On the role of cheap talk in persuasion games*

Mehdi Ayouni†

April 13, 2014

Abstract

In a persuasion problem, an informed agent uses restricted evidence disclosure to communicate with the principal who chooses an outcome. In this paper, we focus on the effect of introducing cheap talk and we show how it can be beneficial for the principal. Without cheap talk, Sher [2011] shows that assuming the principal's utility function is a concave transformation of the agent's utility function, neither randomization nor commitment over the outcome are necessary. We show that with cheap talk, randomization remains unnecessary if the principal's action space is continuous, but is generally needed if it is discrete. In that case, there exists an optimal solution such that every randomization involves only two actions. However, commitment is necessary in both cases if the restriction on evidence disclosure decreases the principal's maximal expected payoff.

Keywords: Cheap talk; Certifiable information; Evidence disclosure; Determinism; Commitment.

1 Introduction

In a persuasion problem, an agent wishes to influence a principal who has to implement an outcome. The agent privately knows the state of the world, also called his *type*, and has *hard evidence* about it. Any certified message that proves a non trivial statement is considered hard evidence. Formally,

*I thank my doctoral advisors Frédéric Koessler (Paris School of Economics – CNRS) and Régis Renault (Université Paris Dauphine) for their valuable guidance and continual help. I also thank the members of THEMA's Theory Group for their suggestions and comments. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

†Thema, Université de Cergy-Pontoise, e-mail: mohamed.ayouni@u-cergy.fr

a piece of evidence is a message proving that the agent belongs to a certain subset of types, as opposed to a *cheap talk* message which is by definition available to all types. The principal ignores the state of the world which is relevant to her decision, but she can interact with the agent before implementing an action. We assume that in the course of a complete interaction only one piece of evidence can be presented. We can interpret this assumption as a time constraint for the principal who can only check a limited amount of evidence before taking an action.

To illustrate this setting, we consider the example of a hiring process. The agent is the applicant who knows his skill level and therefore what kind of jobs he is suited for. The principal is the employer who ignores if the candidate is qualified to take the job he is applying for, so she interviews him before taking a decision. However, the employer cannot possibly check all aspects of the applicant's profile.

Persuasion games usually involve hard evidence disclosure as the only form of communication. The goal of this paper is to study the effect of incorporating cheap talk into these models. We consider the same environment as Sher [2011], where the agent preferences are identical across types. This property makes cheap talk absolutely uninformative if used alone. However, it turns out to be an effective tool for improving the principal's payoff when combined with hard evidence disclosure.

By applying a result of Bull and Watson [2007], who study the introduction of hard evidence in mechanism design models, we show that the *canonical* persuasion game has the following simple structure: a *three-stage communication game* where (i) the agent announces his type, (ii) the principal asks him to show a specific piece of evidence, (iii) he shows a piece of evidence, followed by an *implementation stage* such that if the agent shows the requested piece of evidence, an outcome that depends on the announced type is implemented, otherwise a punishment action is implemented. The main difference between this game and the ones studied in Glazer and Rubinstein [2006] and Sher [2011] is that the agent and the principal exchange cheap talk messages before hard evidence is presented.

Having identified the canonical game form, we establish the necessary and sufficient conditions for the implementation of any outcome function which are expressed as conditions on the values of certain linear programs. The parameters of these linear programs depend on the outcome function, the utility function of the agent and the hard evidence structure.

Sher [2011] studies the no-cheap-talk persuasion game and shows that under a concavity assumption, namely that the principal's utility function is a type-dependent concave transformation of the agent's utility function, the principal needs neither commitment over the implementation stage nor randomization of outcome, for both continuous and discrete action spaces. These results are in fact generalizations of the findings of Glazer and Rubinstein [2006] who considered only binary action spaces, for which the concavity assumption is always satisfied.

In this paper, we address these two properties in persuasion games with cheap talk. We show that under the concavity assumption stated above, there exists an optimal *deterministic* outcome function if the action space is continuous. On the contrary, if it is discrete, we show that such an optimal function does not exist in general. However, there exists an optimal solution such that the support of every outcome contains at most two actions. In addition, these actions are adjacent according to the agent's preferences. In short, randomization is not needed when the action space is continuous and is generally necessary when the action space is discrete, but we still can find a simple form optimal outcome function. We then study the issue of commitment for these deterministic and simple form solutions. It turns out that unlike the no-cheap-talk model, these solutions cannot be implemented with credibility if the principal's maximal expected payoff is strictly decreased by the restriction on evidence disclosure. In other words, the principal needs in that case to commit to the implementation rule in order to enforce these solutions. We also show that when the first-best is unique, its credible implementation is equivalent to its implementation without cheap talk.

In section 2, we present the game and show that it is canonical. In section 3 we establish the necessary and sufficient conditions for an outcome function to be implementable. Then, in section 4, we study the determinism and commitment issues under the concavity assumption. Finally, in section 5 we study the hiring process example.

2 The model

2.1 The environment

We consider a setting with two players: a principal and an agent. There is a finite set of agent types Θ. The agent knows his type θ but the principal knows only the probability distribution of types $\mu \in \Delta(\Theta)$. There is a finite set of hard evidence E. An agent of type θ has access only to the evidence in the subset $E_\theta \subseteq E$. We assume that there exists at least two types with different evidence sets. This is what differentiates hard evidence from cheap talk messages: while it is possible for everybody to claim having a skill, only people who actually master it can back up their statement, through a test for example.

There is a set of actions A available to the principal. An interaction between the agent and the principal has two phases. The first is the communication game which can involve several rounds of cheap talk communication and one round of hard evidence disclosure. The second is the implementation stage in which the principal enforces one action from the set A. The difference between this setting and the one studied in Sher [2011] is that the communication game can involve cheap talk as well.

The agent has a utility function $u : A \to \mathbb{R}$ which is independent of his type. Let a_0 denote an action such that $u(a_0) = \min_{a \in A} u(a)$ whenever the minimum exists[1]. Throughout the paper, a_0 will be called the punishment action and the value of $u(a_0)$ will be set to 0 w.l.o.g. The principal has a utility function $v : \Theta \times A \to \mathbb{R}$ which not only depends on the action she chooses to implement, but also on the type of the agent.

An outcome function $g : \Theta \to \Delta(A)$ is a mapping from types to lotteries over actions. The function g is called *deterministic* if for every type θ, the outcome $g(\theta)$ is simply an action in A.

In line with the models of Glazer and Rubinstein [2004], Glazer and Rubinstein [2006] and Bull and Watson [2007], we assume that in one interaction between the principal and the agent, only one piece of evidence can be produced[2]. This aspect of the model can be interpreted as a time constraint:

[1]It is implicitly assumed that the minimum exists in all the results stated in this paper. But essentially, the results still hold with minor modifications. See Appendix for a study of the case where the infimum is not attained.

[2]This is essentially without loss of generality because if we want to model a limitation to

the principal has to take a decision in limited time so that she can check the agent's claim only partially.

2.2 The game

In this section, we introduce the game of persuasion. The timing of the communication game is as follows:

- Stage 1 : The agent reports his type.
- Stage 2 : The principal asks the agent for a particular piece of evidence.
- Stage 3 : The agent shows a piece of evidence to the principal.

In stage 1, the agent makes a cheap talk claim by reporting a type $\theta \in \Theta$. The principal then asks for a piece of evidence e. Her strategy at stage 2, is represented by $\sigma : \Theta \to \Delta(E)$ where $\sigma(\theta, e)$ is the probability of asking for evidence e given that the agent announced type θ. In stage 3, the agent shows a piece of evidence e', either the requested e or a different one.

At the end of the communication game, the principal has to implement an outcome. We focus on implementation rules such that if the agent shows the requested piece of evidence (i.e. $e' = e$) an outcome $g(\theta)$ which depends only on the reported type θ is implemented, otherwise (i.e. $e' \neq e$) the punishment action a_0 is implemented. Throughout the paper, $\mathbf{G_g}$ denotes the game described above, along with this implementation rule.

As we show in the last part of this section, the game $\mathbf{G_g}$ is *canonical* in the sense that we can restrict attention to it when studying the implementability of a given outcome function $g : \Theta \to \Delta(A)$. Furthermore, this implementation is achieved with truthful reporting in the first stage of the communication game and requires the principal's commitment. In section 4.2 we address the credibility issue where g is considered implementable with credibility if it is the outcome of a Perfect Bayesian Equilibrium (PBE) of a persuasion game where the implementation rule takes a general form, i.e. the outcome depends on the whole path of the communication game. This general game form will be referred to as $\mathbf{G}$ throughout the paper.

