# Governing Through Communication and Intervention

**Doron Levit***

Wharton

November 28, 2014

**Abstract**

This paper considers a principal-agent environment in which a privately informed principal can communicate with the agent and then intervene if the agent ignores her recommendations. The main result shows that communication is more effective with intervention than without it if and only if the cost that intervention imposes on the agent is large relative to the cost that is incurred by the principal. Thus, intervention can in fact harm communication, echoing the common-wisdom that the capacity to make unilateral decisions can discourage trust and cooperation. The analysis sheds light on the effectiveness of different governance arrangements and provides novel predictions about expected patterns of intervention. Applications of the model to corporate governance (optimal board composition, shareholder activism and venture capital) and regulation are discussed in details.

KEYWORDS: Governance, Communication, Intervention, Cheap-Talk.

JEL CLASSIFICATION: C72 D72, D74, D82, D83, G34

*The author is from the University of Pennsylvania, Wharton School, Finance Department. Email: dlevit@wharton.upenn.edu

# Introduction

In a typical principal-agent scenario, the conflict of interests cannot be easily resolved, not even by contracts. As a partial remedy, the principal may retain the right to intervene and unilaterally change the agent's decisions. Intervention, however, requires a non-trivial amount of effort, time and resources, and therefore, it will often be used as a last resort. Instead, the principal would prefer communicating with the agent and convincing him to voluntarily change his decision. If communication is effective, it can obviate the need for intervention. In turn, the effectiveness of communication depends on the common knowledge that if the agent does not follow the principal's recommendation, the principal has the option to intervene.

By nature, communication and intervention are interrelated. Is communication more effective with intervention than without it? If not, then under what conditions intervention harms communication? Can the principal benefit from a commitment not to intervene in the agent's decision? Answering these questions, which is the main objective of this paper, can explain observed patterns of intervention and shed light on the effectiveness of different governance arrangements. Indeed, communication and intervention are the primary mechanisms of governance in a variety of applications: organizations (managers and their subordinates), corporate finance (corporate boards and CEOs, venture capital funds and entrepreneurs, activist investors and firms), regulation (central bank and financial institutions), politics (party leader and members of the political party), diplomacy (crisis prevention) and households (parents and their children).

To study this topic, I consider a principal-agent model in which the agent decides between two non-contractible actions. The action that maximizes the principal's utility is the private information of the principal. The agent trades off the utility of the principal with the additional private benefits he receives from choosing one of these actions. These private benefits create a conflict of interests between the principal and the agent. Before the agent makes his decision, the principal sends the agent a message about the value of each action. This message can be interpreted as the recommendation of the principal. To capture its informal nature, communication is modeled as a strategic transmission of information à la Crawford and Sobel (1982). Based on the message, the agent decides which of the two actions to implement. In particular, the agent is free to ignore the principal's recommendations. The principal observes the agent's decision, and then decides whether to intervene. Intervention imposes costs on the principal, but also on the agent (e.g., the loss of reputation or compensation). If the principal intervenes, she can choose the action that will eventually be implemented, even if it is different from the agent's initial choice. If the principal does not intervene, then the agent's initial decision is unaffected.

Communication is effective if in equilibrium the probability that the agent voluntarily follows the recommendations of the principal is high. Intervention enhances (harms) communication, if communication is more (less) effective when the principal has the option to intervene than when she does not. The first result shows that intervention enhances communication if the cost that intervention imposes on the agent is large relative to the cost that is incurred by the principal. Intuitively, when the cost of intervention is relatively small, the agent understands that the principal is likely to intervene if he does not follow her recommendation. Since intervention imposes a relatively high cost on the agent, the agent would prefer following the principal's recommendation and avoiding intervention, even at the cost of forgoing his private benefits. The credible threat of intervention benefits the principal since it increases her ability to influence the agent without incurring the cost of intervention.

The complementarity between intervention and communication suggests that the lack of intervention is not necessarily an evidence that intervention is ineffective, quite the opposite. Indeed, when cost of intervention is large, the principal will not intervene even if the agent ignores her recommendations. When cost of intervention is small, the principal does not need to intervene since the threat of intervention is sufficient to convince the agent to follow her recommendations. In between, intervention occurs since it is not credible enough to deter the agent from ignoring the principal's recommendations, but it is sufficiently profitable as a corrective tool. Generally, with communication, there is a non-monotonic relationship between the cost of intervention and the likelihood that the option to intervene is exercised.[1]

The second result is somewhat surprising. I show that intervention *harms* communication if the cost that intervention imposes on the agent is small relative to the cost that is incurred by the principal. Since the agent is less likely to follow the principal's recommendations when the principal has the option to intervene, intervention and communication substitute each other. In fact, the ability of the principal to influence the agent through communication can increase with the cost that the principal has to incur in order to intervene. Generally, the principal

[1] Without communication, the treat of intervention cannot be utilized by the principal to instruct the agent to make the right decision, and therefore, the likelihood of intervention decreases with the cost of intervention.

can benefit from a commitment not to intervene in the agent's decisions. Interestingly, this result holds even though the model has the following three features: hold-up problems are deliberately assumed away, the principal has the option to intervene but she does not have to, and intervention imposes a strictly positive cost on the agent.

How can intervention harm communication? Why does the intuition behind the first result no longer hold? In equilibrium, the principal does not reveal her entire private information to the agent when they communicate. Because of their conflict of interests, the principal deliberately manipulates or conceals information that the agent is likely to abuse. By ignoring the principal's recommendations, the agent can elicit additional information from the principal. Indeed, if the agent ignores the principal's recommendation, the principal has to decide whether to intervene. Intervention is an informed decision, and in equilibrium, the principal intervenes only if she is convinced that in the absence of her intervention the outcome will be detrimental. There are two possibilities. First, if the principal does not intervene, the agent must infer that the principal believes that the initial decision does not justify intervention. These are exactly the states in which the agent prefers consuming his private benefits even at the expense of a lower utility for the principal. In other words, non-intervention "confirms" the decision of the agent to ignore the principal's recommendation. Second, if the principal intervenes, the agent infers that the principal believes that the initial decision is detrimental and therefore justifies intervention. Since the agent is also concerned about the principal's utility, intervention in those cases benefits the agent since it "corrects" his initial decision when it is indeed detrimental.

Altogether, by ignoring the principal's recommendations, the agent forces the principal to make a decision which inevitably reveals information she was trying to conceal. The agent trades off the benefit from eliciting additional information with the direct cost that intervention imposes on him. Importantly, since intervention is costly to the principal, the principal behaves as if she is biased toward the agent's initial decision when she decides whether to intervene. This bias is exploited by the agent. In other words, the possibility of intervention creates additional tension between the principal and the agent by providing the agent with opportunities to challenge the principal to back her words with actions. Through this channel intervention harms communication.

The analysis provides a novel explanation for the common wisdom that trust and cooperation can be built when the power to make unilateral decisions is restrained. Indeed, when intervention harms communication, the principal can benefit from a commitment to reduce her capacity to intervene.[2] This result has important implications. For example, it can rationalize the benefit from flat organizations, hands-off management and lax regulation. It can also explain why corporate boards are often too "friendly" to their CEOs (Adams, Hermalin, and Weisbach (2010)), why venture capital funds might leave the entrepreneur with significant control rights in spite of the entrepreneur's inexperience (Kaplan and Stromberg (2004)), and why some activist investors build reputation for working constructively with management as opposed to being adversarial.[3] Moreover, the characterization of the necessary and sufficient conditions under which intervention harms communication, provides predictions about the circumstances under which the benefit from reducing the capacity to intervene is most pronounced.

Intervention requires confrontation, and hence, it may fail. If intervention fails, the agent may act on the new information he learns. To capture this possibility, I extend the model by allowing the agent to revise his initial decision if intervention fails. With option to revise the initial decision, the agent has fewer incentives to follow the principal's recommendations. Indeed, if the principal intervenes and intervention fails, the agent learns that his initial decision is detrimental. But unlike the baseline model, here the agent can himself correct the decision if needed. At the extreme, if intervention is doomed to failure and it only serves as a costly signaling device by the principal ("burning money"), intervention always harms communication.[4] Since intervention conveys information that cannot be otherwise communicated, this result echoes the common observation that the resolution of a conflict is possible only after confrontation occurs. The paradox is that the anticipation for confrontation makes it harder for information to be transmitted through less costly channels of communication. The analysis also provides predictions on the likelihood of observing voluntary revision of the agent's initial decision upon failed intervention.

The analysis in this paper is related to Aghion and Tirole (1997) who study communication and intervention in the presence of a hold-up problem. They focus on the optimal allocation

[2]The commitment not to intervene can at least partly be achieved by planning an option to exit the relationship with the agent. With the option to exit, the principal has fewer incentives to intervene ex-post.

[3]For example, see the website of ValueAct Capital at http://www.valueact.com/strategy.pl

[4]There is another interesting effect that is discussed in details in Section 3: if the principal expects the agent to reverse his decision when intervention fails, from the principal's perspective, intervention becomes more effective. On the margin, the principal is more likely to intervene if her recommendation is ignored, which may or may not hurt the agent.

of authority within the organization, and distinguish between the concepts of formal and real authority. Different from their study, here there is no hold-up problem, communication is strategic, and intervention is costly. Therefore, the value of a commitment not to intervene in the agent's decision is derived from different economic forces.

