or ex post) equilibrium strategy profile $\mu(\cdot)$ such that, for every $t \in \mathcal{T}$, $\rho\big(\mu(t)\big) = t$.*

Definition 3 (Straightforward Implementation)**.** *A reading mechanism with associated reading $\rho(\cdot)$ straightforwardly (interim or ex post) implements $f(\cdot)$ if there exists an (interim or ex post) equilibrium strategy profile $\mu(\cdot)$ such that $\rho\big(\mu(t)\big) = t$, and $\rho_{-i}\big(\mu_{-i}(t_{-i}), m_i\big) = t_{-i}$, for every $t \in \mathcal{T}$, every $i \in N$, and every $m_i \in \mathcal{M}_i$.*

Hence straightforward implementation is more restrictive than accurate implementation. It implies that, if all agents except i use their equilibrium strategy, then the type profile of these non deviators is correctly interpreted. Note also that accurate implementation by an independent reading implies straightforward implementation.

A.2 Necessary and Sufficient Conditions for Implementation

Theorem 6 (Interim Implementation)**.** *There exists a reading mechanism that accurately interim implements $f(\cdot)$ with an independent reading if and only if the following conditions hold for every agent i:*

(i) For every message $m_i \in \mathcal{M}_i$, the set $M_i^{-1}(m_i)$ admits an interim worst case type.

(ii) $M_i(\cdot)$ admits an interim evidence base.

Proof. ($\Leftarrow$) By (ii), we can pick, for each agent i, a one-to-one mapping $e_i : \mathcal{T}_i \to \mathcal{M}_i$ corresponding to an evidence base of i. By (i), we can choose an independent reading $\rho(\cdot)$ such that, for every m_i, $\rho_i(m_i) \in \text{wct}\big(M_i^{-1}(m_i)\big)$ and for every t_i, $\rho_i\big(e_i(t_i)\big) = t_i$. Suppose that every agent i adopts $e_i(\cdot)$ as her strategy in the game defined by the mechanism associated with $\rho(\cdot)$. Then for every t, the mechanism selects the outcome $f\big(\rho(e(t))\big) = f(t)$. Hence, if the strategy profile $e(\cdot)$ is an equilibrium of the game, we have succeeded in accurately implementing $f(\cdot)$. It remains to show that $e(\cdot)$ is indeed an equilibrium. Suppose then that agent i of type t_i deviates with a message $m_i \neq e_i(t_i)$. Then the implemented outcome is $f\big(w_i, t_{-i}\big)$, where $w_i \in \text{wct}\big(M_i^{-1}(m_i)\big)$. But then we know that $v_i(w_i|t_i) \leq v_i(t_i|t_i)$, so the deviation is not profitable for i.

($\Rightarrow$) Let $\rho(\cdot)$ be an independent reading such that the associated mechanism accurately implements $f(\cdot)$, and let $\mu^*(\cdot)$ be the associated equilibrium strategy profile. Then, by definition of accurate implementation, $\rho\big(\mu^*(t)\big) = t$. Consider some message m_i of agent i. The equilibrium condition implies that, for every $t_i \in M_i^{-1}(m_i)$,

$$\begin{aligned} v_i(t_i|t_i) &\geq E\Big(u_i\big(f\big(\rho(m_i, \mu^*_{-i}(t_{-i}))\big); t\big)|t_i\Big) \\ &= E\Big(u_i\big(f\big(\rho_i(m_i), t_{-i}\big); t\big)|t_i\Big) = v_i\big(\rho_i(m_i)|t_i\big), \end{aligned}$$

[7] When types are independent, this restriction has the same flavor as the belief consistency requirement "no signaling what you don't know" of a perfect Bayesian equilibrium (Fudenberg and Tirole, 1991).

where first equality is a consequence of accuracy and independence. Since, by definition of a reading mechanism, $\rho_i\big(m_i)\big) \in M_i^{-1}(m_i)$, this proves that $\rho_i\big(m_i\big) \in \text{wct}\big(M_i^{-1}(m_i)\big)$. This proves (i).