N pieces of evidence instead of one, we would have to replace E_θ by $\{S \subseteq E_\theta \text{ s.t. } |S| \leq N\}$.

Proposition 1. *If g is implemented using a general communication game and a general implementation rule, then it is also implemented in $\boldsymbol{G_g}$ with truthful reporting in stage 1.*

Proof. See Appendix. □

The argument of this proof is split in two steps. First, we note that our framework is a special case of the one studied in Bull and Watson [2007]. Applying their general results, we can already narrow the scope of attention considerably. A general communication game is an extensive form game with three types of nodes :

- Message nodes : one player (principal or agent) sends a message to the other;
- Evidentiary nodes : the agent has to present evidence;
- Terminal nodes : the principal implements an outcome;

such that along every path in the game tree, there is exactly one evidentiary node. This restriction corresponds to the limited evidence disclosure constraint. Bull and Watson [2007] show that if an outcome function is implementable using such a general game then it is also implementable using a three-stage communication game with truthful reporting in stage 1. The timing of the game is as follows:

- Stage 1 : the agent reports his type to the principal.
- Stage 2 : the principal sends a message to the agent.
- Stage 3 : the agent presents a piece of evidence.

The structure of this game is similar to that of our communication game except that instead of directly asking for evidence, the message of stage 2 identifies an information set for the agent in the original extensive form game, more specifically, the one where he has to present evidence.

In the second step, we use the fact that, in our framework, there is only one agent whose preferences are the same across types in order to show that we can restrict attention even further and focus only on the game $\mathbf{G_g}$ for the implementation of a given outcome function g.

3 Implementable outcome functions

In this section, we characterize implementable functions and strategies that implement them. Proposition 1 allows us to focus on simple three-stage games instead of general game forms : the implementation of an outcome function g can be studied within the game $\mathbf{G_g}$ and the problem boils down to finding a strategy σ that allows it. The following lemma characterizes such strategies:

Lemma 1. *σ implements g in $\boldsymbol{G_g}$ with truthful reporting if and only if*

$$\begin{aligned} \forall\theta, \quad & \sigma_{\theta\theta} = 1 \\ \forall\theta,\theta' \quad & \sigma_{\theta'\theta} \leq \frac{u(g(\theta'))}{u(g(\theta))} \end{aligned}$$

where $\sigma_{\theta'\theta} = \sum_{e \in E_{\theta'}} \sigma(\theta, e)$ is the probability for an agent of type θ' to successfully persuade the principal that he is of type θ.

Proof. See Appendix. □

The first set of conditions says that the principal asks only for documents that an agent of the announced type can show. This guarantees that if the agent reports truthfully then the principal will certainly implement the right outcome. The second set of conditions are in fact the incentive compatibility constraints of the agent, which ensure that he reports his type truthfully in the first stage. Actually, truthful reporting in stage 1 is necessary to implement the outcome function in this game, and these conditions make sure that the agent has incentive to tell the truth and that the principal does not make the mistake of asking an agent who reported his true type for evidence he cannot present, which in turn, would induce punishment erroneously.

Using Lemma 1 we can determine the necessary and sufficient conditions for an outcome function g to be implementable. We focus on the strategies satisfying the first set of conditions, i.e. strategies such that the support of $\sigma(\theta, \cdot)$ is contained in E_θ for all types θ, and we study the existence of an incentive compatible strategy among them. Consider an indexing of types in Θ from 1 to n : $\Theta = \{\theta_1, \dots, \theta_n\}$. Let q_k be the size of the evidence set of type θ_k: $q_k = \text{card}(E_{\theta_k})$. E_{θ_k} may then be written as $E_{\theta_k} = \{e_k^1, \dots, e_k^{q_j}\}$. The vector $\sigma(\theta_k, e)|_{e \in E_{\theta_k}}$ describes a point M_k in $\mathbb{R}^{q_k}$. Using this definition, the second set of conditions of Lemma 1 can be interpreted as linear inequalities satisfied by the coordinates of the M_k's for $k \in \{1, \dots, n\}$. From this formulation, we can derive the following result about the implementability of an outcome function g:

Proposition 2. *An outcome function g is implementable if and only if for all $k \in \{1, \ldots, n\}$, the following linear program P_k has a value greater than or equal to 1:*

$$\begin{aligned} \text{Max} \quad & c \cdot x \\ \text{s.t.} \quad & Ax \leq b \\ & x \geq 0, \end{aligned}$$

where $x, c \in \mathbb{R}^{q_k}$, $b \in \mathbb{R}^n$ and A a matrix $n \times q_k$. $\forall l \in \{1, \ldots, q_k\}, \forall j \in \{1, \ldots, n\}$, $c_l = 1$, $b_j = \frac{u(g(\theta_j))}{u(g(\theta_k))}$ and $A_{jl} = \mathbb{1}_{\{e_k^l \in E_{\theta_j}\}}$.

Proof. From Lemma 1, we know that g is implementable if and only if there exists a strategy σ such that

$$\begin{aligned} \forall k, \quad & \sigma_{kk} = 1 \\ \forall k, \forall j, \quad & \sigma_{jk} \leq \frac{u(g(\theta_j))}{u(g(\theta_k))} \end{aligned}$$

For a given $k \in \{1, \ldots, n\}$, let $x \in \mathbb{R}^{q_k}$ denote the vector $\sigma(\theta_k, e)|_{e \in E_{\theta_k}}$, i.e. $x_l = \sigma(\theta_k, e_k^l)$. The condition $\sigma_{kk} = 1$ is then equivalent to the condition $\sum_{l \in \{1, \ldots, q_k\}} x_l = c \cdot x = 1$, where $c \in \mathbb{R}^{q_k}$ and $\forall l, c_l = 1$. Consider the matrix A such that, $\forall l \in \{1, \ldots, q_k\}, \forall j \in \{1, \ldots, n\}$, $A_{jl} = \mathbb{1}_{\{e_k^l \in E_{\theta_j}\}}$. We can then write $\sigma_{jk} = (Ax)_j$ for every j. By defining the vector $b \in \mathbb{R}^n$ such that $b_j = \frac{u(g(\theta_j))}{u(g(\theta_k))}$, we conclude that the set of conditions on σ_{jk} for $j \in \{1, \ldots, n\}$ is equivalent to $Ax \leq b$. So far, we have shown that g is implementable if and only if for every k there exists a vector $x \in \mathbb{R}^{q_k}$, with positive coordinates, such that

$$\begin{aligned} c \cdot x &= 1 \\ Ax &\leq b \end{aligned}$$

If such a vector exists, then the value of P_k is at least 1. Conversely, if x^* is the solution of P_k, with $v = c \cdot x^*$ greater than 1, then the vector $x = \frac{1}{v} x^*$ satisfies the conditions above.

□

The implementability of an outcome function is therefore equivalent to conditions on the values of n linear programs. Each one of these conditions can actually be interpreted[3] as a condition on the value of some zero-sum

[3] I thank Rida Laraki for suggesting this interpretation. It follows from the minmax theorem of Von Neumann and the duality property of linear programming. A proof is given in the appendix.

game: P_k has a value *greater* than 1 if and only if the following zero-sum game Γ_k has a value *smaller* than 1:

- Player 1 (Fictitious): chooses a type j in $\{1, \ldots, n\}$.
- Player 2 (The principal): chooses a document e_k^l among $\{e_k^1, \ldots, e_k^{q_k}\}$.
- If $e_k^l \in E_{\theta_j}$ then player 1 gets $\frac{u(g(\theta_k))}{u(g(\theta_j))} = \frac{1}{b_j}$, otherwise he gets 0.

This condition on the value of Γ_k is equivalent to saying that the principal has a strategy such that an agent of type other than θ_k has no incentive to announce θ_k in the first stage. The interesting feature of this interpretation is that the strategy σ^*, such that for all k, $\sigma^*(\theta_k, \cdot)$ is the optimal strategy of player 2 in Γ_k, is a strategy that implements g (provided that the conditions on values are satisfied).

In the remainder of this section, we focus on implementation in pure strategies. A strategy σ is called *pure* if for every type θ, there exists a piece of evidence e_θ that is requested with certainty if type θ is announced in stage 1, i.e. $\sigma(\theta, e_\theta) = 1$. Pure strategies are particularly interesting because they are easy to implement and to commit to.

Definition 1. An outcome function g is *implementable in pure strategies* if there exists a pure strategy that implements it in $\mathbf{G_g}$.