Following Crawford and Sobel (1982), the literature has considered many variants of their canonical cheap talk model. Starting with Dessein (2002), several papers have studied the trade-off between delegation and strategic communication (e.g., Adams and Ferreira (2007), Agastya et al. (2011), Chakraborty and Yilmaz (2011), and Harris and Raviv (2005, 2008, 2010)). In these models, the principal never delegates decision rights to the biased agent if the agent has no payoff relevant private information. Here, a commitment not to intervene in the agent's decision can be viewed as a form of delegation. However, the agent has no private information that principal does not have, and hence, the motive for "delegation" is different. Moreover, unlike these studies, here the principal cannot perform the task on her own without incurring the cost of intervention, and hence, the alternative to delegation is first communicating with the agent and then retaining the right to intervene. For all of these reasons, the implications of the analysis in this paper are fundamentally different.

Levit (2013) and Shimizu (2008) study models of strategic communication in which the sender has an option to exit the relationship after observing the receiver's decision. In Levit (2013), exit relaxes the conflict of interests between the sender and the receiver by reducing the sensitivity of the sender's payoff to decisions made by the receiver. In Shimizu (2008), exit imposes a large punishment on the receiver. Both studies conclude that exit can enhance the quality of communication.[5] Importantly, exercising the outside option in these models does not change the receiver's decision once the decision is made. By contrast, here the sender's outside option involves changing (or reversing) the receiver's initial decision after the decision is made. The result that intervention can harm communication relies heavily on this assumption. Therefore, the distinction between intervention and exit is crucial. Related, Levit and Malenko (2011) study whether intervention by an uninformed third party can improve the aggregation of information in a nonbinding voting setup. Importantly, in their setup the party who intervenes has no private information to communicate. Here, the main results rely on the assumption that the party who has private information to communicate, is also the one who decides whether

[5] In Levit (2013), the outside option is determined endogenously by a third party, and hence, exit can harm communication if the sender can manipulate the message that the third party receives.

intervene.

Finally, Austen-Smith and Banks (2000) and Kartik (2007) study the conditions under which burning money improves communication. Here, costly intervention can be considered as a form of burning money only if intervention can never be successful. Moreover, different from their analysis, here intervention takes place after the initial stage of communication and after the principal observes the agent's decision. For this reason, burning money never improves the quality of communication in this paper.

This paper is organized as follows. Section 1 presents the baseline model setup, and Section 2 presents the core analysis. Section 3 analyzes a variant of the model in which the agent can voluntarily revise his decision upon failed intervention. In Section 4, I discuss several applications of the model to corporate governance and regulation, and Section 5 concludes. Appendix A gives all proofs of the main results and the Online Appendix gives all supplemental results not in the main text.

# 1 Setup

Consider a principal-agent model in which the payoffs to the principal and the agent depend on action $a \in \{L, R\}$ and random variable $\theta$.[6] Random variable $\theta$ has a continuous probability density function $f$ with full support over $\left[\underline{\theta}, \overline{\theta}\right]$, where $\underline{\theta} < 0 < \overline{\theta}$. For simplicity, I assume that $f$ is symmetric and $\mathbb{E}[\theta] \geq 0$.[7] If action $a \in \{L, R\}$ is implemented then the principal's payoff is given by

$$v(\theta, a) = \theta \cdot \mathbf{1}_{\{a=R\}}, \tag{1}$$

and the agent's payoff is given by $v(\theta + \beta, a)$. Random variable $\beta$ is privately known to the agent, it is independent of $\theta$, and it has a continuous probability density function $g$ with full support over $[0, \overline{\beta}]$.[8] According to (1), the principal's payoff is maximized when action $R$ is implemented if and only if $\theta > 0$. By contrast, the agent's payoff is maximized when action $R$ is implemented if and only if $\theta > -\beta$. Thus, when $\theta \in (-\beta, 0)$ the principal and

[6] In the Online Appendix, I show that similar results hold when the agent chooses from a continuum of actions.

[7] The main results continue to hold when $f$ is asymmetric.

[8] The main results continue to hold when $\beta$ is a common knowledge. Without this assumption, the analysis in Section 3 must involve mixed strategies.

the agent have different preferences. Effectively, $\beta$ captures the intrinsic conflict of interests between the principal and the agent. The agent trades off the principal's utility with his private benefits from action $R$, and the larger is $\beta$, the larger is the agent's bias toward action $R$. The contractual environment is assumed to be incomplete (e.g., Grossman and Hart (1986) and Hart and Moore (1990)), and in this respect, $\beta$ can be interpreted as the residual conflict of interests between the principal and the agent.[9]

The model has four stages. The first stage involves communication between the principal and the agent. I assume that the principal has information about $\theta$ that the agent does not have. For simplicity, the principal privately observes $\theta$ while the agent is uninformed about $\theta$.[10] Based on her private information, the principal sends the agent message $m \in \left[\underline{\theta}, \overline{\theta}\right]$. The principal's information about $\theta$ is non-verifiable, and the content of $m$ does not affect the agent's or the principal's payoffs directly. These assumptions capture the informal nature of communication. In the Online Appendix, I show that similar results hold if the information about $\theta$ is verifiable. I denote by $\rho(\theta)$ the principal's communication strategy, and by $M \subseteq \left[\underline{\theta}, \overline{\theta}\right]$ the set of messages on the equilibrium path.

In the second stage, the agent observes the message $m$ from the principal, and then chooses between the two actions. I denote by $a_A(m, \beta) \in \{L, R\}$ the decision of the agent conditional on observing message $m$ and his private benefit $\beta$.

The key departure of the model from the existing literature on communication, disclosure and persuasion, is the third stage. In the third stage, the principal observes the agent's decision and then decides whether to intervene. I denote by $e(\theta, a_A) = 1$ the principal's decision to intervene and by $e(\theta, a_A) = 0$ her decision not to intervene. Intervention can either fail or succeed. I denote by $\chi = 1$ the event in which intervention succeeds and by $\chi = 0$ the event in which intervention fails. I assume that intervention is successful with probability $\lambda \in (0, 1]$ and it is independent of $\theta$ and $\beta$. If the principal intervenes, whether or

[9] To fix ideas, the model can be given the following interpretation. If $a = R$ investment is undertaken and the project's return is $\theta$. If $a = L$ the project is abandoned and the payoff is zero. The principal would prefer investing as long as the return is positive. The agent has incentives to over-invest, since investment generates access to perks to the extent that is captured by $\beta$. Other applications of the model are discussed in Section 4.

[10] The principal does not have to be uniformly better informed than the agent. The agent may be responsible for many tasks on which the principal has no informational advantage. The model only requires that ex-post the principal has better information than the agent on at least one task. If it is unknown ex-ante on which task the principal will have information, or if it is inefficient to give decision rights on different tasks to different individuals, it would be ex-ante optimal to allocate all decision rights to the agent.

not intervention is successful, the principal incurs a cost $c_P > 0$. Apart from the effort, time and resources that are needed for intervention, parameter $c_P$ can also capture the principal's alterative cost of dealing with the task or her aversion for confrontation. If the principal does not intervene or if intervention fails, the agent's initial decision $a_A$ is implemented. In Section 3, I consider an extension of the model in which the agent can voluntarily revise his initial decision if intervention fails. If the principal intervenes and intervention succeeds, the principal can reverse the agent's decision or keep it in place. I denote by $a_P(\theta, a_A)$ the principal's decision after successful intervention. Either way, if and only if the principal intervenes and intervention succeeds, the agent incurs a cost $c_A > 0$. This cost can be both pecuniary (e.g., loss of compensation) and non-pecuniary (e.g., damaged reputation, loss of authority and the associated embarrassment with affiliates).[11]

Last, in the final period the payoffs are realized and distributed to the principal and the agent. The principal's utility is given by

$$u_P(\theta, a_A, a_P, e, \chi) = \begin{cases} v(\theta, a_P) - c_P & \text{if } e = \chi = 1 \\ v(\theta, a_A) - e \times c_P & \text{else,} \end{cases} \tag{2}$$

and the agent's utility is given by

$$u_A(\theta, a_A, a_P, e, \chi, \beta) = \begin{cases} v(\theta + \beta, a_P) - c_A & \text{if } e = \chi = 1 \\ v(\theta + \beta, a_A) & \text{else.} \end{cases} \tag{3}$$

The principal and the agent are risk-neutral and their preferences, up to $\theta$ and $\beta$, are common knowledge.

# 2 Analysis

Consider the set of Perfect Bayesian Equilibria of the model. The formal definition is given in the Appendix.

