INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

505
2017

Renegotiation proof mechanism design with imperfect type verification

Francisco Silva.

www.economia.puc.cl • ISSN (edición impresa) 0716-7334 • ISSN (edición electrónica) 0717-7593

Renegotiation proof mechanism design with imperfect type verification[*]

Francisco Silva[†]

December 7, 2017

Abstract

I consider the interaction between an agent and a principal who is unable to commit not to renegotiate. The agent's type only affects the principal's utility. The principal has access to a public signal, correlated with the agent's type, which can be used to (imperfectly) verify the agent's report. I define renegotiation proof mechanisms and characterize the optimal one: there is pooling on top - types above a threshold report to be the largest type, while types below the threshold report truthfully - and no regret on top - the mechanism is sequentially optimal for the principal after the agent reports to be the largest type.

JEL classification: D8

Keywords: renegotiation proof, mechanism design, verification

[*] I would like to thank Mustafa Dogan, Nicolas Figueroa, Ju Hu, Samir Mamadehussene, Mallesh Pai, as well as the seminar participants at PUC Chile for their useful comments.

[†] Instituto de Economia, Pontificia Universidad Catolica de Chile, Vicuna Mackenna 4860, Piso 3, Macul, Santiago, Chile. (email: franciscosilva@uc.cl).

1 Introduction

In mechanism design, the principal and the agent(s) are assumed to be able to commit not to renegotiate an agreed upon contract. This ability to commit is, in general, crucial, as mechanisms that are optimal for the principal in a variety of settings are typically not ex-post efficient. But, a lot of times, there does not seem to be a compelling reason for one to think that the players are unable to renegotiate.

Consider the following example. Say that there is a benevolent and risk averse prosecutor (the principal) and an agent who might be innocent or guilty of committing a crime. The prosecutor's preferences are such that she wants to punish the agent, but only if he is guilty, while the agent simply wants to minimize his expected punishment. The principal is also able to receive an exogenous signal (evidence), correlated with the agent's guilt (his type). The principal's most preferred incentive compatible mechanism is a menu of two contracts: a risky contract, which imposes a large punishment if the signal is "bad" but a very small punishment otherwise; and a riskless contract, which imposes a constant punishment in between the previous two (see Silva (2017) and Siegel and Strulovici (2016)). In equilibrium, if the agent is guilty, he takes the riskless contract, but if he is innocent he takes the risky contract. This means that the simple observation that the agent has chosen the risky contract reveals to the principal that the agent is innocent. And yet, the mechanism mandates that the principal punishes the agent heavily if the evidence happens to be "bad". In such circumstances, one must wonder whether the principal would simply follow the previously agreed contract or whether she would approach the agent with a proposal to reduce his sentence, to which the agent would certainly not object to.

The approach that some of the literature on renegotiation proof mechanisms has followed has been to add a "renegotiation proof" constraint to the typical mechanism design problem (Green and Laffont (1987), Forges (1994), Neeman and Pavlov (2013)). While different papers have different definitions, the overall goal of adding the constraint is to guarantee that, if a mechanism is renegotiation proof, then, after the choice of the agent becomes known, the principal does not wish to propose a second alternative mechanism that the agent, at least for some types, prefers over the original one. More rigorously, consider some mechanism $d : M \to X$ - a mapping from some message set M to an outcome set X. Suppose that, in equilibrium, for some type, the agent chooses some $m \in M$. After observing m, imagine that the principal is able to propose the following to the agent: the agent can choose to implement outcome

$d(m)$ or, instead, choose a message $m' \in M'$ with the understanding that the outcome to be implemented will be $d'(m')$. If, for some m, there is a second mechanism $d' : M' \to X$ that the principal strictly prefers to propose after observing m, then d is not renegotiation proof.

One of the drawbacks of the previous literature is that it uses a "one-shot" criterion to determine whether a mechanism is renegotiation proof or not: it might be that d is not renegotiation proof because there is a "blocking" mechanism d', which might itself not be renegotiation proof. But if d' is not renegotiation proof, its validity as a blocking mechanism is put into question. As a result, these these type of constraints end up being too demanding.[1]

The alternative is to explicitly model the dynamic renegotiation game, where the principal is allowed to propose new renegotiation mechanisms indefinitely. However, this raises two issues. First, what model is the right model? Presumably, different models lead to different mechanisms being implemented as it is easy to think of different yet reasonable models to study the same phenomenon. And second, dynamic renegotiation games are typically much harder to solve and are much less tractable than their static mechanism design counterpart.

Strulovici (2017) is one of the few papers that follows the later approach and considers a dynamic renegotiation game, where the principal proposes binding mechanisms in each period until choosing to stop. Because of the difficulty of the problem, he makes several simplifying assumptions. In particular, he assumes that the agent has one of only two possible types and that the principal's utility is independent of the agent's type, as is the case, for example, of a trade framework. He shows that, if the negotiation frictions (the probability that the negotiation is exogenously terminated in each period) are negligible, the mechanism that is implemented is separating - the principal is able to infer the type of the agent - and ex-post efficient.

In this paper, I focus on a special set of mechanism design problems, where the agent's type impacts not his utility but the principal's (the opposite of Strulovici (2017)). This is the case, for example, of a defendant who, regardless of his guilt, wants to minimize his punishment; it is the case of a project manager who, regardless of his skill, wants to maximize the funding his project gets; it may also be the case of an expert who, regardless of his private information, wants the same decision to be made. In order for the principal to be able to separate between the agent's types, I assume

[1] This one shot criterion problem is related to the discussion over farsightedness in the literature on coalition formation. For example, simple notions of the "core" of a game suffer from the same criticism (see Ray (2007) for an overview).

that there is an exogenous public signal (the evidence in the case of the prosecutor) correlated with the agent's type. This signal allows the principal to (imperfectly) verify the claim of the agent and reward or punish him as a result. The environment I study is similar to Ben-Porath, Dekel and Lipman (2014) and to Mylovanov and Zapechelnyuk (2017) except that I focus on the case where there is a single agent: the principal must choose how much of the good to allocate to the agent and has preferences that depend on the agent's type.

I also follow the approach of adding a "renegotiation proof" constraint to the standard mechanism design problem, but I argue that, in this setting, this particular constraint does not have the "one-shot" criterion problem. The difference is that the opportunity to renegotiate is assumed to happen after the signal has been realized. In other words, suppose that, following some mechanism d, the agent has chosen message m and signal s has been realized. After observing m and s, the principal will update her belief about the agent's type. Let d' be the optimal incentive compatible mechanism for the principal given those beliefs and subject to the condition that, if the agent wants to implement d, he can. The "renegotiation proof" constraint imposes that the principal does not strictly want to propose d' after observing m and s.

This change in timing is key in that, once the signal is realized, what the agent finds optimal is independent of his type, unlike what happens before the signal is realized (remember that, while the ex-post utility of the agent is independent of his type, the ex-ante utility is not, because the type is correlated with the signal). So, given d', the agent chooses the same message m' for any type. As a result, receiving m' does not convey any new information to the principal, which, in turn, does not make her want to propose a new mechanism after observing the uninformative message m' chosen by the agent.

In the main part of the text (section 3), I characterize the optimal renegotiation proof mechanism. In contrast to Strulovici (2017), I find that there is no complete separation between the agent's types. In particular, in equilibrium, if the agent is one of the better types (if, for example, his skill is larger than some threshold), then he reports that his type is the best one, while, otherwise, he reports truthfully. So, there is pooling "on top". Furthermore, another feature of the optimal mechanism is that the principal exhibits no regret "on top", i.e. whenever the agent reports that his type is the best, the principal does not want to unilaterally change the outcome imposed by the mechanism, even if she was able to.

For most of the paper, I assume that the exogenous public signal is binary. In section 4, I extend the analysis to consider non-binary signals and show that, at least when there are only two types, the results carry on. In section 5, I show that, unlike the commitment version of the problem, more information might actually be bad for the principal if she cannot commit not to renegotiate.

In section 6, I address the issue of whether it is appropriate or not to allow the principal to propose a renegotiation mechanism after receiving message m but before the realization of signal s. In particular, I describe a dynamic renegotiation game where the principal is always able to propose further and further renegotiation offers to the agent both before and after the signal has arrived and discuss under what conditions would the optimal renegotiation proof mechanism be implemented through such a game.

In section 7, I discuss the related literature.

2 Model

There is one principal and one agent. The agent's private type is given by $\theta \in \{\theta_1, ..., \theta_N\} \equiv \Theta$, where $\theta_n \in \mathbb{R}$ is strictly increasing with n. The prior probability that $\theta = \theta_n$ is denoted by $p_n > 0$. The agent's type affects the distribution of a public random variable $s \in \{0, 1\}$. In particular, let $\pi(\theta) \in (0, 1)$ denote the conditional probability that $s = 1$, given θ. I assume that π is strictly increasing, so that larger values of θ are more likely to generate $s = 1$.

There is a single good labeled $x \in \mathbb{R}$. The agent's utility function is denoted by $u(x)$ and, in addition to being independent of θ, it is continuous, strictly increasing and concave. The principal's utility function is denoted by $v(x, \theta)$. I assume that, for all $\theta \in \Theta$, $v(\cdot, \theta)$ is strictly concave and has a maximum denoted by $x^*(\theta)$. Furthermore, v is assumed to be continuous and to have non-decreasing differences - for any $(x', x) \in \mathbb{R}^2$ such that $x' \geq x$, $\{v(x', \cdot) - v(x, \cdot)\}$ is non-decreasing - which implies that x^* is non-decreasing.[2]

A mechanism is a message set M and a function $d : M \times \{0, 1\} \rightarrow \mathbb{R}$, which maps the message $m \in M$ sent by the agent and the signal s to a decision $d_s(m) \in \mathbb{R}$. Let the set of all such functions be denoted by D^M, for each message set M.[3] Given a

[2] An example is

$$v(x, \theta) = -(x - \theta)^2$$

[3] While I do not consider random mechanisms, it can be shown that the optimal renegotiation proof mechanism is not random due to $u(\cdot)$ being concave and $v(\cdot, \theta)$ being concave for any $\theta \in \Theta$.

mechanism, the agent chooses what message m to send before the realization of the random variable s. A strategy for the agent is a function $\sigma : \Theta \to \Delta M$, where $\sigma(\theta)(m)$ represents the probability that the agent sends message m, when his type is θ. Let Φ^M be the set of all possible strategies when the message set is M.

A system $((M, d), \sigma)$ is the pair composed of the mechanism (M, d) and the strategy σ. System $((M, d), \sigma)$ is incentive compatible (IC) if and only if σ is a Bayes-Nash equilibrium of the game induced by the mechanism (M, d): for all $\theta \in \Theta$ and $m \in M$,

$$\sigma(\theta)(m) > 0 \Rightarrow E(u(d_s(m))|\theta) \geq E(u(d_s(m'))|\theta) \text{ for all } m' \in M$$

Notice that the expectation is taken over s. So, it is assumed that, when the agent chooses a message, he still has not observed the realization of s. Seeing as s is correlated with the agent's type θ, the agent's decision will also depend on it. As a result, while the agent's utility is independent of his type, his expected utility is not.

For each strategy σ, let $E^{\sigma}(v(x, \theta)|m, s)$ denote the expected utility of the principal of choosing x, conditional on message m having been sent and signal s having been realized (so that the expectation is over θ). Notice that, for any σ, and for any pair (m, s), $E^{\sigma}(v(\cdot, \theta)|m, s)$ is strictly concave and has a unique maximizer.

Definition 1 *A system $((M, d), \sigma)$ is renegotiation proof (RP) if, for all $m \in M$ and $s \in \{0, 1\}$,*

$$d_s(m) \geq \arg\max_{x \in \mathbb{R}} E^{\sigma}(v(x, \theta)|m, s)$$

Figure 1 provides a graphical representation of the RP constraint. The idea is that, after message m is sent and signal s is realized, there is no alternative $x \neq d_s(m)$ that makes the agent and the principal better off, given the principal's beliefs. Suppose that

$$d_s(m) < \arg\max_{x \in \mathbb{R}} E^{\sigma}(v(x, \theta)|m, s)$$

so that $d_s(m)$ is to the left of the dotted line in figure 1. Once message m is sent and signal s is realized, there is nothing that prevents the players from agreeing to breaking the previous agreement $d_s(m)$ and switching to x, where

$$x = \arg\max_{x \in \mathbb{R}} E^{\sigma}(v(x, \theta)|m, s)$$

which is at the dotted line in figure 1. But, if $d_s(m)$ is to the right of the dotted line, then there is no other choice x that makes both players better off: for all $x > d_s(m)$, the principal would be made strictly worst given her beliefs, while, for all $x < d_s(m)$, it would be the agent who would be made strictly worst.

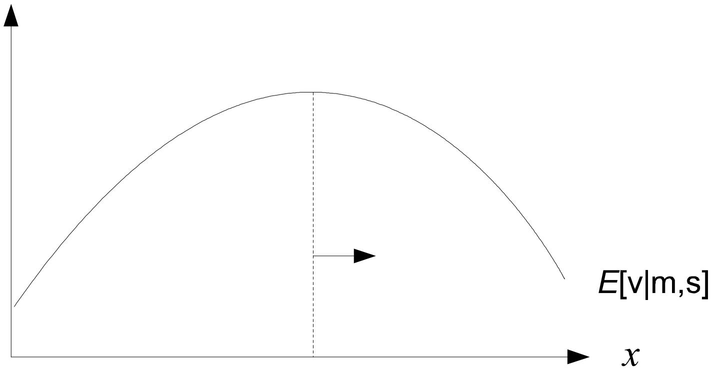

Figure 1: Graphical representation of $E^{\sigma}\left(v\left(x,\theta\right)|m,s\right)$

One can interpret this notion of renegotiation proofness as follows: a mechanism (M,d) is renegotiation proof if, for all $m \in M$ sent with positive probability, after observing signal $s \in \{0,1\}$, the principal does not strictly want to propose a new mechanism (M',d'), where

$$d'\left(m'\right) = \max\left\{\arg\max_{x\in\mathbb{R}}\ E^{\sigma}\left(v\left(x,\theta\right)|m,s\right), d_s\left(m\right)\right\} \text{ for all } m' \in M'$$

In words, d' is a constant mapping that returns either the principal's preferred choice given her beliefs, or the x that was promised to the agent in mechanism d.

