Renegotiation proof mechanism design with imperfect type verification*

Francisco Silva[†]

January 24, 2019

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. The main finding of this paper is that the optimal renegotiation proof mechanism induces pooling at the top, i.e., types above a certain threshold report to be the largest type, while types below the threshold report truthfully.

JEL classification: D8

Keywords: renegotiation proof, mechanism design, verification

*I would like to thank Martin Besfamille, Mustafa Dogan, Selman Erol, Nicolas Figueroa, Ju Hu, Alexandre Janiak, David Kohn, Gaston Llanes, Samir Mamadehussene, Juan Pablo Montero, Mallesh Pai, Vincent Van Kervel, the seminar participants at PUC Chile and at the Econometric Society Winter meeting (2017) and the anonymous referees 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

Normally, in mechanism design, the principal (she) uses the fact that the utility function of the agent (he) depends on his private type in order to be able to separate between them; in general, when the principal proposes a menu of contracts, some of those contracts are chosen by some types of the agent, while other contracts are chosen by other types. However, in many environments, the agent's utility function is independent of his private type (I give several examples below). In these scenarios, proposing a menu of contracts is essentially the same as imposing a single contract, seeing as the agent would always pick the same contract, independently of his type. Thus, there would be no gain in interacting with the agent.

The principal might be able resolve this issue if she has access to some exogenous source of information about the agent's type, because, in that way, she might use that information to "validate" the agent's report.[1] As an example, consider a defendant, who is privately informed about whether he is innocent or guilty, and who interacts with some representative agent of the criminal justice system (a prosecutor or a judge for example). Imagine that it is known that the prosecutor/judge has no other source of information apart from the agent. In that case, the agent will inevitably receive the same treatment regardless of his guilt; if asked about his guilt, the agent would surely claim to be innocent even when he is not. But now suppose that the prosecutor/judge has access to forensic evidence, which, by its nature, is more likely to be incriminating when the agent is guilty than when he is innocent. In addition, imagine that the agent receives the following offer: if the agent confesses to being guilty, he receives some punishment x_c; if he refuses, his punishment will be $\overline{x} > x_c$ if the evidence is incriminatory; and $\underline{x} < x_c$ otherwise. Confronted with this, the agent might react differently depending on his type. In particular, if he knows he is guilty, he is more afraid that the evidence will be incriminating than when he is innocent. Thus, he will be more willing to confess when he is guilty; while the ex-post utility of the agent is type independent, his expected utility is not, which is what enables the principal to gain something from interacting with the agent.

In this paper, I study the optimal mechanisms for a risk averse principal in a setting that has three main properties. First, the agent's private type $\theta \in \mathbb{R}$ does not affect his utility. In particular, the agent simply wants to maximize his reward x: the

[1] In the related literature section, I discuss the literature on hard evidence which also gets around this issue but by assuming that certain types of the agent are able to send verifiable messages.

agent's utility function is $u(x)$, where u is strictly increasing. While the agent's type does not affect the agent's utility, it affects the principal's utility, which is given by $v(x, \theta)$. A good example for a payoff function v that fits the properties of the model is $v(x, \theta) = -(x - \theta)^2$ so that the principal wants to match the reward x of the agent with his type θ. The second property of the model is that the principal has (costless) access to a binary signal $s \in \{0, 1\}$, correlated with the agent's type, which she can use to (imperfectly) verify the agent's claims. Finally, the third property is that the types of the agent that the principal wants to reward the most (which, without loss of generality, are the larger ones) are also the ones that are more likely to generate $s = 1$, i.e., $\Pr\{s = 1|\theta\}$ is increasing with θ.

The interaction between the players is as follows: i) the principal proposes a mechanism, i.e., a menu of contracts, where each contract specifies a reward $x \in \mathbb{R}$ contingent on signal s; ii) the agent privately observes his type and decides on one of the offered contracts; iii) signal s is realized; iv) the agent receives the corresponding reward. In what follows, I describe a few examples for context (in addition to the above example of the defendant).

Example 1: The agent is a local government and his type $\theta \in \mathbb{R}$ represents the quality of the public health services of the region, so that the larger θ is, the poorer the quality is. The principal is the central government which must decide how much funding $x \in \mathbb{R}$ for health services to provide to the local government. The local government wants to maximize the funding it gets, while the central government wants to match that funding with the region's needs. Even though only the local government knows θ, the central government is able to obtain some exogenous signal s about θ by ordering an independent study about the quality of the region's public health services. In the independent study, the public health services are given one of two ratings, namely, good ($s = 0$) and bad ($s = 1$). The probability that there is a good rating is increasing with the actual quality of the public health services of the region. A contract is a pair $(x_0, x_1) \in \mathbb{R}^2$ which specifies the funding provided to the local government in the event that the rating is good (x_0) or bad (x_1).

Example 2: The agent is a project manager who is looking for investment in his project. The project takes two years to be completed and only pays off at the end of the two years. The principal is an investor who must decide how much money $x \in \mathbb{R}$ to invest in the project. The goal of the project manager is to maximize the amount of investment he receives, while the principal's willingness to invest is increasing with the quality of the project $\theta \in \mathbb{R}$. While the agent is the only one who knows the quality of the project θ, at the end of the first year, the investor will be able to inspect the progress of the project and, as a result, obtain a signal correlated with θ, which can be either good ($s = 1$) or bad ($s = 0$). The probability that it is good is increasing with the quality of the project θ. A contract is a pair $(x_0, x_1) \in \mathbb{R}^2$, which can be interpreted as representing an upfront investment of x_0 plus an extra $(x_1 - x_0)$ in the event that the signal after the first year is good (x_1).

Example 3: The agent is one of the victims of a hurricane and is looking to maximize the financial compensation $x \in \mathbb{R}$ he receives from the government to perform repairs. The government wants to pay the victim proportionally to his damages, denoted by θ, which are privately known by the victim. After the agent has made a claim of damages, with some probability he may be audited to check the real extent of the damages ($s = 1$ if the damages are considered severe and $s = 0$ if they are considered light). A contract is a pair $(x_0, x_1) \in \mathbb{R}^2$, where x_1 can be interpreted as representing the upfront payment to the victim, while $x_1 - x_0$ can be interpreted as the fine the victim will have to pay if he is audited by the government and is found to have only suffered light damages.

While I start by characterizing the optimal incentive compatible (IC) mechanism, the main contribution of the paper is to define renegotiation proof (RP) mechanisms and to characterize the optimal renegotiation proof incentive compatible (RPIC) mechanism.

The optimal IC mechanism is such that the principal offers a menu of contracts that have different gaps between the reward that the agent receives if the signal is good and if the signal is bad. This is because the type dependence of the agent's expected utility only comes from the fact that larger types are more confident than lower types that the signal will be good. Therefore, in the optimal IC mechanism, larger types will pick contracts with larger gaps.[2] If there are two types, as in the example of the defendant who might be innocent or guilty, the optimal IC mechanism looks similar to the mechanism I describe above, provided x_c, $\overline{x}$ and $\underline{x}$ are chosen properly; in one contract, there is no gap, so that the reward that is offered is independent of the public signal (in the example, this reward is $-x_c$), while, in the other contract, the reward is

[2]Throughout the paper, I assume that the principal has no other instrument to "incentivize" the agent, because the allocation is only made of a single good, the reward of the agent. However, in mechanism design, it is often the case that the allocation is made of two goods, with one of them entering the agent's utility function linearly (the transfers). In that case, this second good could be used by the principal to provide incentives to the agent without having to distort the first best allocation of the first good.

larger than the constant reward if the signal is "good" ($-\underline{x}$) but smaller if the signal is bad ($-\overline{x}$). Thus, in the context of this example, the optimal IC mechanism looks similar to plea bargaining: by choosing the constant contract, it is as if the agent pleads guilty, while taking the risky contract means taking the risk of going to trial, which can end up with the agent being acquitted or receiving an even larger punishment than when pleading guilty.

The commitment problem that emerges comes from the use of the revelation principle. In the optimal IC mechanism, the agent reports truthfully, i.e., he chooses a different contract for each type. This means that the choice of the contract immediately reveals the type of the agent. In particular, when the agent reveals his type to be the one the principal wants to reward the most, the principal, aware that the agent's type is the one she wants to reward the most, would prefer to choose a larger reward than what had been contractualized. This is especially troubling, because the agent would certainly not oppose such a change, so that the contract would get renegotiated. But, of course, the fact that the agent anticipates this renegotiation destroys any incentives to report truthfully to begin with.

These concerns raise the issue of the extent to which the principal requires commitment power to be able to benefit from the interaction with the agent. If the principal is not able to commit not to renegotiate with the agent, is she still able to do better than by simply imposing a contract? What does the optimal RPIC mechanism look like? Does it still have the same structure as the optimal IC mechanism?

A renegotiation proof mechanism is defined to be such that, for any contract, the principal does not strictly prefer to increase the agreed upon reward once the signal is realized. Using the timing from above, this means that, after signal s has been realized in stage iii), the principal does not strictly prefer to increase reward x_s of the contract the agent has chosen in stage ii). The argument is that, if the principal did want to increase that reward, the agent would not object, so that the original contract would not be implemented.

Characterizing the optimal RPIC mechanism is not as straightforward as characterizing the optimal IC mechanism, because of the principal's limited commitment power. In particular, the revelation principle does not hold. Nevertheless, I find that it is indeed optimal for the principal to offer a menu of different contracts to the agent even when she is unable to commit not to renegotiate, and that the structure of the optimal RPIC mechanism is similar to the optimal IC mechanism: the different contracts that are proposed have different gaps between the reward if the signal is good and if the signal is bad, which induces better types to choose contracts with larger gaps.

However, there is a fundamental difference between the two mechanisms; while the optimal IC mechanism induces the agent to effectively reveal his type by choosing a contract per type, the optimal RPIC mechanism induces pooling at the top: there is a threshold type such that the agent only picks a contract per type if his type is below the threshold, while if it is above the threshold, the agent picks the contract with the largest gap. This implies that after observing that the agent has chosen the contract with the largest gap, the principal will be uncertain of the agent's type. This pooling at the top result contrasts with some of the literature on renegotiation proof mechanism design, in particular with Strulovici (2017), who finds that, when the principal is unable to commit not to renegotiate, the optimal mechanism induces complete type revelation (in his framework with only two types, the high type picks one contract while the low type picks another contract). I discuss this literature in more detail in the related literature section.

The remainder of this paper is structured as follows. In section 2, I present the model. In section 3, I characterize the optimal IC mechanism. In section 4, I formally define RP mechanisms, characterize the optimal RPIC mechanism and discuss its main properties. In section 5, I discuss the related literature in more detail. In appendix A, I extend the model to the case where the agent's type is a continuous random variable. All proofs that are not in the text are in appendix B.

2 Model

2.1 Assumptions

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 $p(\theta_n) > 0$. 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 (so that x can be interpreted as the reward of the agent) 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 nondecreasing differences, i.e., for any $(x',x) \in \mathbb{R}^2$ such that $x' \geq x$, $\{v(x',\cdot) - v(x,\cdot)\}$ is nondecreasing, which implies that x^* is nondecreasing. Finally, I assume that $x^*(\theta_1) < x^*(\theta_N)$ so that $x^*(\cdot)$ is not constant. As mentioned in the introduction, an example is $v(x,\theta) = -(x-\theta)^2$.

2.2 Definitions

A mechanism is a message set M and a function $d : M \times \{0,1\} \to \mathbb{R}$, which maps the message $m \in M$ (also referred to as contract) sent/chosen by the agent and the signal s to a reward $d_s(m) \in \mathbb{R}$.[3] Given a 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.

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, 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. Thus, 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.