When deciding between the alternatives $L$ and $R$, the agent takes into account the prin-

[11] The main results continue to hold if the agent incurs the cost $c_A$ when intervention fails. This alternative assumption effectively implies that $c_A$ is larger, without further consequences.

cipal's message and her expected response to his decision. Suppose the agent chooses action $R$. If the principal does not intervene, the agent's initial decision is not reversed, and the principal's payoff is $\theta$. Suppose the principal intervenes. Intervention is costly, and hence, the principal intervenes only if she intends to reverse the agent's decision. Since intervention is successful with probability $\lambda$, the principal's expected payoff upon intervention is $(1-\lambda)\,\theta - c_P$. It follows, the principal intervenes when $a_A = R$ if and only if $\theta < -\frac{c_P}{\lambda}$. Suppose the agent chooses action $L$. Similarly, if the principal intervenes her expected payoff is $\lambda\theta - c_P$, and if the principal does not intervene her expected payoff is zero. Therefore, the principal intervenes when $a_A = L$ if and only if $\theta > \frac{c_P}{\lambda}$. Overall, the principal intervenes whenever she finds the agent's initial decision detrimental.[12] The next result summarizes these observations.

**Lemma 1** *In any equilibrium, the principal intervenes if and only if $a_A = R$ and $\theta < -\frac{c_P}{\lambda}$, or $a_A = L$ and $\theta > \frac{c_P}{\lambda}$.*

Given the principal's message and intervention policy, the agent follows a threshold decision rule: he is more likely to choose action $R$ when his private benefit $\beta$ is larger.

**Lemma 2** *In any equilibrium and for any message $m \in M$, there is $b(m)$ such that the agent chooses action $L$ if and only if $\beta \leq b(m)$.*

Communication is effective only if in equilibrium the principal reveals information about $\theta$ and the agent conditions his decision on this information with a positive probability. I refer to equilibria with this property as influential.

**Definition 1** *An equilibrium is influential if there exist $\beta_0 \in \left[0, \overline{\beta}\right]$ and $m_1 \neq m_2 \in M$ such that $\mathbb{E}[\theta|m_1] \neq \mathbb{E}[\theta|m_2]$ and $a_A(m_1, \beta_0) \neq a_A(m_2, \beta_0)$, where $\mathbb{E}[\theta|m]$ is the agent's expectations of $\theta$ conditional on observing message $m$ and unconditional on the principal's decision to intervene.*

When the equilibrium is influential, there are at least two different messages the principal sends the agent with a positive probability, these messages convey different information and trigger different decisions by the agent. By contrast, if the equilibrium is non-influential, the

[12] For this reason, similar results would hold if instead the principal could choose $\lambda$ at a cost of $c_P(\lambda)$, where $c_P' > 0$ and $c_P'' > 0$.

agent ignores all messages from the principal. As in any cheap-talk game, there always exists a non-influential equilibrium. The outcome of a non-influential equilibrium is equivalent to assuming no communication between the principal and the agent. In the absence of effective communication, the agent cannot avoid intervention even if he forges his private benefits and chooses action $L$. Since $\Pr[\beta \geq -\mathbb{E}[\theta]] = 1$, the agent prefers action $R$ over $L$ based on his prior beliefs. Therefore, the agent chooses action $R$ for any $\beta$.

**Proposition 1** *A non-influential equilibrium always exists. In any non-influential equilibrium the agent chooses action $R$ with probability one, and the principal intervenes if and only if $\theta < -\frac{c_P}{\lambda}$.*[13]

According to Definition 1, if communication is effective in equilibrium then the principal can influence the agent's decision by sending the appropriate message. Moreover, according to Lemma 2, if there is a message $m_R$ (message $m_L$) that convinces type $\beta_R$ (type $\beta_L$) to choose action $R$ (action $L$), then the same message convinces all types $\beta > \beta_R$ ($\beta < \beta_L$) to choose action $R$ (action $L$) as well. If the equilibrium is influential then it must be $\min_{m \in M} b(m) < \max_{m \in M} b(m)$. Let

$$\begin{aligned} M_R &\equiv \arg\min_{m \in M} b(m) \\ M_L &\equiv \arg\max_{m \in M} b(m). \end{aligned} \tag{4}$$

Since the principal uses her influence in order to maximize her payoff as given by (2), in any influential equilibrium there are exactly two types of messages: messages that maximize the probability that the agent chooses action $R$ ($m \in M_R$), and messages that maximize the probability that the agent chooses action $L$ ($m \in M_L$). Messages in $M_R$ can be interpreted as recommendations to choose $R$, and messages in $M_L$ can be interpreted as recommendations to choose $L$. Based on (2), the principal recommends on action $R$ if $\theta \geq 0$, and on action $L$ if $\theta < 0$.

**Lemma 3** *In any influential equilibrium, $M_L \cup M_R = M$ and $M_L \cap M_R = \varnothing$. Moreover, if $m \in M_L$ then $\theta < 0$ and if $m \in M_R$ then $\theta \geq 0$.*

[13] When $f$ is asymmetric it is possible to have a non-influential equilibrium in which the agent chooses action $L$. Intuitively, if the prior puts a relatively large weight on low values of $\theta$, the agent believes that intervention is less likely when he chooses action $L$. If $c_A$ is sufficiently high, the agent will forgo his private benefit and choose action $L$ in order to avoid the cost of intervention.

An influential equilibrium exists only if the agent finds it in his best interests to follow the principal's recommendations. Suppose the principal sends the agent a message $m \in M_R$. According to Lemma 3, the agent must infer that $\theta \geq 0$. According to Lemma 1, if the agent follows the principal's recommendation and chooses action $R$, the principal will not intervene, and the agent will be able to consume his private benefits. The agent's expected payoff is

$$\mathbb{E}\left[\theta + \beta | m\right]. \tag{5}$$

If the agent ignores the principal's recommendation and chooses action $L$, the principal will intervene whenever $\theta > \frac{c_P}{\lambda}$. The agent's expected payoff in this case is

$$\lambda \Pr\left[\theta > \frac{c_P}{\lambda} | m\right] \left(\mathbb{E}\left[\theta + \beta | \theta > \frac{c_P}{\lambda}, m\right] - c_A\right). \tag{6}$$

Because of his private benefits, the agent always follows the principal's recommendation to implement action $R$. Indeed, since $\beta \geq 0$ and $c_A > 0$, if $m \in M_R$ then (5) is strictly greater than (6).

The challenge of the principal is convincing the agent to choose action $L$. Suppose the principal sends the agent a message $m \in M_L$. According to Lemma 3, the agent must infer that $\theta < 0$. According to Lemma 1, if the agent follows the principal's recommendation and chooses action $L$, the principal will not intervene and the agent's expected payoff will be zero. If the agent ignores the principal's recommendation and chooses action $R$, the principal will intervene whenever $\theta < -\frac{c_P}{\lambda}$. The agent's expected payoff in this case is

$$\begin{aligned} &\Pr\left[\theta \geq -\frac{c_P}{\lambda} | m\right] \mathbb{E}\left[\theta + \beta | \theta \geq -\frac{c_P}{\lambda}, m\right] \\ &+ \Pr\left[\theta < -\frac{c_P}{\lambda} | m\right] \left((1-\lambda)\, \mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{\lambda}, m\right] - \lambda c_A\right). \end{aligned} \tag{7}$$

Indeed, if $\theta > -\frac{c_P}{\lambda}$ the principal does not intervene and the agent's original decision to implement action $R$ remains intact. If $\theta < -\frac{c_P}{\lambda}$ then the principal intervenes, but the agent's decision is reversed only with probability $\lambda$. The agent follows the principal's recommendation to implement action $L$ if and only if (7) is non-positive.

**Proposition 2** *An influential equilibrium always exists. In any influential equilibrium, the principal recommends on action $R$ if and only if $\theta \geq 0$. If the principal recommends on action*

*R, the agent chooses action R with probability one and the principal never intervenes. If the principal recommends on action L, the agent chooses action L if and only if $\beta \le b^*$, where*

$$b^* = c_A \times \frac{\lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right]}{\Pr[\theta < 0] - \lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right]} + \frac{\lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right] \mathbb{E}\left[\theta | \theta < -\frac{c_P}{\lambda}\right] - \Pr[\theta < 0]\, \mathbb{E}[\theta|\theta < 0]}{\Pr[\theta < 0] - \lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right]}. \tag{8}$$

*If the agent follows the recommendation to choose action L then the principal never intervenes. If the agent ignores the recommendation to choose action L, the principal intervenes if and only if $\theta < -\frac{c_P}{\lambda}$.*

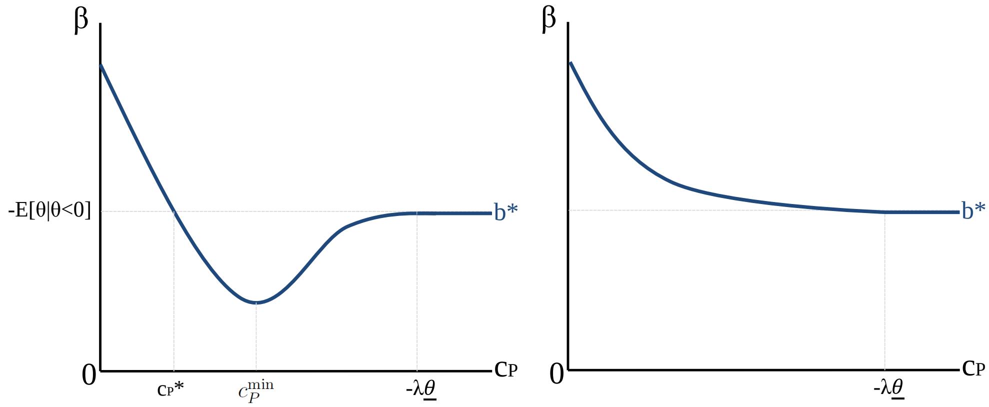
According to Proposition 2, if $\beta > b^*$ the agent prefers ignoring the principal's recommendations even though it increases the likelihood of intervention. By contrast, if $\beta \le b^*$ the agent always follows the principal's recommendation. In those cases, the principal can influence the agent through communication, and intervention is not needed. It follows, if $b^*$ is higher then communication is more effective in equilibrium. In the Online Appendix, I show that influential equilibria Pareto dominate non-influential equilibria. Since influential equilibria always exist, hereafter, I assume that the equilibrium in play is influential.