The fact that a pure strategy maps one piece of evidence with certainty to every type makes it possible to reduce the communication game to a single stage as in the models of Glazer and Rubinstein [2006] and Sher [2011]. Consider an outcome function g and a pure strategy σ that implements it in $\mathbf{G_g}$. In the three-stage communication game, if the agent announces a type θ then the principal asks him for e_θ with certainty, and if he shows e_θ the outcome $g(\theta)$ is implemented, otherwise the outcome a_0 is implemented. It becomes clear then that if we remove the first two stages of the communication game, we can still implement the same outcome function g: let the agent present a piece of evidence e, if $e = e_\theta$ for some type θ in Θ then the principal implements $g(\theta)$, otherwise she implements a_0. This game has the same timing as the game studied in Sher [2011]: a no-cheap-talk communication game followed by an implementation stage. This means that if an outcome function is implementable in pure strategies then it can be implemented without cheap talk.

Notice that if an agent wants to get the outcome $g(\theta)$ for some θ in Θ, he just has to possess evidence e_θ. Therefore if the agent strictly prefers $g(\theta)$ to $g(\theta')$, then the incentive compatibility constraint for type θ' implies that his evidence set $E_{\theta'}$ does not contain e_θ. This property is formalized in the following definition:

Definition 2. An outcome function g is *evidence compatible* if for every type θ there exists a piece of evidence e_θ in E_θ such that:

$$\forall \theta', \text{ if } u(g(\theta')) < u(g(\theta)) \text{ then } e_\theta \notin E_{\theta'}.$$

The evidence compatibility of an outcome function g means that every type θ has one piece of evidence that distinguishes it from all the types with worse outcomes than $g(\theta)$. The previous analysis shows that if an outcome function is implementable in pure strategies then it is evidence compatible.

We conclude the section with the following equivalence result:

Proposition 3. *Let g be an outcome function. The three following statements are equivalent:*

(i) g is implementable in pure strategies.

(ii) g is evidence compatible.

(iii) g is implementable without cheap talk.

Proof. See Appendix. □

4 Optimal outcome functions

The outcome function g is optimal if it maximizes the principal's expected payoff $\sum_{\theta\in\Theta} \mu(\theta)v(g(\theta), \theta)$ among the set of implementable outcome functions F, namely the functions that satisfy the conditions of Proposition 2. The set of evidence compatible functions F_c is a subset of F. As stated in Proposition 3, F_c coincides with the set of functions that are implementable without cheap talk, i.e. using a one-stage communication game where the agent has to simply present a piece of evidence. Sher [2011] studies this particular game form and proves the existence of a *deterministic* outcome function under the assumption that the principal's utility function is a type-dependent concave transformation of the agent's utility function. Furthermore, he shows that such optimal functions are *credible*. In this section, we study these two aspects when the communication game takes its general form, i.e. when the optimal solution is to be found in F instead of the subset F_c.

4.1 Determinism

The determinism result of Sher [2011] holds whether the principal's actions space A is continuous or discrete. However, in the general framework we consider in this paper, there is a difference between these two settings. First, we show that the result still holds for a continuous action space.

Proposition 4. *If the three following conditions are satisfied:*

- *The actions space is* $A = [0, 1]$.
- *The agent's utility function* u *is continuous.*
- *For all* θ, *there exists a concave function* c_θ *such that* $v(\theta, \cdot) = c_\theta(u(\cdot))$[4].

then there exists an optimal deterministic outcome function[5].

Proof. Consider an outcome function g. For a given type θ, the outcome $g(\theta)$ is a distribution over actions. Let $\mathbb{E}_{g(\theta)}(u)$ be the expected utility of an agent under the lottery $g(\theta)$. Because u is continuous over the connected space A, there exists an action $\widehat{g}(\theta) \in A$ such that.

$$u(\widehat{g}(\theta)) = \mathbb{E}_{g(\theta)}(u)$$

This defines a deterministic outcome function $\widehat{g}$. If g is implementable then, using the strategy σ that implements it, we can also implement $\widehat{g}$ (because the agent's expected utilities are identical for both outcome functions and therefore σ satisfies the incentive compatibility constraints for $\widehat{g}$ as well). Now we will compare the principal's utilities under g and $\widehat{g}$ when she faces an agent of type θ.

$$\begin{aligned}
\mathbb{E}_{g(\theta)}(v(\theta, \cdot)) &= \mathbb{E}_{g(\theta)}(c_\theta(u)) \\
&\leq c_\theta(\mathbb{E}_{g(\theta)}(u)) \quad \text{(concavity of } c_\theta\text{)} \\
&\leq c_\theta(u(\widehat{g}(\theta))) = v(\theta, \widehat{g}(\theta)).
\end{aligned}$$

The principal is therefore (weakly) better off not randomizing over actions. The conclusion follows. □

[4]See Appendix for a counterexample when this concavity condition does not hold.

[5]The result holds for any connected action space A provided that for every θ, the maximum of c_θ over $u(A)$ is attained. This condition simply guarantees the existence of optimal solutions.

As we can see in the proof above, the fact that A is connected plays an essential role in the argument. In fact, we can easily find an example where the actions space is discrete and there is no optimal deterministic outcome function. But as we can see in the following proposition, although randomization is needed, we can still find an optimal solution of a simple form.

Proposition 5. *If the two following conditions are satisfied:*

- *The actions space is discrete $A = \{a_0, a_1, a_2, \ldots, a_m\}$.*
- *For all θ, there exists a concave function c_θ such that $v(\theta, \cdot) = c_\theta(u(\cdot))$.*

then there exists an optimal outcome function g such that for every type θ, the outcome $g(\theta)$ is either an action or a lottery over two adjacent actions (when the elements of A are ordered according to the agent's preferences)[6].

Proof. See Appendix. □

The main idea behind this result is that a level of utility for the agent can be achieved through many lotteries and that the choice of a specific lottery among them does not affect his incentives. The principal, on the other hand, prefers the lottery that involves at most the closest two actions to that level of agent's utility because of the concavity assumption.

In conclusion, the determinism result can be extended to the general persuasion game form, with a slight modification in the case of a discrete actions space. But it turns out that the credibility result does not hold and that in general commitment is necessary to the implementation of the optimal solution.

Example 1. Consider a setting with three types $\Theta = \{\theta_1, \theta_2, \theta_3\}$ such that evidence sets are:

$$\begin{aligned} E_{\theta_1} &= \{e_1\} \\ E_{\theta_2} &= \{e_2\} \\ E_{\theta_3} &= \{e_1, e_2\} \end{aligned}$$

Assume that there are three actions $A = \{0, 1, 3\}$. Let the agent's utility function be $u(a) = a$ for all a in A, and the principal's utility function be given by the following table:

[6] The result holds also for any discrete action space A with infinitely many elements, provided that for every θ, the maximum of c_θ over $u(A)$ is attained.

$v(\theta, a)$	0	1	3
θ_1	0	1	0
θ_2	0	1	0
θ_3	0	1	3

Notice that the concavity assumption is satisfied. The best deterministic outcome function in this case is one such that type θ_3 receives the action 3 along with one of the other two types, while the remaining type receives the outcome 1. The utility of the principal for such a function is $V = 4$. This function is not optimal though. The optimal solution assigns action 3 to type θ_3 and the same randomized outcome to types θ_1 and θ_2 such that action 1 has a probability $\frac{3}{4}$ and action 3 has a probability $\frac{1}{4}$. The optimal payoff of the principal is $V = \frac{9}{2}$. This function involves randomization over at most two adjacent actions as described in Proposition 5 and gives a strictly higher payoff than the deterministic ones.

4.2 Credibility

Sher [2011] shows that under the concavity assumption all deterministic optimal outcome functions, which are guaranteed to exist when cheap talk is not allowed, are implementable with credibility. As defined in section 2.2, credible implementation is achieved through a Perfect Bayesian Equilibrium (PBE) in the game **G**. In this section, we show that the introduction of cheap talk into the communication game increases the principal's payoff at the cost of commitment in general.