Finally, the expected utility of the principal under some system $((M,d),\sigma)$ is given by

$$V^M(d,\sigma) \equiv \sum_{\theta\in\Theta}\sum_{m\in M} p(\theta)\,\sigma(\theta)(m)\left(\pi(\theta)\,v(d_1(m),\theta) + (1-\pi(\theta))\,v(d_0(m),\theta)\right)$$

2.3 Preliminary result

Before proceeding to analyzing optimal mechanisms, I start by deriving a property of all IC systems: that the agent's strategy is "monotone".

[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 $\theta \in \Theta$.

Lemma 1 *For any mechanism* (M,d)*, 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 that 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 $\pi(\cdot)/(1-\pi(\cdot))$ is strictly increasing, the statement follows. ■

Lemma 1 implies that there is a certain monotonicity in how the agent reports as a function of his type in an IC system. As an example, imagine that, in some IC system $((M,d),\sigma)$, there are four messages (m_1,m_2,m_3,m_4), each of them sent with positive probability by some type of the agent and with the property that

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

If the system is IC, it has to be that

$$d_0(m_4) < d_0(m_3) < d_0(m_2) < d_0(m_1),$$

because, otherwise, one of the messages would be "dominated" by another, which would imply that it would not be sent by any type with positive probability. Imagine that some type θ finds it optimal to send message m_2. Lemma 1 implies that, for the system $((M,d),\sigma)$ to be IC, it must be that all types larger than θ do not report m_1, while all types smaller than θ do not report m_3 or m_4. Put differently, lemma 1 essentially implies that the larger the agent's type is, the larger is the $d_1(m)$ of the message(s) he sends in an IC system and the lower is the $d_0(m)$.

3 Optimal IC system

A system $((M, d), \sigma)$ is an optimal IC system if it is IC and if there is no other IC system for which the principal's expected utility is larger than $V^M(d, \sigma)$. 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., with probability 1, type θ reports $m = \theta$, for all $\theta \in \Theta$. This means that, without loss of generality, one can think of the optimal IC system as being such that $M = \Theta$ and $\sigma = \sigma^*$, where

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

Let d^* be such that $((\Theta, d^*), \sigma^*)$ is an optimal IC system. The following proposition characterizes d^*.

Proposition 2 *Mapping d^* is such that i) $d_1^*(\theta)$ is (weakly) increasing with θ and $d_0^*(\theta)$ is (weakly) decreasing with θ, ii) for all $n = 1, ..., N-1$, $E(u(d_s^*(\theta_n))|\theta_n) = E(u(d_s^*(\theta_{n+1}))|\theta_n)$, iii) $d_1^*(\theta) \geq d_0^*(\theta)$ for all $\theta \in \Theta$, iv) $d_1^*(\theta_N) > d_1^*(\theta_1) = d_0^*(\theta_1) > d_0^*(\theta_N)$ and v) $d_1^*(\theta_N) \leq x^*(\theta_N)$ and $d_0^*(\theta_N) < x^*(\theta_N)$.*

Proof. i) follows directly from lemma 1 and from the fact that $((\Theta, d^*), \sigma^*)$ is IC. ii) through v) are left for appendix B. ■

Properties i) and iii) taken together imply that larger types pick contracts with larger gaps, i.e., larger differences between the reward should the signal be good ($s = 1$) and bad ($s = 0$). In particular, and using property iv), the contract chosen by the lowest type is flat, and then the difference between d_1 and d_0 increases with the agent's type. This is because what separates the different types of the agent is how confident each one is in that the signal will be good: larger types are more confident than lower types. Thus, larger types are more willing to take the risk of selecting more variable contracts.

As to why the lowest type's contract is flat, one first has to realize that the incentive constraints only bind upwards; i.e., each type does not want to mimic the next largest type (the fact that these constraints bind leads to property ii)). Therefore, reducing the gap of the contract chosen by the lowest type not only increases the incentives for the lowest type to choose it (because the agent is risk averse) but is also directly beneficial for the principal (because she is also risk averse).

In the example of the defendant from the introduction, where there are only two types, the innocent and the guilty, this mechanism looks similar to plea bargaining. The agent has a choice between a risky contract, where he may be either acquitted or receive a large punishment in the event that incriminating evidence is found, and confessing to being guilty and receiving a reduced punishment independent of evidence. In the example of the hurricane victim, the optimal mechanism has a similar interpretation. Suppose again that there are only two types, one who has suffered severe damages and one who has suffered light damages. What is optimal for the principal is to allow the agent to self-report his damages. If he admits that his damages were light, he receives a small compensation package. If he claims that his damages were severe, he is initially granted a larger compensation package but may be audited later, with some probability, to check whether the damages were indeed severe. In the event that the damages are found to be light, the compensation would be taken back. Figure 1 shows a graphical representation of the optimal mechanism when $N = 2$.

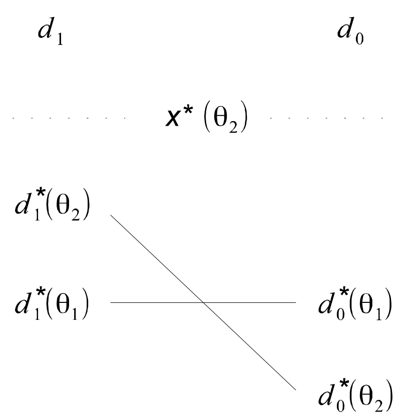

Figure 1: Representation of the optimal IC mechanism d^* when $N = 2$.

An example that fits the model when there are more types is the one of the investor and the project manager. Because the optimal IC mechanism is such that $d_1^*(\theta) \geq d_0^*(\theta)$ for each contract θ, one can interpret each contract offered as an upfront payment of $d_0^*(\theta)$ and then an increase on the investment of $d_1^*(\theta) - d_0^*(\theta) \geq 0$, which is contingent on some positive evaluation of the project. Better projects are more back loaded, with less investment upfront, precisely because their project managers are more willing to take those contracts as they are more confident that they will be evaluated positively later on.

Returning to figure 1, one can see that the rewards paid in the top contract are lower than the principal's preferred reward. Indeed, that is property v) - recall that the principal's preferred reward when $\theta = \theta_N$ is equal to $x^*(\theta_N)$. The argument here is related to the fact that, when the principal chooses $d_1^*(\theta_N)$ and $d_0^*(\theta_N)$, the only constraint that is relevant is that type θ_{N-1} does not mimic type θ_N. This implies that it does not make sense for the principal to choose $d_s^*(\theta_N)$ to be larger than $x^*(\theta_N)$ for any $s = 0, 1$, because lowering that reward would not only improve the principal's expected utility (because it would bring the reward closer to $x^*(\theta_N)$) but would also reduce type θ_{N-1}'s incentives to mimic type θ_N. It also follows that $d_0^*(\theta_N) < x^*(\theta_N)$ because property iv) already establishes that $d_1^*(\theta_N) > d_0^*(\theta_N)$.

Property v) is what motivates the study of renegotiation proof mechanisms. It states that in the optimal IC system $((\Theta, d^*), \sigma^*)$, after receiving message $\theta = \theta_N$ and observing (at least) signal $s = 0$, the principal, after inferring the type of the agent, would prefer to increase his reward from $d_0^*(\theta_N)$ to $x^*(\theta_N)$. Seeing as the agent would not oppose this change, one has to imagine that the two players would renegotiate at that point and would not implement $d_0^*(\theta_N)$. Anticipating this, lower types would be tempted to report to be type θ_N, as they would be confident that the lower reward they were supposed to get, should $s = 0$ be realized, would get renegotiated away.[4] In the next section, I discuss how to construct a mechanism where these renegotiation opportunities do not exist and then characterize the optimal one.

4 Renegotiation proof mechanism design

4.1 Definition of RP systems

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, denoted by $\gamma_s^\sigma(m)$:

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

[4] In the example of the defendant, this means that when the defendant is innocent, he picks a contract that gives him a punishment that the principal would like to reduce at least when $s = 0$. That is what makes the contract renegotiable.

Definition 3 *A system* $((M,d),\sigma)$ *is renegotiation proof (RP) if, for all* $m \in M$ *and* $s \in \{0,1\}$,

$$d_s(m) \geq \gamma_s^\sigma(m)$$

For a system to be RP, it must be that, after any message m is sent and any 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. The argument is that if there was such an alternative x, there would be nothing to prevent the players from switching to it. Whenever $d_s(m) < \gamma_s^\sigma(m)$, the alternative that makes both players better off is $x = \gamma_s^\sigma(m)$, i.e., both players would prefer to increase the reward of the agent. However, if $d_s(m) \geq \gamma_s^\sigma(m)$, while the agent would like to increase his reward, the principal would not, so that they would not agree to renegotiate.[5]

4.2 Characterization of the optimal RPIC system

The goal of this section is to find the optimal renegotiation proof incentive compatible (RPIC) system. The challenge of analyzing RPIC systems is that beliefs matter: the posterior belief that the principal forms after observing the agent's report and the signal determines whether she is willing to renegotiate. As a result, the revelation principle does not follow. However, Bester and Strausz (2001) show that a version of the revelation principle does hold: while one can no longer exclusively focus on the truthful reporting strategy of the agent, one can, without loss of generality, assume that the message space has the same number of elements as the type space ($M = \Theta$), i.e., in the optimal RPIC system, the number of contracts available to the agent does not exceed the number of types.[6]

Let Φ be the set of "pooling at the top" strategies, i.e., Φ is the set of all strategies σ for which there is $n^*(\sigma) \in \{1,...,N-1\}$, $\tau(\sigma) \in [0,1]$ and $\theta_{Top}(\sigma) \in \left\{\theta_{n^*(\sigma)+1},..,\theta_N\right\}$

[5] Another way to think of the optimal RPIC system is as follows. Consider the following game: i) the principal proposes mechanism (M,d), ii) the agent chooses some message $m \in M$, iii) signal $s \in \{0,1\}$ is realized, iv) if she wants, the principal proposes some alternative reward $x \in \mathbb{R}$, v) the agent chooses between rewards $d_s(m)$ and x. Notice that there is no loss of generality in focusing on perfect bayesian equilibria where the principal never proposes any alternative x. Therefore, one can interpret the optimal RPIC system as the corresponding perfect bayesian equilibrium outcome.

[6] 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.

such that

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

In words, if a strategy is a pooling at the top strategy, then there is a threshold type $n^* \geq 1$ and a top message $\theta_{Top} > \theta_{n^*}$ such that, if the agent's type is larger than θ_{n^*}, the agent sends message θ_{Top}; if $\theta_n = \theta_{n^*}$, the agent randomizes between confessing to being type θ_{n^*} and sending message θ_{Top}; if the agent's type is smaller than θ_{n^*}, the agent confesses his type. There is pooling at the top in the sense that types above threshold θ_{n^*} pool on the same message θ_{Top}. In the next proposition, I show that there is an optimal RPIC system where there is pooling at the top.

Let D denote the set of all mappings $d : \Theta \times \{0, 1\} \to \mathbb{R}$.

Proposition 4 *System $\left(\left(\Theta, \widehat{d}\right), \widehat{\sigma}\right)$ is an optimal RPIC system if, of all pairs $(d, \sigma) \in D \times \Phi$, pair $\left(\widehat{d}, \widehat{\sigma}\right)$ maximizes $V^{\Theta}(d, \sigma)$ subject to the following constraints: i) $d_1(\cdot)$ is (weakly) increasing for all θ_n, 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) $d_s\left(\theta_{Top}\left(\sigma\right)\right) = \gamma_s^{\sigma}\left(\theta_{Top}\left(\sigma\right)\right)$ for $s = 0, 1$.*

The previous proposition describes the program that determines the optimal RPIC system. There are two main differences to the previous section. First, there is pooling at the top ($\widehat{\sigma} \in \Phi$), which implies that the report of $\theta_{Top}(\widehat{\sigma})$ induces the largest belief by the principal, i.e., it is when the principal would like to reward the agent the most. The second difference is that the RP constraint only binds at the top, i.e., it is only necessary to impose that, after observing message $\theta_{Top}(\widehat{\sigma})$, the principal does not want to increase the reward of the agent for any signal s. Seeing as that constraint binds, it must hold with equality, which becomes constraint iii). Constraints i) and ii) are similar to the previous section and are the minimal conditions that guarantee incentive compatibility.

