Evidence Based Mechanisms*

Frédéric Koessler† Eduardo Perez-Richet‡

November 3, 2014

Abstract

We study implementation with privately informed agents who can produce evidence. We define evidence based mechanisms as mechanisms such that the designer's contingent plan of action is consistent with available evidence, and characterize the social choice functions that are straightforwardly implementable by such mechanisms. Our results imply that any function that is implementable with transfers is also implementable with sufficient evidence. With private values, the efficient outcome is ex post implementable with budget balanced and individually rational transfers. 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.

*We thank Elhanan Ben Porath, Jeanne Hagenbach, Johanes 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 École 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 (ANR, project NET and project RPIS) for financial support.

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

‡Ecole Polytechnique, e-mail: `eduardo.perez@polytechnique.edu`

1 Introduction

The usual mechanism design approach assumes cheap talk communication: messages have no intrinsic meaning and 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.[1]

The literature on mechanism design with evidence, reviewed below, has mostly focused on providing analytical tools, rather than feasibility results for classical mechanism design environments. By contrast, we provide sufficient conditions for interim (Bayesian) and ex post implementability, and show that they apply to familiar environments like auctions and bilateral trade.[2] These positive implementability results imply that the limits of ex post implementation pointed out by Jehiel, Meyer-Ter-Vehn, Moldovanu, and Zame (2006) do not hold when evidence is available.

Our conditions also allow us to derive an interesting property of mechanism design with evidence. Indeed, one of the most fruitful approaches to mechanism design has been to use transfers as a tool for implementation. However, other tools are available, and the use of evidence is one of them.[3] While different tools can be used as complements, it is also interesting to compare what they can achieve. We show that any social choice function that can be implemented with transfers can also be implemented with sufficient evidence, and some social choice functions can be implemented with evidence but not with transfers. We also show that any social choice function can be implemented with both evidence and transfers (but the transfers needed to do so may have to violate budget balance).

Our approach starts by imposing a constraint on mechanisms with evidence that has gen-

[1] The seminal paper on hard information in mechanism design is Green and Laffont (1986). For communication games, see Grossman (1981) and Milgrom (1981).

[2] In a related paper, Perez-Richet and Tercieux (2014) show that our conditions can be used to prove ex post implementability of stable matchings.

[3] Another tool is costly verification, or audit, as recently illustrated by Ben-Porath, Dekel, and Lipman (2014).

erally been ignored in the literature. It requires the principal to be bound by evidence. For example, a judge may not be able punish a defendant as if he had committed crime C in spite of evidence that he has committed either crime A or B. It may be tempting to use this threat in order to obtain more precision about the crime from the defendant, but evidence that crime C has not been committed should bind the actions of the judge. As another example, consider a policy maker who would like to perform a reform if the majority of the interested parties provide enough support in favor of this reform. Then, the policy maker cannot threaten any single interested party to perform no reform when this party does not show any evidence but all the other parties already provided strong evidence in favor of the reform. We call mechanisms that satisfy this constraint *evidence based mechanisms*. This constraint is similar to the restriction that would be imposed by sequential equilibrium in a game in which the principal would choose her action only after having observed the reports, so it can be interpreted as a sort of immunity requirement against lack of commitment by the principal. Mechanisms that satisfy this constraint may also be considered more legitimate by potential participants. Finally, the use of mechanisms that satisfy this constraint may be externally enforced by courts, as participants in a mechanism may sue the principal for disregarding evidence.

The restriction to evidence based mechanisms allows us to import recent advances in the analysis of communication games with hard information from Hagenbach, Koessler, and Perez-Richet (2014), and to provide tractable conditions for implementation. The key idea is that an evidence based mechanism is characterized by a *reading of the evidence*, that is an interpretation of each message profile that is consistent with its evidentiary content. This is true because any evidence based mechanism simply applies the contingent plan of action of the principal (the social choice function) to a reading of the evidence. Then, implementation requires two types of conditions. First, the evidence structure must be sufficiently rich to provide each informational type of each player with a message that conveys her information. This message must be such that no other informational type of the same player 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. Furthermore, in an evidence based mechanism, this reading must be consistent. 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. By appropriately formulating these intuitive conditions for the interim and ex post cases, we obtain sufficient conditions for interim and ex post implementation.

The first condition implies that a rich evidence structure is useful because it provides the participants with ways of conveying their information. The second condition implies that a rich evidence structure may be harmful in some environments because it provides more deviation opportunities. Whether this is the case depends on incentives. The incentives of a player in the implementation problem can be described by her masquerade relation, which is a binary relation that, for each pair of informational types of a player, (s, t), says whether the player would be better off by convincing the principal that her informational type is s, when it is really t. In fact, the second condition for implementation is satisfied under any evidence structure if and only if no player has a masquerading cycle. It is the tractability of this condition that allows us to easily derive implementability results for particular environments, and the comparison with transfers.

Related Literature. 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 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 et al.

(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) and Deneckere and Severinov (2008) extend this condition to more general mechanism design setups and message structures, while Forges and Koessler (2005) develop a similar condition for communication games. For the following discussion, we will label all these conditions with the term *normality*. An evidence structure satisfies normality if the agents can produce maximal evidence, so that if an agent is in the situation of proving several events $A, B, C, \ldots$ separately, she can also prove the conjunction of all these events $A\&B\&C\ldots$. Under normality, a revelation principle holds in the sense that any implementable social choice function can be truthfully implemented by a mechanism in which participants submit a claim about their type and show the associated maximal evidence. In addition, Deneckere and Severinov (2008) characterizes the incentive constraints that implementable social choice functions must satisfy under normality, and Bull and Watson (2007) shows that when normality does not hold, attention can be restricted to a class of three-stage dynamic mechanisms.

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. But they do not assume normality, so the revelation principle does not hold. 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).

Sher and Vohra (2014) considers the possibility of using evidence in a monopolistic price discrimination framework. Under normality, the optimal mechanism solves 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.

2 The Model

Agents and Preferences. 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 interim belief of agent i about the types of the other agents is given by a distribution $p_i(\cdot|t_i) \in \Delta(\mathcal{T}_{-i})$. 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}.

Messages. 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 $M_i^{-1}(m_i) \equiv \{t_i \in \mathcal{T}_i \,|\, m_i \in M_i(t_i)\}$. $\mathcal{S}_i$ 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.