Definition 3. An outcome function g is *weakly evidence compatible* if

$$\forall \theta, \forall \theta', \text{ if } u(g(\theta')) < u(g(\theta)) \text{ then } E_\theta \not\subset E_{\theta'}.$$

Remark 1. Notice that if an outcome function is evidence compatible then it is weakly evidence compatible. The converse assertion holds if and only if the evidence structure satisfies *normality*[7], i.e. if and only if for every type there exists a maximal piece of evidence for every type :

$$\forall \theta, \exists e_\theta \in E_\theta \text{ s.t. } \forall \theta', \text{ if } e_\theta \in E_{\theta'} \text{ then } E_\theta \subset E_{\theta'}.$$

It is easy to check that if normality is satisfied then the weak evidence compatibility implies evidence compatibility. If normality does not hold then

[7]This condition is called normality by Bull and Watson [2007]. It has also been called the full reports condition by Lipman and Seppi [1995] and the minimal closure condition by Forges and Koessler [2005].

there exists a type $\tilde{\theta}$ such that for every piece of evidence e in $E_{\tilde{\theta}}$, there exists a type θ for which $e \in E_\theta$ and $E_{\tilde{\theta}} \not\subset E_\theta$. Consider two actions $\underline{a}$ and $\overline{a}$ such that $u(\underline{a}) < u(\overline{a})$ and let g be the outcome function defined by:

$$\forall \theta \text{ s.t. } E_{\tilde{\theta}} \not\subset E_\theta, g(\theta) = \underline{a}$$
$$\forall \theta \text{ s.t. } E_{\tilde{\theta}} \subset E_\theta, g(\theta) = \overline{a}$$

First, notice that g is weakly evidence compatible. Then using the consequence of non-normality stated above, we can easily show that g is not evidence compatible. This concludes the proof of the equivalence.

Let F_w denote the set of weakly evidence compatible outcome functions. Lemma 1 implies that all implementable outcome functions are weakly evidence compatible, i.e. $F \subset F_w$. The optimization problem of the principal is:

$$(P) \quad \max_{f \in F} \mathbb{E}(v(\theta, f(\theta)))$$

Let P_w be the optimization problem where outcome functions are only required to be weakly evidence compatible:

$$(P_w) \quad \max_{f \in F_w} \mathbb{E}(v(\theta, f(\theta)))$$

Proposition 6. *Under the assumptions of Proposition 4 (resp. Proposition 5) and the additional two following conditions:*

- *The punishment action is never optimal :* $\forall \theta, 0 \notin \arg\max_t c_\theta(t)$.

- *The value of problem (P) is strictly lower than the value of problem (P_w).*

The optimal solutions of P described in Proposition 4 (resp. Proposition 5) are not implementable with credibility.

Proof. See Appendix. □

Note that the problem P_w is equivalent to the problem where the evidence disclosure is unrestricted, i.e. where the agent of type θ can present as many pieces of evidence as he wants up to the whole set E_θ. In fact, the elements of F_w are exactly the evidence compatible outcome functions in the context of unrestricted evidence disclosure which makes them implementable (by Proposition 3). Furthermore, functions outside the set F_w are not implementable in this setting neither because they violate the incentive compatibility constraints for at least one type.

In other words, Proposition 6 states that if the evidence disclosure restriction decreases the principal's maximal expected payoff then the simple form optimal solutions which are guaranteed to exist under the assumptions of Propositions 4 and 5 cannot be implemented without commitment. This means that introducing cheap talk in this case improves the principal's welfare but requires her commitment. A trade-off is thereby created between optimality and credibility.

Definition 4. An outcome function g is *first-best* if

$$\forall\theta, g(\theta) \in \arg\max_{a\in A} v(\theta, a).$$

If a first-best outcome function is weakly evidence compatible then it is clearly a solution of the problem P_w. The following proposition gives a necessary and sufficient condition for the first-best to be implementable with credibility when it is unique.

Proposition 7. *Assume that there exists a unique first-best outcome function g^*. Then g^* is implementable with credibility if and only if it is evidence compatible.*

Proof. Consider the unique first-best outcome function g^*. Let σ be a strategy that implements g^* with credibility. Consider a type θ. Let $e \in E_\theta$ and $\underline{\theta} = \min\{u(g^*(\theta')) \mid e \in E_{\theta'}\}$ with $u(g^*(\underline{\theta})) < u(g^*(\theta))$. If $\sigma(\theta, e) > 0$, then an agent of type $\underline{\theta}$ has a positive probability of getting $u(g^*(\theta))$ if he deviates. If he gets caught, he can always show e, which guarantees at minimum $u(g^*(\underline{\theta}))$ (because of credibility off equilibrium path). Therefore, the incentive compatibility constraint is violated for type $\underline{\theta}$. This means that if σ implements g^* with credibility, then $\sigma(\theta, e) = 0$ for all evidence e that a lower type has. As a consequence each type has to possess a piece of evidence that no lower type can present, which is exactly the evidence compatibility requirement.

Conversely, if g^* is evidence compatible, then the following strategy is credible and implements the first-best: when evidence e is produced, choose action $g^*(\underline{\theta})$ such that $u(g^*(\underline{\theta})) = \min_{\theta|e\in E_\theta} u(g^*(\theta))$. $\square$

By combining this result with Proposition 3 we see that if a unique first-best is not implementable without cheap talk, then it cannot be implemented with credibility. This is a good illustration of the trade-off between optimality and credibility when cheap talk is introduced: if we do not use cheap talk then the second-best outcome function is credible (see Sher [2011]), but if we allow the use of cheap talk, the principal gets her first-best at the cost of credibility.

Example 2. Consider a setting with five types $\Theta = \{\theta_1, \ldots, \theta_5\}$ such that the evidence sets are:

$$\begin{aligned}
E_{\theta_1} &= E_{\theta_2} = \{\underline{e}\} \\
E_{\theta_3} &= E_{\theta_4} = \{\overline{e}\} \\
E_{\theta_5} &= \{\underline{e}, \overline{e}\}
\end{aligned}$$

Let the actions space be $A = [0,1]$ and the agent's utility function be $u(a) = a$. Assume that for every type θ_i the principal's utility function $v(\theta_i, \cdot)$ is strictly concave and that there exist five ordered actions $0 < a_1 < a_2 < \ldots < a_5$ such that:

$$\forall i, \quad \arg\max_{a \in A} v(\theta_i, a) = \{a_i\}$$

The assumptions of Proposition 4 are satisfied which guarantees the existence of an optimal deterministic solution. Note that any implementable outcome function g has to satisfy the conditions $g(\theta_1) = g(\theta_2)$ and $g(\theta_3) = g(\theta_4)$. The optimization program (P) (restricted to deterministic functions) can be written as follows:

$$\begin{aligned}
\max \quad & [\mu(\theta_1)v(\theta_1, a_{12}) + \mu(\theta_2)v(\theta_2, a_{12})] + [\mu(\theta_3)v(\theta_3, a_{34}) + \mu(\theta_4)v(\theta_4, a_{34})] + \mu(\theta_5)v(\theta_5, a_5) \\
\text{s.t.} \quad & a_5 \leq a_{12} + a_{34}
\end{aligned}$$

If we relax the constraint, we obtain the problem (P_w) which is equivalent to three independent maximization problems with concave objective functions, and the solution is given by $(a^*_{12}, a^*_{34}, a^*_5)$ such that:

$$a_1 \leq a^*_{12} \leq a_2 < a_3 \leq a^*_{34} \leq a_4 < a_5 = a^*_5$$

If $a^*_{12} + a^*_{34} < a^*_5$, Proposition 6 implies that the deterministic solution of (P) is not implementable with credibility. Assume $a_5 \leq a_1 + a_3$ which ensures that the solution of (P_w) is also the solution of (P). Now, consider the following mechanism:

- Stage 1: the agent announces his type.
- Stage 2:
  - if θ_1 or θ_2 is announced ask for $\underline{e}$.
  - if θ_3 or θ_4 is announced ask for $\overline{e}$.
  - if θ_5 is announced, ask for $\underline{e}$ with probability σ and for $\overline{e}$ with probability $1-\sigma$.
- Stage 3:
  - If the requested piece of evidence is presented implement a^*_{12} for types θ_1 and θ_2, a^*_{34} for types θ_3 and θ_4, and $a^*_5 = a_5$ for type θ_5.
  - otherwise implement a_1 (resp. a_3) if the agent wrongfully presented $\underline{e}$ (resp. $\overline{e}$).