The comparative statics of $b^*$ follows directly from (8). First, $b^*$ increases with $c_A$. Intuitively, the principal intervenes only if the agent ignores her recommendations. Therefore, in order to avoid the cost of intervention, the agent is more likely to follow the principal's recommendation when $c_A$ is higher. The next result shows that the comparative statics of $b^*$ with respect to $c_P$ is more subtle.

**Corollary 1** *(i) If $c_A \ge \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$ then $b^*$ is non-increasing with $c_P$. (ii) If $c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$ then $b^*$ strictly increases with $c_P$ if and only if $c_p \in \left(c_P^{\min}, -\lambda\underline{\theta}\right)$, where $c_P^{\min} \in (0, -\lambda\underline{\theta})$ is the unique solution of*

$$\frac{c_P^{\min}}{\lambda} = c_A + b^*\left(c_P^{\min}\right). \tag{9}$$

Figure 1 depicts $b^*$ as a function of $c_P$. At any point above the blue curve the agent ignores the principal's recommendation and chooses action $R$ with probability one. At any point below the blue curve the agent follows the principal's recommendation. It can be seen that if $c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$ then $b^*$ is non-monotonic in $c_P$. One might expect that the agent would follow the recommendations of the principal less often when $c_P$ is higher, since in those instances, the threat of intervention is less credible. Corollary 1 shows that this intuition can be misleading. In particular, it is possible that communication becomes more effective as $c_P$ increases. The next section explains the reasoning behind this result.




$c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$ $\qquad$ $c_A \geq \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$

Figure 1

## 2.1 Is communication more effective with intervention?

To understand the interaction between communication and intervention, I consider a benchmark in which the principal cannot intervene by assumption. As in the baseline model, the principal prefers action $L$ over $R$ if and only if $\theta < 0$. Therefore, upon observing a recommendation to choose action $L$, the agent must infer that $\theta < 0$. According to (3), the agent will choose action $L$ if and only if

$$\beta \leq -\mathbb{E}[\theta|\theta < 0]. \tag{10}$$

According to Proposition 2, the same threshold emerges when intervention is prohibitively costly or entirely ineffective, that is, $\frac{c_P}{\lambda} \geq -\underline{\theta}$.

Communication is less effective with intervention than without intervention if and only if $b^* < -\mathbb{E}[\theta|\theta < 0]$. If $b^* < (>) -\mathbb{E}[\theta|\theta < 0]$ then intervention harms (enhances) communication, since the agent is strictly less (more) likely to follow the principal's recommendation with (without) intervention. The next result follows immediately from the comparison between

$-\mathbb{E}[\theta|\theta<0]$ and $b^*$.

**Proposition 3** *Suppose $\frac{c_P}{\lambda} < -\underline{\theta}$. Intervention harms communication if and only if*

$$c_A < \mathbb{E}[\theta|\theta<0] - \mathbb{E}[\theta|\theta<-c_P/\lambda]. \tag{11}$$

Proposition 3 has two interesting implications. The first one is intuitive: when the cost that intervention imposes on the agent is high relative to the cost that is incurred by the principal, the agent is more likely to follow the principal's recommendation when the principal can intervene. That is, intervention enhances communication. The intuition behind this result is similar to the intuition behind the observation that $b^*$ increases with $c_A$. Figure 2 illustrates that when the distribution of $\theta$ is uniform and symmetric around zero, condition (11) becomes $\frac{c_A}{c_P} < \frac{1}{2\lambda}$. The cutoff $c_P^*$ in the left panel of Figure 1 is the unique value of $c_P$ that satisfies condition (11) with equality.

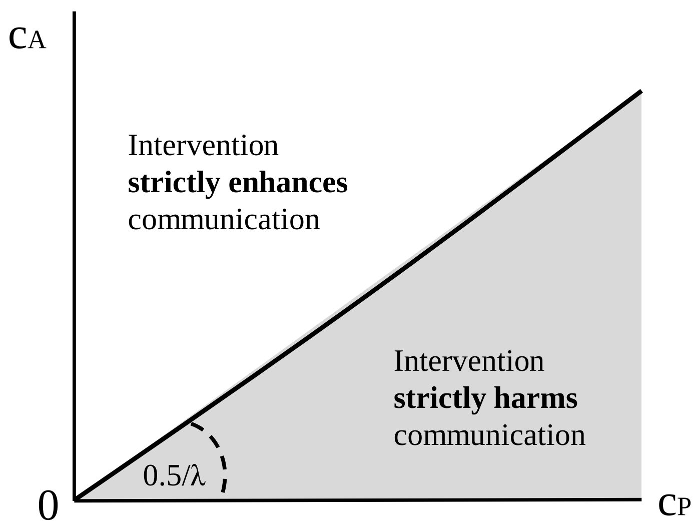



Figure 2

The second implication of Proposition 3 is somewhat surprising. It shows that when condition (11) holds, communication is strictly less effective when the principal can intervene than when intervention is impossible. How can intervention harm communication? To understand the intuition behind this result, note that according to (3), the agent is willing to forgo her private benefits and choose action $L$ if he learns that $\theta$ is sufficiently low. However, the recommendation of the principal does not reveal all information about $\theta$. When the principal recommends the agent to choose action $L$, the agent learns that $\theta < 0$, but he does not know whether $\theta < -\beta$ or $\theta \in [-\beta, 0)$. Indeed, the principal does not reveal the exact value of $\theta$ because if she did, the agent would have chosen action $R$ if $\theta > -\beta$. Instead, by pooling very low realizations of $\theta$ with intermediate realizations of $\theta$, the principal hopes to persuade the agent to choose action $L$ even when $\theta \in [-\beta, 0)$.

The possibility of intervention allows the agent to elicit information that is not otherwise revealed by the principal's recommendations. Since the principal intervenes only if she believes that action $R$ is sufficiently detrimental to justify incurring the costs of intervention, her decision reveals the value of $\theta$ relative to $-\frac{c_P}{\lambda}$. In fact, when deciding whether to intervene, the principal behaves as if she is biased toward action $R$, where the bias is $\frac{c_P}{\lambda}$.[14] The agent exploits this "bias". By ignoring the principal's recommendations, the agent forces her to make an informed decision. In particular, if the principal intervenes then her decision reveals that $\theta < -\frac{c_P}{\lambda}$. The agent learns that ignoring the principal's recommendation to choose action $L$ was a mistake (irrespective of the additional cost $c_A$). In this respect, the principal's decision to intervene "corrects" the agent's initial decision, and benefits him by doing so. At the same time, if the principal does not intervene then her decision reveals that $\theta > -\frac{c_P}{\lambda}$. The agent learns that ignoring the principal's recommendation was justified. Indeed, in these states, the agent prefers consuming his private benefits at the expense of maximizing the principal's utility. In this respect, the principal's decision not to intervene "confirms" the agent's initial decision.

When deciding whether to ignore the principal's decision, the agent also accounts for $c_A$, the direct cost that is imposed by intervention. Given the "correction" and the "confirmation" effects that are discussed above, the agent benefits from the principal's intervention if and only if $\theta < -c_A - \beta$. If $\frac{c_P}{\lambda} \approx c_A + \beta$ then the principal's "bias" coincides with the agent's preferences. As can be seen by (9), the minimum of $b^*$ as a function of $c_P$ is obtained when $\frac{c_P}{\lambda} = c_A + b^*$. Indeed, while the value of the "correction" effect increases with $c_P$, the value of the "confirmation" effect decreases with $c_P$. Overall, when $\frac{c_P}{\lambda} = c_A + b^*$ the agent's benefit from the principal's informed decision whether to intervene is the highest, and hence, the likelihood that the agent follows the principal's recommendations is the lowest. This also explains the

[14] More generally, because intervention is costly, the principal is effectively "biased" against reversing the agent's initial decision.

intuition behind Corollary 1 and the left panel of Figure 1.