To prove (ii), consider the particular case in which $m_i = \mu_i^*(s_i)$ for some type $s_i \in \mathcal{T}_i$. Then $\rho_i\big(m_i, \mu_{-i}^*(t_{-i})\big) = s_i$, by accuracy and independence, and therefore we have shown that s_i is a worst case type of the set certified by $\mu_i^*(s_i)$. The accuracy property also implies that $\mu_i^*(s_i) \neq \mu_i^*(t_i)$ whenever $s_i \neq t_i$. Otherwise, we would have $t_i = \rho_i\big(\mu_i^*(t_i)\big) = \rho_i\big(\mu_i^*(s_i)\big) = s_i$. Therefore, the function $\mu_i^* : \mathcal{T}_i \to \mathcal{M}_i$ defines an evidence base for i. □

It is easy to show that the existence of an evidence base for each agent is necessary for implementation with any mechanism. The worst case type condition, however, is only necessary if we require accurate implementation and independent readings. If the reading is not required to be independent, then ex post instead of interim worst case types could be used (see Theorem 7). To illustrate the importance of accuracy, the following example exhibits a social choice function that is not accurately interim implementable with independent reading, because of a missing interim worst case type, but can nevertheless be implemented by a reading mechanism with independent reading.

Example 4 (Committing to incorrect readings)**.** There are two agents and five alternatives $\mathcal{A} = \{a, b, c, d, e\}$. The set of agent 1's types is $\mathcal{T}_1 = \{t_1^0, t_1^1, t_1^2, t_1^3, t_1^4\}$, and the set of agent 2's types is $\mathcal{T}_2 = \{t_2^1, t_2^2\}$, with a uniform prior probability distribution. Consider the following social choice function:[8]

$f(\cdot,\cdot)$	t_1^0	t_1^1	t_1^2	t_1^3	t_1^4
t_2^1	e	b	a	d	c
t_2^2	e	a	b	c	d

Assume that agent 2's utility is maximized when $f(\cdot)$ is implemented (so that he never has an incentive to deviate), and agent 1's utility function is given by the following table, where the squares indicate the outcomes prescribed by the social choice function:

$u_1(\cdot;\cdot) =$

	t_2^1					t_2^2				
	a	b	c	d	e	a	b	c	d	e
t_1^0	2	2	−1	−1	**0**	2	2	−1	−1	**0**
t_1^1	−1	**0**	−1	2	2	**0**	−1	2	−1	2
t_1^2	**0**	−1	−1	−1	−1	−1	**0**	−1	−1	−1
t_1^3	2	−1	−1	**0**	−1	−1	2	**0**	−1	−1
t_1^4	−1	−1	**0**	−1	−1	−1	−1	−1	**0**	−1

The interim masquerade relations of the agents and the evidence structures are summarized in Figure 2. Agent 1's interim masquerade relation has a cycle. There is an interim evidence base for each agent, but the certifiable set $\{t_1^0, t_1^1, t_1^2, t_1^3\}$ has no interim worst case type. Hence, $f(\cdot)$ is not accurately interim implementable with an independent reading. However, it is implemented with the following independent reading and interim equilibrium strategies, where the red lines correspond to incorrect readings given the equilibrium strategies:

[8] Note that this function satisfies responsiveness, that is, for every $t_i \neq t_i'$, there exists a profile t_{-i} such that $f(t_i, t_{-i}) \neq f(t_i', t_{-i})$.

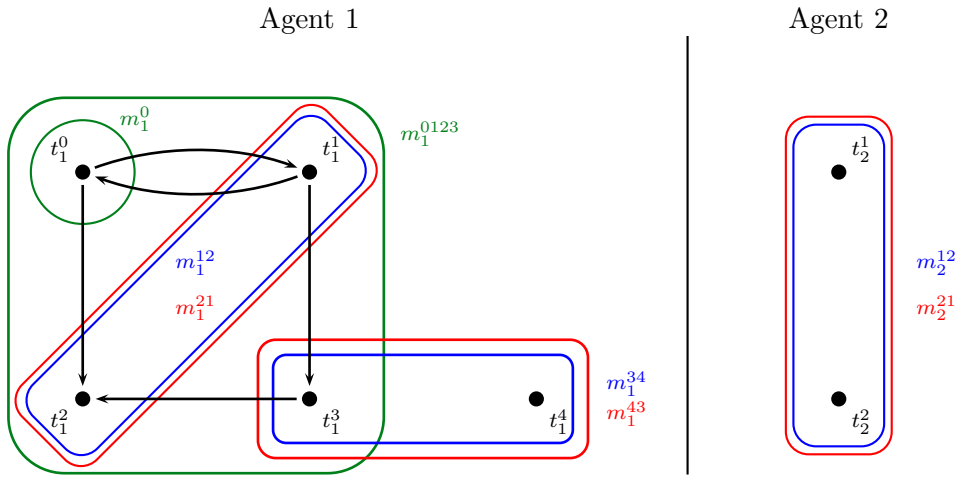


Figure 2: Committing to incorrect readings: interim masquerade relations and evidence structures.