This mechanism implements the optimal deterministic outcome function if and only if there exists $\sigma \in [0,1]$ such that:

$$\begin{aligned}\sigma a^*_5 + (1-\sigma)a_1 &\leq a^*_{12}\\ (1-\sigma)a^*_5 + \sigma a_3 &\leq a^*_{34}\end{aligned}$$

We can simplify the system to find that the existence of such σ is equivalent to:

$$\frac{a^*_{12}-a_1}{a^*_5-a_1} + \frac{a^*_{34}-a_3}{a^*_5-a_3} \geq 1$$

Note that by varying the distribution μ, we can make a^*_{12} (resp. a^*_{34}) take any value in $[a_1, a_2]$ (resp. $[a_3, a_4]$). This means that as long as the following inequality is satisfied, we can find a distribution μ, where all types have positive probability, such that the above mechanism implements the optimal deterministic outcome functions:

$$\frac{a_2-a_1}{a_5-a_1} + \frac{a_4-a_3}{a_5-a_3} > 1$$

There exists an equilibrium where types θ_1 and θ_2 announce θ_1, types θ_3 and θ_4 announce θ_3, and type θ_5 announces his real type in stage 1. Along the equilibrium path, the implemented action is credible : a^*_{12} (resp. a^*_{34}) is optimal given the belief that types θ_1 and θ_2 (resp. θ_3 and θ_4) choose the path leading to it. Off the equilibrium path, an agent presenting the wrong piece of evidence is believed to be the lowest type capable of such deviation (this belief corresponds to the maximal punishment) and the action that is implemented is credible. In conclusion, the deterministic optimal solution in this setting is implementable with credibility. However, if we had $a_{12}^* + a_{34}^* = a_5^*$ then it would be implementable but without credibility, because the implementation would require the use of the punishment action $a_0 = 0$ off equilibrium.

5 Application

Consider a scenario where the principal is an employer and the agent is a job candidate. The employer has several vacancies and needs to find employees in a short period of time. Each job requires a different set of skills which constitute the wanted profile (type). The outputs of these jobs are different and therefore the wage levels the employer is willing to offer are different. The problem is that the candidates disutility of work is the same irrespective of their profile or the job they are doing which means that the employer needs to test each applicant in order to know if he really has the required skill levels to accomplish the job he is applying for. However, because of the time constraint, it is impossible to fully test the applicants. Instead, the employer will only partially examine their skills.

The hiring procedure used in this case can be described by a game $\mathbf{G_g}$: first, the candidate applies for a specific job, thereby announcing that he is of the wanted type, then the employer chooses a test which is equivalent to asking the applicant to present a certain piece of evidence, finally the candidates takes the test, if he passes he is hired for the corresponding wage level, otherwise he is not hired. Using a pure strategy in the second stage is equivalent to telling the applicant about the tests beforehand. In practice, employers do not give such details to applicants.

In the remainder of this section, we focus on an example: The agent, independently of his skills, has a quadratic disutility so that for an hourly wage w, he is willing to work $S(w) = w$ hours. The hourly wage $W(\theta)$ that the employer is willing to offer depends on the skills of the candidate which are represented by his type θ.

In this setting, the utility functions are given by:

- Agent's utility: $u(w) = \frac{w^2}{2}$.
- Principal's utility: $v(w, \theta) = (W(\theta) - w)w$.

The agent's utility function is minimal for a wage $w = 0$. In other words, the principal's punishment action is to choose not to hire the agent. Let there be three types $\Theta = \{(1,0),(0,1),(1,1)\}$. The agent has two characteristics that can take values in $\{0,1\}$. This representation describes the situation where the candidate has two possible skills and where the skill levels are binary: 1 if he masters it and 0 otherwise. Let the set of evidence of each type contain two elements, one for each skill certifying its value. In essence, the evidence structure in this setting is equivalent to that of Example 1. Assume that the principal's willingness to pay for type θ is the sum of the two coordinates, i.e. $W(\theta) = \theta^1 + \theta^2$.

Note that the concavity assumption is satisfied which ensures the existence of an optimal deterministic solution: for every type θ, $v(\theta,\cdot) = c_\theta(u(\cdot))$, where $c_\theta(x) = W(\theta)\sqrt{2x} - 2x$. A deterministic outcome function is given by a vector $w = (w_{01}, w_{10}, w_{11})$ and by rearranging the conditions, we can show that w is implementable if and only if it satisfies $w_{11}^2 \leq w_{01}^2 + w_{10}^2$. In the case of equiprobable types, the principal's program (P) is:

$$\begin{array}{ll} \max & V(w) \\ \text{s.t.} & w_{11}^2 \leq w_{10}^2 + w_{01}^2 \end{array}$$

where $V(w) = (1 - w_{10})w_{10} + (1 - w_{01})w_{01} + (2 - w_{11})w_{11}$ (three times expected payoff). The program (P_w) is simply the unconstrained program and its solution is the first-best given by $w_{01} = w_{10} = \frac{1}{2}$ and $w_{11} = 1$. In order to solve (P), we first check that at the optimum $w_{11}^2 = w_{10}^2 + w_{01}^2$ and $w_{10} = w_{01}$. The optimal outcome function is w^* such that:

- $w_{01}^* = w_{10}^* = \frac{1+\sqrt{2}}{4} > \frac{1}{2}$
- $w_{11}^* = \frac{2+\sqrt{2}}{4} < 1$

We observe that the high type wage is distorded downwards, and the low types wage is distorted upwards. The distortion is similar to informational rent in standard models of adverse selection. However, in adverse selection models, the optimal outcome is better than the first-best for all types, while in this model the high type is worse off. Moreover, it is interesting to notice that the optimal solution is not implementable without cheap talk and therefore the use of cheap talk strictly improve the principal's welfare. Finally, notice that the optimal solution is not credible: the action w_{11}^* has to be implemented at a node where the principal knows that the agent is of type $(1,1)$ but is not optimal given this belief. This is coherent with Proposition 6 as the value of (P_w) is strictly higher than the value of (P) here.

6 Appendix

Proof of Proposition 1. Bull and Watson [2007] showed that if g is implemented by a general mechanism, then it is also implemented by a special three-stage mechanism $M = (\sigma, f)$ with truthful reporting at stage 1. Therefore, for every type θ and every message m, there must exist a piece of evidence e in E_θ such that the outcome $g(\theta)$ is implemented whenever the agent announces θ, the principal sends m and the agent shows e. Formally:

$$\forall\theta, \forall m, \exists e_{\theta,m} \in E_\theta \text{ such that } f(\theta, m, e_{\theta,m}) = g(\theta)$$

For every type θ, let ϕ_θ be a mapping from messages m to evidence $e_{\theta,m}$:

$$\forall m, \phi_\theta(m) \in E_\theta \text{ and } f(\theta, m, \phi_\theta(m)) = g(\theta)$$

We know that M satisfies the incentive compatibility constraints :

$$\forall\theta, \forall\theta', \sum_m \sigma(\theta', m) \max_{e\in E_\theta} u(f(\theta', m, e)) \leq u(g(\theta)).$$

Consider the mechanism $\widehat{M} = (\widehat{\sigma}, \widehat{f})$ defined by:

- $\forall\theta, \forall e,\ \widehat{\sigma}(\theta, e) = \sum_{m\in\phi_\theta^{-1}(e)} \sigma(\theta, m)$.
- $\forall\theta, \forall e,\ \widehat{f}(\theta, e, e) = g(\theta)$.
- $\forall\theta, \forall e' \neq e,\ \widehat{f}(\theta, e, e') = a_0$.

We can easily check that $\forall\theta, \sum_{e\in E_\theta} \widehat{\sigma}(\theta, e) = 1$. As a consequence, the mechanism $\widehat{M}$ is well defined. Notice that this mechanism is equivalent to playing the strategy $\widehat{\sigma}$ in the game G_g. So, in order to prove that $\widehat{\sigma}$ implements g in G_g, we will show that $\widehat{M}$ satisfies the incentive compatibility constraints. First, using the definition of $\widehat{\sigma}$, we have:

$$\forall\theta, \forall\theta', \sum_e \widehat{\sigma}(\theta', e) \max_{e'\in E_\theta} u(\widehat{f}(\theta', e, e')) = \sum_e \sum_{m\in\phi_{\theta'}^{-1}(e)} \sigma(\theta', m) \max_{e'\in E_\theta} u(\widehat{f}(\theta', e, e'))$$

By definition, if $m \in \phi_{\theta'}^{-1}(e)$ then $f(\theta', m, e) = g(\theta') = \widehat{f}(\theta', e, e)$, and for $e' \neq e$, we have $u(f(\theta', m, e')) \geq u(a_0) = u(\widehat{f}(\theta, e, e'))$. Therefore

$$\forall \theta, \forall \theta', \sum_{e} \widehat{\sigma}(\theta', e) \max_{e' \in E_\theta} u(\widehat{f}(\theta', e, e')) \leq \sum_{e} \sum_{m \in \phi_{\theta'}^{-1}(e)} \sigma(\theta', m) \max_{e' \in E_\theta} u(f(\theta', m, e'))$$

The r.h.s term is equal to $\sum_m \sigma(\theta', m) \max_{e \in E_\theta} u(f(\theta', m, e))$ and we finally get:

$$\forall \theta, \forall \theta', \sum_{e} \widehat{\sigma}(\theta', e) \max_{e' \in E_\theta} u(\widehat{f}(\theta', e, e')) \leq u(g(\theta)).$$

Which are exactly the incentive compatibility constraints of $\widehat{M}$. □

Proof of Lemma 1. Given the structure of G_g, the outcome $g(\theta)$ is implemented only if the agent reports type θ in stage 1 and shows the required piece of evidence in stage 3. This means that in order to implement g in G_g, truth telling is necessary in stage 1. As a consequence, if the reported type is θ, the principal can ask only for evidence in E_θ. Otherwise, an agent of type θ who reports truthfully has a positive probability of being asked for evidence he cannot present. A strategy σ satisfying this condition is such that $\forall \theta, \sigma_{\theta\theta} = 1$.