4.3 Pooling at the top

As described above, pooling at the top means that types above a certain threshold pick the same contract. Going back to the examples of the introduction, this would imply that there would be less variability in the types of investment contracts of good projects, compared to those of lesser quality. In the case of the hurricane victim, where the contract is a function of the reported damages, pooling at the top would mean that there would be a cap, so that reported damages above a certain level would receive the same compensation package. And in the case of the defendant, it means that innocent agents are not the only ones who refuse to confess; guilty agents do also with some positive probability.

4.3.1 The argument

The failure of the revelation principle means that, in general, there need not be any connection between an agent's type θ and the message m he chooses. In particular, it need not be that $\theta = m$. Thus, without loss of generality, one can restrict attention to mechanisms where $d_1(m)$ is increasing with m, i.e., $d_1(\theta_{n+1}) \geq d_1(\theta_n)$ for all n. There is no loss of generality because I am simply giving a particular label to the different messages.

Recall that lemma 1 implies that the agent's strategy must be monotone, i.e., larger types choose messages with larger differences between d_1 and d_0. Therefore, if one imposes that $d_1(\cdot)$ is weakly increasing, which, in turn, implies that $d_0(\cdot)$ is weakly decreasing, it follows that, from the point of view of the principal, the messages are ordered according to her preferred reward. This means that

$$\gamma_0^\sigma(\theta_{n+1}) \geq \gamma_1^\sigma(\theta_n) \text{ for all } n$$

simply because, if there is some type θ who is indifferent between sending messages θ_n and θ_{n+1}, all types above θ will certainly not send message θ_n, while all types below θ will certainly not send message θ_{n+1}. If, for some strategy profile σ, one defines message $\theta_{Top}(\sigma)$ as the "top" message, i.e., the one that is sent by the largest types, then it must also be the message after which the principal would like to choose the largest reward.

The crucial property that leads to the pooling at the top property is that the only RP constraint that binds is the one that restricts the top message ($m = \theta_{Top}(\sigma)$), i.e., none of the other RP constraints bind. To see why that is, let us, for the sake of argument, consider only systems for which $d_1(\theta) \geq d_0(\theta)$ for all $\theta \in \Theta$ (I show in the proof of proposition 4 that this is without loss of generality). I argue that if an IC system is such that

$$d_s\left(\theta_{Top}\left(\sigma\right)\right) \geq \gamma_s^\sigma\left(\theta_{Top}\left(\sigma\right)\right) \text{ for } s = 0, 1$$

then, for all $m \in \Theta$ sent with positive probability,

$$d_s\left(m\right) \geq \gamma_s^\sigma\left(m\right) \text{ for } s = 0, 1$$

The argument is easier to understand by considering figure 2. On the left side of figure 2 - part A - I represent, for each signal $s \in \{0, 1\}$, $\gamma_s^\sigma\left(\theta_{Top}\left(\sigma\right)\right)$ and $\gamma_s^\sigma\left(m\right)$ for some $m \neq \theta_{Top}\left(\sigma\right)$ sent with positive probability. As stated above, it must be that $\gamma_1^\sigma\left(\theta_{Top}\left(\sigma\right)\right) \geq \gamma_0^\sigma\left(\theta_{Top}\left(\sigma\right)\right) > \gamma_1^\sigma\left(m\right) \geq \gamma_0^\sigma\left(m\right)$.[7] In B, I add $d_1\left(\theta_{Top}\left(\sigma\right)\right)$ and $d_0\left(\theta_{Top}\left(\sigma\right)\right)$. Because the top message is RP, then $d_s\left(\theta_{Top}\left(\sigma\right)\right) \geq \gamma_s^\sigma\left(\theta_{Top}\left(\sigma\right)\right)$ for $s = 0, 1$. Finally, in C, I add $d_1\left(m\right)$ and $d_0\left(m\right)$. By incentive compatibility, $d_1\left(m\right)$ and $d_0\left(m\right)$ must be "sandwiched" in between $d_1\left(\theta_{Top}\left(\sigma\right)\right)$ and $d_0\left(\theta_{Top}\left(\sigma\right)\right)$. As one can see, this implies that $d_s\left(m\right) \geq \gamma_s^\sigma\left(m\right)$ for $s = 0, 1$, i.e., the RP constraint relative to message m is satisfied.

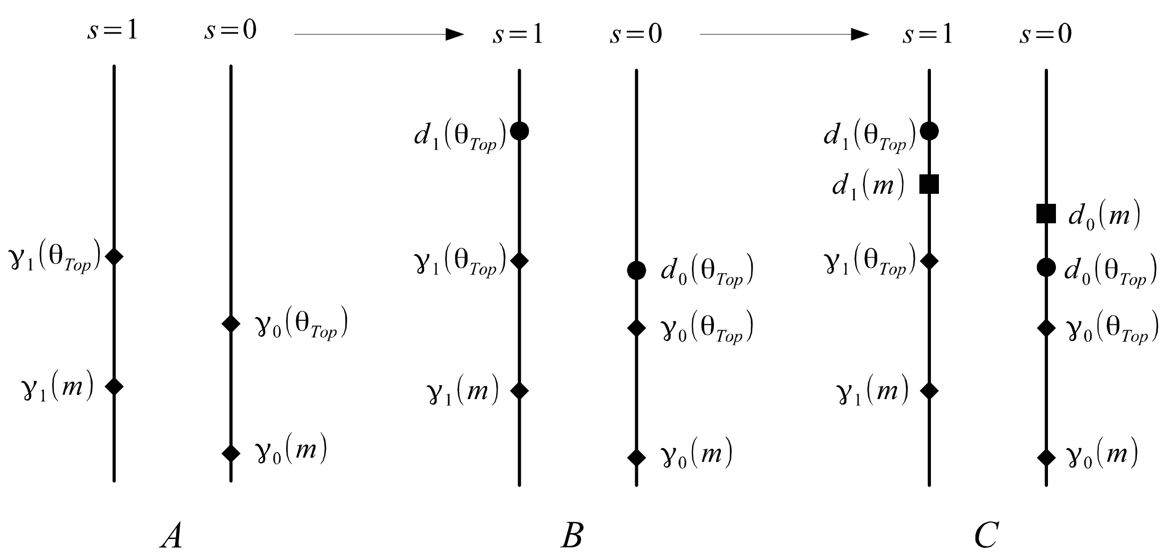

Figure 2: Part C shows that if the top message is RP, then so are the lower messages

What does this have to do with pooling at the top? If all but the top RP constraint do not bind, it follows that the beliefs after each message that is not the top message do

[7] Notice that, for any m, $\gamma_1^\sigma\left(m\right) \geq \gamma_0^\sigma\left(m\right)$, because, conditional on message m, it is more likely that the agent's type is large when $s = 1$ than when $s = 0$.

not matter. The only beliefs that enter the principal's problem are those that emerge after the top message. Therefore, this means that if a type does not send the top message, one can, without loss of generality, simply assume that he reports truthfully, i.e., there is pooling at the top.

Loosely speaking, the idea is that incentive compatibility alone already makes it so that low types receive more than what the principal would like to give them: low types must be given larger rewards than the ones the principal would have preferred so that they will not mimic larger types. Therefore, the requirement that the principal does not want to increase their rewards ex-post ends up not mattering.[8]

4.3.2 Only pooling at the top?

Proposition 4 states that there is pooling at the top, but it leaves open the possibility that there is also some form of pooling at the bottom. If u and v are differentiable, one can find sufficient conditions for which pooling only happens at the top. In what follows, let $\widehat{n}^* \equiv n^*(\widehat{\sigma})$ and $\widehat{\tau} \equiv \tau(\widehat{\sigma})$.

Proposition 5 *Assume that u and v are differentiable, and let*

$$h(x,\theta) \equiv -\frac{\frac{\partial v}{\partial x}(x,\theta)}{u'(x)}$$

i) There cannot be complete pooling at the bottom, i.e., whenever $\widehat{n}^* > 1$,

$$\widehat{d}_1(\theta_{\widehat{n}^*}) > \widehat{d}_1(\theta_1) = \widehat{d}_0(\theta_1) > \widehat{d}_0(\theta_{\widehat{n}^*})^{9}$$

ii) If $\frac{\partial^2 h}{\partial x \partial \theta}(x,\theta) \leq 0$ for all $(x,\theta) \in \mathbb{R} \times \Theta$ and

$$\frac{1}{p(\theta_n)} \frac{\pi(\theta_n) - \pi(\theta_{n-1})}{\pi(\theta_n)\pi(\theta_{n-1})} \tag{1}$$

[8] In Kartik (2009) and Chen (2011), the top types of the agent also pool in their report, as in this paper. 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. By contrast, in this paper, one can show that 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 RP mechanism, it follows that there are many systems where there is no pooling at the top. What I show in the paper is that at least one of the *optimal* RPIC systems exhibits pooling at the top.

[9] The result that $\widehat{d}_1(\theta_1) = \widehat{d}_0(\theta_1)$ does not depend on u and v being differentiable and is also true when $\widehat{n}^* = 1$.

is weakly increasing for all $n > 1$, then $\widehat{d}_1(\theta)$ is strictly increasing for all $\theta \leq \theta_{\widehat{n}^*}$.

Regarding ii), notice that if $v(x,\theta) = -(x-\theta)^2$ and $u(x) = x$, then $\frac{\partial^2 h}{\partial x \partial \theta}(x,\theta) = 0$ for all $(x,\theta) \in \mathbb{R} \times \Theta$. Furthermore, if one assumes that $\pi(\theta_n) = n/N$, then (1) being (weakly) increasing can be written as

$$\frac{p(\theta_{n+1})}{p(\theta_n)} \leq \frac{\left(1-\frac{n}{N}\right)\left(\frac{n}{N}-\frac{1}{N}\right)}{\left(1-\frac{n}{N}-\frac{1}{N}\right)\left(\frac{n}{N}+\frac{1}{N}\right)} \to 1 \text{ as } N \to \infty$$

Therefore, if there are many types, that condition simply becomes that $p(\theta_n)$ be decreasing.

4.4 Only regret at the bottom

The other aspect that is different when one considers renegotiation is the issue of regret by the principal. In the optimal IC system, the principal experiences regret in two ways: it could be that the choice of a particular contract reveals to the principal that the agent's type is high, so that she would rather increase the agent's reward, and it could be that the inference is that the agent's type is low, so that the principal would prefer a lower reward. However, if one allows the players to renegotiate, only the latter type of regret emerges. As stated in proposition 4, this regret has a particular form in the optimal RPIC system, in that it is triggered by all but the top contract, i.e., whenever the agent chooses a contract that is not the top one, he reveals himself to be one of the lowest types, which, in turn, reveals to the principal that the contractualized rewards are "too large". The argument essentially follows from figure 2: only the top RP constraint binds. And the top RP constraint must bind, for otherwise, the optimal IC system would be RPIC.

This result has relevant implications for the investment example of the introduction. If one does not consider renegotiation, one finds that sometimes there is overinvestment, i.e., the investor/principal invests more than what she would have liked ex-post, and sometimes there is underinvestment. However, when one allows the players to renegotiate, there is only overinvestment. And, according to proposition 4, this overinvestment is concentrated on the projects with least quality. Renegotiation then provides an alternative explanation for the apparent prevalence of overinvestment in some markets to some of the current ones based on agency problems (Albuquerque and Wang (2008) or Dow et al. (2005)) and on investors' overconfidence (Malmendier and Tate (2008)).