To conclude, condition (11) reflects the agent's trade-off between the direct cost from intervention and the benefit from the information in the principal's decision to intervene. Importantly, intervention can harm communication only because intervention is an informed decision. Hypothetically, if the principal could commit to intervening whenever the agent ignores her recommendations, then intervention would necessarily enhance communication. Intuitively, with commitment, the principal's decision to intervene does not depend on $\theta$. Therefore, ignoring the recommendations of the principal imposes a direct cost on the agent without providing him the benefit of correction and confirmation.[15]

## 2.2 Is intervention valuable?

If intervention harms communication then the principal can benefit from a commitment not to intervene in the agent's decision. To see why, note that a direct implication of Proposition 2 is that in any influential equilibrium the principal's expected payoff is given by

$$\begin{aligned} W(c_A, c_P) &= \Pr[\theta > 0]\,\mathbb{E}[\theta|\theta > 0] \\ &\quad + \Pr[\beta < b^*(c_A, c_P)] \begin{bmatrix} \Pr[\theta < 0]\,\mathbb{E}[\theta|\theta < 0] \\ -\Pr\left[\theta < -\frac{c_P}{\lambda}\right]\left(\lambda\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right] + c_P\right)\mathbf{1}_{\{c_P < -\lambda\underline{\theta}\}} \end{bmatrix}. \end{aligned} \tag{12}$$

The first term in (12) is the principal's expected utility from her ability to convince the agent to choose action $R$ when $\theta > 0$. The second term is the principal's expected utility from her attempt to convince the agent to choose action $L$ when $\theta < 0$. The term in the brackets captures the possibility that the principal will choose to intervene when her recommendation is ignored.

Expression (12) has several implications. First, since $b^*$ increases with $c_A$ then $W(c_A, c_P)$ increases with $c_A$ as well. That is, the principal is always better off when intervention imposes a higher cost on the agent. Second, the principal gets the highest payoff when $c_P = 0$, in which case, she can simply enforce or threaten to enforce the optimal strategy. Technologically,

[15] Note that with commitment, the agent's expected payoff from choosing action $R$ when the principal recommends on action $L$ is $-\lambda c_A + (1-\lambda)\,\mathbb{E}[\theta + \beta|\theta < 0]$. Therefore, the agent follows the principal's recommendation if and only if $\beta$ is smaller than $\frac{\lambda}{1-\lambda}c_A - \mathbb{E}[\theta|\theta < 0]$, which is strictly greater than the threshold without intervention.

however, it is unlikely that intervention involves no costs. The next result shows that when intervention is costly, the principal can be better off without the option to intervene.

**Proposition 4** *If and only if* $c_A < \mathbb{E}[\theta|\theta<0]-\underline{\theta}$, *there is* $\overline{c}_P \in (c_P^*, -\lambda\underline{\theta})$ *such that* $W(c_A, c_P) < W(c_A, \infty)$ *for all* $c_P \in (\overline{c}_P, -\lambda\underline{\theta})$.

How can the principal be better off without the option to intervene? This is possible only if intervention harms communication. Based on Proposition 3, it is necessary that $c_P > c_P^*$. If $c_P \in (c_P^*, \overline{c}_P]$ then intervention can be preferred by the principal even though it harms communication. The reason is that intervention can partly substitute for communication. However, as $c_P$ increases, intervention becomes more expansive, and hence, less desirable as a substitute for communication. If $c_P > \overline{c}_P$ the principal is better without the option to intervene. These points are illustrated by the left panel of Figure 3, which plots the principal's expected payoff as a function of $c_P$ when $\lambda = 1$, $\theta \sim U[-1,1]$ and $\beta \sim U[0,1]$. The right panel of Figure 3 shows that a commitment not to intervene in the agent's decision is optimal only if $c_A$ is small relative to $c_P$. In this region, not only intervention harms communication, but it is also ineffective on its own. Therefore, the principal would prefer effective communication over ineffective intervention.

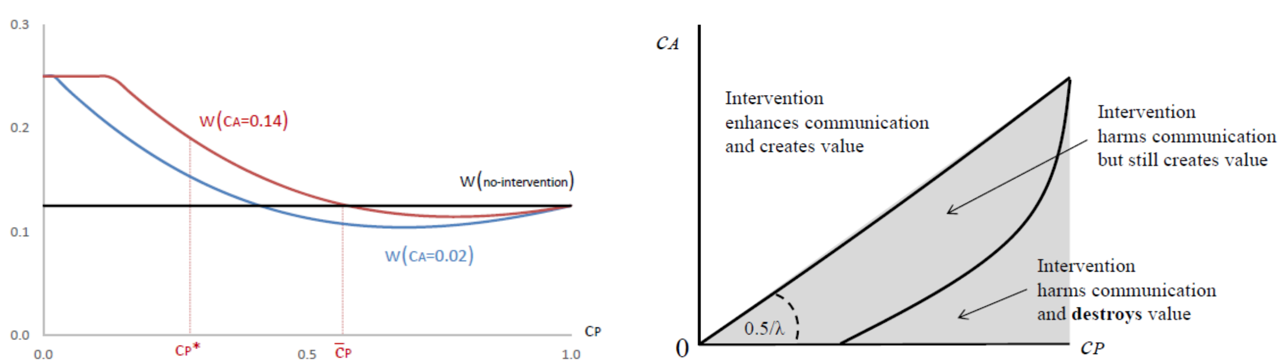



Figure 3

If a commitment not to intervene in the agent's decision is optimal, how can the principal make this commitment credible? In order to have the capacity to intervene, the principal may have to invest resources prior to its interaction with the agent. Therefore, by not making this investment, the principal can commit not to intervene in the agent's decision. Alternatively, the principal can prepare in advance an option to exit her relationship with the agent. Exit provides the principal with an alternative to intervention when the agent ignores her recommendation, and hence, it weakens the principal's incentives to intervene ex-post. Therefore, an option to exit can also be commitment not to intervene in the agent's decision.

## 2.3 When is intervention expected?

In equilibrium, the principal intervenes if the agent ignores her recommendation to choose action $L$, and the benefit from changing the agent's decision from $R$ to $L$ justifies incurring the cost of intervention. The probability of intervention in equilibrium is given by

$$\Pr\left[\beta \geq b^*\right] \times \Pr\left[\theta < -c_P/\lambda\right]. \tag{13}$$

Expression (13) has several implications. First, since $b^*$ increases with $c_A$, the probability of intervention always decreases with $c_A$. Intuitively, when $c_A$ increases the agent has stronger incentives to follow the principals's recommendation in order to avoid the costs that are associated with intervention.[16] Second, note that $c_P$ affects the probability of intervention in two different ways. First, higher $c_P$ reduces the incentives of the principal to intervene since intervention becomes more costly. This force is the reason why the probability of intervention always decreases with $c_P$ when the equilibrium is non-influential. Second, the principal does not need to intervene if the agent follows her recommendation. Therefore, if $b^*$ decreases with $c_P$, the agent is less likely to follow the principal's recommendation, and higher $c_P$ can result with a higher probability of intervention. Generally, the probability of intervention increases with $c_P$ if and only if

$$\frac{1-G\left(b^*\right)}{g\left(b^*\right)} < -\frac{\partial b^*}{\partial c_P}\frac{\lambda F\left(-c_P/\lambda\right)}{f\left(-c_P/\lambda\right)}. \tag{14}$$

Condition (14) demonstrates that the comparative statics of the probability of intervention with respect to $c_P$ can be reversed once communication is considered. When $\frac{1-G(b^*)}{g(b^*)}$ is small, the agent's private benefit is likely to be small. In this range, the principal can effectively influence the agent through communication, and intervention serves only as a threat. Therefore, the probability of intervention is very small. As $c_P$ increases, the threat of intervention becomes

[16] If the equilibrium is non-influential, the agent cannot completely avoid intervention, and hence, the probability of intervention does not necessarily increase with $c_A$.

less credible, and the agent is more likely to ignore the principal's recommendation. Therefore, the principal will have to intervene more often in order to implement action $L$. By contrast, when $\frac{1-G(b^*)}{g(b^*)}$ is large then the agent has large private benefits, and communication is not very effective in the first place. The principal is likely to intervene, and hence, higher $c_P$ results with a lower probability of intervention.

Condition (14) has important implications for empirical work as it suggests that the probability of observed intervention is generally non-monotonic with the cost of intervention. In particular, unobserved intervention is not necessarily an evidence that intervention is ineffective, it may actually suggest that intervention is effective to the extent that its threat is sufficient to induce the agent to follow the recommendations of the principal. Figure 4 illustrates this point by plotting the probability of intervention in equilibrium as a function of $c_P$ when $\lambda = 1$, $\theta \sim U[-1, 1]$ and $\beta \sim U[0, 1]$.