Definition 1 (Own Type Certifiability). *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 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 players.[4]

Mechanisms and Implementation. 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 $\mu : \mathcal{T} \to \mathcal{M}$ is an interim equilibrium[5] 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 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 player i, and every message $m_i \in M_i(t_i)$,

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

[4]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).

[5]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.

Hence an ex post equilibrium is an interim equilibrium of the same game for every structure of interim beliefs.

Definition 2 (Interim and Ex Post Implementation)**.** *We say that a mechanism $g(\cdot)$ interim (respectively, ex post) implements the social choice function $f(\cdot)$ if there exists an interim (respectively, ex post) equilibrium $\mu(\cdot)$ of the game generated by $g(\cdot)$, such that $g\big(\mu(t)\big) = f(t)$ for every $t \in \mathcal{T}$.*

Evidence Based Mechanisms. A mechanism is evidence based 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.

Definition 3 (Evidence Based Mechanism)**.** *A mechanism $g(\cdot)$ is evidence based if $g(m) \in f\big(R(m)\big)$ for every message profile $m \in \mathcal{M}$.*

The main reason for being interested in these mechanism is that it will be easy to characterize the social choice functions that can be implemented by well behaved evidence based mechanisms. However, these mechanisms also have properties that may make them more desirable in practice. First they are simple. Second, a non evidence based mechanism is not immune to commitment issues, since it sometimes requires the designer to take an action that she knows to be suboptimal given the information she has received. Third, evidence based mechanism may be perceived as more legitimate. And finally, non evidence based mechanisms may in some cases be challenged in courts: for example, an agent may sue an institution that uses a non evidence based mechanism for treating him in a way that is not compatible with its mission (the social choice function) given the evidence.

The outcome function of an evidence based mechanism is completely pinned down by a reading of the evidence. Indeed, the outcome function $g(\cdot)$ of an evidence based 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 evidence based mechanism, but different readings may generate the same mechanism if the social function is not one to one.

We will say that an evidence based mechanism *truthfully implements* $f(\cdot)$ if it reads the evidence correctly on the equilibrium path of the corresponding equilibrium.

Definition 4 (Truthful Implementation)**.** *An evidence based mechanism with associated reading $\rho(\cdot)$ truthfully (interim or ex post) implements $f(\cdot)$ if there exists an (interim or ex post) equilibrium strategy profile $\mu(\cdot)$ such that, for every $t \in \mathcal{T}$, $\rho\big(\mu(t)\big) = t$.*

Our sufficient conditions for ex post implementation become necessary if we restrict attention to *straightforward implementation*, a more stringent condition than truthful implementation. It requires the existence of an equilibrium μ^* such that the equilibrium message $\mu_i^*(t_i)$ is read as t_i regardless of the messages sent by agents other than i.

Definition 5 (Straightforward Implementation)**.** *An evidence based 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 truthful implementation. It implies that, if all players except i use their equilibrium strategy, then the type profile of these non deviators is correctly interpreted. Note also that truthful implementation by an independent reading implies straightforward implementation.

3 The Masquerade Relation: Describing the Incentives

The Masquerade. We start by characterizing the ex post payoff for player i to masquerade as another type s_i when she is really of type t_i, under a social choice function $f(\cdot)$. This is given by the following ex post masquerading payoff function

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

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

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

These payoff functions represent the incentives of players 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 player 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. We call the relation that defines this graph a *masquerade relation*.

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 $v_i(s_i|t_i) > v_i(t_i|t_i)$. For the *ex post masquerade relation*, we say that t_i wants to 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 (ex post or interim) masquerade relation, we will use the notation $\rightarrow$.

We can use this relation to define a *worst-case type* 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 follows, respectively for the interim and ex post masquerade relations

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

and

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

Graphically, a worst case type is a type in $\mathcal{S}_i$ with no incoming arrow 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 $\rightarrow$ on $\mathcal{T}_i$ admits a cycle $(t_i^1, \ldots, t_i^k)$ if $t_i^1 \rightarrow t_i^2 \rightarrow \cdots t_i^k \rightarrow t_i^1$. Figure 1 illustrates the fact that cycles in the masquerade relation can preclude the existence of worst case types. The link between acyclicity and worst case types is in fact deeper as shown by the following lemma.

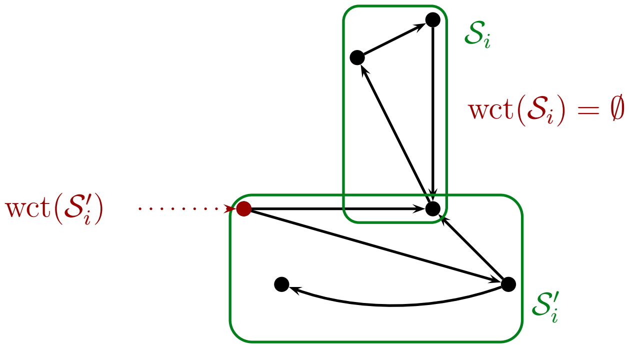


Figure 1: Masquerade relation and worst case types.

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

$$t_i \to s_i \;\Rightarrow\; w(s_i) > w(t_i). \tag{WR}$$

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.

Evidence Base. An *evidence base* is a base of messages that a player can use to certify each of her possible types. To achieve that, it must be the case that the message used to certify type t_i could not be profitably used by another type s_i. Hence, t_i must be a worst case type of the set certified by this message.

Definition 6 (Evidence Base)**.** *An evidence base for player 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*

$t_i \in \mathrm{wct}\big(M_i^{-1}\big(e_i(t_i)\big)\big)$ *for every* t_i.

Thus, an evidence base provides each type t_i of player i with a message $e_i(t_i)$ that certifies a set from which which no type of player i would like to masquerade as t_i, that is, $M_i^{-1}(e_i(t_i)) \subseteq \{s_i \in T_i : s_i \stackrel{\mathfrak{M}}{\not\to} t_i\}$ for every $t_i \in T_i$.

Example 1 (Evidence Base)**.** As an illustration, consider a player i with three possible types, $\mathcal{T}_i = \{t^1, t^2, t^3\}$, whose masquerade relation is given by $t^1 \xrightarrow{\mathfrak{M}} t^2 \xrightarrow{\mathfrak{M}} t^3$. 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. ◇

When *own type certifiability* holds, the collection of messages $m_i(t_i)$ that certify the singletons $\{t_i\}$, for all $t_i \in \mathcal{T}_i$, forms an evidence base, regardless of which social choice function $f(\cdot)$ is being considered. In general, however, an evidence base is linked to the masquerade relation, and therefore to the social choice function $f(\cdot)$.