Interpreted in this way, this definition of a renegotiation proof system resembles Neeman and Pavlov (2013), discussed in the introduction. However, it does not have the one shot criterion problem. Recall that the one shot criterion problem was that d' might itself not be renegotiation proof. In particular, it might be that, once d' is offered and the agent chooses some m', this conveys additional information to the principal, which might make her want to propose a second mechanism (M'',d''). However, in this framework, this is not a problem. Notice that once signal s becomes commonly known, there is no way of separating between the agent's types. In particular, given any mechanism proposed after the signal has been realized, if the agent finds it optimal to choose m' for some type, he finds it optimal for any type. So, when the principal chooses mechanism d', the fact that the agent chooses some m' does not convey any new information to her. Therefore, the one shot criterion critique does not apply provided d' is the best possible mechanism that the principal can choose, given her beliefs after observing m and s, which is the case precisely because, after s has been realized, there is no way of separating between the agent's types.

This approach is related with the literature on complete information renegotiation models, where the "renegotiation proof constraint" is simply that the mechanism be (ex-post) efficient (Maskin and Moore (1999), Neeman and Pavlov (2013)). Even though there is incomplete information in my framework, once s is realized, it is as if there is complete information because the agent's type does not impact his preferences.

2.1 Applications

2.1.1 Allocation problems

Consider the case where one of the players (the principal) has some resource or good that he wants to allocate to an applicant (the agent). For example, a prosecutor who must decide how to allocate a punishment to a defendant, a government who must decide how many units of a public good to allocate to a particular city, an aid agency that must decide the amount of resources devoted to a specific project, an investor who must choose how much money to invest on a certain firm. In all of these cases, the principal cares about the type θ of the agent: whether the defendant is guilty or innocent, whether the city is in real need of public goods, whether the project can produce results, whether the firm is likely to be profitable. The larger θ is, the more resources the principal wants to allocate to the agent. On the other hand, the agent simply wants to maximize the amount of resources he gets from the principal (or minimize his punishment in the case of the defendant), regardless of his type. In all of these cases, one would suspect that it would be possible for the principal to obtain some exogenous information about the agent's type, which is captured by signal s.

These allocation problems are somewhat similar to those studied by Ben-Porath, Dekel and Lipman (2014) and by Mylovanov and Zapechelnyuk (2017) except that, in these papers, the principal has one unit to allocate to one of many agents, while in this paper, the principal must choose how many units to allocate to a single agent.

2.1.2 Decision Maker and Expert

Consider the case where there is a decision maker (the principal) who must make a decision $x \in \mathbb{R}$. The consequences of choosing each $x \in \mathbb{R}$ depend on a random variable θ that the decision maker does not observe. The decision maker is able to hire an expert (the agent) who knows θ but is biased: regardless of θ, the expert wants the decision maker to choose as large of x as possible. The decision maker, through some other means, is able to get some imperfect information about θ, which is captured by signal s.

3 Characterization of the optimal mechanism

3.1 IC systems

I first start by deriving a property of all incentive compatible systems: that the agent's strategy profile is "monotone".

Lemma 2 *For any system $((M,d),\sigma)$, and for any $m \in M$ and $m' \in M$ such that $d_1(m) \geq d_1(m')$, if there is $\widehat{\theta} \in \mathbb{R}_+$ such that*

$$E\left(u\left(d_s(m)\right)|\widehat{\theta}\right) = E\left(u\left(d_s(m')\right)|\widehat{\theta}\right)$$

then

$$\begin{cases} E\left(u\left(d_s(m)\right)|\theta\right) \geq E\left(u\left(d_s(m')\right)|\theta\right) \text{ for all } \theta > \widehat{\theta} \\ E\left(u\left(d_s(m)\right)|\theta\right) \leq E\left(u\left(d_s(m')\right)|\theta\right) \text{ for all } \theta < \widehat{\theta} \end{cases}$$

where both inequalities are strict if $d_1(m) > d_1(m')$.

Proof. If $d_1(m) = d_1(m')$, for $\widehat{\theta}$ to exist, it must be that $d_0(m) = d_0(m')$, so the statement follows trivially. If $d_1(m) > d_1(m')$, then for $\widehat{\theta}$ to exist it must be that

$$\frac{\pi\left(\widehat{\theta}\right)}{1-\pi\left(\widehat{\theta}\right)} = \frac{u\left(d_0(m')\right) - u\left(d_0(m)\right)}{u\left(d_1(m)\right) - u\left(d_1(m')\right)}$$

Given that the function $\frac{\pi(\cdot)}{1-\pi(\cdot)}$ is strictly increasing, the statement follows. ■

Lemma 2 is useful in that it implies a certain monotonicity in how the agent reports as a function of his type in an IC system. In particular, take any IC system $((M, d), \sigma)$ such that every message that is sent with positive probability is distinct, i.e. if there is $m \in M$ and $m' \in M$ such that $\sigma(\theta)(m) > 0$ for some $\theta \in \Theta$ and $\sigma(\theta')(m') > 0$ for some $\theta' \in \Theta$, then $(d_0(m), d_1(m)) \neq (d_0(m'), d_1(m'))$. Lemma 2 implies that the larger the agent's type is the larger is $d_1(m)$ of the message(s) that he sends.

In Figure 2, I represent three possible strategy profiles assuming that

$$d_1(m_4) > d_1(m_3) > d_1(m_2) > d_1(m_1)$$

and $N = 3$. The last profile cannot be a part of an IC system because, if type θ_3 randomizes between messages m_3 and m_1, and, so, is indifferent between them, it cannot be that type θ_2 prefers to send message m_2.

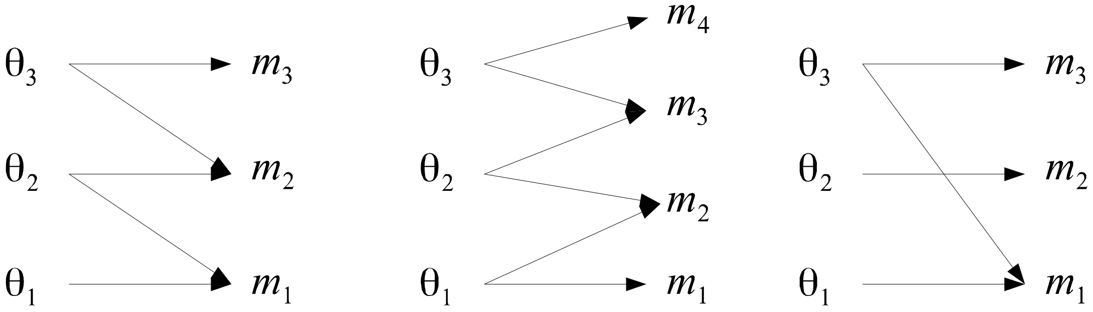

Figure 2: Example of three strategy profiles. The profile on the right cannot be a part of an IC system.

3.2 Optimal IC system

The problem of finding an optimal IC system can be made simpler by appealing to the revelation principle, which states that there is an optimal IC system such that the agent reports truthfully, i.e., $M = \Theta$ and $\sigma = \sigma^*$, where

$$\sigma^*(\theta)(m) = \begin{cases} 1 \text{ if } m = \theta \\ 0 \text{ otherwise} \end{cases} \quad \text{for all } m \in \Theta \text{ and } \theta \in \Theta$$

Let

$$V(d,\sigma) \equiv \sum_{n=1}^{N} p_n \sum_{\widehat{n}=1}^{N} \sigma(\theta_n)(\theta_{\widehat{n}})\left(\pi(\theta_n) v\left(d_1(\theta_{\widehat{n}}), \theta_n\right) + (1-\pi(\theta_n)) v\left(d_0(\theta_{\widehat{n}}), \theta_n\right)\right)$$

denote the principal's expected utility under mechanism (Θ, d) and when the agent reports according to σ. Notice that

$$V(d,\sigma^*) \equiv \sum_{n=1}^{N} p_n \left(\pi(\theta_n) v\left(d_1(\theta_n), \theta_n\right) + (1-\pi(\theta_n)) v\left(d_0(\theta_n), \theta_n\right)\right)$$

Proposition 3 *System $((\Theta, d^*), \sigma^*)$ is an optimal IC system if*

$$d^* \in \arg\max_{d \in D^\Theta} \left\{V(d,\sigma^*) \ \textit{subject to } i) \textit{ and } ii)\right\}$$

where i)

$$d_1(\cdot) \ \textit{is (weakly) increasing}$$

and ii) for all $n = 1, ..., N-1$,

$$E\left(u\left(d_s(\theta_n)\right)|\theta_n\right) = E\left(u\left(d_s(\theta_{n+1})\right)|\theta_n\right)$$

Proof. The problem of finding the optimal IC system, by definition, involves maximizing $V(d,\sigma^*)$ subject to all incentive constraints, $N-1$ per type, that prevent each type from mimicking any other type. By Lemma 2, one can add constraint $i)$ to this program without constraining it further. In order to prove proposition 3, I consider a relaxed program where, in addition to constraint i), I only consider the incentive constraint which prevents each type from mimicking the next largest type: type θ_n does not want to mimic type θ_{n+1}. In the appendix, I show that, in any solution of the relaxed program, this incentive constraint holds with equality (condition ii)) - the intuition is that the principal wants to reward larger types, so what constrains her is that lower types might want to pretend to be larger. Therefore, because conditions i) and ii) together imply that the system is IC, it follows that the solution of the relaxed problem is the optimal IC system. ■

Proposition 4 lists some of the properties of the optimal IC mechanism.

Proposition 4 *The optimal IC system* $((\Theta, d^*), \sigma^*)$ *is such that*

i.

$$d_1^*(\theta) \geq d_0^*(\theta) \text{ for all } \theta \in \Theta$$

ii.

$$d_1^*(\theta_N) \leq x^*(\theta_N) \text{ and } d_0^*(\theta_N) < x^*(\theta_N)$$

iii.

$$d_1^*(\theta_1) = d_1^*(\theta_0)$$

Proof. See appendix. ■

Part i) establishes that the agent is rewarded if $s = 1$, because of the positive correlation between the agent's type and the signal. Part ii) of the proposition is particularly important in that it establishes that the optimal IC system is not renegotiation proof, because the principal would prefer to increase $d_0^*(\theta_N)$ to $x^*(\theta_N)$. Part iii) states that the lowest type receives a constant outcome.

Figure 3 represents the optimal IC mechanism when $N = 2$.

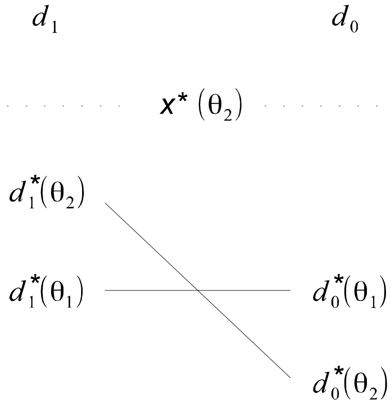

Figure 3: Representation of the optimal IC mechanism d^* when $N = 2$.

The case of $N = 2$ can be helpful in gaining some intuition on why system $((\Theta, d^*), \sigma^*)$ is optimal. Imagine first that the principal selects a mechanism that does not discriminate with respect to the message sent by the agent, i.e. $d(\theta_2) = d(\theta_1)$. In that case, she would choose $d_1(\theta_2) > d_0(\theta_2)$ simply because, when $s = 1$, it is more likely that the agent's type is θ_2, which makes the principal more willing to choose a larger x. Consider the following change to the mechanism: imagine that the principal allows the agent to admit that he is the low type ($\theta = \theta_1$), and, if he does, the principal chooses a constant $d_s(\theta_1)$ for any s such that

$$u(d_s(\theta_1)) = \pi(\theta_1) u(d_1(\theta_2)) + (1 - \pi(\theta_1)) u(d_0(\theta_2))$$

In words, if the agent admits that he is the low type, he receives a constant lottery that leaves him exactly indifferent to reporting to being the high type. The principal is happy with this change as she is risk averse ($v(\cdot, \theta)$ is strictly concave for all θ). In fact, even if the principal was risk neutral, she would approve of this change provided the agent is risk averse.

Figure 4 shows the optimal IC mechanism when $N = 3$ where one can see that the level of risk is smaller as the type of the agent becomes smaller.

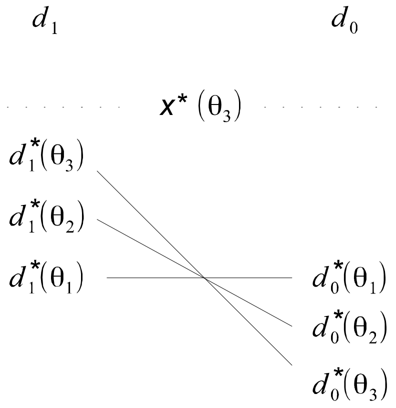

Figure 4: Representation of the optimal IC mechanism d^* when $N = 3$.

3.3 Optimal RPIC system

The challenge of analyzing RP systems is that beliefs matter: the posterior belief that the principal forms after observing the agent's report and the signal determines whether or not she is willing to change her promised decision. As a result, the revelation principle does not follow.