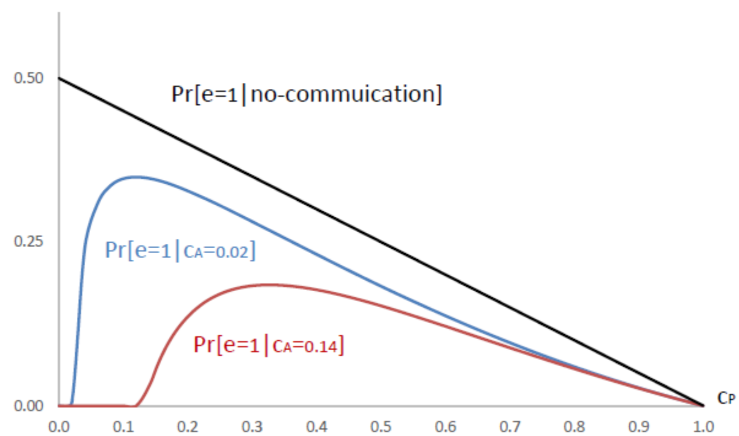



Figure 4

# 3 Learning from (failed) intervention attempts

Since intervention is confrontational by nature, the principal's unilateral attempt to change the outcome can fail ($\lambda < 1$). If intervention fails, the agent may change his initial decision in response to the new information he learns about $\theta$. Therefore, generally, there are two interlarded channels through which information is transmitted: direct communication and signaling through costly intervention. To study the interaction between these two channels, I consider the following addition to the baseline model: I assume that if the principal intervenes but intervention fails ($e = 1$ and $\chi = 0$), the agent has the option to revise his initial decision with no additional costs. I denote the agent's final decision by $a_F \in \{L, R\}$. I assume the agent cannot revise his initial decision if the principal does not intervene. Essentially, the principal can always intervene if she has not done so already. In the Online Appendix, I show that the influential equilibria that exist under this assumption, continue to exist even if it is relaxed.

Consider an influential equilibrium of the modified setup.[17] As in Section 2, the principal recommends on action $R$ if and only if $\theta \geq 0$, the agent always follows a recommendation to implement action $R$, and the principal never intervenes if the agent follows her recommendations. The main departure from the analysis in Section 2 arises when the agent ignores the principal's recommendation to implement action $L$. Let $\mu_R$ be the probability that $a_F = R$ in those circumstances. The principal faces the following trade-off: if the principal intervenes she gets $(1-\lambda)\,\mu_R\theta - c_P$, and otherwise she gets $\theta$. Therefore, the principal intervenes if and only if $\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}$. Note that the principal's decision to intervene depends on $\mu_R$. In particular, lower $\mu_R$ implies that intervention is more likely to succeed, and therefore, the principal has stronger incentives to intervene.

In equilibrium, $\mu_R$ has to reflect the principal's beliefs about the circumstances under which the agent revises his initial decision. These beliefs depend on the agent's initial decision, since this decision is informative about $\beta$. Indeed, the agent's private benefits not only affect his initial decision but also the incentives to revise it. Since the agent expects the principal to intervene whenever $\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}$, the agent revises his decision from $R$ to $L$ if and only if $\beta \leq -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right]$. In equilibrium, the incentives of the agent to revise his initial decision have to be consistent with the principal's beliefs about $\beta$. The next result fully characterizes all influential equilibria.

**Proposition 5** *An influential equilibrium always exists. In any influential equilibrium there are $b^{**} > 0$ and $\mu_R^{**} \in (0, 1]$ such that following hold:*

$(i)$ *The principal recommends on action $R$ if and only if $\theta \geq 0$. If the principal recommends on action $R$, the agent chooses action $R$ with probability one and the principal never*

[17] The complete analysis of non-influential equilibria is given in the Online Appendix. Similar to the analysis in Section 2, if the equilibrium is non-influential then messages from the principal have no effect on the agent's decision, before or after intervention. However, when the agent has the option to revise his initial decision, an equilibrium can be influential even if the agent ignores the message when making his initial decision, but responds to the message when deciding whether to reverse the initial decision. In the Online Appendix, I show this type of equilibria do not exit, and therefore, all influential equilibria must satisfy Definition 1.

*intervenes. If the principal recommends on action $L$, the agent chooses action $L$ if and only if $\beta \leq b^{**}$. If the agent follows the recommendation to choose action $L$, the principal never intervenes. If the agent ignores the recommendation to choose action $L$, the principal intervenes if and only if $\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}}$. Upon failed intervention, the agent revises his initial decision with probability $1-\mu_R^{**}$.*

(ii) *There is $\hat{c}_A > 0$ such that:*

(a) *If $c_A \geq \hat{c}_A$ then $b^{**} = b^*$ and $\mu_R^{**} = 1$.*

(b) *If $c_A < \hat{c}_A$ then*

$$b^{**} = c_A \times \frac{\lambda \Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}}\right]}{\Pr\left[-\frac{c_P}{1-(1-\lambda)\mu_R^{**}} < \theta < 0\right]} - \mathbb{E}\left[\theta | -\frac{c_P}{1-(1-\lambda)\mu_R^{**}} < \theta < 0\right] \quad (15)$$

*and $\mu_R^{**} < 1$, where $\mu_R^{**} = \phi\left(\mu_R^{**}, b^{**}\right)$ and*

$$\phi(x, y) \equiv \Pr\left[\beta > -\mathbb{E}\left[\theta | \theta < -\frac{c_P}{1-(1-\lambda)x}\right] | \beta > y\right]. \quad (16)$$

According to Proposition 5, if $c_A \geq \hat{c}_A$ then $b^{**} = b^*$ and $\mu_R^{**} = 1$. Intuitively, when $c_A$ is large, the agent ignores the principal's recommendation and risks costly intervention only if his private benefits are large. When the agent's private benefits are large, the additional information in the principal's decision to intervene is not sufficient to convince the agent to revise the initial decision and forgo his private benefits. In this range, the equilibrium coincides with the equilibrium in Proposition 2. By contrast, if $c_A < \hat{c}_A$ then learning has a dramatic effect on the nature of equilibrium. In this range, the agent revises his initial decision with positive probability, and the threshold below which the agent follows the recommendation to choose action $L$, as reflected by (15), is generally different from the one in Proposition 2.

**Corollary 2** *There is $c_A^* \in (0, \hat{c}_A]$ such that $c_A \leq c_A^* \Rightarrow b^{**} < b^*$.*

Corollary 2 implies that costly signaling and communication *substitute* each other when $c_A \leq c_A^*$. Intuitively, when $c_A$ is small, the agent will consider ignoring the principal's recommendation and risking costly intervention even if his private benefits from action $R$ are relatively small. Therefore, if the principal intervenes, the agent may find it optimal to reverse his initial decision, as intervention provides a stronger signal that action $L$ is optimal. Since the agent's option to "correct" his initial decision is "in the money", the agent finds intervention to be less costly overall. For this reason, the agent has fewer incentives to follow the principal's recommendation, and communication becomes less effective.[18]

If $c_A \in (c_A^*, \hat{c}_A]$ then it is possible to construct examples in which $b^{**} > b^*$. In those cases, costly signaling and communication *complement* each other. Intuitively, when $\mu_R^{**} < 1$ then all else equal, intervention is more effective from the principal's point of view. Indeed, there is a positive probability that the agent's decision is changed even if intervention fails. Therefore, the principal is more likely to intervene. When $c_A$ is not too low ($c_A > c_A^*$), this effect can increase the incentives of the agent to follow the principal's recommendation in order to avoid the cost of intervention.

In the baseline model, assuming $\lambda = 0$ was equivalent to assuming that the principal cannot intervene. Indeed, since intervention is costly but completely ineffective, the principal never intervenes even though technically she could. Here, however, intervention also signals information about $\theta$ upon which the agent can act. According to Proposition 5, if $\lambda = 0$ then the principal intervenes if and only if $\theta < -\frac{c_P}{1-\mu_R^{**}}$, where $\mu_R^{**} < 1$ and

$$b^{**} = -\mathbb{E}\left[\theta | -\frac{c_P}{1-\mu_R^{**}} < \theta < 0\right].^{19} \tag{17}$$

Note that (17) is smaller than $-\mathbb{E}[\theta|\theta < 0]$, and hence, relative to a benchmark in which intervention is ruled out by assumption (e.g., $c_P \to \infty$), communication is less effective with the option to "burn money". Therefore, intervention is beneficial only if it allows the principal to directly affect the final decision. The signaling role of intervention, on its own, only harms communication. More generally, in the Online Appendix, I show that when $\lambda > 0$ intervention harms communication if and only if $c_P$ is large relative to $c_A$.

[18] Interestingly, when $c_A < c_A^*$ the possibility of voluntarily revision increases the likelihood of intervention in equilibrium. Indeed, voluntarily revision not only increases the effectiveness of intervention from the principal's point of view ($\mu_R^{**} < 1$), but it also decreases the agent's incentives to follow the principal's recommendation ($b^{**} < b^*$).

[19] If $\lambda = 0$ then an influential equilibrium in which the principal never intervenes and $b^{**} = -\mathbb{E}[\theta|\theta < 0]$ always exists. However, in the Online Appendix, I show that this equilibrium does not survive the Grossman and Perry (1986) criterion.

### 3.1 When is voluntary revision expected?

Finally, the model has implications for the likelihood that in equilibrium the agent voluntarily revises his initial decision. Proposition 5 shows that if $c_A \geq \hat{c}_A$ the agent never revises his initial decision, and if $c_A < \hat{c}_A$ the agent revises his decision from $R$ to $L$ if and only if

$$b^{**} \leq \beta < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R^{**}}\right]. \tag{18}$$

The probability that the agent voluntarily revises his decision conditional on failed intervention is $1-\mu_R^{**}$.