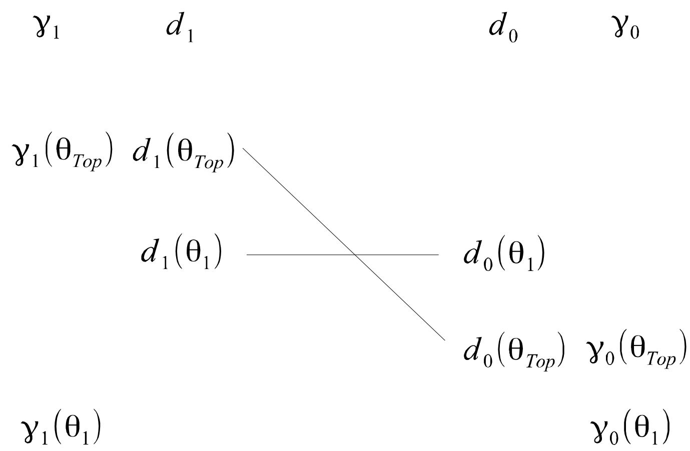

Figure 3: The optimal RPIC system when $N = 2$.

Figure 3 illustrates the optimal RPIC system when $N = 2$, which, as discussed above, can be used to discuss, among others, the defendant example. Message $m = \theta_{Top}$ is sent by type θ_2 with probability 1 and by type θ_1 with some probability $\tau \in (0, 1)$. This means that, upon receiving message $m = \theta_2$, the principal will not be certain of the type of the agent. Therefore, when she observes that $s = 1$, she will be more convinced that the agent's type is θ_2 (the innocent type), which will make her want to give the agent a larger reward (a smaller punishment). Seeing as the mechanism is sequentially optimal at the top, the optimal RPIC system has the principal do exactly that. When the principal observes message $m = \theta_1$, she infers that the agent's type is θ_1 (the guilty type). While the optimal RPIC system specifies that the reward that follows that message is constant, it must be larger than the reward that the principal would prefer to give type θ_1 to ensure that the low type has enough incentives to confess to being the low type: because the principal infers that the agent is guilty, she would like to renege on her promise of leniency and increase her punishment. Naturally, she cannot do this as the rights of confessing agents are generally protected.[10]

4.5 The threshold type

Proposition 4 provides a two-step description of how to find the optimal RPIC system. The first step involves finding the optimal mechanism, given any threshold (n^*, τ). Let that mechanism be denoted by $d(n^*, \tau)$. The second step is about optimizing over

[10] In the United States, the rights of confessing agents are protected under rule 11 of the federal rules of criminal procedure.

all the thresholds and finding the optimal threshold $(\widehat{n}^*, \widehat{\tau})$. Notice that the expected utility of the principal, given some threshold (n^*, τ) and mechanism $d(n^*, \tau)$ is given by

$$A(n^*, \tau) + B(n^*, \tau)$$

where

$$A(n^*, \tau) \equiv \max_{x_0, x_1} \left\{ \sum_{n=n^*+1}^{N} p(\theta_n) g(\theta_n, x_1, x_0) + (1-\tau) p(\theta_{n^*}) g(\theta_{n^*}, x_1, x_0) \right\}$$

and

$$B(n^*, \tau) \equiv \left\{ \begin{array}{c} \sum_{n=1}^{n^*-1} p(\theta_n) \left(g(\theta_n, d_1(n^*, \tau)(\theta_n), d_0(n^*, \tau)(\theta_n))\right) + \\ \tau p(\theta_{n^*}) \left(g(\theta_{n^*}, d_1(n^*, \tau)(\theta_{n^*}), d_0(n^*, \tau)(\theta_{n^*}))\right) \end{array} \right\}$$

and where

$$g(\theta_n, x_1, x_0) \equiv \pi(\theta_n) v(x_1, \theta_n) + (1 - \pi(\theta_n)) v(x_0, \theta_n)$$

If one assumes that u and v are differentiable, then the optimal threshold $(\widehat{n}^*, \widehat{\tau})$ is such that

$$\frac{\partial A}{\partial \tau}(\widehat{n}^*, \widehat{\tau}) + \frac{\partial B}{\partial \tau}(\widehat{n}^*, \widehat{\tau}) = 0$$

By the envelope theorem, we have that

$$\frac{\partial A}{\partial \tau}(\widehat{n}^*, \widehat{\tau}) = -p(\theta_{\widehat{n}^*}) g(\theta_{\widehat{n}^*}, \gamma_1(\widehat{n}^*, \widehat{\tau}), \gamma_0(\widehat{n}^*, \widehat{\tau})) \tag{2}$$

where

$$(\gamma_1(n^*, \tau), \gamma_0(n^*, \tau)) = \arg\max_{x_0, x_1} \left\{ \sum_{n=n^*+1}^{N} p(\theta_n) g(\theta_n, x_1, x_0) + (1-\tau) p(\theta_{n^*}) g(\theta_{n^*}, x_1, x_0) \right\}$$

while

$$\frac{\partial B}{\partial \tau}(\widehat{n}^*, \widehat{\tau}) = p(\theta_{\widehat{n}^*}) g\left(\theta_{\widehat{n}^*}, \widehat{d}_1(\theta_{\widehat{n}^*}), \widehat{d}_0(\theta_{\widehat{n}^*})\right) + \tag{3}$$

$$\lambda(\widehat{n}^*, \widehat{\tau}) \left[\begin{array}{c} \pi(\theta_{\widehat{n}^*}) u'(\gamma_1(\widehat{n}^*, \widehat{\tau})) \frac{\partial \gamma_1}{\partial \tau}(\widehat{n}^*, \widehat{\tau}) + \\ (1 - \pi(\theta_{\widehat{n}^*})) u'(\gamma_0(\widehat{n}^*, \widehat{\tau})) \frac{\partial \gamma_0}{\partial \tau}(\widehat{n}^*, \widehat{\tau}) \end{array} \right] \tag{4}$$

where $\lambda(n^*, \tau)$ represents the multiplier associated with the following incentive constraint:

$$\begin{aligned} & \pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_{n^*}\right)\right) \\ = & \pi\left(\theta_{n^*}\right) u\left(\gamma_1\left(n^*, \tau\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(\gamma_0\left(n^*, \tau\right)\right) \end{aligned}$$

constraint.

Analyzing these two expressions reveals the trade-off that determines the choice of the optimal threshold. Raising the threshold makes more types confess, which has two effects. On the one hand, this is good for the principal because she is risk averse: for any (n^*, τ), the principal would prefer type θ_{n^*} to take contract $\left(d_1\left(n^*, \tau\right)\left(\theta_{n^*}\right), d_0\left(n^*, \tau\right)\left(\theta_{n^*}\right)\right)$ over contract $\left(\gamma_1\left(n^*, \tau\right), \gamma_0\left(n^*, \tau\right)\right)$, because, while the agent is indifferent between the two, the former has less variance, i.e.,

$$\gamma_1\left(n^*, \tau\right)-\gamma_0\left(n^*, \tau\right)>d_1\left(n^*, \tau\right)\left(\theta_{n^*}\right)-d_0\left(n^*, \tau\right)\left(\theta_{n^*}\right).$$

This means that that $(2)+(3)>0$.

However, if more types confess, the beliefs held by the principal after observing that the top contract has been chosen become larger. Therefore, there will be an increase in the top contract's rewards: $\gamma_s\left(n^*, \tau\right)$ is increasing with τ for $s=0,1$, which, in turn, implies that the rewards of all other contracts must also be larger, by incentive compatibility. Seeing as the types below the threshold are types that receive rewards that are above the principal's preferred rewards (as discussed in the previous section), this ends up being negative for the principal, which is why $\lambda\left(\widehat{n}^*, \widehat{\tau}\right)<0$ and $(4)<0$.[11]

Finally, one can also show that the optimal threshold is always interior, because, whenever $(n^*, \tau)=(1,0)$, then $(4)=0$. This means that the lowest type always confesses with positive probability.

[11] In the proof of proposition 4 (step 3), I show that constraint

$$\begin{aligned} & \pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_{n^*}\right)\right) \\ = & \pi\left(\theta_{n^*}\right) u\left(\gamma_1\left(n^*, \tau\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(\gamma_0\left(n^*, \tau\right)\right) \end{aligned}$$

can be replaced by

$$\begin{aligned} & \pi\left(\theta_{n^*}\right) u\left(d_1\left(\theta_{n^*}\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(d_0\left(\theta_{n^*}\right)\right) \\ \geq & \pi\left(\theta_{n^*}\right) u\left(\gamma_1\left(n^*, \tau\right)\right)+\left(1-\pi\left(\theta_{n^*}\right)\right) u\left(\gamma_0\left(n^*, \tau\right)\right) \end{aligned}$$

and is binding, which proves that $\lambda\left(n^*, \tau\right)<0$ for any threshold (n^*, τ).

4.6 Comparative statics

By rearranging some terms, one can write the principal's expected utility under the optimal RPIC system as $\widehat{A}+\widehat{B}$, where

$$\widehat{A} \equiv \sum_{n=1}^{N} p\left(\theta_n\right) g\left(\theta_n, \gamma_1\left(\widehat{n}^*, \widehat{\tau}\right), \gamma_0\left(\widehat{n}^*, \widehat{\tau}\right)\right)$$

and

$$\widehat{B} \equiv \left\{ \begin{array}{l} \sum_{n=1}^{\widehat{n}^*-1} p\left(\theta_n\right)\left(g\left(\theta_n, d_1\left(\widehat{n}^*, \widehat{\tau}\right)\left(\theta_n\right), d_0\left(\widehat{n}^*, \widehat{\tau}\right)\left(\theta_n\right)\right)-g\left(\theta_n, \gamma_1\left(\widehat{n}^*, \widehat{\tau}\right), \gamma_0\left(\widehat{n}^*, \widehat{\tau}\right)\right)\right)+ \\ \widehat{\tau} p\left(\theta_{\widehat{n}^*}\right)\left(g\left(\theta_{\widehat{n}^*}, d_1\left(\widehat{n}^*, \widehat{\tau}\right)\left(\theta_{\widehat{n}^*}\right), d_0\left(\widehat{n}^*, \widehat{\tau}\right)\left(\theta_{\widehat{n}^*}\right)\right)-g\left(\theta_{\widehat{n}^*}, \gamma_1\left(\widehat{n}^*, \widehat{\tau}\right), \gamma_0\left(\widehat{n}^*, \widehat{\tau}\right)\right)\right) \end{array} \right\}$$

Dividing the principal's expected utility in this manner is useful in that it makes clear the influence of risk aversion in the ability of the principal to benefit from interacting with the agent. Notice that $\widehat{A}$ represents the payoff the principal would get if, instead of allowing the agent to choose, she was to impose contract $\left(\gamma_1\left(\widehat{n}^*, \widehat{\tau}\right), \gamma_0\left(\widehat{n}^*, \widehat{\tau}\right)\right)$. If the principal was not able to offer more than a single contract, she would not want to offer $\left(\gamma_1\left(\widehat{n}^*, \widehat{\tau}\right), \gamma_0\left(\widehat{n}^*, \widehat{\tau}\right)\right)$ but rather a contract with smaller rewards. In particular, she would like to choose contract

$$\left(x_0', x_1'\right)=\arg\max_{x_0, x_1} E\left(g\left(\theta_n, x_1, x_0\right)\right)$$

The reason why she prefers to offer a menu of different contracts comes from $\widehat{B}$. $\widehat{B}$ represents the gain for the principal of having each type $\theta_n \leq \theta_{\widehat{n}^*}$ confess rather than pool with the larger types. As discussed above, because the principal is risk averse, she benefits from having the agent confess and receive a less variant contract. I show below that the agent being more risk averse is also beneficial for the principal for similar reasons. If the principal knows the agent is risk averse, she is able to reduce the expected reward the agent receives when he chooses to confess his type and still have the agent willing to do so, provided that she also reduces the variance of that contract.[12]

[12] By this argument, it follows that if both the principal and the agent were risk neutral, the optimal RPIC system would simply consist of the principal imposing one contract. I discuss this case in more detail in Silva (2018).