Let $\widehat{\Phi} \subseteq \Phi^\Theta$ be the set of "pooling on top" strategies, i.e. $\widehat{\Phi}$ is the set of all strategies

$\sigma \in \Phi^{\Theta}$ for which there is $n^* (\sigma) = 1, ..., N$ and $\tau (\sigma) \in [0, 1]$ such that

$$\begin{cases} \sigma(\theta_n)(\theta_N) = 1 \text{ for all } n > n^*(\sigma) \\ \sigma(\theta_{n^*})(\theta_N) = \tau(\sigma) \\ \sigma(\theta_{n^*})(\theta_{n^*}) = 1 - \tau(\sigma) \\ \sigma(\theta_n)(\theta_n) = 1 \text{ for all } n < n^*(\sigma) \end{cases}$$

In words, if $M = \Theta$ and $\sigma \in \widehat{\Phi}$, then there is $n^* \geq 1$ such that, if the agent's type is larger than θ_{n^*}, the agent claims to be of type θ_N - the largest possible type; if $\theta_n = \theta_{n^*}$, the agent randomizes between confessing to be type θ_{n^*} and claiming to be type θ_N; if the agent's type is smaller than θ_{n^*}, the agent confesses his type. Figure 5 shows an example where $N = 5$ and $n^* = 3$.

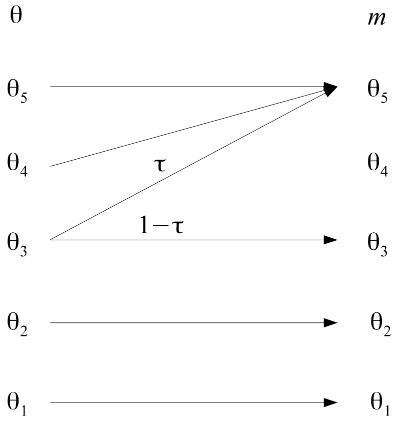

Figure 5: Example of a strategy profile $\sigma \in \widehat{\Phi}$ when $N = 5$ and $n^* = 3$.

Proposition 5 *System* $\left(\left(\Theta, \widehat{d}\right), \widehat{\sigma}\right)$ *is an optimal RPIC system if*

$$\left(\widehat{d}, \widehat{\sigma}\right) \in \arg \max_{(d,\sigma) \in D^{\Theta} \times \widehat{\Phi}} \left\{V(d, \sigma) \ \textit{subject to i), ii) and iii)}\right\}$$

where i)

$$d_1(\cdot) \ \textit{is (weakly) increasing}$$

and

$$d_1(\theta_n) = d_1(\theta_N) \ \textit{for all} \ n > n^*(\sigma)$$

ii) for all $n = 1, ..., N-1$,

$$E\left(u\left(d_s\left(\theta_n\right)\right)|\theta_n\right) = E\left(u\left(d_s\left(\theta_{n+1}\right)\right)|\theta_n\right)$$

and iii) for all $s = 0, 1$,

$$d_s\left(\theta_N\right) = \arg\max_{x\in\mathbb{R}}\ E^{\sigma}\left(v\left(x,\theta\right)|m = \theta_N, s\right)$$

In the optimal RPIC system, the agent either confesses his type or reports to be the largest type. He chooses the latter option only if his type is sufficiently large - there is "pooling on top". As a result, a report of θ_N induces the largest belief by the principal. Condition iii) states that, after that report, and conditional on the observed signal s, the mechanism $\widehat{d}$ chooses the principal's sequentially optimal choice - the RP constraint binds on top. So, if the principal observes the top message, she never regrets the choice she makes, unlike what happened in the second best system. Notice that, despite the RP constraint being a "message-by-message" constraint, the only message for which it binds is the top message. Finally, condition ii) states that each type is indifferent between reporting truthfully and reporting to be of the next largest type.

Figure 6 shows an example of the optimal RPIC system when $N = 2$. I use the following notation:

$$\gamma_s\left(m\right) \equiv \arg\max_{x\in\mathbb{R}}\ E^{\sigma}\left(v\left(x,\theta\right)|m, s\right)$$

Message θ_2 is sent by type θ_2 with probability 1 and by type θ_1 with some probability $\tau \in (0,1)$. As a result, the sequentially optimal x that follows message θ_2 - $\gamma_s\left(\theta_2\right)$ - depends on the signal s. The system is such that the principal's preferred choice is implemented after the top message has been sent.

Below, I provide a sketch of the proof of proposition 5. The detailed proof can be found in the appendix.

Proof (Sketch). There are 5 steps to the proof:

Step 1: $M = \Theta$

The first difficulty of characterizing the optimal RPIC system is that, in principle, the message set M can be arbitrarily large. However, because the model only contemplates one agent, it fits into the conditions for which the result of Bester and Strausz

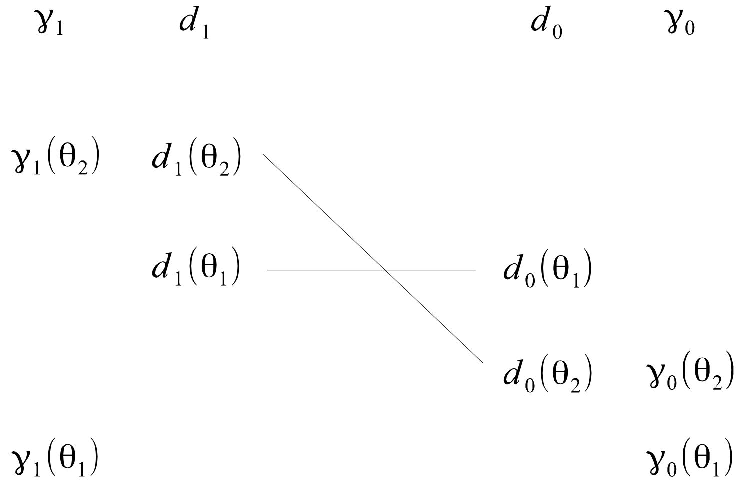

Figure 6: The optimal RPIC system when $N = 2$.

(2001) holds.[4] Therefore, by Bester and Strausz (2001), it follows that there is a RPIC system where $M = \Theta$.

Step 2: $\widehat{d_1}(m) \geq \widehat{d_0}(m)$ for any $m \in M$.

I show this result in the appendix, but the intuition is the following. Conditional on receiving any given message, the principal would prefer to select a larger x after signal $s = 1$ than after signal $s = 0$ simply because s and θ are positively correlated. What I show in the appendix is that this desire by the principal is not in conflict with constraints IC or RP. In particular, if there was some message m' such that $\widehat{d_1}(m') < \widehat{d_0}(m')$, the principal would do better by increasing $\widehat{d_1}(m')$ and decreasing $\widehat{d_0}(m')$ while preserving incentive compatibility and renegotiation proofness.

Step 3: The RP constraint only (possibly) binds at $m = \theta_N$.

Take any IC system such that message $m = \theta_N$ is the "top" message - of all messages that are sent with positive probability, it is the one with the largest d_1. In the appendix, I show that if message $m = \theta_N$ is RP, i.e. if

$$d_s(\theta_N) \leq \arg\max_{x \in \mathbb{R}} \; E^{\sigma}(v(x,\theta)|m = \theta_N, s) \text{ for } s = 0, 1$$

[4] In Bester and Strausz (2001), the principal can commit to a decision $x \in X$, which then constrains a second decision $y \in F(x)$ that the principal cannot commit to. Both x and y then enter the principal's utility function. My model can be interpreted as follows: the principal first commits to a decision $d_s(m)$ for some signal s and some message m. After the signal s and the message m are realized, the principal can choose any $x \leq d_s(m)$, and only the latter choice impacts her utility.

then the whole system is RP, i.e. for all $m \in \Theta$ sent with positive probability,

$$d_s(m) \leq \arg\max_{x \in \mathbb{R}} \; E^{\sigma}(v(x,\theta)|m,s) \text{ for } s = 0,1$$

The argument is easier to understand by looking at Figure 7:

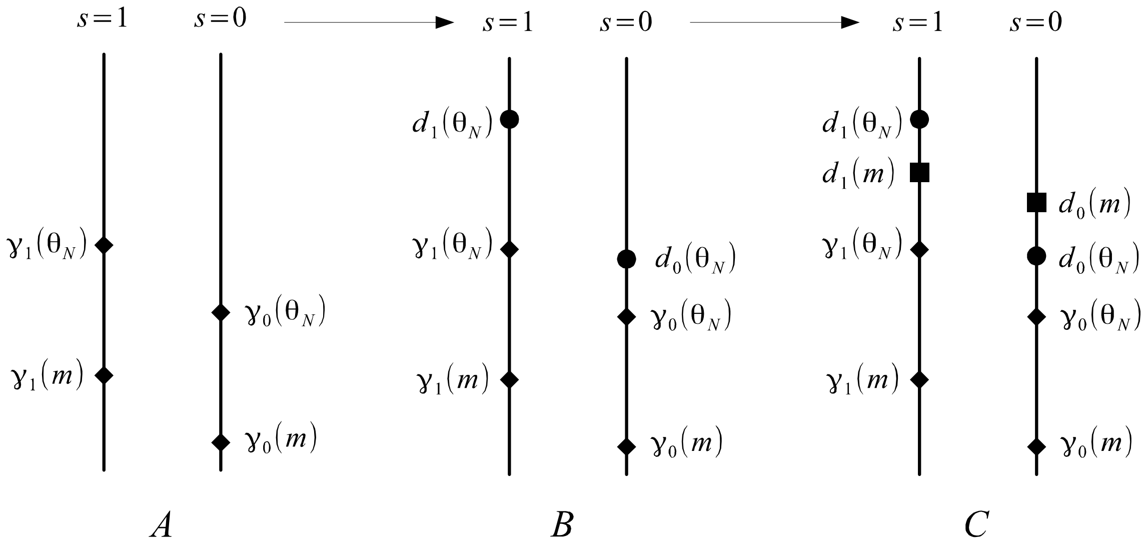

Figure 7: Part C shows that if the top message is RP, then so are the lower messages

On the left side of Figure 7 - part A - I represent, for each signal s, $\gamma_s(\theta_N)$ and $\gamma_s(m)$ for some m sent with positive probability. Because $m = \theta_N$ is the "top" message, it follows that $\gamma_1(\theta_N) \geq \gamma_0(\theta_N) > \gamma_1(m)$. In B, I add $d(\theta_N)$. Because the top message is RP, then $d_s(\theta_N) \geq \gamma_s(\theta_N)$ for $s = 0,1$. Furthermore, $d_1(\theta_N) > d_0(\theta_N)$ by step 2. Finally, in C, I add $d(m)$. By incentive compatibility, $d_1(m)$ and $d_0(m)$ must be "sandwiched" in between $d_1(\theta_N)$ and $d_0(\theta_N)$, which implies that message m must also be RP.

Step 3 implies that the only beliefs that matter are the ones after the top message. So, without loss of generality, one can focus on strategy profiles where the agent either sends the top message or confesses his type. In a way, the revelation principle - that the agent reports truthfully - only applies to types that do not send the top message. That is why there is an optimal RPIC system $\left(\left(\Theta, \widehat{d}\right), \widehat{\sigma}\right)$ where $\widehat{\sigma} \in \widehat{\Phi}$, i.e. there is pooling on top.

Step 4: Each type is indifferent to sending the message sent by the next largest type.

Seeing as I have already established that there is an optimal RPIC system where $M = \Theta$ and $\widehat{\sigma} \in \widehat{\Phi}$, what is left is to choose $\widehat{d} \in D^{\Theta}$ in addition to some n^* and some $\tau \in [0,1]$ in order to maximize the principal's expected utility. There are two constraints: IC and RP. The IC condition can be represented as three separate constraints: a) that d_1 be increasing, b) that each type does not want to send the message that the next largest type is sending, and c) that each type does not want to send the message that the next lowest type is sending. The RP constraint is simply a condition that applies only to the top message.

Consider a relaxed version of the problem where c) is eliminated. In the appendix, I show that, in the solution of the relaxed problem, b) always holds with equality: each type is indifferent to sending the message sent by the next largest type. The argument can be followed using part C of figure 7. Take the type sending message m and say that the next largest message is m_N. If the type sending message m was strictly better than sending message m_N, the principal could lower $d_0(m)$ to some x_0 and do strictly better provided that $x_0 > \gamma_0(m)$.

Given that b) holds with equality, c) is satisfied, so the solution of the relaxed problem is also the solution of the non-relaxed problem.

Step 5: The RP constraint holds with equality at $m = \theta_N$.

Consider the relaxed problem of step 4. Lowering $d_s(m_N)$ all the way to $\gamma_s(m_N)$ has no downside: it strictly increases the principal's expected utility by definition, and it reduces the incentives of the type sending the next lowest message to send message m_N. In fact, it is because the RP constraint always binds that the optimal IC system is not renegotiation proof. ■

Finally, the statement of proposition 5 makes it easy to compare the optimal RPIC system with two other systems of interest.

First, notice that if $n^* = N$, then the agent reports truthfully for any type. In that case, the optimal RPIC system would be ex-post efficient just like in Strulovici (2017). If that was the case, the mechanism would be such that

$$d_s(\theta) = x^*(\theta_N) \text{ for all } \theta \in \Theta \text{ and } s \in \{0,1\}$$

simply because the system would have to be sequentially optimal on top and incentive compatible. This is clearly not a good mechanism and is, for example, worse than the one discussed below. So, one must conclude that the optimal RPIC is not ex-post efficient.

Second, if $n^* = 1$ and $\tau = 1$, then the agent reports to be the largest type regardless of his actual type. In that case, the mechanism is equivalent to cheap talk. In particular, if the principal did not have commitment power, it can be shown that there is no informative equilibrium, so that the best that the principal can do is to choose x based on the signal but ignoring the agent's report.[5] Seeing as, in general, neither n^* nor τ are 1, one can conclude that the optimal RPIC system does better than the cheap talk alternative.