An evidence base defines a base of messages that can be used in a fully revealing strategy of an interim equilibrium. For ex post implementation, we will need a base of messages that can be used as evidence by an agent regardless of the types of others. If such a base exists, we call it a *universal evidence base*.

Definition 7 (Universal Evidence Base)**.** *A universal evidence base for player 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 $t_i \in \bigcap_{t_{-i} \in \mathcal{T}_{-i}} \mathrm{wct}\big(M_i^{-1}\big(e_i(t_i)\big)\big|t_{-i}\big)$ for every t_i.*

Whenever own type certifiability is satisfied for a player, her message correspondence admits a universal evidence base, regardless of the social choice function under consideration.

4 Interim 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 evidence based mechanisms, the principal can only punish a deviator with a consistent but skeptical reading, that is by attributing the 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, that is $t_i \in \text{wct}\big(M_i^{-1}(m_i)\big)$. Therefore, for such skeptical readings to be available, the existence of worst case types is necessary. These intuitions are formalized in the following theorem that characterizes the social choice functions that can be truthfully implemented by an evidence based mechanism with an independent reading.

Theorem 1 (Interim Implementation)**.** *There exists an evidence based mechanism that truthfully implements $f(\cdot)$ with an independent reading if and only if the following conditions hold for every player i:*

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

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

The sufficiency proof is by construction. The idea is to pick, for each agent i, a function $e_i : \mathcal{T}_i \to \mathcal{M}_i$ corresponding to an evidence base of her message correspondence, and to construct a mechanism such that the message $e_i(t_i)$ is correctly read as t_i, and it is an equilibrium strategy profile for all agents to use the strategy $e_i(\cdot)$. The latter is achieved by reading each out-of-equilibrium message m_i as a worst case type of $M_i^{-1}(m_i)$.

In fact, it is easy to show that the existence of an evidence base for each player is necessary for implementation with any mechanism. The worst case type condition, however, is only necessary if we require truthful 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 the next section). To illustrate the importance of truthfulness, the following example exhibits a social choice function that is not truthfully implementable with independent reading, because of a missing worst case type, but can be implemented non truthfully by an evidence based mechanism with an independent reading.

Example 2 (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:[6]

$f(\cdot,\cdot)$	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:

$$u_1(\cdot;\cdot) = \begin{array}{cccccc} & & & t_2^1 & & \\ & a & b & c & d & e \\ t_1^0 & 2 & 2 & -1 & -1 & \boxed{0} \\ t_1^1 & -1 & \boxed{0} & -1 & 2 & 2 \\ t_1^2 & \boxed{0} & -1 & -1 & -1 & -1 \\ t_1^3 & 2 & -1 & -1 & \boxed{0} & -1 \\ t_1^4 & -1 & -1 & \boxed{0} & -1 & -1 \end{array} \qquad \begin{array}{cccccc} & & & t_2^2 & & \\ & a & b & c & d & e \\ t_1^0 & 2 & 2 & -1 & -1 & \boxed{0} \\ t_1^1 & \boxed{0} & -1 & 2 & -1 & 2 \\ t_1^2 & -1 & \boxed{0} & -1 & -1 & -1 \\ t_1^3 & -1 & 2 & \boxed{0} & -1 & -1 \\ t_1^4 & -1 & -1 & -1 & \boxed{0} & -1 \end{array}$$

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

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

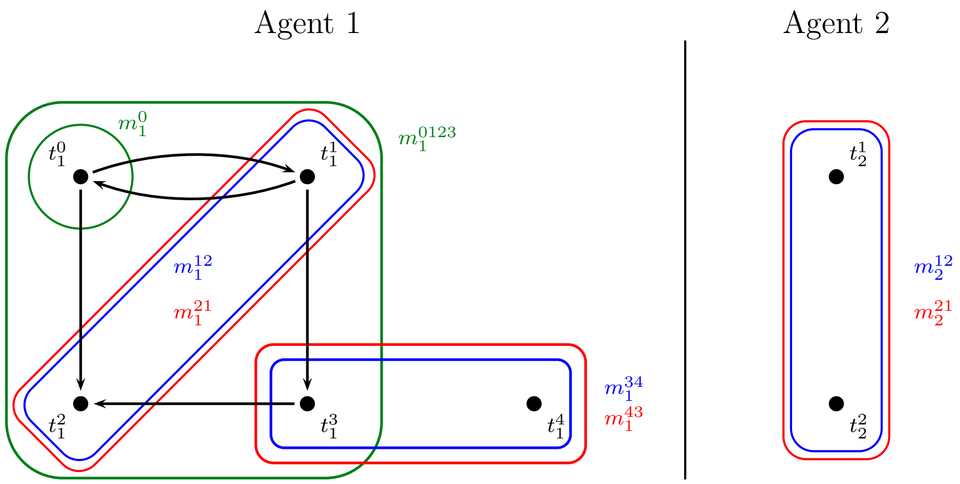


Figure 2: Committing to incorrect readings: masquerade relations and evidence structures.

$$
\begin{array}{ccccc|ccccc}
 & \mu_1 & & \rho_1 & & & \mu_2 & & \rho_2 & \\
t_1^0 & \longmapsto & m_1^0 & \longmapsto & t_1^0 & t_2^1 & \longmapsto & m_2^{12} & \longmapsto & t_2^2 \\
t_1^1 & \longmapsto & m_1^{12} & \longmapsto & t_1^2 & t_2^2 & \longmapsto & m_2^{21} & \longmapsto & t_2^1 \\
t_1^2 & \longmapsto & m_1^{21} & \longmapsto & t_1^1 & & & & & \\
t_1^3 & \longmapsto & m_1^{34} & \longmapsto & t_1^4 & & & & & \\
t_1^4 & \longmapsto & m_1^{43} & \longmapsto & t_1^3 & & & & & \\
 & & m_1^{0123} & \longmapsto & t_1^3 & & & & &
\end{array}
$$

The intuition is that, by committing to incorrect readings, the principal can emulate the use of inconsistent punishments while remaining within the boundaries of evidence based 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 truthfully 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.[7] $\diamond$

[7]Note that the conclusion does not change if we modify the evidence structure so as to satisfy the normality

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.