μ_1		ρ_1		μ_2		ρ_2	
t_1^0	$\longmapsto$	m_1^0	$\longmapsto$	t_1^0	t_2^1	$\longmapsto$	m_2^{12}	$\longmapsto$	t_2^2
t_1^1	$\longmapsto$	m_1^{12}	$\longmapsto$	t_1^2	t_2^2	$\longmapsto$	m_2^{21}	$\longmapsto$	t_2^1
t_1^2	$\longmapsto$	m_1^{21}	$\longmapsto$	t_1^1					
t_1^3	$\longmapsto$	m_1^{34}	$\longmapsto$	t_1^4					
t_1^4	$\longmapsto$	m_1^{43}	$\longmapsto$	t_1^3					
		m_1^{0123}	$\longmapsto$	t_1^3					

The intuition is that, by committing to incorrect readings, the principal can emulate the use of inconsistent punishments while remaining within the boundaries of reading mechanisms. To see that, note that, given the masquerade relation of agent 1, the key is to dissuade the use of the message m_1^{0123}. This cannot be done accurately because of the cycle. In the mechanism described above, m_1^{0123} is interpreted as t_1^3, which should make t_1^1 willing to use this message. The trick is that the principal is voluntarily misinterpreting the equilibrium messages of agent 2, so agent 1 with type t_1^1, expects that the outcome implemented by the principal when she pretends to be t_1^3 and the true type of agent 2 is t_2^2 will be $f(t_1^3, t_2^2) = f(t_1^4, t_2^1) = c$. Thus, this is as if the principal attributed the message m_1^{0123} to t_1^4, which no type in $M_1^{-1}(m^{0123})$ wants to masquerade as. The principal cannot do that directly because such a reading would not be consistent with evidence. But she can emulate that outcome by misreading evidence from agent 2 on the equilibrium path.[9] ◇

For ex post implementation, we weaken the necessary properties of the mechanisms used in the characterization since we require straightforward implementation instead of accurate implementation with an independent reading.

Theorem 7 (Ex Post Implementation)**.** *There exists a reading mechanism that straightforwardly ex post implements $f(\cdot)$ if and only if the following conditions hold for every agent i:*

(i) For every $t_{-i} \in \mathcal{T}_{-i}$, and every message $m_i \in \mathcal{M}_i$, the set $M_i^{-1}(m_i)$ admits a worst case type given t_{-i}.

(ii) $M_i(\cdot)$ admits an evidence base.

[9]Note that the conclusion does not change if we modify the evidence structure so as to satisfy the normality condition of Bull and Watson (2007), Deneckere and Severinov (2008), and Forges and Koessler (2005). For example, if we complete the above evidence structure with messages certifying the singletons, the allocation $f(\cdot)$ is still implementable with the above readings and messaging strategies, but is not accurately implementable. Interestingly, $f(\cdot)$ is then implemented without asking maximal evidence to the agents: if the designer asks each agent to completely certify his type, then $f(\cdot)$ cannot be implemented with a reading mechanism.

Proof. ($\Leftarrow$) This part of the proof is the same as the proof of Theorem 1. The straightforwardness property is satisfied by construction of $\rho(\cdot)$.

($\Rightarrow$) Let $\rho(\cdot)$ be a reading such that the associated mechanism straightforwardly implements $f(\cdot)$, and let $\mu^*(\cdot)$ be the associated ex post equilibrium strategy profile. Consider some message m_i of agent i. The equilibrium condition implies that, for every $t_{-i} \in \mathcal{T}_{-i}$, and every $t_i \in M_i^{-1}(m_i)$, and

$$\begin{aligned} v_i(t_i|t_i;t_{-i}) &\geq E\Big(u_i\big(f\big(\rho(m_i,\mu^*_{-i}(t_{-i}))\big);t\big)|t_i\Big) \\ &= E\Big(u_i\big(f\big(\rho_i(m_i,\mu^*_{-i}(t_{-i})),t_{-i}\big)\big);t\big)|t_i\Big) = v_i\big(\rho_i(m_i,\mu^*_{-i}(t_{-i}))|t_i\big), \end{aligned}$$

where the second line comes from the straightforward implementation property. Since, by definition of a reading mechanism, $\rho_i(m_i\mu^*_{-i}(t_{-i})) \in M_i^{-1}(m_i)$, this proves that $\rho_i(m_i,\mu^*_{-i}(t_{-i})) \in \text{wct}\big(M_i^{-1}(m_i)|t_{-i}\big)$. This proves (i).