**Corollary 3** *The probability of voluntary revision conditional on failed intervention is decreasing with $c_A$.*

Intuitively if $c_A$ is too high, the agent would prefer implementing action $L$ in the first place, instead of choosing $R$, triggering costly intervention, and only then reversing the initial decision if intervention fails.[20]

# 4 Applications

## 4.1 Optimal board structure

In a typical corporation, the CEO runs the company on a daily basis, but the board of directors sets the strategy, approves major decisions, and has the right to replace the CEO. Directors can use their business, legal or finance expertise, in order to advise the CEO. But directors can also force the CEO to comply with the board's policy if needed. Since board intervention requires coordination among directors, the cost of intervention can be proxied by the board size, independence, diversity or busyness. Board intervention also imposes costs on the CEO by either affecting his compensation or reputation, which can be proxied by tenure or age.

[20] The 2014 campaign of the hedge fund activist Daniel S. Loeb against Sotheby's is an example of voluntarily revision. In short, Dan Loeb had been running a proxy fight to get himself and two other nominees on the Sotheby's board of directors. He was hampered in that effort by Sotheby's poison pill, which restricted him from buying more shares. He sued Sotheby's to make it get rid of the pill. He lost that lawsuit, and a couple of days later, Sotheby's caved, agreeing to put Loeb's nominees on the board (see "Dan Loeb and Sotheby's Are Friends Now" BloombergView, May 5 2014.)

The analysis therefore sheds light on the optimal board composition, suggesting that friendly boards, boards that have a low capacity of intervention, can be optimal.

## 4.2 Shareholder activism

In a typical campaign, the activist investor buys a sizeable stake in a public company and then engages with the management or the board of directors, expressing its dissatisfaction or view of how the company should be managed.[21] Occasionally, if the company refuses to comply with the activist's demand, the activist ends up litigating or lunching a proxy fight in order to gain board seats, and thereby, force its ideas on the company. Running a proxy fight requires the activist to reach-out to other shareholders in order to win their vote. The probability of success depends on ownership of other institutional investors, support of proxy advisory firms, and other corporate governance measures of the company. A proxy fight can damage incumbent directors' public image and result with their resignation. The analysis highlights the factors that might affect the success of the campaign. In particular, the analysis suggests that the threat of intervention can be counter-productive. In order to limit her ability to intervene, the activist may consider targeting companies in which coordination with other investors is harder (e.g., dispersed ownership) filing schedule 13-G (instead of schedule 13-D), or build reputation for working constructively with management.

**Shareholder activism and board structure**

The model provides an interesting link between shareholder activism and board structure. Intervention is always optimal ex-post, but as was discussed in details, it may be suboptimal ex-ante. In those cases, we can expect boards to be friendly to CEOs as a form of a commitment not to intervene. Ex-post, activist investors may identify instances in which the CEO ignores the recommendations of the board, and yet the board does not intervene. In those instances, the activist may launch a campaign to replace board members with more aggressive and demanding directors, effectively breaking-up the board's commitment not to intervene. Under this interpretation, shareholder activism can generate positive returns since the ex-post optimal decision is eventually taken. However, since a commitment not to intervene in the CEO's decisions is valuable, shareholder activism can be undesired nevertheless.

[21] Activist investors have a market-wide perspective on assets valuation that corporate boards often lack.

## 4.3 Venture capital

The interaction between entrepreneurs and venture capital investors is another application of the model. The general partners of a venture capital fund often have expertise that complements the technological skills of the entrepreneur\founder. This expertise may be due to the general partners' experience as former executives of companies in related industries or as investors in other ventures. Their expertise may also reflect their relationships with other portfolio companies, investment bankers, lawyers and accountants. The general partners can use their expertise and advise the founder how to smoothly transit the start-up from the development phase to the production phase, recruit key employees, build relationships with costumers and so on. In many cases, the general partners are board members and hold various voting and liquidation rights. If the founder is not focused on profit maximization, the general partners can exercise their power and liquidate the start-up. In this respect, the analysis in this paper suggests that VC funds can benefit from holding fewer control rights.

## 4.4 Regulation

The tension between a regulator (e.g., the central bank) and a regulatee (e.g., financial institutions) often circles around the existence and magnitude of externalities that the regulatee imposes on the economy. Regulators may have information about the extent of these externalities. For example, central banks often have private information about macro economic indicators and the aggregate state of the financial system. The central bank can use this information in order to evaluate the consequences of excessive leverage or risk-taking by financial institutions. The central bank may use its information to propose polices that limit the contribution or exposure of various financial institutions to systematic risk. If these institutions refuse to comply with the proposed polices, the central bank can intervene by forcing changes on their balance sheet (both on the asset side and on the liability side) or trigger changes in senior management. The analysis in this paper suggests that regulator can be more effective if its capacity for intervention is limited. In this respect, lax regulation can be optimal.

# 5 Concluding remarks

In this paper I consider a principal-agent model in which a privately informed principal can communicate with the agent and then intervene if the agent ignores her recommendation. The main result shows that intervention enhances communication if and only if the cost that intervention imposes on the agent is high relative to the cost that is incurred by the principal. In particular, the possibility of intervention creates additional tension between the principal and the agent: it gives the agent the opportunity to challenge the principal to back her words with actions. Thus, somewhat surprisingly, intervention can harm communication, and in this respect, intervention can be counter-productive. This result echoes the common-wisdom that the capacity to make unilateral decisions can discourage trust and cooperation, and it implies that the principal can benefit from a commitment to reduce her capacity to intervene.

The analysis also demonstrates that since intervention conveys information that cannot be otherwise communicated, the resolution of a conflict is possible only after confrontation occurs. Importantly, the anticipation for confrontation makes it harder for information to be transmitted through less costly channels of communication.

The analysis can be applied to organizations, corporate finance, regulation, politics, diplomacy and households. The development of these applications is left for future research.

# References


[1] Adams, R. B. and D. Ferreira, 2007, A theory of friendly boards, *Journal of Finance* 62, 217-250.

[2] Adams, R. B., B. E. Hermalin, and M. S. Weisbach, 2010, The role of boards of directors in corporate governance: A conceptual framework and survey, *Journal of Economic Literature* 48, 58-107.

[3] Agastya M., P.K. Bag and I. Chakraborty (2011), Communication and Authority with a Partially Informed Exper, The RAND Journal of Economics Volume 45, Issue 1, pages 176–197, Spring 2014

[4] Aghion, P., and J. Tirole.1997, Formal and real authority in organizations, *Journal of Political Economy* 105, 1-29.

[5] Austen-Smith, D., and J. S. Banks, 2000, Cheap Talk and Burned Money, *Journal of Economic Theory*, 91(1),1-16.

[6] Chakraborty, A., and B. Yilmaz, 2011, Authority, consensus and governance, Working paper.

[7] Crawford, V. P., and J. Sobel, 1982, Strategic Information Transmission, *Econometrica* 50, 1431-1451.

[8] Dessein, W., 2002, Authority and Communication in Organizations, *Review of Economic Studies,* 69, 811–838.

[9] Grossman, S., and O. Hart, 1986, The costs and benefits of ownership: A theory of vertical and lateral integration, *Journal of Political Economy* 94, 691-719.

[10] Grossman, S. J., and M. Perry 1986, Perfect Sequential Equilibrium, *Journal of Economic Theory* 39 (June 1986): 97-119.

[11] Harris, M. and A. Raviv, 2005, Allocation of decision-making authority, *Review of Finance* 9, 353–383.

[12] Harris, M. and A. Raviv, 2008, A Theory of Board Control and Size, *Review of Financial Studies*, 21, 1797-1832.

[13] Harris, M. and A. Raviv, 2010, Control of corporate decisions: Shareholders vs. management. *Review of Financial Studies* 23:4115–47

[14] Hart, O., and J. Moore, 1990, Property rights and the theory of the firm, *Journal of Political Economy* 98, 1119-1158.

[15] Kaplan, S. and P. Stromberg, 2003, Financial contracting theory meets the real world: An empirical analysis of venture capital contracts, *Review of Economic Studies* 70, 281–315.

[16] Kartik N., 2007, A Note On Cheap Talk and Burned Money, *Journal of Economic Theory*, September 2007.

[17] Levit D., 2013, Soft Shareholder Activism, Working paper.

[18] Levit D., and N. Malenko, 2011, Nonbinding Voting for Shareholder Propoals, *Journal of Finance*, 66(5), 1579-1614.

[19] Shimizu, T., 2008. Cheap talk with the exit option: a model of exit and voice, Working paper.

# A Appendix

This Appendix includes the proofs of the results in the main text. All the supplemental results and extensions of the baseline model are provided in the Online Appendix. I use the following definition for equilibrium in the baseline model.