Proposition 6 *Let $u : \mathbb{R} \rightarrow \mathbb{R}$ be any concave utility function and let $\xi : \mathbb{R} \rightarrow \mathbb{R}$ be any strictly concave function. The expected utility of the principal under the optimal RPIC system is smaller when the agent's utility function is u then when it is $\xi(u)$.*

5 Related Literature

5.1 The setting

The environment studied in this paper, where the agent's utility function is independent of his type, has been studied before in different contexts. First, there is a literature on hard evidence that considers the same type of preferences but, instead of assuming that there is an exogenous public signal which allows the principal to differentiate between the different types of the agent, it assumes that the message set of the agent is type dependent (for example, Green and Laffont (1986), Bull and Watson (2007), Glazer and Rubinstein (2004,2006), Hart, Kremer and Perry (2017)). In a separate paper (Silva (2017)), I show that these models are equivalent to assuming that the signal of this paper is perfectly correlated with the agent's type but is privately observed by the agent, who can then verifiably present it to the principal. Those assumptions, coupled with the assumption that I make that better types are more likely to draw the "good" signal imply that the principal would be able to implement her most preferred allocation through a simple unravelling argument, even if she did not have any commitment power.

Ben-Porath, Dekel and Lipman (2014) and Mylovanov and Zapechelnyuk (2017) study a similar problem in that there is a principal who cares about the type of the agent, agents have type independent utility functions, there are no transfers and there is an exogenous signal correlated with the agent's type. However, they focus on the case where the principal has commitment power, while the largest portion of this paper is devoted to studying limited commitment power. 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.

Finally, Siegel and Strulovici (2016) study a similar environment but assume that the agent only has two types, in addition to only considering the case where the principal is able to commit. In Silva (2018), I consider an application of this problem to N agents but assume that each agent only has two types and that the principal is risk neutral.

5.2 Renegotiation proof mechanism design

5.2.1 The one-shot problem

RP mechanisms have been studied in the contexts of complete and incomplete information. If there is complete information, 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 RP, 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 that might impact the desire of the other player(s) to renegotiate.

Some of the literature has addressed this issue by adding an RP constraint to the typical mechanism design problem (Green and Laffont (1987), Forges (1994), Neeman and Pavlov (2013), Goltsman (2011), Beshkar (2016)). While different papers have different definitions, the overall goal of adding the constraint is to guarantee that if a mechanism is RP, 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 (M, d). 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 (M', d') that the principal strictly prefers to propose after observing m, then (M, d) is not RP.

One of the drawbacks of the previous literature is that it uses a "one-shot" criterion to determine whether a mechanism is RP or not. In particular, it might be that (M, d) is not RP because there is a "blocking" mechanism (M', d'), which might itself not be RP, i.e., after observing m' it might be that the principal wishes to propose some other renegotiation mechanism (M'', d''). However, if (M', d') is not RP, its validity as a blocking mechanism is put into question. As a result, these type of constraints end up being too demanding.[13]

5.2.2 Overcoming the one-shot criterion problem

One way to overcome the one-shot criterion problem is to explicitly model renegotiation as an infinite game, where the principal can always propose a renegotiation mechanism before implementing a contract. Strulovici (2017) does exactly this; he explicitly models a dynamic renegotiation game, where the principal proposes binding mechanisms in each period until choosing to stop. In particular, he characterizes the set of perfect bayesian equilibria in the case that negotiation frictions (the probability that the negotiation is exogenously terminated in each period) are negligible.

I follow a different approach. As described in section 4, I also add an RP constraint to the standard mechanism design problem. That constraint essentially guarantees that once the agent has chosen a message m and signal s has been realized, the principal does not want to propose a different mechanism. Thus, essentially, the difference of my approach to the more classical approach is that the opportunity to renegotiate comes after signal s has been realized. 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). Using the notation from above, given a second renegotiation mechanism (M', d'), the agent chooses the same message m' for any type, provided signal s has already been realized. As a result, receiving m' does not convey any new information to the principal, which, in turn, does not make her want to propose some new mechanism (M'', d''), i.e., (M', d') is not renegotiated.[14]

[13] 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).

[14] 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.

5.2.3 Comparison to Strulovici (2017)

The main result in Strulovici (2017) is that the mechanism that is implemented is separating (each type reports truthfully; when there are only two types, the high type picks one contract while the low type picks a different contract) and ex-post efficient. This would mean that, in equilibrium, the principal would get to know with certainty (or close to it) the real type of the agent. However, in my setting, while the mechanism is still ex-post efficient, I find that there is no complete separation; for example, when there are only two types, the low type randomizes between choosing the high type contract and the low type contract. The reason for the contrast is that, in Strulovici (2017), 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 that succeed in screening between the agent's types. In a way, the agent's private information always matters to the renegotiation, so that as long as the principal is uncertain about the agent's type, she always benefits from renegotiating. In my paper, that impetus does not exist once the signal has been realized, which is when the renegotiation is assumed to take place; once the signal becomes publicly known, the agent's decision becomes independent of his type, so further separation becomes impossible.

One question the reader might have, following the description of Strulovici (2017), concerns what in my model prevents the principal from renegotiating with the agent after receiving his message but before the public signal is realized. The answer to this depends very much on the timing of events. Imagine that the public signal is certain to arrive at some period t, say $t = 5$. The best mechanism the principal can hope to implement is the optimal RPIC mechanism, because any mechanism that is not RP would be renegotiated once the signal has arrived at period $t = 5$. However, the concern might be that the optimal RPIC mechanism may not be implementable because it might get renegotiated away before period $t = 5$. In fact, if the principal was to propose the optimal RPIC mechanism at period $t = 1$, it might be the case that at period $t = 2$, she would like to renegotiate it for reasons that are similar to Strulovici (2017): as long as there is uncertainty with respect to the agent's type, the principal might benefit from proposing further and further renegotiation offers to the agent. However, to prevent this, what the principal can do to implement the optimal RPIC mechanism is to wait precisely until period $t = 5$ to propose the optimal RPIC mechanism to the agent. In that way, there would be no time to renegotiate. Once the agent responds to the offer of the principal, the signal is realized, which destroys any desire of the principal to renegotiate further. Thus, it is not as if the principal does not want to renegotiate the RPIC mechanism before the signal arrives; it is more that she waits until the last minute before the signal is realized to propose it, precisely to prevent those renegotiation opportunities from arising.[15]

6 Appendix

6.1 Appendix A - the continuum case

In this section, I extend the model to the case where the agent's type θ is a continuum random variable. Assume that $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$, where $0 < \underline{\theta} < \overline{\theta} < 1$, and that, without loss of generality, $\pi(\theta) = \theta$ for all $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$.

6.1.1 Optimal IC system

Let us start by considering the problem of finding the optimal IC system. Once again by the revelation principle, I only consider direct mechanisms with truthful reporting. The only difference to section 3 is that one must replace the "discrete" incentive constraints by continuous ones. Notice that system $((\Theta, d), \sigma^*)$ is incentive compatible if and only if, for all $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$,

$$\theta \in \arg \max_{\theta' \in \left[\underline{\theta}, \overline{\theta}\right]} \theta u\left(d_1\left(\theta'\right)\right) + (1-\theta)\, u\left(d_0\left(\theta'\right)\right)$$

For any mapping d, let

$$\widetilde{u}_d(\theta) \equiv \theta u\left(d_1(\theta)\right) + (1-\theta)\, u\left(d_0(\theta)\right)$$

[15] This logic is somewhat similar to Evans and Reiche (2015), where it is shown that if one considers a fixed renegotiation length, the optimal IC mechanism can be implemented. Thus, the fact that the mechanism can be renegotiated does not matter. In my paper, it is as if renegotiation can always occur, so the optimal IC system cannot be implemented, because it would be renegotiated after the signal has been realized. However, because the public signal arrives at a specific moment of time, the renegotiation that could occur before the signal is known will not occur.

Following Myerson (1981), it follows that system $((\Theta, d), \sigma^*)$ is incentive compatible if and only if i) for all $\theta \in [\underline{\theta}, \overline{\theta}]$,

$$\widetilde{u}_d(\theta) = \int_{\underline{\theta}}^{\theta} (u(d_1(z)) - u(d_0(z)))\, dz + \widetilde{u}(\underline{\theta})$$

and ii) $u(d_1(\cdot)) - u(d_0(\cdot))$ is (weakly) increasing.

Following Milgrom and Segal (2002), it follows that one can replace conditions i) and ii) by a) for all $\theta \in [\underline{\theta}, \overline{\theta}]$,

$$\widetilde{u}_d'(\theta) = u(d_1(\theta)) - u(d_0(\theta)) \text{ for all } \theta \in [\underline{\theta}, \overline{\theta}]$$

and b) $\widetilde{u}_d'(\cdot)$ is (weakly increasing).

Notice that, given $\widetilde{u}_d$, one can find the corresponding incentive compatible allocation (d_1, d_0) as follows: for all $\theta \in [\underline{\theta}, \overline{\theta}]$,

$$d_0(\theta) = u^{-1}(\widetilde{u}_d(\theta) - \theta \widetilde{u}_d'(\theta))$$

and

$$d_1(\theta) = u^{-1}(\widetilde{u}_d(\theta) + (1-\theta)\widetilde{u}_d'(\theta))$$

Therefore, the problem of finding the optimal mapping d for which system $((\Theta, d), \sigma^*)$ is the optimal IC system becomes one of finding the function $\widetilde{u}$ that maximizes

$$\int_{\underline{\theta}}^{\overline{\theta}} p(\theta)\, g\left(\theta, u^{-1}(\widetilde{u}(\theta) + (1-\theta)\widetilde{u}'(\theta)), u^{-1}(\widetilde{u}(\theta) - \theta\widetilde{u}'(\theta))\right) d\theta \tag{5}$$

subject to

$$\widetilde{u}' \text{ is weakly increasing} \tag{6}$$

Formulating the problem with continuous types facilitates the analysis of whether there are any conditions for which the optimal d^* is separating, i.e., d_1^* is strictly increasing. For that to happen, it would have to be that constraint (6) is not binding. In the following proposition, I show that, in general, there are no such conditions, i.e., there is always some pooling in the optimal IC system.

Proposition 7 *If $h(\cdot,\theta)$ is strictly monotone for all $\theta \in \Theta$, then the mapping d' that maximizes (5) violates constraint (6).*

6.1.2 Optimal RPIC system

One of the difficulties involved with a continuum of types is that the result in Bester and Strausz (2001) is not certain to extend to continuous types. So, there might be loss of generality in assuming that $M = \Theta$. However, if one does assume that $M = \Theta$, all the steps of proposition 4 still follow. The only difference is that, just like in the optimal IC system, one has to replace the local incentive constraints by their continuous counterpart.