Theorem 2 (Interim Implementation – Acyclicity)**.** *There exists an evidence based mechanism that truthfully implements $f(\cdot)$ with an independent reading under any message structure that satisfies own type certifiability if and only if, for every player i, the interim masquerade relation $\xrightarrow{\mathfrak{M}}$ is acyclic on $\mathcal{T}_i$.*

Indeed, own type certifiability ensures that the evidence base condition is satisfied. Then to truthfully implement $f(\cdot)$ with an independent reading, it is necessary and sufficient to satisfy the worst case type condition. By Lemma 1 , we know that it is satisfied for every certifiable subset if and only if the masquerade relation is acyclic. Note that this condition also characterizes the set of social choice functions that are implementable when every subset of types is certifiable..

5 Ex Post Implementation

The results in this section parallel those for interim implementation. Ex post implementation is interesting because it is robust to changes in the information structure of the agents. These ideas are rigorously formulated in Bergemann and Morris (2005).

Ex post implementation will require the use of skeptical interpretations of vague messages from i for every realization of t_{-i}. Hence, we will rely on the existence of *ex post* worst case

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 truthfully 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 an evidence based mechanism.

types. Furthermore, to communicate her information, a player needs the existence, for each of her types, of a message that other types are not both willing and able to use. Therefore, we need a universal evidence base.[8]

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

Theorem 3 (Ex Post Implementation)**.** *There exists an evidence based mechanism that straightforwardly ex post implements $f(\cdot)$ if and only if the following conditions hold for every player 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 a universal evidence base.

The sufficiency part of the proof is by construction, and the intuition behind it is very similar to the case of interim implementation. The universal evidence base of an agent provides a natural candidate for her ex post equilibrium strategy, so we construct a mechanism that reads each message from this evidence base correctly. 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 an ex post worst case type for the deviator. It is easy to see that such readings make unilateral deviation non profitable ex post.

Finally, it is useful to characterize the social choice functions that can be implemented by any message structure that satisfies own type certifiability.

Theorem 4 (Ex Post Implementation – Acyclicity)**.** *There exists an evidence based mechanism that truthfully implements $f(\cdot)$ under any message structure that satisfies own type certifiability if and only if, for every player i, and every $t_{-i} \in \mathcal{T}_{-i}$, the masquerade relation $\xrightarrow{\mathfrak{M}[t_{-i}]}$ is acyclic on $\mathcal{T}_i$.*

[8]It is easy to show that the existence of a universal evidence base for each player is in fact a necessary condition for ex post implementation by any mechanism.

We conclude this section by a few remarks on the links between interim and ex post implementation. First, the ex post equilibrium that ex post implements a social choice function is also an interim equilibrium for any structure of beliefs $p_1, \ldots, p_n$. Therefore ex post implementability by an evidence based mechanism implies interim implementability by an evidence based 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 evidence based mechanism, we provide an example such that the conditions of Theorem 1 and Theorem 3 are not satisfied, but truthful interim implementation by an evidence based mechanism is possible.

Example 3 (Truthful 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 a (universal) 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 1 and Theorem 3 are not satisfied. Truthful 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

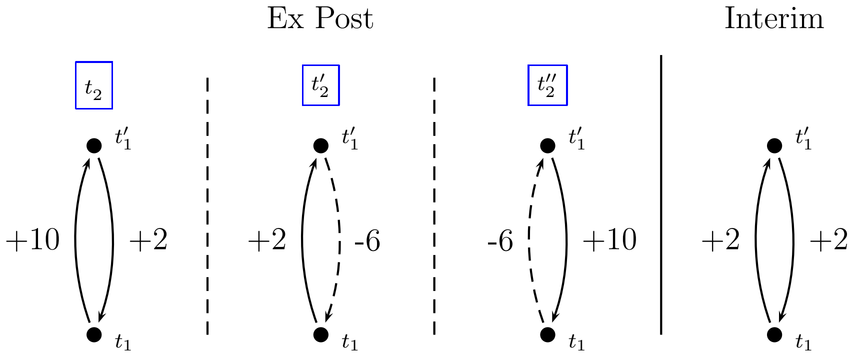


Figure 3: Truthful interim implementation without independence or straightforwardness – ex post and interim masquerade relations of agent 1.

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. $\diamond$

6 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 agent i. 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 evidence based. By the revelation principle, if a social choice function is (interim or ex post) implementable for a given evidence free message structure, then it is (interim or ex post) implementable by a direct mechanism, so we can restrict attention to direct mechanisms, and the following incentive compatibility conditions are necessary and sufficient conditions for (respectively) interim and ex post implementability

Definition 8 (Evidence Free Incentive Compatibility)**.** *A social choice function satisfies interim incentive compatibility if, for every agent i and every $t_i, s_i \in \mathcal{T}_i$*

$$v_i(s_i|t_i) \leq v_i(t_i|t_i) \tag{IIC}$$

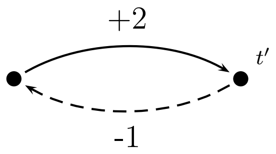


Figure 4: Masquerade relation of Example 4 – implementation with evidence is possible, but not with transfers.

It 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 interim transfer functions $\hat{\tau}_i : \mathcal{T}_i \to \mathbb{R}$, and the corresponding modified masquerading payoffs

$$V_i(s_i|t_i; t_{-i}) = v_i(s_i|t_i; t_{-i}) + \tau_i(s_i; t_{-i}),$$

and

$$V_i(s_i|t_i) = v_i(s_i|t_i) + \hat{\tau}_i(s_i).$$

Evidence vs Transfers. We start with a simple example showing that evidence based mechanisms can sometimes achieve implementation in situations where transfers cannot.

Example 4 (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

$$u(a, t) = u(a', t') = 0 \qquad u(a', t) = 2 \qquad u(a, t') = -1.$$

Therefore t wants to masquerade as t', but t' does not want to masquerade as t, hence the masquerade relation, represented in Figure 4, is acyclic, and the social choice function is (interim or ex post) implementable with evidence. It is not (interim or ex post) implementable with transfers, because any transfer that is sufficient to discourage t from claiming t' makes t' claim to be t. $\diamond$

In fact, every social choice function that is implementable with transfers can be implemented by an evidence based 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, wanted to masquerade as this type, then this incentive would be aggravated by the transfer differential. 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 5. *If $f(\cdot)$ is interim (respectively ex post) implementable with transfers and no evidence, it is also interim (respectively ex post) implementable by an evidence based mechanism under any evidence structure such that each $M_i(\cdot)$ admits an evidence base (respectively, a universal evidence base). Furthermore, there exist social choice functions that are implementable by an evidence based mechanism under any evidence structure such that each $M_i(\cdot)$ admits an evidence base, but not with transfers.*

Examining the link between these two forms of implementation through the lens of Rochet (1987) and his cyclical monotonicity condition gives an interesting perspective.

