Evidence Based Mechanisms
Supplementary Appendix

Frédéric Koessler* Eduardo Perez-Richet†

June 18, 2016

This supplementary material contains proofs and additional results that complement the paper "Evidence Based Mechanisms". We start by presenting additional definitions (Section A) to provide necessary and sufficient conditions for interim and ex-post implementation (Section B). In Section C we show that any social choice function can be implemented by a reading mechanism if the designer can combine evidence and transfers. In Section D we show that in multiple-object auctions individual rationality and efficiency may generate cycles in the ex post masquerade relations of the agents.

A 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[1] 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 player i is given by the function

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

For the *interim masquerade relation*, we say that t_i wants to masquerade as s_i, denoted by

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

†École Polytechnique, e-mail: eduardo.perez@polytechnique.edu

[1]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.

$t_i \xrightarrow{\mathfrak{M}} s_i$, 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

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

An *interim 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. 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 $t_i \in \mathrm{wct}\big(M_i^{-1}\big(m_i\big)\big)$.

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

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

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.[2]

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 A.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 A.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 players 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.

B Necessary and Sufficient Conditions for Implementation

Theorem B.1 (Interim Implementation)**.** *There exists a reading mechanism that accurately interim implements $f(\cdot)$ with an independent reading if and only if the following conditions*

[2]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).

hold for every player 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 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 accurately 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 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 $t_i \in M_i^{-1}(m_i)$,

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

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 $\rho_i\big(m_i\big) \in \mathrm{wct}\big(M_i^{-1}(m_i)\big)$. 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 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. $\square$

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 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 B.2). To illustrate the importance of accuracy, the following example exhibits a social choice function that is not accurately interim implementable with independent reading,

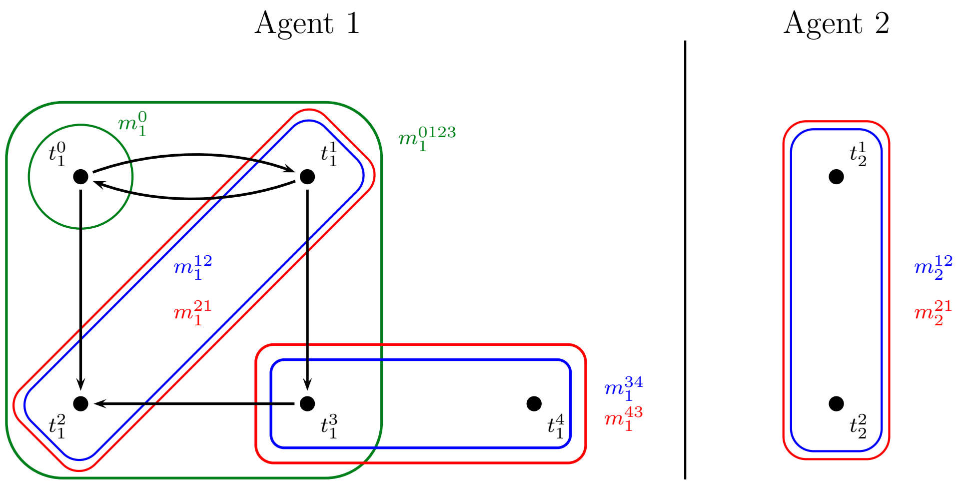

Figure B.1: Committing to incorrect readings: interim masquerade relations and evidence structures.

because of a missing interim worst case type, but can nevertheless be implemented by a reading mechanism with independent reading.

Example B.1 (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:[3]

$f(\cdot,\cdot)$	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:

$$
u_1(\cdot;\cdot) \;=\; \begin{array}{c|ccccc}
 & \multicolumn{5}{c}{t_2^1} \\
 & a & b & c & d & e \\
t_1^0 & 2 & 2 & -1 & -1 & \boxed{0} \\
t_1^1 & -1 & \boxed{0} & -1 & 2 & 2 \\
t_1^2 & \boxed{0} & -1 & -1 & -1 & -1 \\
t_1^3 & 2 & -1 & -1 & \boxed{0} & -1 \\
t_1^4 & -1 & -1 & \boxed{0} & -1 & -1
\end{array}
\qquad
\begin{array}{c|ccccc}
 & \multicolumn{5}{c}{t_2^2} \\
 & a & b & c & d & e \\
t_1^0 & 2 & 2 & -1 & -1 & \boxed{0} \\
t_1^1 & \boxed{0} & -1 & 2 & -1 & 2 \\
t_1^2 & -1 & \boxed{0} & -1 & -1 & -1 \\
t_1^3 & -1 & 2 & \boxed{0} & -1 & -1 \\
t_1^4 & -1 & -1 & -1 & \boxed{0} & -1
\end{array}
$$

The interim masquerade relations of the agents and the evidence structures are summarized in Figure B.1. Agent 1's interim masquerade relation has a cycle. There is an interim evidence

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

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:

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

The intuition is that, by committing to incorrect readings, the principal can emulate the use of inconsistent punishments while remaining within the boundaries of 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.[4] $\diamond$

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 B.2 (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 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}.

[4] 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.

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

Proof. ($\Rightarrow$) 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 an 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 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. $\square$

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 con-

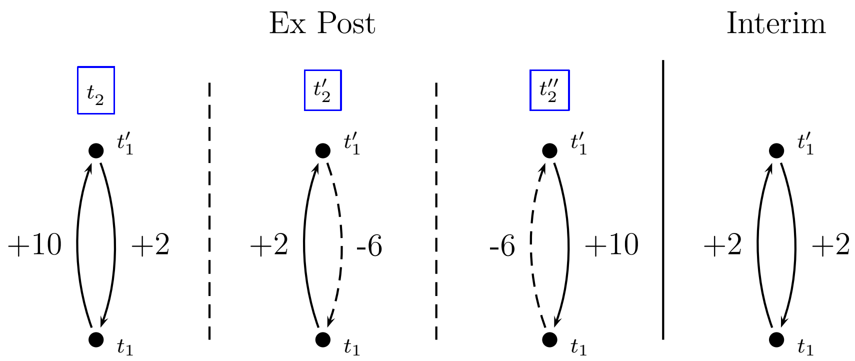

Figure B.2: Accurate interim implementation without independence or straightforwardness – ex post and interim masquerade relations of agent 1.

ditions of Theorem B.1 and Theorem B.2 are not satisfied, but accurate interim implementation by a reading mechanism is possible.

Example B.2 (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 B.2 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 B.1 and Theorem B.2 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. $\diamond$

C 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 suppress 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 a reading mechanism as long as an evidence base is available, hence, in particular, if own type certifiability is satisfied.

Theorem C.1. *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, 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.*

Proof. 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. □

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.

D Multiple-Object Auctions

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 D.1 (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 $\big\{\emptyset, \{1\}, \{2\}, \{1,2\}\big\}$. Agent 1 has information, encoded in the type set $\mathcal{T}_1 = \{s_1, t_1\}$, while

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 D.1: Two Multi-Object Auction Environments

agent 2 has no information. We consider two payoff environments, for which the valuations of the different bundles are given in Table D.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. $\diamond$

References

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.

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

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.