4 Non-binary signal

So far in the paper, I have assumed that $s \in \{0, 1\}$. This assumption plays a key role in proving Lemma 2, which allows me to focus on monotone strategies. As a result, if the signal is not binary, the problem of finding the optimal renegotiation proof system becomes considerably more complicated. Nevertheless, if $N = 2$, it is possible to show that the analogous to proposition 5 holds.

Let the support of s be some finite set S and let $f(s|\theta)$ denote the conditional probability of s given $\theta \in \{\theta_1, \theta_2\}$. Assume that $\frac{f(\cdot|\theta_2)}{f(\cdot|\theta_1)}$ is strictly increasing.

Proposition 6 *System* $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$ *is an optimal RPIC system if*

$$\left(\widetilde{d}, \widetilde{\sigma}\right) \in \arg \max_{(d,\sigma)\in D^{\Theta}\times\widehat{\Phi}} \left\{V(d,\sigma) \;\; \textit{subject to } i) \textit{ and } ii)\right\}$$

where i)

$$E\left(u\left(d_s\left(\theta_1\right)\right)|\theta_1\right) = E\left(u\left(d_s\left(\theta_2\right)\right)|\theta_1\right)$$

and ii) for all $s \in S$*,*

$$d_s\left(\theta_2\right) = \arg\max_{x\in\mathbb{R}} \;\; E^{\sigma}\left(v\left(x,\theta\right)|m = \theta_2, s\right)$$

Before proving proposition 6, it is convenient to characterize system $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$.

[5] If the principal cannot commit, and given the monotone structure on the reporting profile induced by incentive compatibility, each type of the agent prefers to report the largest message as it must lead to a larger x for any signal s.

Proposition 7 *If* $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$ *solves the program of proposition* 6*, then*

a) $\widetilde{d}_s(\theta_2)$ *is strictly increasing with* s*,*

b) $\widetilde{d}_s(\theta_1)$ *is constant with* s*.*

Proof. a) follows due to condition ii) in the statement of proposition 6 together with the assumption that $\frac{f(\cdot|\theta_2)}{f(\cdot|\theta_1)}$ is strictly increasing. b) follows because of the risk aversion assumption on both the principal and the agent. The complete proof is provided in the appendix. ■

As is clear by proposition 7, the optimal RPIC system when the signal is not binary is very much like the one characterized in the previous section: if the agent chooses to report to be the largest type, his reward depends on whether the signal supports his claim; if he confesses to be the lowest type, he receives a constant reward. Notice that, using proposition 7, it is easy to confirm that system $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$ is indeed RPIC.[6] So, in order to prove proposition 6, it is enough to show that the principal prefers system $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$ to any other RPIC system.

Proof of Proposition 6. From Bester and Strausz (2001), it follows that one only needs to consider systems where $M = \Theta$. Nevertheless, for convenience, let $M = \{m_0, m_1, m_2\}$ and take some RPIC system $((M, d), \sigma)$ such that m_0 is not sent ($\sigma(\theta)(m_0) = 0$ for any $\theta \in \Theta$). Without loss of generality, assume that $\Pr^\sigma\{\theta = \theta_2 | m_2\} \geq \Pr^\sigma\{\theta = \theta_2 | m_1\}$. I show that system $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$ is (weakly) preferred by the principal to system $((M, d), \sigma)$. I do this by building successive systems that continuingly improve the principal's expected utility until reaching the final system $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$.

Consider system $((M, d^1), \sigma)$ where i) $d^1(m_2)$ is sequentially optimal for the principal, i.e.

$$d^1(m_2) \in \arg\max_{x:S\to\mathbb{R}} \sum_{s\in S} f^\sigma(s|m_2)\, E^\sigma(v(x_s, \theta)\,|m_2, s)$$

[6] System $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$ is RP because, for all $s \in S$,

$$\widetilde{d}_s(\theta_1) > \min_{s\in S}\left\{\widetilde{d}_s(\theta_2)\right\} > x^*(\theta_1)$$

It is IC because the low type is indifferent between the two reports, and because the expected utility of reporting $m = \theta_2$ is larger for type θ_2 :

$$\sum_{s\in S}(f(s|\theta_2) - f(s|\theta_1))\, u\left(\widetilde{d}_s(\theta_2)\right) \geq 0$$

(which follows from See and Chen (2008)).

and ii)

$$d^1(m_1) \in \arg\max_{x:S\to\mathbb{R}} \left\{ \begin{array}{c} \sum_{s\in S} f^{\sigma}(s|m_1)\, E^{\sigma}(v(x_s,\theta)\,|m_1,s) \\ s.t.\ E(u(x_s)\,|\theta_1) \geq E(u(d_s^1(m_2))\,|\theta_1) \end{array} \right\}$$

where $f^{\sigma}(s|m)$ represents the probability that s is realized given that the message sent by the agent is m. In words, ii) simply means that the principal maximizes the expected utility she gets from when the agent sends message m_1, conditional on the low type's expected utility of sending message m_1 being larger than sending message m_2. For completeness, say that $d^1(m_0) = d^1(m_1)$.

By construction, system $((M,d^1),\sigma)$ is preferred by the principal to system $((M,d),\sigma)$ - the principal is certainly better off after message m_2, while after message m_1 she is also better off, because

$$E\left(u\left(d_s^1(m_2)\right)|\theta_1\right) \leq E(u(d_s(m_2))\,|\theta_1)$$

Consider system $((M,d^1),\sigma^1)$, where $\sigma^1 = \sigma$ except that

$$\sigma^1(\theta_1)(m_1) = \sigma^1(\theta_1)(m_1) - v$$

while

$$\sigma^1(\theta_1)(m_0) = v$$

where $v \geq 0$ is such that $\Pr^{\sigma^1}\{\theta = \theta_2|m_2\} = \Pr^{\sigma^1}\{\theta = \theta_2|m_1\}$ (see figure 8). Strategy σ^1 decreases the probability that the low type sends message m_1 and increases the probability it sends message m_0 in such a way that the posterior beliefs after messages m_1 and m_2 are equal. System $((M,d^1),\sigma^1)$ gives the same expected utility to the principal as does system $((M,d^1),\sigma)$.

Consider system $((M,d^2),\sigma^1)$ where i) $d^2(m_2)$ is sequentially optimal for the principal, and ii)

$$d^2(m_n) \in \arg\max_{x\in S\to\mathbb{R}} \left\{ \begin{array}{c} \sum_{s\in S} f^{\sigma^1}(s|m_n)\, E^{\sigma^1}(v(x_s,\theta)\,|m_n,s) \\ s.t.\ E(u(x_s)\,|\theta_1) \geq E(u(d_s^2(m_2))\,|\theta_1) \end{array} \right\}$$

for $n = 0,1$. The mechanism d^2 does the same as mechanism d^1 except that it also deals with message m_0, which is now sent with positive probability. Once again, it

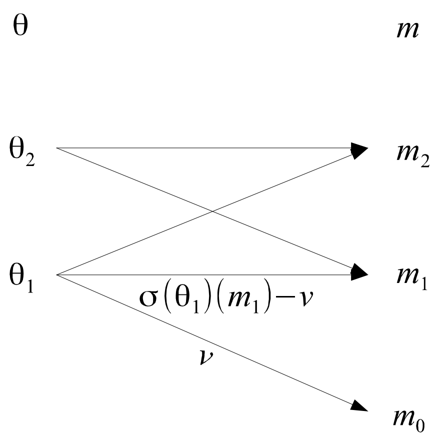

Figure 8: Representation of σ^1

follows that system $\left(\left(M, d^2\right), \sigma^1\right)$ is preferred by the principal to system $\left(\left(M, d^1\right), \sigma^1\right)$. Furthermore, it follows that, not only are the beliefs after messages m_1 and m_2 equal, but also $d^2\left(m_2\right) = d^2\left(m_1\right)$. Therefore, there is an equivalent system $\left(\left(M, d^2\right), \sigma^2\right)$ where σ^2 is such that $\sigma^2\left(\theta_2\right)\left(m_2\right) = 1$, $\sigma^2\left(\theta_1\right)\left(m_2\right) = 1 - v$ and $\sigma^2\left(\theta_1\right)\left(m_2\right) = v$ (see figure 9).

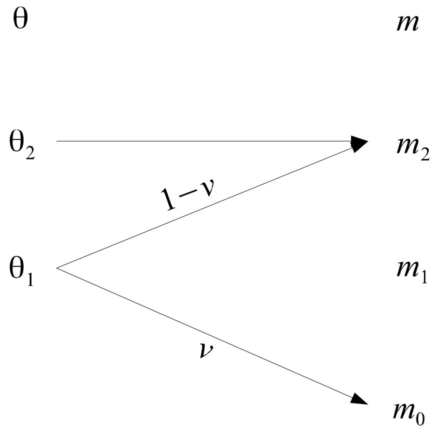

Figure 9: Representation of σ^2

In system $\left(\left(M, d^2\right), \sigma^2\right)$ messages m_2 and m_1 are merged so that one ends up with a system that is directly comparable to system $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$. In particular, seeing as

$$E\left(u\left(d_s^2\left(m_0\right)\right)|\theta_1\right) = E\left(u\left(d_s^2\left(m_2\right)\right)|\theta_1\right)$$

it follows that system $\left(\left(\Theta, \widetilde{d}\right), \widetilde{\sigma}\right)$ is (weakly) preferred to system $\left(\left(M, d^2\right), \sigma^2\right)$ by the principal, which completes the proof. ■

5 The impact of more information

In contexts of limited commitment power, the principal might actually be made worst off by having access to more (or better) information. In particular, it is known that, when the principal has no commitment power, being able to gather information from some source independent from the agent may harm her, because it may make meaningful communication with the agent harder (see, for example, Lai (2014) or Ishida and Shimizu (2016)). That is also the case for renegotiation proof systems as I illustrate with the following example.

Example 8 *Consider the case where $N = 2$, $\theta_2 = 2$, $\theta_1 = 1$, $u(x) = x$ and*

$$v(x, \theta) = \begin{cases} -(x-\theta)^2 & \text{if } x \geq 1.4027 \\ -\infty & \text{if } x < 1.4027 \end{cases}$$

so that it is as if there is a lower bound of 1.4027 on the set of x that the principal can choose from.

Assume that $s \in \{0_A, 0_B, 1\}$ and consider the following distribution: if $\theta = \theta_2$, the probability that $s = 1$ is equal to 0.5, and the probability that $s = 0_A$ is equal to $0.25 + \varepsilon$ for some $\varepsilon \geq 0$; if $\theta = \theta_1$, the probability that $s = 1$ is equal to 0.2, and the probability that $s = 0_A$ is equal to 0.4. Figure 10 illustrates.

The idea is that ε represents the quality of the signal for the principal. If $\varepsilon = 0$, it is as if there are only two possible signal realizations for s - 1 and 0 - but when ε increases, signals 0_A and 0_B become and more distinguishable. In figure 11, I show the results of comparing the expected utility for the principal from the optimal RPIC system for different values of ε.

As figure 11 illustrates, increases in ε do not always lead to an increase in the expected utility of the principal: more information might make the principal worst. To better understand this result, it is convenient to start by thinking about the full commitment problem.

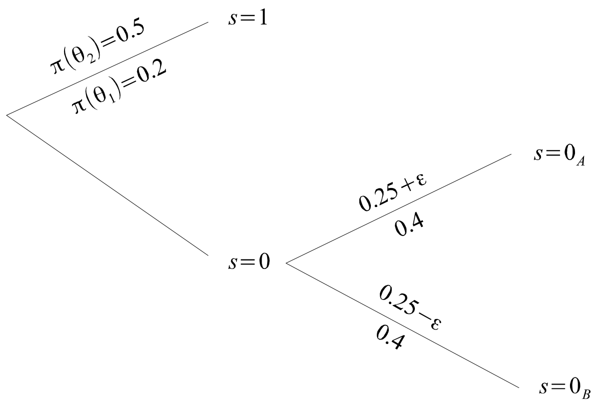

Figure 10: Example where $s \in \{0_A, 0_B, 1\}$.

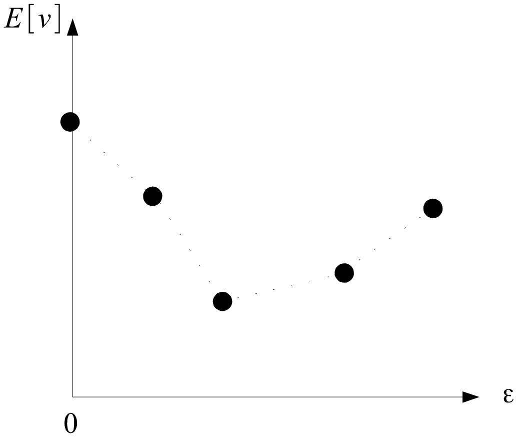

Figure 11: The picture shows the expected utility of the principal for 5 distinct levels of ε, each separated by 0.01.

If the principal had commitment power, more information could not make her worse. To see this, imagine that $\varepsilon = 0$ so that the optimal IC mechanism is such that, for any message, the outcome after signal 0_A and 0_B is the same. When ε increases from 0, the principal is able to choose whether to change the outcome after either signal, but is free not to. So, at worst, she is left with the same expected utility as when $\varepsilon = 0$.