Definition 9 (Cyclical Monotonicity)**.** *A function $v : \mathcal{T}_i \times \mathcal{T}_i \to \mathbb{R}$ is cyclically monotone, if, for every finite sequence $\left(t_i^1, \ldots, t_i^k\right)$,*

$$\sum_{\ell=1}^{k} v\left(t_i^{\ell+1}, t_i^{\ell}\right) - v\left(t_i^{\ell}, t_i^{\ell}\right) \leq 0,$$

where, by convention, $t_i^{k+1} = t_i^1$.

Rochet (1987) shows that the ex post masquerade payoff function with transfers $V_i(\cdot; t_{-i})$ satisfies (EPIC) for some transfer function $\tau_i(\cdot)$ if and only if the corresponding ex post masquerading payoff function without transfers $v_i(\cdot; t_{-i})$ satisfies cyclical monotonicity. Similarly, the interim masquerade payoff function $V_i(\cdot)$ satisfies (IIC) if and only the corresponding interim masquerading payoff function without transfers $v_i(\cdot)$ is cyclically monotone.

Hence, another way to arrive at Theorem 5 would be to notice that cyclical monotonicity of a masquerading payoff function implies acyclicity of the corresponding masquerade relation. The reciprocal implication does not hold, as evidenced by the masquerade relation of Example 4 depicted in Figure 4. The link between the two notions is clarified by the following result.

Lemma 2 (Acyclicity and Cyclical Monotonicity)**.** *The masquerade relation associated with a masquerading payoff function $v : \mathcal{T}_i \times \mathcal{T}_i \to \mathbb{R}$ is acyclic if and only if there exists a strictly positive scalar function $\lambda : \mathcal{T}_i \to \mathbb{R}_+^*$ such that the function defined by $v^\lambda(s_i, t_i) = \lambda(t_i)v(s_i, t_i)$ is cyclically monotone.*

The necessity of cyclical monotonicity is analog to the first step in the rationalizabiliry theorem of Afriat (1967). The best demonstration we found is in lemma 1 of Geanakoplos (2013). The sufficiency part is easy (see Appendix A). Lemma 2 can be useful to prove that a masquerade relation is acyclic.[9]

Evidence and Transfers. It may be interesting for a designer to combine evidence and transfers. When complemented with evidence, the role of transfers becomes quite different. In the usual context, transfers must annihilate any incentive for the agent to make false claims about her type in the direct mechanism. When evidence is available, the role of transfers is to modify the incentive structure so as to make the masquerade relation acyclic. That is, transfers must make lies tractable for the designer, so that she can read the evidence skeptically.

We show that, in this case, any social choice function can be implemented by an evidence based mechanism as long as an evidence base is available, hence, in particular, if own type certifiability is satisfied.

Theorem 6. *For any social choice function $f(\cdot)$, there exist interim transfer functions $\hat{\tau}_i : \mathcal{T}_i \to \mathbb{R}$ and ex post transfer functions $\tau_i : \mathcal{T}_i \to \mathbb{R}$ for $i = 1, \ldots, n$ such that, for every i,*

[9] See Section 10, for an example.

the masquerade relations associated with the interim and ex post masquerading payoff functions with transfer $V_i(s_i|t_i)$ and $V_i(s_i|t_i;t_{-i})$ are acyclic.

The idea of the proof is extremely simple: if any ex post difference in transfers is sufficiently large as to overcome any difference in payoff from changes in the chosen action, then transfers govern the masquerading payoffs. Then all types try to obtain the highest transfer, and the worst case type of any subset of types is the one with the lowest transfer.

This result shows that the association of unlimited transfers with evidence is powerful. In practice, however, transfers may be constrained in many ways: budget balance, individual rationality or distributional concerns. The next sections shed some light on what can be done while imposing the first two of these constraints.

7 Efficient Mechanism Design with Private Values

For the remainder of the paper we focus on ex post implementation. In this section, we consider the problem of implementing an efficient social choice function, so that

$$f(t) \in \arg\max_{a\in\mathcal{A}} \sum_i u_i\big(a;t\big).$$

For that purpose, we allow the mechanism designer to use both evidence and transfers, and we restrict ourselves to ex post implementation. Given a social choice function $f(\cdot)$, an ex post transfer scheme τ_i for $i=1,\ldots,n$ 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 individual rationality possible and budget balance possible to satisfy together, we assume that, for every type profile t, $\sum_i v_i(t_i|t) \geq 0$.

We assume private values, that is $u_i(a;t) = u_i(a;t_i)$. In this case, we show that if a universal evidence base is available for each player, then any efficient social choice function is ex post implementable by an evidence based mechanism with any transfer scheme. In particular, it is possible to choose the transfer scheme so as to satisfy individual rationality and budget balance, and even extract full surplus.

Theorem 7. *Assume own type certifiability or, more generally, that a universal evidence base is available for each player. Then, under private values, it is possible to straightforwardly ex post implement any efficient social choice function with any transfer scheme by an evidence based mechanism. In particular, the transfer scheme can be chosen to satisfy individual rationality and budget balance, and even extract full surplus.*

The proof of this result is very simple and is related to the classical Vickrey-Clarkes-Groves mechanism. To see that, 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 substracting her externality on other participants. Then i's ex post 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 ex post masquerade relation given t_{-i} which is therefore acyclic by Lemma 1. We can conclude with Theorem 4.

In a way, this result is almost a corollary of Theorem 5. Since, under private values, an efficient social choice function can be implemented with transfers by a VCG mechanism, then it can also be implemented with evidence and no transfers. The value added of Theorem 7 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. We also provide an example of a multiple-good auction in which full surplus extraction cannot be achieved.

8 Auctions

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

Single-Object Auctions. 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[10] 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 8 (Single-Object Auctions)**.** *Assume own type certifiability or, more generally, that a universal evidence base is available for each player. Then, any individually rational auction is straightforwardly ex post implementable by an evidence based mechanism. In particular, the fully extractive auction is implementable.*

The proof consists in showing that the ex post masquerade relations of the agents are acyclic. Fixing the type of other players, the type set of a player can be partitioned between winning types and losing types. Losing types might want to masquerade as winning types, but individual rationality of the auction implies that winning types do not want to masquerade as losing types.