In addition, σ must be such that the agent has no incentive to lie about his type. The expected utility of an agent of type θ' when she reports type θ is equal to $\sigma_{\theta'\theta} u(g(\theta))$. In particular, the expected utility of truth telling for an agent of type θ is $u(g(\theta))$ as long as the strategy satisfies the first condition. Therefore, the agents have incentive to tell the truth if $\forall \theta, \forall \theta', \sigma_{\theta'\theta} u(g(\theta)) \leq u(g(\theta'))$.

Conversely, if σ satisfies $\forall \theta, \forall \theta', \sigma_{\theta'\theta} u(g(\theta)) \leq u(g(\theta'))$, then truth telling in stage 1 is a best response for the agent and it leads to the implementation of g.

Finally, we show that the second set of conditions as written in Lemma 1 does not involve division by zero. This is true because we can assume without loss of generality that for all θ, $u(g(\theta)) > u(a_0) = 0$. Actually, solving the problem where $u(g(\theta)) = 0$ for some types θ in a subset $\Theta_0 \subseteq \Theta$ is equivalent to solving the problem where those types and their evidence are removed, i.e. the set of types is $\Theta \setminus \Theta_0$ and the evidence set is $E \setminus \cup_{\theta \in \Theta_0} E_\theta$. Consider the original problem. For every $\theta' \in \Theta_0$ the conditions will be $\forall \theta \in \Theta \setminus \Theta_0 \sigma_{\theta'\theta} = 0$. Therefore, the principal cannot ask any type outside Θ_0 for evidence that any type in this subset can present. If σ is a solution of the original problem, then its restriction to $\Theta \setminus \Theta_0$ is a solution of the reduced problem. Conversely, if σ is a solution of the reduced problem, let σ' be equal to σ on $\Theta \setminus \Theta_0$ and for every type θ in Θ_0, let $\sigma'(\theta, e_\theta) = 1$ for some $e_\theta \in E_\theta$. We can easily check that σ' is a solution of the original program. □

Proof of Proposition 3. Let g be an outcome function. Recall the three statements:

(i) g is implementable in pure strategies.

(ii) g is evidence compatible.

(iii) g is implementable without cheap talk.

In order to prove the equivalence, we will show the following implications: (i)⇒(ii)⇒(iii)⇒(i).

(i)⇒(ii) A pure strategy σ is such that $\forall \theta, \exists e_\theta \in E$ such that $\sigma(\theta, e_\theta) = 1$. If σ implements g then it must satisfy the first set of conditions in Lemma 1, which yields $\forall \theta, e_\theta \in E_\theta$. Thus, $\sigma_{\theta'\theta} = 1$ if $e_\theta \in E_{\theta'}$ and $\sigma_{\theta'\theta} = 0$ otherwise. It follows from the second set of conditions that $e_\theta \notin E_{\theta'}$ whenever $u(g(\theta')) < u(g(\theta))$, i.e. g is evidence compatible.

(ii)⇒(iii) For every type θ there exists a piece of evidence e_θ in E_θ such that:

$$\forall \theta', \text{ if } u(g(\theta')) < u(g(\theta)) \text{ then } e_\theta \notin E_{\theta'}.$$

Consider the following mechanism: the agent shows a piece of evidence and then the principal implements $g(\theta)$ if she observes e_θ for some θ, and a_0 if she observes any evidence outside the set $\{e_\theta | \theta \in \Theta\}$. The best response of an agent of type θ will be to show e_θ and this no-cheap-talk mechanism implements g.

(iii)⇒(i) g is implementable without cheap talk. Such a mechanism is defined by a mapping $h : E \to \Delta(A)$. The principal implements the outcome $h(e)$ if the agent shows the piece of evidence e. The strategy of an agent of type θ is given by a probability distribution over evidence in E_θ, denoted $\xi(\theta)$. This mechanism implements g if there is a best response ξ to h such that for every type θ, $g(\theta, \cdot) = \sum_{e \in E_\theta} \xi(\theta, e) h(e, \cdot)$. As ξ is a best response, if $\xi(\theta, e) > 0$ then $u(h(e)) = \max_{e' \in E_\theta} u(h(e'))$. The utility of the outcome $g(\theta)$:

$$
\begin{aligned}
u(g(\theta)) &= \mathbb{E}_{g(\theta)}(u) \\
&= \sum_{e \in E_\theta} \xi(\theta, e)\mathbb{E}_{h(e)}(u) \\
&= \max_{e \in E_\theta} u(h(e))
\end{aligned}
$$

For every θ, choose a piece of evidence $e_\theta \in E_\theta$ from the support of $\xi(\theta)$, i.e. e_θ such that $\xi(\theta, e) > 0$. Consider the game $\mathbf{G_g}$ with the pure strategy σ such that for every type θ, $\sigma(\theta, e_\theta) = 1$. We have

$$
\forall \theta, \quad u(g(\theta)) = \max_{e \in E_\theta} u(h(e)) = u(h(e_\theta))
$$

Therefore, if two types θ and θ' are such that $u(g(\theta')) < u(g(\theta))$ then $e_\theta \notin E_{\theta'}$ which ensures that $\sigma_{\theta'\theta} = 0 < \frac{u(g(\theta'))}{u(g(\theta))}$. In conclusion, the pure strategy σ implements g.

□

Proof of Proposition 5. Reorder the action space $A = \{a_0, a_1, \ldots, a_m\}$ according to the agent's preferences, i.e. $0 = u(a_0) < u(a_1) < u(a_2) < \ldots < u(a_m)$. Consider an implementable outcome function g such that for some type θ, the utility level $u(g(\theta))$ cannot be achieved using one action, i.e. there exists l such that $u(a_l) < u(g(\theta) < u(a_{l+1})$. Given that the agent is indifferent between all lotteries with expected utility $u(g(\theta)$, the principal can choose the one that maximizes her utility. Because of the concavity assumption, the principal is better off choosing to randomize only over a_l and a_{l+1}. The conclusion follows.

□

Proof of Proposition 6. We give the proof for deterministic optimal outcome functions under the assumptions of Proposition 4. The argument can be simply applied for the optimal solutions of Proposition 5 using the following transformation: if $A = \{a_1, \ldots, a_m\}$ is ordered according to the agent's preferences, i.e. $0 = u(a_0) < u(a_1) < u(a_2) < \ldots < u(a_m)$, we can obtain an equivalent problem with a connected actions space defined by:

$$
\tilde{A} = \bigcup_{l \in \{1, \ldots, m\}} \Delta(\{a_{l-1}, a_l\}).
$$

Under the assumptions of Proposition 4, let g be an optimal deterministic outcome function. Assuming g is implementable with credibility and the value of (P_w) is strictly higher than the value of (P), we prove a contradiction.

The assumption that g is implementable with credibility means by definition that it is the outcome of a PBE in the game $\mathbf{G}$. Let σ be the strategy that the principal uses at stage 2 in this PBE. The goal is to define an outcome function $\tilde{g}$ and a strategy $\tilde{\sigma}$ by slightly modifying g and σ, such that the principal strictly prefers $\tilde{g}$ to g and $\tilde{\sigma}$ implements $\tilde{g}$ without credibility, i.e. in the game $\mathbf{G}_{\tilde{\mathbf{g}}}$, thereby contradicting the optimality of g.