The same does not happen if we consider renegotiation proof systems. In this case, when ε increases, the principal is no longer able to choose the same outcomes as she was when $\varepsilon = 0$ - if the principal becomes convinced that the agent's type is larger, she has no choice but to increase x. Conditional on receiving the top message θ_2, the fact that the principal has more information is a good thing, because the optimal RPIC is sequentially optimal on top. But, changes after message θ_2 force the principal to change *what happens after message θ_1 in order for the system to be incentive compatible. In this particular example, the principal has no choice but to increase the (constant) x that is implemented after receiving message θ_1, which is detrimental to her. As a result, the overall impact of having better information can actually be negative.*

6 The renegotiation game

Recall that my definition of renegotiation proofness essentially says that if a system is renegotiation proof, the principal should not want to propose an alternative renegotiation mechanism after observing the message m sent by the agent and after observing signal s. This definition does not have the one-shot problem because, once the signal is revealed, the choice of the agent is no longer informative, which eliminates any desire by the principal to renegotiate the alternative mechanism. However, the reader might wonder whether this is an appropriate definition. In particular, what is preventing the principal from proposing an alternative mechanism after observing m but before observing s? Below, I describe a simple renegotiation game that implements the optimal RPIC mechanism even though the principal is able to make several renegotiation offers to the agent before and after the signal is realized.

Consider the following renegotiation game:

In period 1, the principal proposes a mechanism (M, d^1) where $d^1 : M \times \{0,1\} \rightarrow \mathbb{R}$. Given d^1, the agent chooses $m_1 \in M$. The choice of the agent binds the players in the sense that it is necessary that both players agree in a different outcome for $d^1(m_1)$ not to be implemented. In particular, for any period $t \leq T$, the principal proposes mechanism $(\{M \cup \{r\}\}, d^t)$ where $d^t : \{M \cup \{r\}\} \times \{0,1\} \rightarrow \mathbb{R}$ and $d^t(r) = d^{t-1}(m_{t-1})$ - the agent has the choice of rejecting (r) the new alternatives that the principal offers and sticking with what has been agreed to in the previous period. After observing d^t, the agent chooses $m_t \in \{M \cup \{r\}\}$.

At the end of period T, signal $s \in \{0,1\}$ is realized and is publicly available. This means that at period $T+1$, s is known so that the mechanism that the principal proposes is only a function of the message chosen by the agent: the principal proposes $d^{T+1} : \{M \cup \{r\}\} \rightarrow \mathbb{R}$ such that $d^{T+1}(r) = d^T_s(m_T)$. After observing d^{T+1}, the agent chooses his message $m_{T+1} \in \{M \cup \{r\}\}$. The difference to the periods before T is that, at the end of period $T+1$, the principal is able to choose between implementing decision

$d^{T+1}(m_{T+1})$, which would end the game and return a utility of $v\left(d^{T+1}(m_{T+1}), \theta\right)$ for the principal (given θ) and of $u\left(d^{T+1}(m_{T+1})\right)$ for the agent, or to proceed to the following period. In each period $t > T+1$, the timing is the same: the principal proposes a mechanism $d^t : \{M \cup \{r\}\} \to \mathbb{R}$ such that $d^t(r) = d^{t-1}(m_{t-1})$, the agent chooses $m_t \in \{M \cup \{r\}\}$ and the principal chooses whether to implement $d^t(m_t)$. The only difference is that, should $d^t(m_t)$ be implemented, the payoffs are discounted by $\delta_p \in (0,1)$ in the case of the principal and by $\delta_a \in (0,1)$ in the case of the agent so that the payoff vector would be

$$\left(\delta_p^{t-T-1} v\left(d^t(m_t), \theta\right), \delta_a^{t-T-1} u\left(d^t(m_t)\right)\right)$$

It can easily be shown that all perfect Bayesian equilibria of this game implement the optimal RPIC mechanism discussed in the previous section. There are two parts to the argument. First, consider what happens after the signal is realized, at the end of period T. From then on, it is impossible for the principal to further separate between the agent's types. So, given that there is discounting, the best mechanism that the principal can offer is a constant mechanism of either the sequentially optimal choice of the principal or her last promise to the agent - whichever is largest - and then immediately implement it. If $T = 1$, then, in period 1, the principal, anticipating what will happen once the signal is revealed, simply offers the optimal RPIC mechanism so that it does not get renegotiated at period $T+1$. If $T > 1$, the problem is no different. It is best for the principal to wait until period T to make a mechanism offer (by making "bad" offers that do not tie her hands in the first $T-1$ periods) as making offers before period T only increases the risk that she will want to renegotiate them away in the following periods.[7] Therefore, one can conclude that, in this game, even though the principal has the opportunity to make several offers to the agent before the signal is realized, she chooses not to and the optimal RPIC mechanism is implemented.

Of course, had the game been different this would no longer necessarily be the case. For example, if the public signal arrived randomly, then the principal would have an added incentive to propose a proper mechanism earlier, which might get renegotiated away in the following periods, should the signal not get realized. As a result, in that case, the principal would not be able to do as well as with the optimal RPIC system. Nevertheless, it should be clear that, even in that case, the principal would not want to

[7] Evans and Reiche (2015) consider a similar problem except that the game ends at period T without any signal being realized. The authors focus on finding the set of all mechanisms that can be proposed at period 1 and do not get renegotiated.

implement an ex-post efficient mechanism, in contrast to Strulovici (2017). Given the renegotiation opportunities that exist after the signal has been realized, the best ex-post efficient mechanism that the principal can hope to implement is the one described in section 4 when $n^* = N$ - a renegotiation proof mechanism (given my definition) with truthful reporting. But, as discussed in section 4, that is worst for the principal than the cheap talk alternative where the principal ignores the agent's report, so that, if nothing else, the principal would prefer to go with that mechanism instead, which, by definition, elicits no information from the agents.

7 Related literature

Renegotiation proof mechanisms have been studied in contexts of complete and incomplete information. If there is complete information, the problem is simplified by the fact that there are no different types for the same player who might want different things. Therefore, notions of renegotiation proofness are tied together with ex-post Pareto efficiency (Maskin and Moore (1999) and Neeman and Pavlov (2013)). In particular, if nothing else, if a mechanism is renegotiation proof, then it must be efficient. If not, agents would simply somehow settle on something that made them all better off. Adding incomplete information complicates the problem in that expressing a willingness to renegotiate reveals information which might impact the desire to renegotiate of the other player(s).

Green and Laffont (1987) discuss how to model renegotiation proof mechanisms in a context with multiple agents. In their paper, a mechanism is renegotiation proof if no agent wishes to change their report after observing everyone else's report. Forges (1994) and Neeman and Pavlov (2013) differ from Green and Laffont (1987) in that agents are not only allowed to choose a different report but they are also able to propose a different mechanism altogether. However, as discussed, they run into the one shot criterion problem, which makes their renegotiation proof requirement too demanding. Goltsman (2011) and Beshkar (2016) use one-shot renegotiation proof definitions to study the hold-up problem and the role of arbitration in trade agreements respectively.

Strulovici (2017) studies a renegotiation game similar to the one of the previous section with two differences: first, there is no signal in his paper, so that the game starts at period $T + 1$, and second, should the principal choose to proceed to the next period, rather than having discounted payoffs, the author assumes that there is a probability that the currently agreed upon contract is implemented. He then studies the case where such probability (the negotiation frictions as he puts it) is arbitrarily small. Seeing as he assumes that the agent's type affects the agent's utility, as opposed to the principal's as in this paper, his problem of finding the set of perfect Bayesian equilibria of the renegotiation game becomes far more complex than mine.

The mechanism that is implemented in Strulovici (2017) is also "posterior efficient" like this paper's, but there is complete separation between the (two) agent's types: the principal proposes a mechanism with two options at period 1, the agent chooses differently depending on his type, and the principal immediately implements the mechanism. The driving force for the complete separation result is that, should there not be complete separation and should the negotiation frictions be small, there would be an impetus for the principal to propose further mechanisms which succeed in screening between the agent's types. However, that impetus does not exist in my paper once the signal has been realized, because, when that happens, the agent's decision becomes independent of his type and that is why, in my paper, the optimal RPIC mechanism implies partial but not full separation between the agent's types.

There is also a literature that studies the impact of assuming that players cannot commit not to renegotiate in long-term relationships (Laffont and Tirole (1990), Hart and Tirole (1988), Battaglini (2007), Maestri (2017)), as opposed to a short-term relationship like in this paper. The idea is to model the interaction between two players who, at the beginning of a long relationship, may write a long-term contract but may not commit to renegotiate it in future periods. The renegotiation protocol is typically one-shot - one of the players proposes an amendment to the active contract, which, if accepted, produces immediate effects.

In the optimal RPIC system, there is pooling on top - the larger types report to be the largest type, while the lowest types report truthfully. This type of equilibrium is similar to those found in Kartik (2009) or Chen (2011), where the top types of the agent also pool. These papers extend the classic cheap talk framework of Crawford and Sobel (1982) to include costs of lying in the case of the former, and a probability that either the sender or the receiver are naive in the latter. They provide an explanation for the phenomenon of sender's exaggeration - self interested senders exaggerate their claims even though their bias is known by the receiver.

By contrast, in this paper, the only cheap talk equilibrium is uninformative - the principal ignores the agent's report and decides based solely on the signal. And, even if the principal has some commitment power and can implement any renegotiation proof mechanism, it follows that there are many systems where there is no pooling on top. What I show in the paper is that, at least one of the *optimal* renegotiation proof systems exhibits pooling on top, because the fact that the agent's strategy is monotone implies that the renegotiation proof constraint only binds the top message.

The setting that I study is similar to the one of Ben-Porath, Dekel and Lipman (2014) and of Mylovanov and Zapechelnyuk (2017) in that there is a principal who cares about the type of the agent, agents whose utility is independent of type, no transfers and an exogenous signal correlated with the agent's type. In terms of the setting, there are two main differences. First, both papers consider a problem where the principal chooses one of the many agents to allocate one unit of a good, while I focus on the case where there is a single agent and the principal chooses how many units of a good to allocate to him. Second, they have different assumptions with respect to the verification technology: Ben-Porath, Dekel and Lipman (2014) assume that, at a cost, the principal can get to know the type of a given agent, while Mylovanov and Zapechelnyuk (2017) assume that only the chosen agent can be verified. In addition to this, both papers assume that the principal has commitment power, while the largest portion of this paper is devoted to studying limited commitment.

8 Appendix

8.1 Proof of Proposition 3 (continued)

Proof. Consider the "relaxed" problem where the only two constraints considered are i) $d_1(\theta)$ is increasing; and ii) for all $n = 1, ..., N-1$,

$$\pi(\theta_n) u(d_1(\theta_n)) + (1-\pi(\theta_n)) u(d_0(\theta_n)) \geq \pi(\theta_n) u(d_1(\theta_{n+1})) + (1-\pi(\theta_n)) u(d_0(\theta_{n+1}))$$

I show that, in any solution of the relaxed problem, ii) must hold with equality: for all $n = 1, ..., N-1$,

$$\pi(\theta_n) u(d_1(\theta_n)) + (1-\pi(\theta_n)) u(d_0(\theta_n)) = \pi(\theta_n) u(d_1(\theta_{n+1})) + (1-\pi(\theta_n)) u(d_0(\theta_{n+1}))$$

Suppose not. Then, there is some type θ_n such that

$$\pi(\theta_n) u(d_1(\theta_n)) + (1-\pi(\theta_n)) u(d_0(\theta_n)) > \pi(\theta_n) u(d_1(\theta_{n+1})) + (1-\pi(\theta_n)) u(d_0(\theta_{n+1}))$$

By i), it follows that $d_1(\theta_{n+1}) \geq d_1(\theta_n)$ and so $d_0(\theta_{n+1}) < d_0(\theta_n)$.

Assume first that $x^*(\theta_{n+1}) > d_0(\theta_{n+1})$. Then, the principal would be better off by increasing $d_0(\theta_{n+1})$ and still satisfy ii), which is a contradiction to optimality of the relaxed problem. Assume instead that $x^*(\theta_{n+1}) \leq d_0(\theta_{n+1})$. This implies that $x^*(\theta_n) < d_0(\theta_n)$. As a result, the principal would prefer to lower $d_0(\theta_n)$ and still satisfy ii), which is a contradiction to optimality of the relaxed problem. Therefore, ii) holds with equality. ■

8.2 Proof of Proposition 4

Recall that $((\Theta, d^*), \sigma^*)$ solves the relaxed problem described in proposition 3, where the only constraints are that $d_1(\cdot)$ is increasing and that each type does not want to mimic the next largest type. For convenience, I refer to the first constraint by $C1$ and the second one by $C2$.