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

Hence a masquerading cycle can only comprise winning types. But the only reason a winning type might want to masquerade as another winning type is to pay a lower price, therefore the masquerade relation among winning types is governed by the price function, which rules out cycles.

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 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 player 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.[11]

Jehiel et al. (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.

Multiple Objects: Examples with Cycles. With multiple objects, as we show with the following examples, individually rational and efficient auctions may generate cycles in the ex post masquerade relations of the agents. This is important for several reasons. First, it shows that, even when evidence bases are available, evidence based 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. In the first example below, full extraction cannot be achieved, but efficiency and individual rationality can be achieved by foregoing an information rent. In the second example, individual rationality and efficiency cannot be achieved together.

Example 5 (Two Multiple-Objects Auctions)**.** Consider auction environments with two agents, and two goods. The set of possible bundles that can be allocated to an agent is $\{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

[11] 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.

Environment 1

	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$\mathfrak{u}_1(s_1)$	0	$\boxed{7}$	3	10
$\mathfrak{u}_2(s_1)$	0	5	$\boxed{4}$	9
$\mathfrak{u}_1(t_1)$	0	10	$\boxed{2}$	12
$\mathfrak{u}_2(t_1)$	0	$\boxed{15}$	1	16

Environment 2

	$\emptyset$	$\{1\}$	$\{2\}$	$\{1,2\}$
$\mathfrak{u}_1(s_1)$	0	$\boxed{2}$	8	10
$\mathfrak{u}_2(s_1)$	0	0	$\boxed{9}$	9
$\mathfrak{u}_1(t_1)$	0	8	$\boxed{5}$	13
$\mathfrak{u}_2(t_1)$	0	$\boxed{9}$	1	10

Table 1: Two Multi-Object Auction Environments

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.

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. $\diamond$

9 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 evidence based. We show that ex post implementation by an evidence based mechanism is possible. As in the case of auctions, this result illustrates how ex post implementation of non trivial social choice functions with multidimensional types and interdependent valuations 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 ex post masquerading payoffs of the seller and the buyer are given by

$$v_1(s_1|t_1, t_2) = \tau_1(s_1, t_2) + \big(1 - \lambda(s_1, t_2)\big)\varsigma(t_1, t_2),$$
$$v_2(s_2|t_2, t_1) = \tau_2(t_1, s_2) + \lambda(t_1, s_2)\beta(t_1, t_2).$$

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

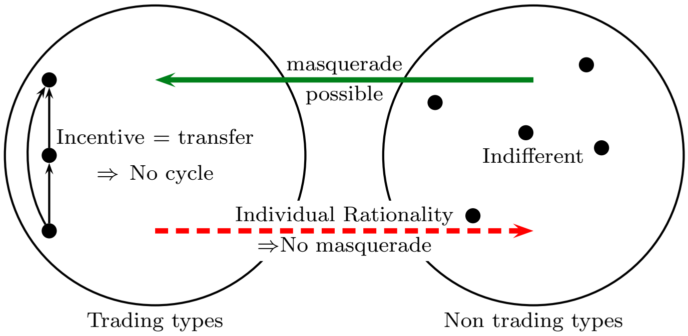
$$\lambda(t_1, t_2) = 1 \;\Rightarrow\; \tau_1(t_1, t_2) \geq \varsigma(t_1, t_2) \;\text{ and }\; \tau_2(t_1, t_2) \geq -\beta(t_1, t_2)$$
$$\lambda(t_1, t_2) = 0 \;\Rightarrow\; \tau_1(t_1, t_2) = \tau_2(t_1, t_2) = 0.$$

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:

$$\tau_1(t_1, t_2) = \lambda(t_1, t_2)\big\{\varsigma(t_1, t_2) + \alpha^s(t_1, t_2)\mathcal{G}(t_1, t_2)\big\}$$
$$\tau_2(t_1, t_2) = -\lambda(t_1, t_2)\big\{\beta(t_1, t_2) - \alpha^b(t_1, t_2)\mathcal{G}(t_1, t_2)\big\},$$

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 9 (Bilateral Trade)**.** *Any efficient, budget balanced and individually rational trading rule is straightforwardly ex post implementable by an evidence based mechanism as long as a universal evidence base is available for each player.*



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

Figure 5: Bilateral Trade – intuition for the proof of Theorem 9.

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 9 illustrates this intuition.

10 Biased Experts

In this section, we consider an environment where transfers are not available. The designer is a decision maker who is advised by multiple experts with heterogeneous preferences, and commits to using their information according to a contingent policy plan (the social choice function). In the absence of evidence, this is the framework of the delegation problem studied by Holmström (1984). We provide conditions on the preferences of the experts under which a contingent plan can be ex post implemented by an evidence based mechanism. Our setup adapts the single-expert game theoretic model of Hagenbach et al. (2014) to a mechanism design environment with multiple experts.

The experts have multidimensional information in $\mathcal{T}_i \subseteq \mathbb{R}^{K_i}$. The principal takes a multidimensional action in $\mathcal{A} \subseteq \mathbb{R}^K$. Her contingent policy plan is given by the social choice function $f(\cdot)$. We assume that all agents have quadratic preferences of the form[12]

$$u_i(a;t) = -\|a - g_i(t)\|^2,$$

where $g_i : \mathcal{T} \to \mathbb{R}^K$ is agent i's preferred action, and $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^K$. Note that the preferred actions of an expert is allowed to depend on the information of other experts. We seek conditions on the $g_i(\cdot)$ functions that ensure that $f(\cdot)$ is ex post implementable. To do this, we provide a sufficient condition on the bias functions that lead to acyclic ex post masquerade relations.

This characterization requires a few notations. We define, for each i, the disagreement correspondence $D_i^f : f(\mathcal{T}) \to \mathbb{R}^K$, by $D_i^f(a) = \{g_i(t) : t \in \mathcal{T} \text{ and } f(t) = a\}$. Hence, for a given realizable action a, $D_i^f(a)$ describes the set of actions that expert i prefers for all information states t such that $a = f(t)$.