The concavity assumption tells us that:

$$\forall \theta, \exists c_\theta \text{ concave s.t. } v(\theta, \cdot) = c_\theta(u(\cdot)) \tag{1}$$

For every level t of agent's utility, let $\Theta_t = \{\theta \mid u(g(\theta)) = t\}$ the set of types of utility t according to g. The credible implementation implies that at every terminal node, the principal chooses the action that maximizes her expected payoff given her belief about the types. As a consequence, for every nonempty Θ_t:

$$t \in \arg\max_s \sum_{\theta \in \Theta_t} \mu_\theta c_\theta(s)$$

The problem (P_w) is defined in this context by:

$$\begin{aligned} \max \quad & \sum_{\theta \in \Theta} \mu_\theta c_\theta(x_\theta) \\ \text{s.t.} \quad & \forall \theta, \theta' \quad \mathbb{1}_{\{E_\theta \subset E_{\theta'}\}}(x_\theta - x_{\theta'}) \leq 0 \end{aligned}$$

Consider (P_w^t) the restricted version of (P_w) to types in Θ_t:

$$\begin{aligned} \max \quad & \sum_{\theta \in \Theta_t} \mu_\theta c_\theta(x_\theta) \\ \text{s.t.} \quad & \forall \theta, \theta' \in \Theta_t \quad \mathbb{1}_{\{E_\theta \subset E_{\theta'}\}}(x_\theta - x_{\theta'}) \leq 0 \end{aligned}$$

Notice that there exists a nonempty Θ_t such that the vector $x = (t, t, \ldots, t)$ is not a solution of (P_w^t), because otherwise g would be a solution of (P_w), which would imply that its value is equal to the value of (P). From the concavity assumption and the convexity of the set of (P_w^t)'s feasible solutions, we get:

$$\begin{aligned} \forall \lambda > 0, \exists x^{\lambda} \quad \text{s.t.} \quad & \forall \theta \in \Theta_t, \quad |x_\theta^\lambda - t| \leq \lambda \\ \text{and} \quad & \sum_{\theta \in \Theta_t} \mu_\theta c_\theta(x_\theta) > \sum_{\theta \in \Theta_t} \mu_\theta c_\theta(t) \end{aligned}$$

For every λ, define an outcome function g_λ such that:

$$\begin{aligned} \forall \theta \in \Theta_t, \quad & u(g_\lambda(\theta)) = x_\theta^\lambda \\ \forall \theta \notin \Theta_t, \quad & g_\lambda(\theta) = g(\theta) \end{aligned}$$

The function g_λ is obtained as a slight modification of g and it gives a strictly higher payoff to the principal. We will now prove that for some λ, g_λ is implementable using the fact that g is implementable with credibility.

First, for every type θ define $\underline{u}_\theta$ the highest utility level that an agent of type θ can achieve if he deviates and is asked to provide evidence he does not possess:

$$\underline{u}_\theta = \max_{e \in E_\theta} \min_{\theta' | e \in E_{\theta'}} u(g^*(\theta'))$$

where g^* is the first-best outcome function. Because punishment is assumed to never be optimal, we obtain that for all types θ, the level of utility $\underline{u}_\theta$ is strictly positive. For every utility level s, let the path that types in Θ_s follow be labeled s and the principal's strategy at the second stage be labeled σ_s. The credible implementation implies that:

$$\forall \theta, (s - \underline{u}_\theta)\sigma_s(E_\theta) \leq u(g(\theta)) - \underline{u}_\theta$$

In the game $\mathbf{G}_{\mathbf{g}_\lambda}$, for every positive ϵ define a strategy σ^ϵ as follows: For every θ outside the set Θ_t, let $\sigma^\epsilon(\theta, \cdot)$ be identical to $\sigma_s(\cdot)$ where s is the label of the path that type θ follows. For every θ and θ' in Θ_t such that $x_{\theta'}^\lambda < x_\theta^\lambda$ we know that there exists a piece of evidence $e_{\theta \backslash \theta'}$ in E_θ that is not in $E_{\theta'}$ because of the feasibility of x in (P_w^t). For all such cases, let $\sigma^\epsilon(\theta, e_{\theta \backslash \theta'}) = \epsilon$. For every other piece of evidence e, let $\sigma^\epsilon(\theta, e) = (1 - k\epsilon)\sigma(\theta, e)$ where k is the number of $e_{\theta \backslash \theta'}$s with ϵ weight. We then get for every θ in Θ_t and θ' in Θ_s:

$$\text{If } s = t \quad \text{then} \quad \sigma^{\epsilon}_{\theta'\theta} \leq 1 - \epsilon < \frac{u(g(\theta'))}{u(g(\theta))} = 1$$

$$\text{If } s < t \quad \text{then} \quad \sigma^{\epsilon}_{\theta'\theta} \leq \sigma_t(E_{\theta'}) + k\epsilon, \text{ with } \sigma_t(E_{\theta'}) < \frac{u(g(\theta'))}{u(g(\theta))}$$

$$\text{If } s > t \quad \text{then} \quad \sigma^{\epsilon}_{\theta\theta'} = \sigma_s(E_{\theta}) < \frac{u(g(\theta))}{u(g(\theta'))}$$

These inequalities are strict because of the way σ^{ϵ} is defined and the fact that $\underline{u}_{\theta}$ is always strictly positive. For well chosen positive ϵ and λ, we can see that these inequalities still hold for the outcome function g_{λ} as well, meaning that it can be implemented using σ^{ϵ}.

All along, it was assumed that all types in Θ_t follow the same path in the game $\mathbf{G}$, but the argument still holds when they follow separate paths as well. We would just have to define $\sigma^{\epsilon}(\theta, \cdot)$ using the strategy on the path followed by type θ.

□

Study of the case where $\mathbf{a_0}$ does not exist. This happens when $\inf_{a\in A} u(a)$ is not attained. If $\inf_{a\in A} u(a) = -\infty$ then the situation is equivalent to having full evidence disclosure because the punishment can be as big as the principal wants. Formally, g is implementable if and only if there exists a strategy σ and $\alpha \in \mathbb{R}$ such that

$$\forall \theta, \theta' \quad \sigma_{\theta'\theta} u(g(\theta)) - (1 - \sigma_{\theta'\theta})\alpha \leq u(g(\theta'))$$

First, note that if $E_{\theta} \subseteq E_{\theta'}$ then $\sigma_{\theta'\theta}$ is necessarily equal to 1 which implies $u(g(\theta)) \leq u(g(\theta'))$. Consider the following strategy : if the agent reports type θ, ask for all evidence in E_{θ} with the same probability. Then $\forall \theta, \theta'$, if $E_{\theta} \not\subseteq E_{\theta'}$, the above inequality is satisfied for α large enough. Because we have a finite number of such inequalities, we can take the largest α to satisfy all of them. We conclude that if $\inf_{a\in A} u(a) = -\infty$, g is implementable if and only if

$$\forall \theta, \theta' \quad E_{\theta'} \subseteq E_{\theta} \Longrightarrow u(g(\theta')) \leq u(g(\theta))$$

If on the other hand $\inf_{a\in A} u(a)$ is finite, we can set it to 0 w.l.o.g and denote by a_{ϵ} an action such that $u(a_{\epsilon}) = \epsilon$ for all $\epsilon > 0$. By continuity of u, such action always exists. In this case, g is implementable if and only if there exists a strategy σ such that

$$\begin{aligned} \exists \epsilon > 0 \text{ s.t } \forall \theta, \theta' \qquad & \sigma_{\theta'\theta} u(g(\theta)) + (1 - \sigma_{\theta'\theta})\epsilon \leq u(g(\theta')) \\ \Leftrightarrow \forall \theta, \theta' \qquad & \text{if } u(g(\theta')) < u(g(\theta)) \text{ then } \sigma_{\theta'\theta} u(g(\theta)) < u(g(\theta')) \end{aligned}$$

We conclude that a strategy σ implements g iff

$$\forall \theta, \theta', \text{ if } u(g(\theta')) < u(g(\theta)) \text{ then } \sigma_{\theta'\theta} < \frac{u(g(\theta'))}{u(g(\theta))}$$

Implementation results follow from Lemma 1, where, in this context, certain inequalities are replaced with strict inequalities. The subsequent results still hold but have to be modified accordingly.

□

Zero-sum game interpretation of Proposition 2. Consider the zero-sum game Γ_k defined as follows, having a value $w \leq 1$:

- Player 1 (Fictitious): chooses a type j in $J = \{1, \ldots, k-1\}$.
- Player 2 (The principal): chooses a document e_k^l in $\{e_k^1, \ldots, e_k^{q_k}\}$. Let $L = \{1, \ldots, q_k\}$.
- If $e_k^l \in E_{\theta_j}$ then player 1 gets $\frac{u(g(\theta_k))}{u(g(\theta_j))} = \frac{1}{b_j}$, otherwise he gets 0.