Proposition 4.i)

$$d_1^*(\theta) \geq d_0^*(\theta) \text{ for all } \theta \in \Theta$$

Proof. Suppose not and let $\theta_{\widehat{n}} \in \Theta$ be the largest $\theta \in \Theta$ such that $d_1^*(\theta_{\widehat{n}}) < d_0^*(\theta_{\widehat{n}})$. Let

$$z(\theta) \equiv \pi(\theta) d_1^*(\theta) + (1 - \pi(\theta)) d_0^*(\theta)$$

Consider the following alternative mechanism d' where

a) for all $\theta > \theta_{\widehat{n}}$,

$$d'(\theta) = d^*(\theta)$$

b) for all $\theta \leq \theta_{\widehat{n}}$,

$$d_1'(\theta_n) = d_1'(\theta_{\widehat{n}}) = \min\{d_1'(\theta_{\widehat{n}+1}), z(\theta_{\widehat{n}})\}$$

and

$$d_0'(\theta) = \frac{z(\theta) - \pi(\theta) d_1'(\theta)}{(1 - \pi(\theta))}$$

I first show that $z(\cdot)$ is decreasing for $\theta \leq \theta_{\widehat{n}}$. Take any $\theta_n < \theta_{n+1} \leq \theta_{\widehat{n}}$. If $d^*(\theta_n) = d^*(\theta_{n+1})$, the statement trivially follows. If $d^*(\theta_n) \neq d^*(\theta_{n+1})$, it follows that

$$\frac{\pi(\theta_n)}{(1 - \pi(\theta_n))} = \frac{u(d_0^*(\theta_n)) - u(d_0^*(\theta_{n+1}))}{u(d_1^*(\theta_{n+1})) - u(d_1^*(\theta_n))} \leq \frac{d_0^*(\theta_n) - d_0^*(\theta_{n+1})}{d_1^*(\theta_{n+1}) - d_1^*(\theta_n)}$$

where the last inequality follows because u is concave and because

$$d_1^*(\theta_n) < d_1^*(\theta_{n+1}) < d_0^*(\theta_{n+1}) < d_0^*(\theta_n)$$

As a result, it follows that

$$z(\theta_n) \geq \pi(\theta_n) d_1^*(\theta_{n+1}) + (1 - \pi(\theta_n)) d_0^*(\theta_{n+1}) > z(\theta_{n+1}) \tag{1}$$

Notice also that

$$\begin{aligned} d_0'(\theta_n) - d_0'(\theta_{n+1}) &= \frac{z(\theta_n) - \pi(\theta_n) d_1'(\theta_n)}{(1 - \pi(\theta_n))} - \frac{z(\theta_{n+1}) - \pi(\theta_{n+1}) d_1'(\theta_{n+1})}{(1 - \pi(\theta_{n+1}))} \\ &= \frac{(1 - \pi(\theta_n))^{-1}}{(1 - \pi(\theta_{n+1}))} \left(\begin{array}{c} z(\theta_n)(1 - \pi(\theta_{n+1})) - z(\theta_{n+1})(1 - \pi(\theta_n)) \\ + (\pi(\theta_{n+1}) - \pi(\theta_n)) d_1'(\theta_{n+1}) \end{array} \right) \end{aligned}$$

By (1), it follows that

$$z(\theta_n)(1 - \pi(\theta_{n+1})) - z(\theta_{n+1})(1 - \pi(\theta_n)) \geq (\pi(\theta_n) - \pi(\theta_{n+1})) d_1^*(\theta_{n+1})$$

which implies that

$$d_0'(\theta_n) - d_0'(\theta_{n+1}) \geq \frac{1 - \pi(\theta_n)}{1 - \pi(\theta_{n+1})} (\pi(\theta_{n+1}) - \pi(\theta_n)) (d_1'(\theta_{n+1}) - d_1^*(\theta_{n+1})) > 0$$

because

$$d_1'(\theta_{n+1}) > d_1^*(\theta_{n+1})$$

so that $d_0'(\cdot)$ is decreasing for $\theta \leq \theta_{\widehat{n}}$.

System $((\Theta, d'), \sigma^*)$ satisfies $C1$ by definition. It also satisfies $C2$ because, for $n < \widehat{n}$,

$$E(u(d_s'(\theta_n)) | \theta_n) \geq E(u(d_s'(\theta_{n+1})) | \theta_n)$$

which follows because $d_0'(\cdot)$ is decreasing for $\theta \leq \theta_{\widehat{n}}$, while

$$E(u(d_s'(\theta_{\widehat{n}})) | \theta_{\widehat{n}}) \geq E(u(d_s'(\theta_{\widehat{n}+1})) | \theta_{\widehat{n}})$$

because

$$E(u(d_s'(\theta_{\widehat{n}})) | \theta_{\widehat{n}}) \geq E(u(d_s^*(\theta_{\widehat{n}})) | \theta_{\widehat{n}})$$

Finally, notice that under d', for every $\theta \leq \theta_{\widehat{n}}$, the expected x of reporting truthfully is the same (and equal to $z(\theta)$) but the risk is smaller than with d^*, because

$$d_1^*(\theta) < d_1'(\theta) \leq d_0'(\theta) < d_0^*(\theta)$$

As a result, it follows that

$$E\left(v\left(d_s'(\theta), \theta\right) | \theta\right) > E\left(v\left(d_s^*(\theta), \theta\right) | \theta\right)$$

because $v(\cdot, \theta)$ is strictly concave, which means that system $((\Theta, d'), \sigma^*)$ is strictly preferred by the principal to system $((\Theta, d^*), \sigma^*)$, which is a contradiction. ■

Proposition 4.ii)

$$d_1^*(\theta_N) \leq x^*(\theta_N) \text{ and } d_0^*(\theta_N) < x^*(\theta_N)$$

Proof. First, I show that $d_0^*(\theta_N) \leq x^*(\theta_N)$. Suppose not. In that case, if one considers an alternative mechanism d' such that $d' = d^*$ except that $d_0'(\theta_N) = x^*(\theta_N)$, we get a contradiction in that system $((\Theta, d'), \sigma^*)$ would satisfy $C1$ and $C2$ and would be strictly preferred by the principal to system $((\Theta, d^*), \sigma^*)$ because $v(\cdot, \theta)$ is strictly concave for all $\theta \in \Theta$.

Now, I show that $d_1^*(\theta_N) \leq x^*(\theta_N)$. Let $\widehat{n} \geq 1$ be such that $d^*(\theta_n) = d^*(\theta_{\widehat{n}})$ for all $n \geq \widehat{n}$. In that case, consider the following alternative mechanism d'' where, for all $\theta \in \Theta$ and $s = 0, 1$,

$$d_s''(\theta) = \begin{cases} d_s^*(\theta) \text{ if } \theta < \theta_{\widehat{n}} \text{ or if } (\theta \geq \theta_{\widehat{n}} \text{ and } s = 0) \\ \max\left\{d_1^*\left(\theta_{\widehat{n}-1}^*\right), x^*(\theta_N)\right\} \text{ if } \theta \geq \theta_{\widehat{n}} \text{ and } s = 1 \end{cases}$$

(where it is assumed that $d_1^*(\theta_0) < x^*(\theta_N)$). Once again, system $((\Theta, d''), \sigma^*)$ satisfies $C1$ and $C2$ and is strictly preferred by the principal to system $((\Theta, d^*), \sigma^*)$ because $v(\cdot, \theta)$ is strictly concave for all $\theta \in \Theta$.

It follows that if the statement is not true, that $d_1^*(\theta_N) = d_0^*(\theta_N) = x^*(\theta_N)$. This means that $d_s^*(\theta) = x^*(\theta_N)$ for all $\theta \in \Theta$ and $s = 0, 1$. Consider the alternative mechanism d''' where, for all $\theta \in \Theta$ and for $s = 0, 1$,

$$d_s'''(\theta) = \arg\max_{x \in \mathbb{R}} E\left(v(x, \theta) | s\right)$$

Notice that $d_s'''(\theta)$ is independent of θ and is such that $d_1'''(\theta) \neq d_0'''(\theta)$. As a result, it follows that system $((\Theta, d'''), \sigma^*)$ satisfies $C1$ and $C2$ and is strictly preferred by the principal to system $((\Theta, d^*), \sigma^*)$. ■

Proposition 4.iii)

$$d_1^*(\theta_1) = d_0^*(\theta_1)$$

Proof. Suppose not so that $d_1^*(\theta_1) > d_0^*(\theta_1)$. Consider the alternative mechanism d' where $d' = d^*$ except that

$$d_1'(\theta_1) = d_0'(\theta_1) = \pi(\theta_1) d_1^*(\theta_1) + (1 - \pi(\theta_1)) d_0^*(\theta_1) < d_1^*(\theta_1)$$

It follows that system $((\Theta, d'), \sigma^*)$ satisfies $C1$, it satisfies $C2$ because u is concave and is strictly preferred by the principal to system $((\Theta, d^*), \sigma^*)$ because $v(\cdot, \theta_1)$ is strictly concave, which is a contradiction. ■

8.3 Proof of Proposition 5

Before proving each of the steps from the main text, it is important to go through a number of results that are used throughout the proof.

The first thing to notice is that if there are two distributions F and F' over Θ such that

$$\max[\text{supp}[F]] \leq \min[\text{supp}[F']]$$

then

$$\arg\max_{x \in \mathbb{R}} E[v(x, \theta) | F] \leq \arg\max_{x \in \mathbb{R}} E[v(x, \theta) | F']$$

This observation allows me to show that any two non-distinct messages can be merged:

Lemma 5.1. *If there is an RPIC system $((M, d), \sigma)$ such that there are two messages $m' \in M$ and $m'' \in M$ such that $d(m') = d(m'')$, then system $((M, d), \sigma')$ is also RPIC, where $\sigma' = \sigma$ except that*

$$\sigma'(\theta_n)(m') = \sigma(\theta_n)(m') + \sigma(\theta_n)(m'') \text{ for all } n$$

and

$$\sigma'(\theta_n)(m'') = 0$$

Proof. Take any RPIC system $((M, d), \sigma)$ and any two messages $m' \in M$ and $m'' \in M$

such that $d(m') = d(m'') \equiv (\widehat{x}_1, \widehat{x}_0)$. Let

$$x'_s \equiv \arg\max_{x\in\mathbb{R}} \; E^{\sigma}\left(v\left(x,\theta\right)|m',s\right)$$

and

$$x''_s \equiv \arg\max_{x\in\mathbb{R}} \; E^{\sigma}\left(v\left(x,\theta\right)|m'',s\right)$$

for $s = 0, 1$. Because the system is RP, $\widehat{x}_s \geq \max\{x'_s, x''_s\}$.

For $s = 0, 1$, let

$$\begin{aligned}\widetilde{x}_s &\equiv \arg\max_{x\in\mathbb{R}} \; E^{\sigma'}\left(v\left(x,\theta\right)|m',s\right)\\ &= \arg\max_{x\in\mathbb{R}} \; \{\varkappa E^{\sigma}\left(v\left(x,\theta\right)|m',s\right) + (1-\varkappa) E^{\sigma}\left(v\left(x,\theta\right)|m'',s\right)\}\end{aligned}$$

for some $\varkappa \in [0,1]$. I claim that $\widetilde{x}_s \leq \max\{x'_s, x''_s\}$ which proves the statement.

Suppose not, so that $\widetilde{x}_s > \max\{x'_s, x''_s\}$ for some $s = 0, 1$. Because $E^{\sigma}\left[v\left(\cdot,\theta\right)|m,s\right]$ is strictly concave for any (m, s), it follows that

$$E^{\sigma}\left[v\left(\widetilde{x}_s,\theta\right)|m',s\right] < E^{\sigma}\left[v\left(\max\{x'_s, x''_s\},\theta\right)|m',s\right]$$

and that

$$E^{\sigma}\left[v\left(\widetilde{x}_s,\theta\right)|m'',s\right] < E^{\sigma}\left[v\left(\max\{x'_s, x''_s\},\theta\right)|m'',s\right]$$

which is a contradiction to $\widetilde{x}_s$ being sequentially optimal under profile σ', after message m' and signal s. ■

Step 1: *By Bester and Strausz (2001), there is an optimal RPIC system where $M = \Theta$. Without loss of generality, in what follows I assume that $M = \Theta$.*

Step 2: *If system $((\Theta, d), \sigma)$ is an optimal RPIC, then, for any $m \in \Theta$ such that there is $\theta \in \Theta$ where $\sigma(\theta)(m) > 0$, $d_1(m) \geq d_0(m)$.*

Proof. Suppose not, so that there is an optimal RPIC system $((\Theta, d), \sigma)$ such that

$$\widetilde{M} \equiv \{m \in \Theta : d_1(m) < d_0(m) \text{ and } \sigma(\theta)(m) > 0 \text{ for some } \theta \in \Theta\}$$

is non-empty. The proof shows that there is an alternative RPIC system $((\Theta, d'), \sigma)$ that the principal strictly prefers to $((\Theta, d), \sigma)$.

Description of d' :

Let

$$\theta' \equiv \begin{cases} \theta_N \text{ if } \widetilde{M} = \Theta \\ \min\left\{\theta \in \Theta : \sigma(\theta, m) > 0 \text{ for some } m \notin \widetilde{M}\right\} \text{ if } \widetilde{M} \subset \Theta \end{cases}$$

and

$$\theta'' \equiv \max\left\{\theta \in \Theta : \sigma(\theta, m) > 0 \text{ for some } m \in \widetilde{M}\right\}$$

Notice that $\theta' \geq \theta''$ and that, if $\widetilde{M} = \Theta$, then $\theta' = \theta''$. Likewise,

$$m' \in \begin{cases} \arg\max_{m \in \widetilde{M}} \ d_1(m) \text{ if } \widetilde{M} = \Theta \\ \arg\min_{m \in \Theta \setminus \widetilde{M}} \ d_1(m) \text{ if } \widetilde{M} \subset \Theta \end{cases}$$

and

$$m'' \in \arg\max_{m \in \widetilde{M}} \ d_1(m)$$

Finally, let z' (z'') denote the certainty equivalent of the agent when his type is $\theta = \theta'$ (θ'') :

$$u(z') = \pi(\theta') u(d_1(m')) + (1 - \pi(\theta')) u(d_0(m'))$$

and

$$u(z'') = \pi(\theta'') u(d_1(m'')) + (1 - \pi(\theta'')) u(d_0(m''))$$

For all $m \in \Theta$ and $s = 0, 1$,

$$d'_s(m) = \begin{cases} d_s(m) \text{ if } m \notin \widetilde{M} \\ z \text{ if } m \in \widetilde{M} \end{cases}$$

where $z = \min\{z', z''\}$.

System $((\Theta, d'), \sigma)$ is RPIC:

I start by showing that system $((\Theta, d'), \sigma)$ is IC. If $\widetilde{M} = \Theta$, then the statement follows trivially. Suppose, instead, that $\widetilde{M} \subset \Theta$.