Theorem 10. *Assume own type certifiability or, more generally, that a universal evidence base is available for each player. Then, the social choice function $f(\cdot)$ can be straightforwardly implemented by an evidence based mechanism whenever, for each player i, there exists a convex function $\phi_i : \mathbb{R}^K \to \mathbb{R}$, and for every $a \in f(\mathcal{T})$, $D_i^f(a) \in \partial\phi_i(a)$, where $\partial\phi_i : \mathbb{R}^K \rightrightarrows \mathbb{R}^K$ is the subdifferential of $\phi(\cdot)$.*

The intuition for this proposition is better interpreted in terms of the biases $b_i(t) = g_i(t) - f(t)$. The bias function of an expert can be seen as a vector field on the set of realizable actions, each arrow pointing from an action in the contingent plan of the designer toward the preferred action of the expert. Then the condition means that the biases cannot derive from a function that is too concave. In a one dimensional framework, it means that the biases can be centrifugal, or mildly centripetal, but not strongly centripetal. For a more detailed interpretation of the condition, we refer the reader to Hagenbach et al. (2014).

[12] Our results can be easily adapted to the case where the utility is a quadratic function of the form $u_i(a;t) = -(a - g_i(t))' \Omega_i (a - g_i(t))$, for some symmetric positive semidefinite matrix Ω_i.

11 Conclusion

In an environment where agents have access to evidence, we have defined a new class of mechanisms in which the designer must base every decision on received evidence. While restrictive, this class of mechanisms offers a solution to many classical mechanism design problems that are problematic under cheap talk. This is subject to having a rich evidence structure (say, own type certifiability). When this is the case, our results show that the set of incentive compatible 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. Our characterization of implementable social choice functions can also provide a basis for optimal mechanism design with evidence. In particular, if we assume that all subsets of types are certifiable, a contingent action plan is incentive compatible if and only if it creates no masquerading cycles.

Appendix

A Proofs

Proof of Lemma 1. (i) $\Leftrightarrow$ (ii). Suppose that the masquerade relation has a cycle $t^1 \to \cdots \to t^k \to t^1$. Then the set $\{t^1, \cdots, t^k\}$ has no worst case type, hence (i) $\Rightarrow$ (ii). Now suppose that a set $\mathcal{S}_i$ has no worst case type, and pick a type s_i^1 from this set. By the absence of worst case type, there must be a type $s_i^2 \neq s_i^1$ in $\mathcal{S}_i$ such that $s_i^2 \to s_i^1$. In the same fashion, there must be a type $s_i^3 \neq s_i^2$ in $\mathcal{S}_i$ such that $s_i^3 \to s_i^2$. Either $s_i^3 = s_i^1$ and we have found a cycle, or $s_i^3 \neq s_i^1$ and we can continue in the same fashion. Because $\mathcal{S}_i$ is finite, this process will eventually reach a type s_i^k such that $s_i^k = s_i^\ell$ for some $\ell < k$. But then $s_i^k \to s_i^{k-1} \to \cdots \to s_i^\ell = s_i^k$ forms a cycle. Hence (ii) $\Rightarrow$ (i).

(ii) $\Leftrightarrow$ (iii). Suppose that there exists a weak representation $w(\cdot)$ of the masquerade relation, then for any $\mathcal{S}_i \subseteq \mathcal{T}_i$, $\emptyset \neq \arg\min_{s_i \in \mathcal{S}_i} w(s_i) \subseteq \mathrm{wct}\big(\mathcal{S}_i\big)$. Hence (iii) $\Rightarrow$ (ii). Then, suppose that (ii) holds and define the function $w(\cdot)$ as follows. For every $s_i \in \mathrm{wct}\big(\mathcal{T}_i\big)$, let $w(s_i) = 0$. Let

$\mathcal{T}_i^1 = \mathcal{T}_i \setminus \mathrm{wct}\big(\mathcal{T}_i\big)$. For every $s_i \in \mathrm{wct}\big(\mathcal{T}_i^1\big)$, let $w(s_i) = 1$. Let $\mathcal{T}_i^2 = \mathcal{T}_i^1 \setminus \mathrm{wct}\big(\mathcal{T}_i^1\big)$. For every $s_i \in \mathrm{wct}\big(\mathcal{T}_i^2\big)$, let $w(s_i) = 2$. Continuing in this fashion defines an algorithm that constructs the function $w(\cdot)$ on $\mathcal{T}_i$: indeed, the process ends because $\mathcal{T}_i$ is finite, and it exhausts all the elements of $\mathcal{T}_i$ because of (ii). Now suppose that $t_i \to s_i$. Then s_i cannot be a worst case type of any set that includes both s_i and t_i, and therefore the algorithm described above must reach t_i (strictly) before it reaches s_i. Hence $w(t_i) < w(s_i)$, and this proves that (ii) $\Rightarrow$ (iii). $\square$

Proof of Theorem 1. ($\Leftarrow$) By (ii), we can pick, for each player 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, $\rho_i(m_i) \in \mathrm{wct}\big(M_i^{-1}(m_i)\big)$ and for every t_i, $\rho_i\big(e_i(t_i)\big) = t_i$. Suppose that every player 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 truthfully implementing $f(\cdot)$. It remains to show that $e(\cdot)$ is indeed an equilibrium. Suppose then that player 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 $w_i \in \mathrm{wct}\big(M_i^{-1}(m_i)\big)$. 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 truthfully implements $f(\cdot)$, and let $\mu^*(\cdot)$ be the associated equilibrium strategy profile. Then, by definition of truthful implementation, $\rho\big(\mu^*(t)\big) = t$. Consider some message m_i of agent i. The equilibrium condition implies that, for every $t_i \in M_i^{-1}(m_i)$,

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

where first equality is a consequence of truthfulness and independence. Since, by definition of an evidence based mechanism, $\rho_i\big(m_i)\big) \in M_i^{-1}(m_i)$, this proves that $\rho_i\big(m_i\big) \in \mathrm{wct}\big(M_i^{-1}(m_i)\big)$. 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 truthfulness 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 truthfulness 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. $\square$

Proof of Theorem 3. ($\Rightarrow$) We construct the reading as follows. For every i, let $e_i : \mathcal{T}_i \to \mathcal{M}_i$ be a one-to-one mapping associated with a universal evidence base of player 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 player 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 players but i use this strategy profile, then a message m_i of player 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 player 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}, player i does not gain from the deviation if her true type is t_i. Finally, the straightforwardness property is satisfied by construction of $\rho(\cdot)$.