Proposition 8 *If $M = \Theta$, there is an optimal RPIC system $\left(\left(\Theta, \widehat{d}\right), \widehat{\sigma}\right)$ such that i) there is pooling at the top, i.e., every type θ above some threshold θ^* reports the same message θ_{Top}, ii) the top contract is sequentially optimal, i.e.,*

$$\widehat{d}_s(\theta_{Top}) = \arg\max_{x\in\mathbb{R}} E^{\widehat{\sigma}}\left(v(x,\theta)\,|m = \theta_{Top}, s\right)$$

and iii) for all $\theta \leq \theta^*$,

$$\widehat{d}_1(\theta) = u^{-1}\left(\widetilde{u}(\theta) + (1-\theta)\,\widetilde{u}'(\theta)\right) \text{ and } \widehat{d}_0(\theta) = u^{-1}\left(\widetilde{u}(\theta) - \theta\widetilde{u}'(\theta)\right)$$

where $\widetilde{u} : [\underline{\theta}, \theta^*] \to \mathbb{R}$ maximizes

$$\int_{\underline{\theta}}^{\theta^*} p(\theta)\, g\left(\theta, u^{-1}\left(\widetilde{u}(\theta) + (1-\theta)\,\widetilde{u}'(\theta)\right), u^{-1}\left(\widetilde{u}(\theta) - \theta\widetilde{u}'(\theta)\right)\right) d\theta$$

subject to a) $\widetilde{u}'$ being weakly increasing, b)

$$u^{-1}\left(\widetilde{u}(\theta) + (1-\theta)\,\widetilde{u}'(\theta)\right) \leq \widehat{d}_1(\theta_{Top})$$

and c)

$$\widetilde{u}(\theta^*) = \theta u\left(\widehat{d}_1(\theta_{Top})\right) + (1-\theta)\, u\left(\widehat{d}_0(\theta_{Top})\right)$$

As proposition 8 shows, the main insights of the optimal RPIC system still remain true when the agent's type is a continuous random variable: there is still pooling at the top, and there is only regret at the bottom.

6.2 Appendix B - Proofs

6.2.1 Proof of Proposition 2

Proposition 2.ii. For all $n = 1, ..., N-1$,

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

Proof. Notice that d^* maximizes $V^\Theta(d, \sigma)$ subject to all incentive constraints. By lemma 1, one can add the constraint that imposes that $d_1(\cdot)$ is increasing without constraining the problem further. Consider a relaxed version of this problem, where the principal maximizes $V^\Theta(d, \sigma)$ subject only to the following constraints: $C1$) $d_1(\cdot)$ is increasing, and $C2$) for all $n = 1, ..., N-1$,

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

In words, constraint $C2$ states that each type does not want to mimic the next largest type. I start by showing that, in any solution $\widetilde{d}$ of the relaxed problem, $C2$ must hold with equality: for all $n = 1, ..., N-1$,

$$\begin{aligned} & \pi(\theta_n) u\left(\widetilde{d}_1(\theta_n)\right) + (1-\pi(\theta_n)) u\left(\widetilde{d}_0(\theta_n)\right) \\ = & \pi(\theta_n) u\left(\widetilde{d}_1(\theta_{n+1})\right) + (1-\pi(\theta_n)) u\left(\widetilde{d}_0(\theta_{n+1})\right) \end{aligned} \tag{7}$$

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

$$\begin{aligned} & \pi(\theta_n) u\left(\widetilde{d}_1(\theta_n)\right) + (1-\pi(\theta_n)) u\left(\widetilde{d}_0(\theta_n)\right) \\ > & \pi(\theta_n) u\left(\widetilde{d}_1(\theta_{n+1})\right) + (1-\pi(\theta_n)) u\left(\widetilde{d}_0(\theta_{n+1})\right) \end{aligned}$$

By $C1$, it follows that $\widetilde{d}_1(\theta_{n+1}) \geq \widetilde{d}_1(\theta_n)$ and so $\widetilde{d}_0(\theta_{n+1}) < \widetilde{d}_0(\theta_n)$.

Assume first that $x^*(\theta_{n+1}) > \widetilde{d}_0(\theta_{n+1})$. Then, the principal would be better off by increasing $\widetilde{d}_0(\theta_{n+1})$ and still satisfy $C2$, which is a contradiction to optimality of the relaxed problem. Assume instead that $x^*(\theta_{n+1}) \leq \widetilde{d}_0(\theta_{n+1})$. This implies that $x^*(\theta_n) < \widetilde{d}_0(\theta_n)$. As a result, the principal would prefer to lower $\widetilde{d}_0(\theta_n)$ and still satisfy $C2$, which is again a contradiction to optimality of the relaxed problem. Therefore, $C2$ holds with equality.

The argument is complete by noticing that the solution of the relaxed problem is in fact the solution of the original problem, i.e., $\widetilde{d} = d^*$, because if $\widetilde{d}$ is such that $C1$

and (7) hold, then system $\left(\left(\Theta, \tilde{d}\right), \sigma^*\right)$ is IC. ■

Proposition 2.iii)

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

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

$$z(\theta_n) \geq \pi(\theta_n) d_1^*(\theta_{n+1}) + (1 - \pi(\theta_n)) d_0^*(\theta_{n+1}) > z(\theta_{n+1}) \tag{8}$$

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 (8), 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 gap 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

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

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 3.iv)

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

Proof. I start by showing that $d_1^*(\theta_1) = d_0^*(\theta_1)$. 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.

As for the second part of the statement, notice that $d^*_1(\theta_N) \geq d^*_1(\theta_1)$ by $C1$. If $d^*_1(\theta_N) = d^*_1(\theta_1)$, then this would mean that $d^*_s(\theta)$ would be independent of s and θ - the principal would simply impose a certain reward. That system would be worse than system d'', where

$$d''_s(\theta) = \arg\max_{x \in \mathbb{R}} \; E(v(x, \theta) | s) \;\; \text{for } s = 0, 1$$

for all $\theta \in \Theta$, which is trivially IC. So, the statement follows. ■

Proposition 3.v)

$$d^*_1(\theta_N) \leq x^*(\theta_N) \;\; \text{and} \;\; d^*_0(\theta_N) < x^*(\theta_N)$$

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, then $d^*_1(\theta_N) = d^*_0(\theta_N) = x^*(\theta_N)$. 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 $s = 0, 1$,

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

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 $((\Theta, d^*) , \sigma^*)$. ■

6.2.2 Proof of Proposition 4

Preliminaries: 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 4.1. *If there is a 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

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

and

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

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

For $s=0,1$, let

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

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

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

and that

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

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

As discussed in the text, by Bester and Strausz (2001), there is an optimal RPIC system where $M=\Theta$, so that, without loss of generality, in what follows I assume that $M=\Theta$.

I prove proposition 4 in four steps.

Step 1: *If system $\left(\left(\Theta,d\right),\sigma\right)$ is an optimal RPIC system, then, for any $m\in\Theta$ such that there is $\theta\in\Theta$ where $\sigma\left(\theta\right)\left(m\right)>0$, $d_1\left(m\right)\geq d_0\left(m\right)$.*

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

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

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

Description of d' :

Let

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

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 $m \in \widetilde{M}$,

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

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 $m \in \widetilde{M}$,

$$\sum_{n=1}^{N} p\left(\theta_n\right)\sigma\left(\theta_n\right)\left(m\right)\left(\begin{array}{c}\pi\left(\theta_n\right)v\left(d_1\left(m\right),\theta_n\right)+\\ \left(1-\pi\left(\theta_n\right)\right)v\left(d_0\left(m\right),\theta_n\right)\end{array}\right) < \sum_{n=1}^{N} p\left(\theta_n\right)\sigma\left(\theta_n\right)\left(m\right)v\left(z,\theta_n\right) \tag{9}$$

which proves the statement.

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

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

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

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

Notice that the LHS of (9) can be written as the sum of A and B, where

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

and

$$B=\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)\left(\pi\left(\theta_{n}\right)-\pi\left(\widehat{\theta}\right)\right)\left(v\left(d_{1}\left(m\right),\theta_{n}\right)-v\left(d_{0}\left(m\right),\theta_{n}\right)\right)$$

Let

$$\widehat{B}=\frac{B}{\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)}$$

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,

$$\widehat{B}\leq\left[\frac{\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)\left(\pi\left(\theta_{n}\right)-\pi\left(\widehat{\theta}\right)\right)}{\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)}\right]\left[\frac{\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)h\left(\theta_{n}\right)}{\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)}\right]=0$$

(see lemma 2.1. in See and Chen (2008)), and so $B\leq0$.

Notice that

$$A<\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)v\left(\pi\left(\widehat{\theta}\right)d_{1}\left(m\right)+\left(1-\pi\left(\widehat{\theta}\right)\right)d_{0}\left(m\right),\theta_{n}\right)$$

because $v\left(\cdot,\theta_{n}\right)$ is strictly concave for any $\theta\in\Theta$.

Furthermore, we have that

$$\arg\max_{x\in\mathbb{R}}\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)v\left(x,\theta_{n}\right)<d_{1}\left(m\right)<z\leq\widehat{z}\leq\pi\left(\widehat{\theta}\right)d_{1}\left(m\right)+\left(1-\pi\left(\widehat{\theta}\right)\right)d_{0}\left(m\right)$$

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

$$\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)v\left(\pi\left(\widehat{\theta}\right)d_{1}\left(m\right)+\left(1-\pi\left(\widehat{\theta}\right)\right)d_{0}\left(m\right),\theta_{n}\right)\leq\sum_{n=1}^{N}p\left(\theta_{n}\right)\sigma\left(\theta_{n}\right)\left(m\right)v\left(z,\theta_{n}\right)$$

which implies (9). ■

Step 2 is divided into two parts:

Step 2a: *If system* $((\Theta, d), \sigma)$ *is IC, is such that* $d_1(m) \geq d_0(m)$ for any $m \in \Theta$, *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 2a 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 2b, in this case, beliefs only matter after the top message θ_N.

Step 2b: *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)

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

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

$$\tau = 1 - \sigma(\theta_{n^*})(m_N)$$

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 contracts to choose from, so it follows that system $((\Theta, d'), \sigma')$ is IC. And by Step 2a) 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 2 implies that there is an optimal system where there is pooling at the top: notice that, without loss of generality, for all $\sigma \in \Phi$, $\theta_{Top}(\sigma)$ described in the text can be defined to be equal to $m_N = \theta_N$. It also 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

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

ii)

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

iii)

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

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 a) a monotonicity condition stating that $d_1(m)$ is increasing, b) an "upper" incentive constraint, stating that the lowest type sending each message does not want to send the following one, c) a "lower" incentive constraint, stating that the largest type sending each message does not want to send the preceding one, and d) a renegotiation proof condition that applies only to the largest message $m = \theta_N$.

Formally,

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

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

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

and

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

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

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

while condition c) can be written as

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

and, for all $n=1,\ldots,n^*-1$,

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

Finally, the RP condition d) can be stated as

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

and

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

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

Step 3: *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

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

(which implies that $\widehat{\tau}>0$), 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

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

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 3 implies that the solution of Γ' satisfies b), which makes it also the solution of Γ. This means that i) and ii) of the statement of proposition 4 have been proved. What is left is iii).

Step 4: *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

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

and

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

Proof. Suppose not. Consider first the case where

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

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' than with d; satisfies a) and d) by definition and is strictly preferred by the principal, which is a contradiction to optimality.

Suppose instead that

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

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' than 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. ■

Step 4 shows iii) of the statement of proposition 4.