Now, consider the particular case where $m_i = \mu_i^*(s_i)$ for some type $s_i \in \mathcal{T}_i$. Then $\rho_i(m_i,\mu^*_{-i}(t_{-i})) = s_i$, by the straightforwardness property, and therefore we have shown that s_i is a worst case type of the set certified by $\mu_i^*(s_i)$ given t_{-i}. The straightforwardness property also implies that $\mu_i^*(s_i) \neq \mu_i^*(t_i)$ whenever $s_i \neq t_i$. Otherwise, we would have $t_i = \rho_i\big(\mu_i^*(t_i),\mu^*_{-i}(t_{-i})\big) = \rho_i\big(\mu_i^*(s_i),\mu^*_{-i}(t_{-i})\big) = s_i$. Therefore, the function $\mu_i^* : \mathcal{T}_i \to \mathcal{M}_i$ defines an evidence base for i. □

Ex post implementability by a reading mechanism implies interim implementability by a reading mechanism. However, the reading used for ex post implementation, even if it satisfies straightforwardness, may not satisfy independence. To illustrate the relations between ex post and interim implementation by reading mechanisms, we provide an example such that the conditions of Theorem 6 and Theorem 7 are not satisfied, but accurate interim implementation by a reading mechanism is possible.

Example 5 (Accurate interim implementation without independence or straightforwardness)**.** Consider two agents. The type sets are $\mathcal{T}_1 = \{t_1, t_1'\}$ and $\mathcal{T}_2 = \{t_2, t_2', t_2''\}$. The common prior is that the types of the two agents are independently and uniformly distributed over their respective supports. We assume that the only certifiable sets of agent 2 are the singletons $\{t_2\}$, $\{t_2'\}$ and $\{t_2''\}$, so that there is no need to incentivize full revelation from agent 2. The certifiable sets for agent 1 are the singletons, $\{t_1\}$ and $\{t_1'\}$, and the set $\{t_1, t_1'\}$, so that there exists an evidence base, but agent 1 needs to be incentivized to provide precise information. For simplicity, we denote the messages by the sets they certify.

The ex post masquerading relations of agent 1 and her interim masquerading relation are given in Figure 3 with intensities. There is an ex post cycle when the type of agent 2 is t_2, and there is an interim cycle. Therefore the conditions of Theorem 6 and Theorem 7 are not satisfied. Accurate interim implementation is possible with the following reading:

$$\rho_1\big(\{t_1,t_1'\},\{t_2\}\big) = \rho_1\big(\{t_1,t_1'\},\{t_2'\}\big) = t_1 \qquad \text{and} \qquad \rho_1\big(\{t_1,t_1'\},\{t_2''\}\big) = t_1'.$$

Indeed, if the type of agent 2 is t_2'', the uninformative message $\{t_1, t_1'\}$ of agent 1 is read as t_1', which is an ex post worst case type. Hence she has no incentive to be vague conditionally on the type of agent 2 being t_2''. Agent 1 cannot be given ex post incentives if the type of agent 2 is t_2, but the designer can dissuade her from being vague by pooling this event with the event in which agent 2 has type t_2'. The expected masquerading gain conditional of agent 2 not being of type t_2 is +6 for a t_1 type masquerading as a t_1' type, and -2 for a t_1' type masquerading as a t_1 type, an therefore interpreting the vague message as t_1 is dissuasive. ◇

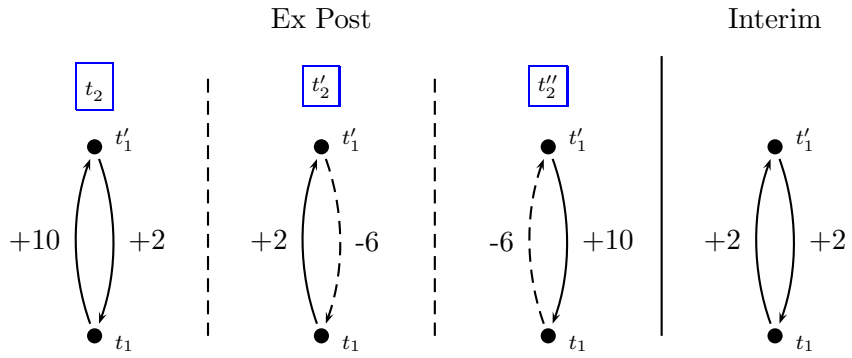


Figure 3: Accurate interim implementation without independence or straightforwardness – ex post and interim masquerade relations of agent 1.