($\Leftarrow$) 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 an evidence based 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 a universal evidence base for i. $\square$

Proof of Theorem 5. The social choice function is interim implementable with transfers if and only if $V_i(s_i|t_i)$ satisfies (IIC), that is if and only if there exists transfer functions $\hat{\tau}_i : \mathcal{T}_i \to \mathbb{R}$ such that, for every player i, and for every pair of types s_i, t_i,

$$v_i(s_i|t_i) - v_i(t_i|t_i) \leq \hat{\tau}_i(t_i) - \hat{\tau}_i(s_i).$$

Hence, $-\hat{\tau}_i(\cdot)$ is a weak representation of the masquerade relation of player i, and therefore, by Lemma 1, condition (i) of Theorem 1 is satisfied for i (and worst case types are the types who receive the highest interim expected transfer). Since $M_i(\cdot)$ admits an evidence base, condition (ii) is also satisfied for i. Because this holds for every player i, Theorem 1 allows us to conclude.

Similarly, for ex post implementation, (EPIC) implies that $-\tau_i(\cdot; t_{-i})$ is a weak representation of the ex post masquerade relation of player i given t_{-i}, allowing us to conclude by Lemma 1 and Theorem 3.

The second part of the theorem is proved by Example 4 both for the interim and the ex post case. $\square$

Proof of Lemma 2. The necessity part can be found in Geanakoplos (2013, Lemma 1). For the sufficiency part, suppose that the function $v^\lambda(s_i, t_i) = \lambda(t_i)v(s_i, t_i)$ is cyclically monotone, and consider a finite sequence $(t_i^1, \cdots, t_i^k)$. By cyclical monotonicity, we have $\sum_{\ell=1}^k \lambda(t_i^\ell)\big\{v\big(t_i^{\ell+1}, t_i^\ell\big) - v\big(t_i^\ell, t_i^\ell\big)\big\} \leq 0$, where, by convention, $t_i^{k+1} = t_i^1$. But for at least some $\ell \in \{1, \cdots, k\}$, $\lambda(t_i^\ell)\big\{v\big(t_i^{\ell+1}, t_i^\ell\big) - v\big(t_i^\ell, t_i^\ell\big) \leq 0$, and, since $\lambda(t_i^\ell) > 0$, we have $v\big(t_i^{\ell+1}, t_i^\ell\big) \leq v\big(t_i^\ell, t_i^\ell\big)$, implying that the sequence $(t_i^1, \cdots, t_i^k)$ cannot constitute a masquerading cycle. $\square$

Proof of Theorem 6. Let $\Delta = \max_{s_i \neq t_i} |v_i(s_i|t_i) - v_i(t_i|t_i)|$. Then denote all the possible types of i by $t_i^1, \cdots, t_i^m$, and let $\tau_i(t_i^\ell) = (\ell - 1)\Delta$. That makes the masquerading payoff $V_i\big(t_i^\ell|t_i^k\big)$ increasing in ℓ for every k since $V_i\big(t_i^{\ell+1}|t_i^k\big) - V_i\big(t_i^\ell|t_i^k\big) = \Delta + v_i\big(t_i^{\ell+1}|t_i^k\big) - v_i\big(t_i^\ell|t_i^k\big) \geq 0$. Hence the corresponding masquerade relation is acyclic. $\square$

Proof of Theorem 8. 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 4. $\square$

Proof of Theorem 9. We show that these trading rules lead to acyclic ex post 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 ex post 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. $\square$

Proof of Theorem 10. The subdifferentials of convex functions satisfy a cyclical monotonicity condition, as shown in Rockafellar (1972) for example. Therefore it must be the case that, for every finite sequence $(t_i^1, \cdots, t_i^k)$, and every t_{-i}

$$\sum_{\ell=1}^{k} \Big\langle g_i\big(t_i^\ell, t_{-i}\big), f\big(t_i^{\ell+1}, t_{-i}\big) - f\big(t_i^\ell, t_{-i}\big) \Big\rangle \leq 0,$$

where by convention $t_i^{k+1} = t_i^1$. But then, we must have

$$\begin{aligned}
\sum_{\ell=1}^{k} \left\{ v_i\big(t_i^{\ell+1} | t_i^\ell, t_{-i}\big) - v_i\big(t_i^\ell | t_i^\ell, t_{-i}\big) \right\} &= -\sum_{\ell=1}^{k} \left\{ \Big\| f\big(t_i^{\ell+1}, t_{-i}\big) - g_i\big(t_i^\ell, t_{-i}\big) \Big\|^2 - \Big\| f\big(t_i^\ell, t_{-i}\big) - g_i\big(t_i^\ell, t_{-i}\big) \Big\|^2 \right\} \\
&= 2 \sum_{\ell=1}^{k} \Big\langle g_i\big(t_i^\ell, t_{-i}\big), f\big(t_i^{\ell+1}, t_{-i}\big) - f\big(t_i^\ell, t_{-i}\big) \Big\rangle \\
&\qquad - \sum_{\ell=1}^{k} \left\{ \Big\| f\big(t_i^{\ell+1}, t_{-i}\big) \Big\|^2 - \Big\| f\big(t_i^\ell, t_{-i}\big) \Big\|^2 \right\} \\
&= 2 \sum_{\ell=1}^{k} \Big\langle g_i\big(t_i^\ell, t_{-i}\big), f\big(t_i^{\ell+1}, t_{-i}\big) - f\big(t_i^\ell, t_{-i}\big) \Big\rangle \\
&\leq 0
\end{aligned}$$

Therefore, the ex post masquerading payoff of player i given t_{-i} satisfies cyclical monotonicity, and we can conclude by Lemma 2 and Lemma 1. $\square$

References

Afriat, S. N. (1967): "The Construction of a Utility Function from Expenditure Data," *International Economic Review*, 8, 67–77.

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

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.

Geanakoplos, J. (2013): "Afriat from Maxmin," Mimeo.

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.

Holmström, B. (1984): "On the Theory of Delegation," in *Bayesian Models in Economic Theory*, Elsevier Science Ltd.

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.

Perez-Richet, E. and O. Tercieux (2014): "Evidence Based Matching," Working paper.

Rochet, J.-C. (1987): "A Necessary and Sufficient Condition for Rationalizability in a Quasilinear Context," *Journal of Mathematical Economics*, 16, 191–200.

Rockafellar, R. T. (1972): *Convex Analysis*, Princeton University Press, Princeton NY.

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.

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