6.2.3 Proof of Proposition 5

Part i) has two parts:

Proposition 5.i.1. $\widehat{d}_1(\theta_1) = \widehat{d}_0(\theta_1)$

Proof. This result follows for the same argument as with the optimal IC system: the only constraint than binds the choice of $\widehat{d}_1(\theta_1)$ and of $\widehat{d}_0(\theta_1)$ is the incentive constraint that states that type $\theta = \theta_1$ does not want to mimic type $\theta = \theta_2$. So, giving type θ_1, the only type who might send message $m = \theta_1$, a constant reward both increases the principal's expected utility (because she is risk averse) and increases the incentives for the lowest type to report to being the lowest type (because he is also risk averse). ■

Proposition 5.i.2. *If* $\widehat{n}^* > 1$, then

$$\widehat{d}_1(\theta_{\widehat{n}^*}) > \widehat{d}_1(\theta_1) = \widehat{d}_0(\theta_1) > \widehat{d}_0(\theta_{\widehat{n}^*})$$

Proof. Suppose not, so that

$$\widehat{d}_1(\theta_n) = \widehat{d}_0(\theta_n) = k$$

for any $n \leq \widehat{n}^*$. By proposition 4, it follows that k is such that

$$u(k) = \pi(\theta_{\widehat{n}^*})\, u(\widehat{\gamma}_1) + (1 - \pi(\theta_{\widehat{n}^*}))\, u(\widehat{\gamma}_0)$$

where

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

Notice that $\widehat{\gamma}_1 > k > \widehat{\gamma}_0$. Consider the alternative system $((\Theta, d'), \widehat{\sigma})$, where $d' = \widehat{d}$ except that

$$d_1'(\theta_n) = k + \varepsilon \text{ and } d_0'(\theta_n) = k - \delta(\varepsilon)$$

for all $n \leq \widehat{n}^*$, where $\delta(\varepsilon)$ is such that

$$u(k) = \pi(\theta_{\widehat{n}^*}) u(k+\varepsilon) + (1 - \pi(\theta_{\widehat{n}^*})) u(k - \delta(\varepsilon)) \tag{10}$$

for some $\varepsilon \geq 0$. If ε is sufficiently small, then $k + \varepsilon < \widehat{\gamma}_1$, so that system $((\Theta, d'), \widehat{\sigma})$ would be RPIC.

Notice that, if $\varepsilon = 0$, the two systems are equal. Let V^ε denote the expected utility of the principal, under system $((\Theta, d'), \widehat{\sigma})$, for a given $\varepsilon \geq 0$. Notice that

$$\begin{aligned} \frac{dV^\varepsilon}{d\varepsilon} &= \sum_{n=1}^{\widehat{n}^*-1} p(\theta_n) \left(\pi(\theta_n) \frac{\partial v}{\partial x}(k+\varepsilon, \theta_n) - (1-\pi(\theta_n)) \frac{\partial v}{\partial x}(k - \delta(\varepsilon), \theta_n) \delta'(\varepsilon) \right) + \\ & \widehat{\tau} p(\theta_{\widehat{n}^*}) \left(\pi(\theta_{\widehat{n}^*}) \frac{\partial v}{\partial x}(k+\varepsilon, \theta_{\widehat{n}^*}) - (1-\pi(\theta_{\widehat{n}^*})) \frac{\partial v}{\partial x}(k - \delta(\varepsilon), \theta_{\widehat{n}^*}) \delta'(\varepsilon) \right) \end{aligned}$$

Using (10), we have that

$$\delta'(0) = \frac{\pi(\theta_{\widehat{n}^*})}{1 - \pi(\theta_{\widehat{n}^*})}$$

so that

$$\begin{aligned} \frac{dV^\varepsilon}{d\varepsilon}(0) &= \sum_{n=1}^{\widehat{n}^*-1} p(\theta_n) \frac{\partial v}{\partial x}(k, \theta_n) \left(\pi(\theta_n) - (1-\pi(\theta_n)) \frac{\pi(\theta_{\widehat{n}^*})}{1-\pi(\theta_{\widehat{n}^*})} \right) + \\ & \widehat{\tau} p(\theta_{\widehat{n}^*}) \frac{\partial v}{\partial x}(k, \theta_{\widehat{n}^*}) \left(\pi(\theta_{\widehat{n}^*}) - (1-\pi(\theta_{\widehat{n}^*})) \frac{\pi(\theta_{\widehat{n}^*})}{1-\pi(\theta_{\widehat{n}^*})} \right) \end{aligned}$$

Recall that $k > \gamma_0 > x^*(\theta_n)$ for all $\theta_n \leq \theta_{\widehat{n}^*}$ so that $\frac{\partial v}{\partial x}(k, \theta_n) < 0$ for all $\theta_n \leq \theta_{\widehat{n}^*}$. This implies that

$$\pi(\theta_n) < (1 - \pi(\theta_n)) \frac{\pi(\theta_{\widehat{n}^*})}{1 - \pi(\theta_{\widehat{n}^*})}$$

for all $\theta_n < \theta_{\widehat{n}^*}$, which means that that $\frac{dV^\varepsilon}{d\varepsilon}(0) > 0$, so that there is a sufficiently small $\varepsilon > 0$ for which system $((\Theta, d'), \widehat{\sigma})$ is both RPIC and better for the principal than the optimal RPIC system, which is a contradiction. ■

Proof of Proposition 5.ii.. Following proposition 4, mapping $\widehat{d} : \{\theta_1, ..., \theta_{\widehat{n}^*}\} \times$

$\{0,1\} \to \mathbb{R}$ maximizes

$$\sum_{n=1}^{\widehat{n}^*-1} p(\theta_n)\left(\pi(\theta_n) v\left(d_1(\theta_n),\theta_n\right) + (1-\pi(\theta_n)) v\left(d_0(\theta_n),\theta_n\right)\right) +$$
$$\widehat{\tau} p(\theta_{\widehat{n}^*})\left(\pi(\theta_{\widehat{n}^*}) v\left(d_1(\theta_{\widehat{n}^*}),\theta_{\widehat{n}^*}\right) + (1-\pi(\theta_{\widehat{n}^*})) v\left(d_0(\theta_{\widehat{n}^*}),\theta_{\widehat{n}^*}\right)\right)$$

subject to the following local incentive constraints: for all $n < \widehat{n}^*$

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

and

$$\begin{aligned} & \pi(\theta_{\widehat{n}^*}) u\left(d_1(\theta_{\widehat{n}^*})\right) + (1-\pi(\theta_{\widehat{n}^*})) u\left(d_0(\theta_{\widehat{n}^*})\right) \\ = \; & \pi(\theta_{\widehat{n}^*}) u\left(\widehat{\gamma}_1\right) + (1-\pi(\theta_{\widehat{n}^*})) u\left(\widehat{\gamma}_0\right) \end{aligned}$$

and the monotonicity constraint:

$$d_1(\theta_1) \leq d_1(\theta_2) \leq \ldots \leq d_1(\theta_{\widehat{n}^*}) \leq \widehat{\gamma}_1$$

Start by noticing that $\widehat{\gamma}_1 > \widehat{d}_1(\theta_{\widehat{n}^*})$, so that constraint $d_1(\theta_{\widehat{n}^*}) \leq \widehat{\gamma}_1$ is not binding. I show that, under the stated conditions, the solution of the relaxed problem, where all monotonicity constraints are eliminated, satisfies the monotonicity constraint, which shows the result, i.e., if the solution of the relaxed problem is $\widetilde{d}$, I show that $\widetilde{d}_1(\theta_n)$ is strictly increasing with n, for all $n \leq \widehat{n}^*$.

Let $\lambda_n \geq 0$ be Lagrange multiplies associated with each of the incentive constraints. I start by showing that, for all $n \leq \widehat{n}^*$,

$$\lambda_n = \sum_{n'=1}^{n} p(\theta_{n'}) \xi_{n'} h\left(\widetilde{d}_1(\theta_{n'}),\theta_{n'}\right) \frac{\pi(\theta_{n'})}{\pi(\theta_n)} \tag{11}$$

where

$$\xi_n = \begin{cases} 1-\widehat{\tau} \text{ if } n = \widehat{n}^* \\ 1 \text{ if } n < \widehat{n}^* \end{cases}$$

I show this result by induction. Notice that the first order condition (Foc) with respect to $d_1(\theta_1)$ can be written as:

$$p(\theta_1)\frac{\partial v}{\partial x}\left(\widetilde{d}_1(\theta_1),\theta_1\right)+\lambda_1 u'\left(\widetilde{d}_1(\theta_1)\right)=0 \Leftrightarrow \lambda_1 = p(\theta_1)\,h\left(\widetilde{d}_1(\theta_1),\theta_1\right)$$

Now, take any $n<\widehat{n}^*$ and assume that

$$\lambda_{n-1}=\sum_{n'=1}^{n-1}p(\theta_{n'})\,h\left(\widetilde{d}_1(\theta_{n'}),\theta_{n'}\right)\frac{\pi(\theta_{n'})}{\pi(\theta_{n-1})} \tag{12}$$

The Foc with respect to $d_1(\theta_n)$ can be written as

$$p(\theta_n)\pi(\theta_n)\frac{\partial v}{\partial x}\left(\widetilde{d}_1(\theta_n),\theta_n\right)+\lambda_n\pi(\theta_n)u'\left(\widetilde{d}_1(\theta_n)\right)=\lambda_{n-1}\pi(\theta_{n-1})u'\left(\widetilde{d}_1(\theta_n)\right)$$

After replacing (12), one gets (11). Finally, if one considers the Foc with respect to $d_1(\theta_{\widehat{n}^*})$, we get that

$$p(\theta_{\widehat{n}^*})(1-\widehat{\tau})\pi(\theta_{\widehat{n}^*})\frac{\partial v}{\partial x}\left(\widetilde{d}_1(\theta_{\widehat{n}^*}),\theta_{\widehat{n}^*}\right)+\lambda_{\widehat{n}^*}\pi(\theta_{\widehat{n}^*})u'\left(\widetilde{d}_1(\theta_{\widehat{n}^*})\right)=\lambda_{\widehat{n}^*-1}\pi(\theta_{\widehat{n}^*-1})u'\left(\widetilde{d}_1(\theta_{\widehat{n}^*})\right)$$

By replacing λ_{n^*-1}, we get (11) evaluated at $n=\widehat{n}^*$.

Notice that, for any $0<n\leq\widehat{n}^*$, the Foc with respect $d_0(\theta_n)$ is given by

$$\begin{aligned}&p(\theta_n)\xi_n(1-\pi(\theta_n))\frac{\partial v}{\partial x}\left(\widetilde{d}_0(\theta_n),\theta_n\right)+\lambda_n(1-\pi(\theta_n))u'\left(\widetilde{d}_0(\theta_n)\right)\\ =\;&\lambda_{n-1}(1-\pi(\theta_{n-1}))u'\left(\widetilde{d}_0(\theta_n)\right)\end{aligned}$$

Using (11), one can rewrite it as

$$h\left(\widehat{d}_1(\theta_n),\theta_n\right)-h\left(\widehat{d}_0(\theta_n),\theta_n\right)=\frac{\kappa_n}{\xi_n}\sum_{n'=1}^{n-1}p(\theta_{n'})\xi_{n'}\pi(\theta_{n'})\,h\left(\widehat{d}_1(\theta_{n'}),\theta_{n'}\right) \tag{13}$$

where

$$\kappa_n\equiv\frac{\pi(\theta_n)-\pi(\theta_{n-1})}{p(\theta_n)(1-\pi(\theta_n))\pi(\theta_n)\pi(\theta_{n-1})}>0$$

Notice that the conditions of the statement imply that κ_n is (weakly) increasing, while, by definition ξ_n is (weakly) decreasing. Therefore, κ_n/ξ_n is weakly increasing. Furthermore, because the system is RP, it follows that $h\left(\widehat{d}_1(\theta_n),\theta_n\right)>0$ for all $n\leq\widehat{n}^*$. This means that the right hand side of (13) is strictly increasing with n. If

$\frac{\partial^2 h}{\partial x \partial \theta_n}(x, \theta_n) \leq 0$ for all (x, θ_n), then it must be that, for all $n < \widehat{n}^*$,

$$0 \leq \widehat{d}_1(\theta_n) - \widehat{d}_0(\theta_n) < \widehat{d}_1(\theta_{n+1}) - \widehat{d}_0(\theta_{n+1})$$

Given that the local incentive constraints imply that either

$$\widehat{d}_1(\theta_{n+1}) \geq \widehat{d}_1(\theta_n) \text{ or } \widehat{d}_0(\theta_{n+1}) \geq \widehat{d}_0(\theta_n)$$

it follows that $\widehat{d}_1(\theta_n)$ is strictly increasing with n, for all $n \leq \widehat{n}^*$. ■

6.3 Proof of Proposition 6

I start by showing the following statement.