The matrix Λ of this game is therefore given by $\forall j \in J, l \in L, \Lambda_{jl} = \frac{\mathbb{1}_{\{e_k^l \in E_{\theta_j}\}}}{b_j} = \frac{A_{jl}}{b_j}$. The von Neumann theorem ensures that there exists $X^* \in \Delta(J)$ and $Y^* \in \Delta(L)$ such that:

$$\forall l \in L, (X^*\Lambda)_l \geq w \text{ and } \forall j \in j, (\Lambda Y^*)_j \leq w$$

If $w = 0$, then $\forall j \in j, (\Lambda Y^*)_j \leq w$ translates into $\Lambda Y^* = 0$ because all terms are positive. It is easy to see that in this case P_k is feasible for every $x = \alpha Y^*$ with $\alpha > 0$, and that its value is unbounded.

Otherwise, $0 < w \leq 1$, let $x^* = \frac{1}{w} Y^*$ and $y^* = \frac{1}{w} X^*$. The following program D_k is the dual of P_k:

$$\begin{aligned} \text{Min} \quad & y \cdot b \\ \text{s.t.} \quad & yA \geq c \\ & y \geq 0 \end{aligned}$$

We check that P_k is feasible for x^* and D_k is feasible for y^*, and that $y^* \cdot b = c \cdot x^* = \frac{1}{w}$. The value of P_k is therefore $v = \frac{1}{w} \geq 1$.

Conversely, if there exists a zero column of A, i.e. $\exists l_0$ s.t. $A_{jl_0} = 0, \forall j$, then P_k is feasible for every x such that $x_{l_0} = \alpha > 0$ and $x_l = 0, \forall l \neq l_0$. Therefore, P_k's value is unbounded. Let $Y^* \in \Delta(L)$ be such that $Y_{l_0} = 1$. We then have $\Lambda Y^* = 0$ and the value of Γ_k is 0.

If A has no zero column, then we check that P_k is feasible for $x = 0$ and D_k is feasible for y large enough. P_k has therefore a value v and there exists $x^* \geq 0$ and $y^* \geq 0$ such that $y^* \cdot b = c \cdot x^* = v$. Assuming that $v \geq 1$, define $X^* = \frac{1}{v} y^*$ and $Y^* = \frac{1}{v} x^*$. We check that $X^* \in \Delta(J)$, $Y^* \in \Delta(L)$ and that:

$$\forall l \in L, (X^* \Lambda)_l \geq \frac{1}{v} \text{ and } \forall j \in J, (\Lambda Y^*)_j \leq \frac{1}{v}$$

Γ_k's value is therefore $w = \frac{1}{v} \leq 1$.

□

On the concavity assumption of Proposition 4. The existence of an optimal deterministic outcome function relies upon the concavity assumption stated in Proposition 4:

$$\forall \theta, \exists c_\theta \text{ concave such that } \forall a \in A, v(\theta, a) = c_\theta(u(a))$$

In this section, we construct an example where this condition is violated and find a randomized outcome function that improves the principal's payoff compared to deterministic ones.

Consider the following framework: $A = [0, 1]$, $u(a) = a$ and $\Theta = \{\theta_1, \theta_2, \theta_3\}$ with $E_{\theta_1} = \{e_1\}$, $E_{\theta_2} = \{e_2\}$ and $E_{\theta_3} = \{e_1, e_2\}$. The principal utility functions have the following shape:

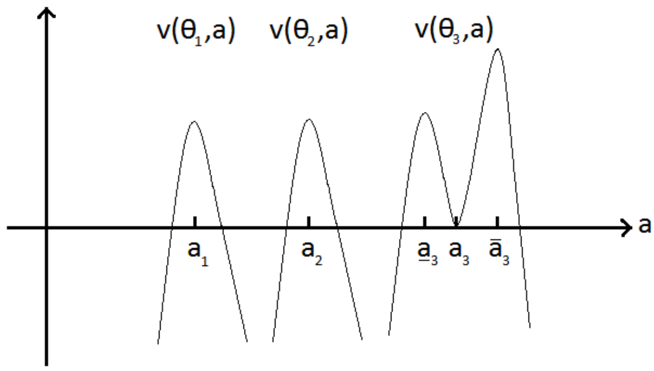

Figure 1: Principal utility functions

where $a_3 = a_1 + a_2$ and $v(\theta_2, a) = v(\theta_1, a - (a_2 - a_1))$. Let the probability distribution of types be μ such that $\mu(\theta) = \frac{1}{3}$ for all $\theta \in \Theta$. Note that the assumptions of Proposition 4 are all satisfied except for the concavity assumption, which $v(\theta_3, \cdot)$ violates. Our goal is to show that in this case, the principal can strictly increase his payoff by randomizing over actions. First, we determine the best deterministic outcome function, and then we find a non-deterministic function that gives a strictly higher payoff to the principal.

An outcome function g is implementable if and only if $g(\theta_1)+g(\theta_2) \geq g(\theta_3)$ and $g(\theta_3) = \max_i g(\theta_i)$. We know that $a_3 = a_1 + a_2$ which guarantees that the outcome function $\tilde{g}$ given by $\tilde{g}(\theta_i) = a_i$ for $i \in \{1, 2\}$ and $\tilde{g}(\theta_3) = \underline{a}_3$ is implementable. Let's show that $\tilde{g}$ is the best deterministic outcome function. Consider an implementable deterministic outcome function g, and let $V(g) = \frac{1}{3}\sum_{i=1}^{3} v(\theta_i, g(\theta_i))$ denote the expected payoff of g. If $g(\theta_1) \leq a_1$ and $g(\theta_2) \leq a_2$, then the implementability condition and the shape of the principal's utility functions ensure that $V(g) \leq V(\tilde{g})$. If $g(\theta_1) > a_1$ or $g(\theta_2) > a_2$, the only benefit would be to allow $g(\theta_3)$ to be larger than a_3. Set $g(\theta_3) = a_3 + t$ for $t > 0$. We can show that for a fixed t, it is optimal to set $g(\theta_1) = a_1 + \frac{t}{2}$ and $g(\theta_2) = a_2 + \frac{t}{2}$. For such a configuration, we get:

$$V(g) = V(\tilde{g}) + \frac{1}{3}(\gamma(t) - 2\psi(\frac{t}{2}))$$

where

$$\begin{aligned}
v(\theta_1, a_1 + x) &= v(\theta_1, a_1) - \psi(x) \quad \forall x \geq 0 \\
v(\theta_2, a_2 + x) &= v(\theta_2, a_2) - \psi(x) \quad \forall x \geq 0 \\
v(\theta_3, a_3 + x) &= v(\theta_3, \underline{a}_3) + \gamma(x) \quad \forall x \geq 0
\end{aligned}$$

We conclude that $\tilde{g}$ is the best deterministic outcome function if $\gamma(t) \leq 2\psi(\frac{t}{2})$. Now, consider the non-deterministic outcome function $\overline{g}$ given by $\overline{g}(\theta_i) = a_i$ for $i \in \{1, 2\}$, $\overline{g}(\theta_3) = \underline{a}_3$ with probability α and $\overline{a}_3$ with probability $1 - \alpha$. For α such that $a_3 = \alpha\underline{a}_3 + (1 - \alpha)\overline{a}_3$, the function $\overline{g}$ is implementable and $V(\overline{g}) > V(\tilde{g})$.

□

References

Jesse Bull and Joel Watson. Hard evidence and mechanism design. *Games Econ. Behav.*, 58:75–93, 2007.

F. Forges and F. Koessler. Communication equilibria with partially verifiable types. *Journal of Mathematical Economics*, 41(7):793–811, 2005.

Jacob Glazer and Ariel Rubinstein. On optimal rules of persuasion. *Econometrica*, 72(6):1715–1736, November 2004.

Jacob Glazer and Ariel Rubinstein. A study in the pragmatics of persuasion: a game theoretical approach. *Theoretical Economics*, 1:395–410, 2006.

Barton L. Lipman and D. Seppi. Robust inference in communication games with partial provability. *Journal of Economic Theory*, 66:370–405, 1995.

Itai Sher. Credibility and determinism in a game of persuasion. *Games Econ. Behav.*, 71:409–419, 2011.