References

Ben-Porath, E., E. Dekel, and B. L. Lipman (2014): "Optimal Allocation with Costly Verification," *American Economic Review*.

——— (2017): "Mechanisms with Evidence: Commitment and Robustness," *mimeo*.

Ben-Porath, E. and B. L. Lipman (2012): "Implementation with Partial Provability," *Journal of Economic Theory*, 147, 1689–1724.

Bergemann, D. and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 6, 1771–1813.

——— (2011): "Robust Implementation in General Mechanisms," *Games and Economic Behavior*, 71, 261–281.

Bull, J. and J. Watson (2007): "Hard Evidence and Mechanism Design," *Games and Economic Behavior*, 58, 75–93.

Crawford, V. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

Dasgupta, P. and E. Maskin (2000): "Efficient Auctions," *Quarterly Journal of Economics*, 115, 341–388.

Deneckere, R. and S. Severinov (2008): "Mechanism Design with Partial State Verifiability," *Games and Economic Behavior*, 64, 487–513.

Dziuda, W. (2011): "Strategic Argumentation," *Journal of Economic Theory*, 146, 1362–1397.

Forges, F. and F. Koessler (2005): "Communication Equilibria with Partially Verifiable Types," *Journal of Mathematical Economics*, 41, 793–811.

Fudenberg, D. and J. Tirole (1991): "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53, 236–260.

Giovannoni, F. and D. J. Seidmann (2007): "Secrecy, Two-Sided Bias and the Value of Evidence," *Games and Economic Behavior*, 59, 296–315.

Glazer, J. and A. Rubinstein (2004): "On Optimal Rules of Persuasion," *Econometrica*, 72, 1715–1736.

——— (2006): "A Study in the Pragmatics of Persuasion: a Game Theoretical Approach," *Theoretical Economics*, 1, 395–410.

Green, J. R. and J.-J. Laffont (1986): "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 53, 447–56.

Grossman, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24, 461–483.

Grossman, S. J. and O. D. Hart (1980): "Disclosure Laws and Takeover Bids," *Journal of Finance*, 35, 323–334.

Hagenbach, J., F. Koessler, and E. Perez-Richet (2014): "Certifiable Pre-Play Communication: Full Disclosure," *Econometrica*, 82, 1093–1131.

Hart, S., I. Kremer, and M. Perry (2017): "Evidence Games: Truth and Commitment," *The American Economic Review*, 107, 690–703.

Jehiel, P., M. Meyer-Ter-Vehn, B. Moldovanu, and W. R. Zame (2006): "The Limits of Ex Post Implementation," *Econometrica*, 3, 585–610.

Kartik, N. and O. Tercieux (2012): "Implementation With Evidence," *Theoretical Economics*, 7, 323–355.

Lipman, B. L. and D. J. Seppi (1995): "Robust Inference in Communication Games with Partial Provability," *Journal of Economic Theory*, 66, 370–405.

Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23–38.

Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391.

Milgrom, P. and J. Roberts (1986): "Relying on the Information of Interested Parties," *Rand Journal of Economics*, 17, 18–32.

Myerson, R. B. and M. A. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29, 265–281.

Okuno-Fujiwara, A., M. Postlewaite, and K. Suzumura (1990): "Strategic Information Revelation," *Review of Economic Studies*, 57, 25–47.

Seidmann, D. J. and E. Winter (1997): "Strategic Information Transmission with Verifiable Messages," *Econometrica*, 65, 163–169.

Sher, I. (2011): "Credibility and Determinism in a Game of Persuasion," *Games and Economic Behavior*, 71, 409–419.

——— (2014): "Persuasion and Dynamic Communication," *Theoretical Economics*, 9, 99–136.

Sher, I. and R. Vohra (2014): "Price Discrimination Through Communication," *Theoretical Economics*, forthcoming.

Shin, H. S. (2003): "Disclosures and Asset Returns," *Econometrica*, 71, 105–133.

Singh, N. and D. Wittman (2001): "Implementation with Partial Verification," *Review of Economic Design*, 6, 63–84.

Strausz, R. (2016): "Mechanism Design with Partially Verifiable Information," *mimeo*.

Wolinsky, A. (2003): "Information Transmission when the Sender's Preferences are Uncertain," *Games and Economic Behavior*, 42, 319–326.