Assume first that $z = z' \leq z''$. In this case, type $\theta = \theta'$ is indifferent between m' and m'' so system $((\Theta, d'), \sigma)$ is IC because $d_1(m') \geq z'$. If, on the contrary, $z = z'' < z'$, then type $\theta = \theta'$ strictly prefers m' to m''. Given that $d_1(m') \geq z'$, it follows that all types $\theta \geq \theta'$ do not strictly prefer to report m''. It also follows that type $\theta = \theta''$ has the same utility under system $((\Theta, d'), \sigma)$ that he did under system $((\Theta, d), \sigma)$. As a result, and because system $((\Theta, d), \sigma)$ is IC, he does not want to deviate to any

$m \notin \widetilde{M}$. Finally, it follows that all types $\theta \leq \theta''$ also do not strictly prefer to report $m \notin \widetilde{M}$ because $d_1(m') \geq z''$, so the system $((\Theta, d'), \sigma)$ is IC.

Given that $((\Theta, d), \sigma)$ is RP, it follows that, for $s = 0, 1$ and for $m \in \widetilde{M}$,

$$\begin{aligned}
\arg\max_{x\in\mathbb{R}} \; E^{\sigma}\left(v(x,\theta)\,|m, s\right) &\leq \arg\max_{x\in\mathbb{R}} \; E^{\sigma}\left(v(x,\theta)\,|m'', s\right) \\
&\leq \arg\max_{x\in\mathbb{R}} \; E^{\sigma}\left(v(x,\theta)\,|m'', s=1\right) \\
&\leq d_1(m'') \\
&< z
\end{aligned}$$

Therefore, it follows that system $((\Theta, d'), \sigma)$ is RP.

The principal strictly prefers $((\Theta, d'), \sigma)$ **to** $((\Theta, d), \sigma)$:

I show that, for any $m \in \widetilde{M}$,

$$\sum_{n=1}^{N} p_n \sigma(\theta_n)(m)\left(\pi(\theta_n) v(d_1(m), \theta_n) + (1 - \pi(\theta_n)) v(d_0(m), \theta_n)\right) < \sum_{n=1}^{N} p_n \sigma(\theta_n)(m) v(z, \theta_n) \tag{2}$$

which proves the statement.

Take one such $m \in \widetilde{M}$ and let $\widehat{\theta} \in [\theta_1, \theta_N]$ be such that

$$\pi\left(\widehat{\theta}\right) = \frac{\sum_{n=1}^{N} p_n \sigma(\theta_n)(m)\,\pi(\theta_n)}{\sum_{n=1}^{N} p_n \sigma(\theta_n)(m)} \leq \pi(\theta'')$$

Likewise, let $\widehat{z} \in \mathbb{R}$ be such that

$$\begin{aligned}
u(\widehat{z}) &= \pi\left(\widehat{\theta}\right) u(d_1(m)) + \left(1 - \pi\left(\widehat{\theta}\right)\right) u(d_0(m)) \\
&\geq \pi\left(\widehat{\theta}\right) u(d_1(m'')) + \left(1 - \pi\left(\widehat{\theta}\right)\right) u(d_0(m'')) \\
&\geq u(z'') \\
&\geq u(z)
\end{aligned}$$

Notice that the LHS of (2) can be written as the sum of A and B, where

$$A = \sum_{n=1}^{N} p_n \sigma(\theta_n)(m)\left(\pi\left(\widehat{\theta}\right) v(d_1(m), \theta_n) + \left(1 - \pi\left(\widehat{\theta}\right)\right) v(d_0(m), \theta_n)\right)$$

and

$$B=\sum_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)\left(\pi\left(\theta_n\right)-\pi\left(\widehat{\theta}\right)\right)\left(v\left(d_1\left(m\right),\theta_n\right)-v\left(d_0\left(m\right),\theta_n\right)\right)$$

Let

$$\widehat{B}=\frac{B}{\sum\limits_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)}$$

and

$$h\left(\theta_n\right)\equiv v\left(d_1\left(m\right),\theta_n\right)-v\left(d_0\left(m\right),\theta_n\right)$$

Notice that h is non-increasing because v has non-decreasing differences. Therefore,

$$\widehat{B}\leq\left[\frac{\sum\limits_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)\left(\pi\left(\theta_n\right)-\pi\left(\widehat{\theta}\right)\right)}{\sum\limits_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)}\right]\left[\frac{\sum\limits_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)h\left(\theta_n\right)}{\sum\limits_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)}\right]=0$$

(see lemma 2.1. in See and Chen (2008)), and so $B\leq 0$.

Notice that

$$A<\sum_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)v\left(\pi\left(\widehat{\theta}\right)d_1\left(m\right)+\left(1-\pi\left(\widehat{\theta}\right)\right)d_0\left(m\right),\theta_n\right)$$

because $v\left(\cdot,\theta_n\right)$ is strictly concave for any $\theta\in\Theta$.

Furthermore, we have that

$$\arg\max_{x\in\mathbb{R}}\sum_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)v\left(x,\theta_n\right)<d_1\left(m\right)<z\leq\widehat{z}\leq\pi\left(\widehat{\theta}\right)d_1\left(m\right)+\left(1-\pi\left(\widehat{\theta}\right)\right)d_0\left(m\right)$$

where the last inequality follows because u is concave. This, together with the fact that $v\left(\cdot,\theta_n\right)$ is strictly concave for any $\theta\in\Theta$, implies that

$$\sum_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)v\left(\pi\left(\widehat{\theta}\right)d_1\left(m\right)+\left(1-\pi\left(\widehat{\theta}\right)\right)d_0\left(m\right),\theta_n\right)\leq\sum_{n=1}^{N}p_n\sigma\left(\theta_n\right)\left(m\right)v\left(z,\theta_n\right)$$

which implies (2). ■

Step 3 is divided into two parts:

Step 3a: *If system* $((\Theta, d), \sigma)$ *is IC and there is* $m_N \in \Theta$ *such that* i)

$$\sigma(\theta_N)(m_N) > 0$$

ii)

$$d_1(m_N) > d_1(m) \text{ for all } m \text{ such that } \sigma(\theta)(m) > 0 \text{ for some } \theta \in \Theta$$

and iii)

$$d_s(m_N) \geq \arg\max_{x \in \mathbb{R}} E^{\sigma}(v(x, \theta) | m_N, s) \text{ for } s = 0, 1$$

then system $((\Theta, d), \sigma)$ *is RP.*

Proof. Notice that

$$\arg\max_{x \in \mathbb{R}} E^{\sigma}(v(x, \theta) | m, s) \leq \arg\max_{x \in \mathbb{R}} E^{\sigma}(v(x, \theta) | m_N, s = 0) \leq d_0(m_N) \leq d_s(m)$$

for any $m \in \Theta$ and for $s = 0, 1$. ■

Step 3a is particularly useful in that it allows me to apply the revelation principle to non-top messages. The reason that the revelation principle does not hold in an environment with limited commitment is that beliefs matter. But, as I show in Step 3b, in this case, beliefs only matter after the top message.

Step 3b: *For any optimal RPIC system* $((\Theta, d), \sigma)$*, there is another RPIC system* $((\Theta, d'), \sigma')$ *that the principal is indifferent to, where*

i)

$$\sigma'(\theta_n)(m) = \begin{cases} 1 \text{ if } m = m_N \\ 0 \text{ if } m \neq m_N \end{cases} \text{ if } n > n^*$$

ii)

$$\sigma'(\theta_n)(m) = \begin{cases} \tau \text{ if } m = m_N \\ 1 - \tau \text{ if } m = m_n \\ 0 \text{ if } m \neq m_n, m_N \end{cases} \text{ if } n = n^*$$

iii)

$$\sigma'(\theta_n)(m) = \begin{cases} 1 \text{ if } m = m_n \\ 0 \text{ if } m \neq m_n \end{cases} \text{ if } n < n^*$$

for some $n^* = 1, ..., N$*,* $\tau \in [0, 1]$ *and* $m_N \in \Theta$.

Proof. Take any *optimal RPIC system* $((\Theta, d), \sigma)$ and, without loss of generality, assume that there is a unique "top" message m_N:

$$\sigma(\theta_N)(m_N) > 0$$

and

$$d_1(m_N) > d_1(m) \text{ for all } m \in \Theta$$

Let n^* be the index of the smallest type to send message m_N with a positive probability:

$$n^* = \min \{n : \sigma(\theta_n)(m_N) > 0\}$$

and let

$$\tau = \sigma(\theta_{n^*})(m_N)$$

Define d' as follows:

i)

$$d'(m_n) = d(m_N) \text{ for all } n > n^*$$

ii)

$$d'(m_n) = d(\widehat{m}_n) \text{ for all } n \leq n^*$$

where

$$\widehat{m}_n \in \arg \max_{m:\sigma(\theta_n)(m)>0} \pi(\theta_n) v(d_1(m), \theta_n) + (1 - \pi(\theta_n)) v(d_0(m), \theta_n)$$

Notice that in system $((\Theta, d'), \sigma')$ the agent has the same expected utility for any type $\theta \in \Theta$ than under system $((\Theta, d), \sigma)$. Furthermore, there are less distinct lotteries to choose from, so it follows that system $((\Theta, d'), \sigma')$ is IC. And by Step 3a) it is RP. Finally, the principal (weakly) prefers system $((\Theta, d'), \sigma')$ because, for all $\theta \in \Theta$,

$$\begin{aligned} & \sum_{m \in \Theta} \sigma(\theta)(m)(\pi(\theta) v(d_1(m), \theta) + (1 - \pi(\theta)) v(d_0(m), \theta)) \\ \leq & \sum_{m \in \Theta} \sigma'(\theta)(m)(\pi(\theta) v(d'_1(m), \theta) + (1 - \pi(\theta)) v(d'_0(m), \theta)) \end{aligned}$$

■

Step 3 implies that the problem of finding a strategy profile that is a part of an optimal RPIC system can be reduced to the simpler problem of finding $n^* = 1, ..., N$ and $\tau \in [0, 1]$. In particular, it follows that RPIC system $\left(\left(\Theta, \widehat{d}\right), \widehat{\sigma}\right)$ is an optimal

RPIC system provided that

$$\widehat{\sigma}(\theta_n)(m) = \begin{cases} 1 \text{ if } m = \theta_N \\ 0 \text{ if } m \neq \theta_N \end{cases} \quad \text{if } n > \widehat{n}^*$$

ii)

$$\widehat{\sigma}(\theta_n)(m) = \begin{cases} \widehat{\tau} \text{ if } m = \theta_N \\ 1 - \widehat{\tau} \text{ if } m = \theta_n \\ 0 \text{ if } m \neq \theta_n, \theta_N \end{cases} \quad \text{if } n = \widehat{n}^*$$

iii)

$$\widehat{\sigma}(\theta_n)(m) = \begin{cases} 1 \text{ if } m = \theta_n \\ 0 \text{ if } m \neq \theta_n \end{cases} \quad \text{if } n < \widehat{n}^*$$

and that $\left(\widehat{d}, \widehat{n}^*, \widehat{\tau}\right)$ solves the following program, labeled as Γ.

The principal chooses (d, n^*, τ) in order to maximize her expected utility subject to i) a monotonicity condition stating that $d_1(m)$ is increasing, ii) an "upper" incentive constraint, stating that the lowest type sending each message does not want to send the following one, iii) a "lower" incentive constraint, stating that the largest type sending each message does not want to send the preceding one, and iv) a renegotiation proof condition that applies only to the largest message $m = \theta_N$.

Formally,

$$\begin{aligned} \widehat{V}(d, n^*, \tau) \;=\; & \sum_{n=n^*+1}^{N} p_n \left(\pi(\theta_n) v(d_1(\theta_N), \theta_n) + (1 - \pi(\theta_n)) v(d_0(\theta_N), \theta_n)\right) + \\ & p_{n^*} \left[\begin{array}{c} \tau \left(\pi(\theta_{n^*}) v(d_1(\theta_N), \theta_{n^*}) + (1 - \pi(\theta_{n^*})) v(d_0(\theta_N), \theta_{n^*})\right) + \\ (1 - \tau)\left(\pi(\theta_{n^*}) v(d_1(\theta_{n^*}), \theta_{n^*}) + (1 - \pi(\theta_{n^*})) v(d_0(\theta_{n^*}), \theta_{n^*})\right) \end{array} \right] + \\ & \sum_{n=1}^{n^*-1} p_n \left(\pi(\theta_n) v(d_1(\theta_n), \theta_n) + (1 - \pi(\theta_n)) v(d_0(\theta_n), \theta_n)\right) \end{aligned}$$

Condition a) can be stated as

$$\begin{cases} d_1(\theta_n) = d_1(\theta_N) \text{ for all } n > n^* \\ d_1(\theta_n) \geq d_1(\theta_{n-1}) \text{ for all } n = 2, ..., n^* + 1 \end{cases}$$

Condition b) can be written as

$$\begin{aligned}
& \tau\left[\pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_N\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_N\right)\right)\right] \\
\geq \; & \tau\left[\pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_{n^*}\right)\right)\right]
\end{aligned}$$

and

$$\begin{aligned}
& (1-\tau)\left[\pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_{n^*}\right)\right)\right] \\
\geq \; & (1-\tau)\left[\pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_{n^*-1}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_{n^*-1}\right)\right)\right]
\end{aligned}$$

and, for all $n=2,...,n^*-1$,

$$\begin{aligned}
& \pi\left(\theta_n\right) u\left(d_1\left(\theta_n\right)\right)+\left(1-\pi\left(\theta_n\right)\right) u\left(d_0\left(\theta_n\right)\right) \\
\geq \; & \pi\left(\theta_n\right) u\left(d_1\left(\theta_{n-1}\right)\right)+\left(1-\pi\left(\theta_n\right)\right) u\left(d_0\left(\theta_{n-1}\right)\right)
\end{aligned}$$

while condition c) can be written as

$$\begin{aligned}
& (1-\tau)\left[\pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_{n^*}\right)\right)\right] \\
\geq \; & (1-\tau)\left[\pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_N\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_N\right)\right)\right]
\end{aligned}$$

and, for all $n=1,...,n^*-1$,

$$\begin{aligned}
& \pi\left(\theta_n\right) u\left(d_1\left(\theta_n\right)\right)+\left(1-\pi\left(\theta_n\right)\right) u\left(d_0\left(\theta_n\right)\right) \\
\geq \; & \pi\left(\theta_n\right) u\left(d_1\left(\theta_{n+1}\right)\right)+\left(1-\pi\left(\theta_n\right)\right) u\left(d_0\left(\theta_{n+1}\right)\right)
\end{aligned}$$

Finally, the RP condition d) can be stated as

$$d_1\left(\theta_N\right) \geq \arg\max_{d_1 \in \mathbb{R}}\left\{\sum_{n=n^*+1}^{N} p_n \pi\left(\theta_n\right) v\left(d_1, \theta_n\right)+\tau p_{n^*} \pi\left(\theta_{n^*}\right) v\left(d_1, \theta_{n^*}\right)\right\}$$

and

$$d_0\left(\theta_N\right) \geq \arg\max_{d_0 \in \mathbb{R}}\left\{\sum_{n=n^*+1}^{N} p_n\left(1-\pi\left(\theta_n\right)\right) v\left(d_0, \theta_n\right)+\tau p_{n^*}\left(1-\pi\left(\theta_{n^*}\right)\right) v\left(d_0, \theta_{n^*}\right)\right\}$$

Consider the relaxed problem Γ' that is equal to Γ except that b) is eliminated.