Lemma 6.1. *For any*

$$x_1' > x_1'' \geq x_0'' > x_0'$$

such that

$$\pi u(x_1') + (1-\pi) u(x_0') = \pi u(x_1'') + (1-\pi) u(x_0'')$$

for some $\pi \in (0,1)$*, then*

$$\pi \xi(u(x_1')) + (1-\pi) \xi(u(x_0')) < \pi \xi(u(x_1'')) + (1-\pi) \xi(u(x_0''))$$

Proof. Let η be such that

$$\eta u(x_1') + (\pi - \eta) u(x_0') = \pi u(x_1'')$$

and notice that

$$\eta \in (\max\{0, 2\pi - 1\}, \pi)$$

It follows that

$$\begin{aligned} \xi(u(x_1'')) &= \xi\left(\frac{\eta}{\pi} u(x_1') + \frac{(\pi - \eta)}{\pi} u(x_0')\right) \\ &> \frac{\eta}{\pi} \xi(u(x_1')) + \frac{(\pi - \eta)}{\pi} \xi(u(x_0')) \end{aligned}$$

and

$$\begin{aligned}\xi\left(u\left(x_0''\right)\right) &= \xi\left(\frac{(\pi-\eta)}{(1-\pi)}u\left(x_1'\right)+\frac{(1-2\pi+\eta)}{(1-\pi)}u\left(x_0'\right)\right)\\ &> \frac{(\pi-\eta)}{(1-\pi)}\xi\left(u\left(x_1'\right)\right)+\frac{(1-2\pi+\eta)}{(1-\pi)}\xi\left(u\left(x_0'\right)\right)\end{aligned}$$

Therefore,

$$\begin{aligned}&\eta\xi\left(u\left(x_1'\right)\right)+(\pi-\eta)\,\xi\left(u\left(x_0'\right)\right)+(\pi-\eta)\,\xi\left(u\left(x_1'\right)\right)+(1-2\pi+\eta)\,\xi\left(u\left(x_0'\right)\right)\\ <\;&\pi\xi\left(u\left(x_1''\right)\right)+(1-\pi)\,\xi\left(u\left(x_0''\right)\right)\end{aligned}$$

which implies that

$$\pi\xi\left(u\left(x_1'\right)\right)+(1-\pi)\,\xi\left(u\left(x_0'\right)\right)<\pi\xi\left(u\left(x_1''\right)\right)+(1-\pi)\,\xi\left(u\left(x_0''\right)\right)$$

■

Consider the program described in proposition 4, assuming that the agent's utility function is u. Solving that program returns a threshold $(\widehat{n}^*,\widehat{\tau})$ and a mapping $\widehat{d}$. In step 3 in the proof of proposition 4, I show that one can replace the "equality" incentive constraints in the optimal program

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

by

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

Therefore, threshold $(\widehat{n}^*,\widehat{\tau})$ and $\widehat{d}$ would still satisfy all constraints of the program if the agent's utility function was $\eta(u)$, by lemma 6.1. and by the fact that, whenever $\widehat{d}\left(\theta_n\right)\neq\widehat{d}\left(\theta_{n+1}\right)$,

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

But it will not be optimal, because not all incentive constraints will not hold with equality, a condition for optimality. Therefore, it must be that the principal's expected utility is larger if the agent's utility function is given by $\eta(u)$.

6.3.1 Proof of Proposition 7

Recall that, in order to find the optimal IC system, one must find the function $\widetilde{u}$ that maximizes

$$\int_{\underline{\theta}}^{\overline{\theta}} \varrho\left(\theta, \widetilde{u}\left(\theta\right), \widetilde{u}'\left(\theta\right)\right) d\theta$$

where

$$\varrho\left(\theta, \widetilde{u}, \widetilde{u}'\right) \equiv p\left(\theta\right) g\left(\theta, u^{-1}\left(\widetilde{u} + \left(1-\theta\right)\widetilde{u}'\right), u^{-1}\left(\widetilde{u} - \theta\widetilde{u}'\right)\right)$$

subject to $\widetilde{u}'$ being weakly increasing. Consider the relaxed problem where the constraint is disregarded and let the solution of that problem be denoted by $\widehat{u}$. Seeing as there is no explicit boundary condition, it must be that

$$\frac{\partial\varrho}{\partial\widetilde{u}'}\left(\theta, \widehat{u}\left(\underline{\theta}\right), \widehat{u}'\left(\underline{\theta}\right)\right) = \frac{\partial\varrho}{\partial\widetilde{u}'}\left(\theta, \widehat{u}\left(\overline{\theta}\right), \widehat{u}'\left(\overline{\theta}\right)\right) = 0$$

Given that

$$\frac{\partial\varrho}{\partial\widetilde{u}'}\left(\theta, \widetilde{u}, \widetilde{u}'\right) = p\left(\theta\right)\theta\left(1-\theta\right)\left(-h\left(u^{-1}\left(\widetilde{u} + \left(1-\theta\right)\widetilde{u}'\right), \theta\right) + h\left(u^{-1}\left(\widetilde{u} - \theta\widetilde{u}'\right), \theta\right)\right)$$

it follows that

$$\widehat{u}'\left(\underline{\theta}\right) = \widehat{u}'\left(\overline{\theta}\right) = 0$$

Therefore, for the solution of the relaxed problem to satisfy the disregarded constraint, it would have to be that $\widehat{u}'\left(\theta\right) = 0$ for all $\theta \in \Theta$, i.e., $\widehat{d}_1\left(\theta\right) = \widehat{d}_0\left(\theta\right)$ for all $\theta \in \Theta$. Such an allocation is clearly not optimal among all incentive compatible allocations, as it is worse than an allocation that also does not depend on θ but does depend on the signal s.

6.4 Proof of Proposition 8

Steps 1 and 2 of the proof of proposition 4 still hold true, so that there is pooling at the top and the RP constraints of all non-top messages do not bind. Let θ^* denote the threshold type, so that all types $\theta > \theta^*$ send message θ_{Top}, while each type $\theta \leq \theta^*$ sends message $m = \theta$. The principal wants to choose θ^* and d in order to maximize her expected utility subject to the following constraints: i) the RP constraint

$$d_1(\theta_{Top}) \geq \arg\max_{x\in\mathbb{R}} \int_{\theta^*}^{\overline{\theta}} \theta v(x,\theta)\, d\theta$$

and

$$d_0(\theta_{Top}) \geq \arg\max_{x\in\mathbb{R}} \int_{\theta^*}^{\overline{\theta}} (1-\theta)\, v(x,\theta)\, d\theta$$

ii) the monotonicity constraint:

$$d_1(\theta_{Top}) \geq d_1(\theta^*)$$

and $d_1(\theta)$ is (weakly) increasing for all $\theta \leq \theta^*$, iii) the incentive constraint below θ^*:

$$\theta \in \arg\max_{\theta' \in [\underline{\theta},\theta^*]} \theta u(d_1(\theta')) + (1-\theta)\, u(d_0(\theta'))$$

for all $\theta \leq \theta^*$ and iv) the incentive constraint at θ^*:

$$\theta^* u(d_1(\theta^*)) + (1-\theta^*)\, u(d_0(\theta^*)) = \theta^* u(d_1(\theta_{Top})) + (1-\theta^*)\, u(d_0(\theta_{Top}))$$

Consider the relaxed version of this problem where iv) is replaced by iv')

$$\theta^* u(d_1(\theta^*)) + (1-\theta^*)\, u(d_0(\theta^*)) \geq \theta^* u(d_1(\theta_{Top})) + (1-\theta^*)\, u(d_0(\theta_{Top}))$$

Notice that any d for which i-iv' hold is such that $d_s(\theta) > x^*(\theta)$ for all $\theta < \theta^*$. Therefore, it must be that the solution of the relaxed version is such that iv' holds with equality: if not, the principal could decrease all the rewards below θ^*, which would improve her expected utility, and still not violate constraints i)-iii), which would be a contradiction. Therefore, the solution of the relaxed problem is also the solution of the original problem.

Considering the relaxed version of the problem is useful in that it makes clear that the RP constraint binds:

$$\widehat{d}_1(\theta_{Top}) = \arg\max_{x\in\mathbb{R}} \int_{\theta^*}^{\overline{\theta}} \theta v(x,\theta)\, d\theta$$

and

$$\widehat{d}_0(\theta_{Top}) = \arg\max_{x\in\mathbb{R}} \int_{\theta^*}^{\overline{\theta}} (1-\theta)\, v(x,\theta)\, d\theta$$

This is because having $\widehat{d}_s(\theta_{Top})$ for any s larger than what is sequential optimal for the principal only makes constraint iv' harder to satisfy. Therefore, there is only regret at the bottom.

The rest of the statement of proposition 8 follows by taking the same steps as in the case with commitment power and replacing the incentive compatibility condition iii) by the one that is on the text:

$$\widehat{d}_1(\theta) = \widetilde{u}(\theta) + (1-\theta)\,\widetilde{u}'(\theta) \text{ and } \widehat{d}_0(\theta) = \widetilde{u}(\theta) - \theta\widetilde{u}'(\theta)$$

where $\widetilde{u} : [\underline{\theta}, \theta^*] \to \mathbb{R}$ maximizes the program described on the statement of the proposition.

References

[1] Albuquerque, Rui, & Neng Wang (2008). Agency conflicts, investment, and asset pricing. *The Journal of Finance*, 63(1), 1-40.

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

[3] Ben-Porath, Elchanan, Eddie Dekel & Barton L. Lipman (2014). Optimal allocation with costly verification. *American Economic Review*, 104(12), 3779-3813.

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

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

[6] Bull, Jesse, & Joel Watson (2007). Hard evidence and mechanism design. *Games and Economic Behavio*r, 58(1), 75-93.

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

[8] Crawford, Vincent P., & Joel Sobel (1982). Strategic information transmission. *Econometrica*, 1431-1451.

[9] Dow, James, Gary Gorton & Arvind Krishnamurthy (2005). Equilibrium investment and asset prices under imperfect corporate control. *American Economic Review*, 95(3), 659-681.

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

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

[12] Glazer, Jacob, & Ariel Rubinstein (2004). On optimal rules of persuasion. *Econometrica*, 72(6), 1715-1736.

[13] Glazer, Jacob, & Ariel Rubinstein (2006). A study in the pragmatics of persuasion: A game theoretical approach. *Theoretical Economics*, 395-410.

[14] Green, Jerry R., & Jean-Jacques Laffont (1986). Partially verifiable information and mechanism design. *The Review of Economic Studies*, 53(3), 447-456.

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

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

[17] Hart, Oliver, & Jean Tirole (1988). Contract renegotiation and Coasian dynamics. *The Review of Economic Studies*, 55(4), 509-540.

[18] Hart, Sergiu, Ilan Kremer & Motty Perry (2017). Evidence games: truth and commitment. *The American Economic Review*, 107(3), 690-713.

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

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

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

[22] Malmendier, Ulrike, & Geoffrey Tate (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of financial Economics*, 89(1), 20-43.

[23] Maskin, Eric & John Moore (1999). Implementation and renegotiation. *Review of Economic Studies*, 39-56.

[24] Milgrom, Paul, & Ilya Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2), 583-601.

[25] Myerson, Roger B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1), 58-73.

[26] Mylovanov, Tymofiy, and Andriy Zapechelnyuk (2017). Optimal allocation with ex post verification and limited penalties. *American Economic Review,* 107(9), 2666-94.

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

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

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

[30] Siegel, Ron & Bruno Strulovici (2018). Judicial mechanism design. *Working paper.*

[31] Silva, Francisco (2018). If we confess our sins. *International Economic Review,* forthcoming.

[32] Silva, Francisco (2017). The importance of commitment power in games with imperfect evidence. *Working paper.*

[33] Strulovici, Bruno (2017). Contract Negotiation and the Coase Conjecture. *Econometrica*, 85(2), 585-616.