Step 4: *There is a solution* $\left(\widehat{d},\widehat{n}^*,\widehat{\tau}\right)$ *of the program* Γ' *such that*

$$\pi\left(\theta_{n^*}\right)u\left(\widehat{d}_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right)u\left(\widehat{d}_0\left(\theta_{n^*}\right)\right)=\pi\left(\theta_{n^*}\right)u\left(\widehat{d}_1\left(\theta_N\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right)u\left(\widehat{d}_0\left(\theta_N\right)\right)$$

and, for all $n=1,...,\widehat{n}^*-1$,

$$\pi\left(\theta_n\right)u\left(\widehat{d}_1\left(\theta_n\right)\right)+\left(1-\pi\left(\theta_n\right)\right)u\left(\widehat{d}_0\left(\theta_n\right)\right)=\pi\left(\theta_n\right)u\left(\widehat{d}_1\left(\theta_{n+1}\right)\right)+\left(1-\pi\left(\theta_n\right)\right)u\left(\widehat{d}_0\left(\theta_{n+1}\right)\right)$$

Proof. Take any solution of Γ' and denote it by $\left(\widehat{d},\widehat{n}^*,\widehat{\tau}\right)$. Suppose that c) holds strictly. If

$$\begin{aligned}&\left(1-\widehat{\tau}\right)\left[\pi\left(\theta_{n^*}\right)u\left(\widehat{d}_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right)u\left(\widehat{d}_0\left(\theta_{n^*}\right)\right)\right]\\ >\;&\left(1-\widehat{\tau}\right)\left[\pi\left(\theta_{n^*}\right)u\left(\widehat{d}_1\left(\theta_N\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right)u\left(\widehat{d}_0\left(\theta_N\right)\right)\right]\end{aligned}$$

(which implies that $\widehat{\tau}<1$), then there is mapping $d':\Theta\times\{0,1\}$ such that $d'=\widehat{d}$ except that $d_0'\left(\theta_{n^*}\right)$ is such that

$$\begin{aligned}&\left(1-\widehat{\tau}\right)\left[\pi\left(\theta_{n^*}\right)u\left(\widehat{d}_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right)u\left(d_0'\left(\theta_{n^*}\right)\right)\right]\\ =\;&\left(1-\widehat{\tau}\right)\left[\pi\left(\theta_{n^*}\right)u\left(\widehat{d}_1\left(\theta_N\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right)u\left(\widehat{d}_0\left(\theta_N\right)\right)\right]\end{aligned}$$

Given that

$$x^*\left(\theta_{n^*}\right)\leq\widehat{d}_0\left(\theta_N\right)\leq d_0'\left(\theta_{n^*}\right)<\widehat{d}_0\left(\theta_{n^*}\right)$$

it follows that the principal strictly prefers the alternative $\left(d',\widehat{n}^*,\widehat{\tau}\right)$ to $\left(\widehat{d},\widehat{n}^*,\widehat{\tau}\right)$ which contradicts the optimality of the latter.

If, for some $n=1,...,n^*-1$,

$$\pi\left(\theta_n\right)u\left(\widehat{d}_1\left(\theta_n\right)\right)+\left(1-\pi\left(\theta_n\right)\right)u\left(\widehat{d}_0\left(\theta_n\right)\right)>\pi\left(\theta_n\right)u\left(\widehat{d}_1\left(\theta_{n+1}\right)\right)+\left(1-\pi\left(\theta_n\right)\right)u\left(\widehat{d}_0\left(\theta_{n+1}\right)\right)$$

then there is mapping $d':\Theta\times\{0,1\}$ such that $d'=\widehat{d}$ except that $d_0'\left(\theta_n\right)$ is such that

$$\pi\left(\theta_n\right)u\left(\widehat{d}_1\left(\theta_n\right)\right)+\left(1-\pi\left(\theta_n\right)\right)u\left(d_0'\left(\theta_n\right)\right)=\pi\left(\theta_n\right)u\left(\widehat{d}_1\left(\theta_{n+1}\right)\right)+\left(1-\pi\left(\theta_n\right)\right)u\left(\widehat{d}_0\left(\theta_{n+1}\right)\right)$$

Given that

$$x^*\left(\theta_n\right)\leq\widehat{d}_0\left(\theta_{n+1}\right)\leq d'\left(\theta_n\right)<\widehat{d}_0\left(\theta_n\right)$$

it follows that the principal strictly prefers the alternative $\left(d',\widehat{n}^*,\widehat{\tau}\right)$ to $\left(\widehat{d},\widehat{n}^*,\widehat{\tau}\right)$ which contradicts the optimality of the latter.

Thus, one concludes that c) must bind in any solution of Γ'. Finally, if the optimal $\widehat{\tau} = 1$, then it is a solution to choose $\widehat{d}_s(\theta_{n^*}) = \widehat{d}_s(\theta_N)$ for $s = 0, 1$ (among others). ■

Step 5: *In any solution* $\left(\widehat{d}, \widehat{n}^*, \widehat{\tau}\right)$ *of the program* Γ' *such that* $\widehat{d}(\theta_N) \neq \widehat{d}(\theta_{\widehat{n}^*})$, it must be that

$$\widehat{d}_1(\theta_N) = \arg\max_{d_1 \in \mathbb{R}} \left\{ \sum_{n=n^*+1}^{N} p_n \pi(\theta_n) v(d_1, \theta_n) + \widehat{\tau} p_{n^*} \pi(\theta_{n^*}) v(d_1, \theta_{n^*}) \right\}$$

and

$$\widehat{d}_0(\theta_N) = \arg\max_{d_0 \in \mathbb{R}} \left\{ \sum_{n=n^*+1}^{N} p_n (1 - \pi(\theta_n)) v(d_0, \theta_n) + \widehat{\tau} p_{n^*} (1 - \pi(\theta_{n^*})) v(d_0, \theta_{n^*}) \right\}$$

Proof. Suppose not. Consider first the case where

$$\widehat{d}_0(\theta_N) > \arg\max_{d_0 \in \mathbb{R}} \left\{ \sum_{n=n^*+1}^{N} p_n (1 - \pi(\theta_n)) v(d_0, \theta_n) + \widehat{\tau} p_{n^*} (1 - \pi(\theta_{n^*})) v(d_0, \theta_{n^*}) \right\} \equiv \widehat{x}_0$$

Consider the alternative mechanism d' where d' is identical to $\widehat{d}$ except that

$$d'_0(\theta_N) = \widehat{x}_0$$

The new mechanism satisfies c), because reporting θ_N is less appealing with d' then with d; and satisfies a) and d) by definition. The fact that mechanism d' is preferred by the principal is a contradiction to x being optimal.

Suppose instead that

$$\widehat{d}_1(\theta_N) > \arg\max_{d_1 \in \mathbb{R}} \left\{ \sum_{n=n^*+1}^{N} p_n \pi(\theta_n) v(d_1, \theta_n) + \widehat{\tau} p_{n^*} \pi(\theta_{n^*}) v(d_1, \theta_{n^*}) \right\} \equiv \widehat{x}_1$$

Consider the alternative mechanism d' where d' is identical to $\widehat{d}$ except that

$$d'_1(\theta_N) = \max\left\{\widehat{x}_1, \widehat{d}_1(\theta_{n^*})\right\}$$

The new mechanism satisfies c), because reporting θ_N is less appealing with d' then with d; and satisfies a) and d) by definition. Mechanism d' is strictly preferred by the principal due to the strict concavity of v and the fact that $\widehat{x}_1 \leq d'_1(m_N) < \widehat{d}_1(m_N)$, which is a contradiction to $\widehat{d}$ being optimal. ■

8.4 Proof of Proposition 7

Proof. Condition ii) in the statement of proposition 6 implies that there is a strictly increasing mapping from the posterior belief that $\theta = \theta_2$, denoted by $\Pr^{\widetilde{\sigma}}(\theta = \theta_2 | m = \theta_2, s)$, and $\widetilde{d}_s(\theta_2)$. Notice that

$$\overset{\widetilde{\sigma}}{\Pr}(\theta = \theta_2 | m = \theta_2, s) = \frac{p_2 f(s|\theta_2)}{p_2 f(s|\theta_2) + p_1 \tau(\widetilde{\sigma}) f(s|\theta_1)}$$

for some $\tau(\widetilde{\sigma}) \in [0,1]$. Given that $\frac{f(\cdot|\theta_2)}{f(\cdot|\theta_1)}$ is strictly increasing, then $\Pr^{\widetilde{\sigma}}(\theta = \theta_2 | m = \theta_2, s)$ is also strictly increasing, which proves a).

As for b), suppose that $\widetilde{d}_s(\theta_1)$ is not constant with s and consider mechanism d', where $d'(\theta_2) = \widetilde{d}(\theta_2)$ but

$$u(d'_s(\theta_1)) = \sum_{s' \in S} f(s'|\theta_1) u\left(\widetilde{d}_{s'}(\theta_1)\right) \text{ for all } s \in S$$

System $((\Theta, d'), \widetilde{\sigma})$ satisfies i) and ii) and is strictly preferred by the principal, because $v(\cdot, \theta)$ is strictly concave for all $\theta \in \Theta$, u is concave and

$$x^*(\theta_1) \leq \min_{s \in S}\left\{\widetilde{d}_s(\theta_2)\right\} < d'_s(\theta_1)$$

which would be a contradiction. ■

References

[1] Battaglini, M. (2007). Optimality and renegotiation in dynamic contracting. *Games and economic behavior*, 60(2), 213-246.

[2] Ben-Porath, E., Dekel, E., & Lipman, B. L. (2014). Optimal allocation with costly verification. *American Economic Review*, 104(12), 3779-3813.

[3] Beshkar, M. (2016). Arbitration and renegotiation in trade agreements. *Journal of Law, Economics, and Organization*, 32 (3), 586-619.

[4] Bester, H., & Strausz, R. (2001). Contracting with imperfect commitment and the revelation principle: the single agent case. *Econometrica*, 69(4), 1077-1098.

[5] Chen, Y. (2011). Perturbed communication games with honest senders and naive receivers. *Journal of Economic Theory*, 146(2), 401-424.

[6] Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, 1431-1451.

[7] Evans, R., & Reiche, S. (2015). Contract design and non-cooperative renegotiation. *Journal of Economic Theory*, 157, 1159-1187.

[8] Forges, F. (1994). Posterior efficiency. *Games and Economic Behavior*, 6(2), 238-261.

[9] Green, J. R., & Laffont, J. J. (1987). Posterior implementability in a two-person decision problem. *Econometrica*, 69-94.

[10] Goltsman, M. (2011). Optimal information transmission in a holdup problem. *The RAND Journal of Economics*, 42(3), 495-526.

[11] Hart, O. D., & Tirole, J. (1988). Contract renegotiation and Coasian dynamics. *The Review of Economic Studies*, 55(4), 509-540.

[12] Ishida, J., & Shimizu, T. (2016). Cheap talk with an informed receiver. *Economic Theory Bulletin*, 4(1), 61-72.

[13] Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies*, 76(4), 1359-1395.

[14] Laffont, J. J., & Tirole, J. (1990). Adverse selection and renegotiation in procurement. *The Review of Economic Studies*, 57(4), 597-625.

[15] Lai, E. K. (2014). Expert advice for amateurs. *Journal of Economic Behavior & Organization*, 103, 1-16.

[16] Maestri, L. (2017). Dynamic contracting under adverse selection and renegotiation. *Journal of Economic Theory*, 171, 136-173.

[17] Maskin, E., & Moore, J. (1999). Implementation and renegotiation. *Review of Economic Studies*, 39-56.

[18] Mylovanov, T. & Zapechelnyuk (2017): Optimal allocation with ex-post verification and limited penalties. *American Economic Review,* forthcoming.

[19] Neeman, Z., & Pavlov, G. (2013). Ex post renegotiation-proof mechanism design. *Journal of Economic Theory*, 148(2), 473-501.

[20] Ray, D. (2007). A game-theoretic perspective on coalition formation. *Oxford University Press.*

[21] See, C. T., & Chen, J. (2008). Inequalities on the variances of convex functions of random variables. *J. Inequal. Pure and Appl. Math*, 9(3), 1-5.

[22] Siegel, R. & Strulovici, B. (2016): Improving criminal trials by reflecting residual doubt: multiple verdicts and plea bargains. *Working paper.*

[23] Silva, F. (2017). If we confess our sins. *Working paper.*

[24] Strulovici, B. (2017): Contract Negotiation and the Coase Conjecture. *Econometrica*, 585-616.