**Definition 2** *A Perfect Bayesian Equilibrium is a set of messages $M$, the principal's communication strategy $\rho^*(\theta) : [\underline{\theta}, \overline{\theta}] \to [\underline{\theta}, \overline{\theta}]$, the agent's decision making strategy $a_A^*(m, \beta) : [\underline{\theta}, \overline{\theta}] \times [0, \overline{\beta}] \to \{L, R\}$, the principal's intervention strategy $e^*(\theta, a_A) : [\underline{\theta}, \overline{\theta}] \times \{L, R\} \to \{0, 1\}$, and the principal's implementation strategy upon intervention $a_P^*(\theta, a_A) : [\underline{\theta}, \overline{\theta}] \times \{L, R\} \to \{L, R\}$, such that the following conditions are satisfied:*

*(i) For any $\theta$, $\rho^*(\theta) \in \arg\max_{m \in M} \mathbb{E}[u_P(\theta, a_A^*(m, \beta), a_P^*(\theta, a_A), e^*(\theta, a_A), \chi)]$, where the expectations are taken with respect to $\beta$ and $\chi$.*

*(ii) For any $m \in M$, $a_A^*(m, \beta) \in \arg\max_{\hat{a} \in \{L,R\}} \mathbb{E}[u_A(\theta, \hat{a}, a_P^*(\theta, \hat{a}), e^*(\theta, \hat{a}), \chi, \beta) | \rho^*(\theta) = m]$, where the agent's conditional expectation of $\theta$ is consistent with Bayes' rule.*

*(iii) For any $\theta$ and $a_A \in \{L, R\}$, $e^*(\theta, a_A) \in \arg\max_{\hat{e} \in \{0,1\}} \mathbb{E}[u_P(\theta, a_A, a_P^*(\theta, a_A), \hat{e}, \chi)]$ where the expectations are taken with respect to $\chi$.*

*(iv) For any $\theta$ and $a_A$, if $e^*(\theta, a_A) = 1$ then $a_P^*(\theta, a_A) \in \arg\max_{\hat{a} \in \{L,R\}} \mathbb{E}[u_P(\theta, a_A, \hat{a}, 1, \chi)]$ where the expectations are taken with respect to $\chi$.*

*(v) $M = \left\{m \in [\underline{\theta}, \overline{\theta}] : \text{there exists } \theta \in [\underline{\theta}, \overline{\theta}] \text{ such that } \rho^*(\theta) = m\right\}$.*

## A.1 Proofs of Section 2

**Proof of Lemma 2.** Suppose $m_0 \in M$. Based on Lemma 1, $e = 1$ if and only if $a_A = R$ and $\theta < -\frac{c_P}{\lambda}$, or $a_A = L$ and $\theta > \frac{c_P}{\lambda}$. If $a_A = R$ the agent's expected utility is

$$\begin{aligned}
&\Pr\left[\theta \geq -\frac{c_P}{\lambda} | m_0\right] \mathbb{E}\left[\theta + \beta | \theta \geq -\frac{c_P}{\lambda}, m_0\right] \\
&+ \Pr\left[\theta < -\frac{c_P}{\lambda} | m_0\right] \left[(1-\lambda)\,\mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{\lambda}, m_0\right] - \lambda c_A\right],
\end{aligned} \tag{19}$$

and if $a_A = L$, his expected utility is

$$\Pr\left[\theta > \frac{c_P}{\lambda}|m_0\right]\lambda\left(\mathbb{E}\left[\theta+\beta|\theta > \frac{c_P}{\lambda}, m_0\right] - c_A\right). \tag{20}$$

Comparing the two terms and using some algebra, $a_A = R$ if and only if $\beta \geq b(m_0)$, where

$$\begin{aligned} b(m_0) &= c_A \frac{\Pr\left[\theta < -\frac{c_P}{\lambda}|m_0\right] - \Pr\left[\theta > \frac{c_P}{\lambda}|m_0\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}|m_0\right]} \\ &\quad - \frac{\frac{1-\lambda}{\lambda}\mathbb{E}[\theta|m_0] + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}|m_0\right]\mathbb{E}\left[\theta| - \frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}, m_0\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}|m_0\right]}. \end{aligned} \tag{21}$$

■

**Proof of Proposition 1.**

Based on the proof of Lemma 2, given message $m$, the agent chooses action $R$ if and only if $\beta \geq b_m$ where $b_m$ is given by (21). When the equilibrium is non-influential, no message is informative about $\theta$, and hence, for any $m$, $b_m$ can be rewritten as

$$\begin{aligned} b^{NI} &= c_A \times \frac{\Pr\left[\theta < -\frac{c_P}{\lambda}\right] - \Pr\left[\theta > \frac{c_P}{\lambda}\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right]} \\ &\quad - \frac{\frac{1-\lambda}{\lambda}\mathbb{E}[\theta] + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right]\mathbb{E}\left[\theta| - \frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right]}. \end{aligned} \tag{22}$$

Since $f$ is symmetric and $\mathbb{E}[\theta] \geq 0$, then $\mathbb{E}\left[\theta| - \frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right] \geq 0$ and $\Pr\left[\theta < -\frac{c_P}{\lambda}\right] - \Pr\left[\theta > \frac{c_P}{\lambda}\right] \leq 0$. Therefore, $b^{NI} \leq 0$. ■

**Proof of Lemma 3.** According to Lemma 1, $e(\theta, a_A) = 1$ if and only if $a_A = R$ and $\theta < -\frac{c_P}{\lambda}$, or $a_A = L$ and $\theta > \frac{c_P}{\lambda}$. Based on Lemma 2, $\Pr[a_A(\beta, m) = R] = 1 - G(b(m))$, where $b(m)$ is given by (21). Thus, if the principal sends message $m$ then her expected utility is:

$$\mathbb{E}[u_P(\theta, a_A, a_P, e, \chi)|\theta] = \begin{cases} (1-G(b(m)))((1-\lambda)\theta - c_P) & \textit{if } \theta < -\frac{c_P}{\lambda} \\ (1-G(b(m)))\theta & \textit{if } -\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda} \\ (1-G(b(m)))\theta + G(b(m))(\lambda\theta - c_P) & \textit{if } \frac{c_P}{\lambda} < \theta \end{cases} \tag{23}$$

Therefore, if $\theta > 0$ the principal chooses $m \in \arg\min_{m\in M} b(m)$ and if $\theta < 0$ the principal chooses $m \in \arg\max_{m\in M} b(m)$. ■

**Proof of Proposition 2.** Consider any influential equilibrium. According to Lemma 2, for any $m_0 \in M$ the agent chooses action $L$ if and only if $\beta \leq b(m_0)$ where $b(m_0)$ is given by (21). According to Definition 1, $\min_{m\in M} b(m) < \max_{m\in M} b(m)$, and both $M_R$ and $M_L$ are not empty. As was argued in the discussion that preceded Proposition 2, if $m_0 \in M_R$ then $b(m_0) < 0$, that is, the agent follows the principal's recommendation and chooses action $R$ with probability one. Suppose $m_0 \in M_L$. According to Lemma 3, $\theta < 0$. Therefore, (21) can be rewritten as

$$\begin{aligned} b(m_0) \quad = \quad & c_A \times \frac{\lambda \Pr\left[\theta < -\frac{c_P}{\lambda}|m_0\right]}{1-\lambda\Pr\left[\theta < -\frac{c_P}{\lambda}|m_0\right]} \\ & + \frac{\lambda \Pr\left[\theta < -\frac{c_P}{\lambda}|m_0\right]\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}, m_0\right] - \mathbb{E}\left[\theta|m_0\right]}{1-\lambda\Pr\left[\theta < -\frac{c_P}{\lambda}|m_0\right]}. \end{aligned} \tag{24}$$

Since $b(m_0) = \max_{m\in M} b(m)$ for all $m_0 \in M_L$, then $b(m_0)$ is invariant to $m_0$. Since $m_0 \in M_L$ if and only if $\theta < 0$, some algebra and the integration over all $m_0 \in M_L$ show that $b(m_0) = b^*$. Therefore, in any influential equilibrium, the agent follows the principal's recommendation and chooses action $L$ if and only if $\beta \leq b^*$. Note that the intervention policy follows from Lemma 1.

We now show that an influential equilibrium always exists. Consider an equilibrium in which the principal sends message $m_R$ if $\theta \geq 0$ and message $m_L \neq m_R$ otherwise. As was argued in the discussion that preceded Proposition 2, the agent always follows the principal's recommendation if he observes message $m_R$. Since $m = m_L$ if and only if $\theta < 0$, (21) evaluated at $m_L$ can be rewritten as $b^*$. Thus, the agent follows the principal's recommendation to implement action $L$ if and only if $\beta \leq b^*$. Given the agent's expected behavior, it is in the best interest of the principal to follow the proposed communication strategy. Finally, note that

$$\begin{aligned} & \lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right]\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right] - \Pr\left[\theta < 0\right]\mathbb{E}\left[\theta|\theta < 0\right] \\ = \quad & -\lambda\Pr\left[-\frac{c_P}{\lambda} < \theta < 0\right]\mathbb{E}\left[\theta| -\frac{c_P}{\lambda} < \theta < 0\right] - (1-\lambda)\Pr\left[\theta < 0\right]\mathbb{E}\left[\theta|\theta < 0\right] > 0, \end{aligned}$$

and hence, $b^* > 0$. So this equilibrium is indeed influential. ■

**Proof of Corollary 1 .** Consider several properties of $b^*(c_P)$ as a function of $c_P$. First, $b^*(0) = \frac{\lambda}{1-\lambda}c_A - \mathbb{E}[\theta|\theta<0]$. Second, $b^*(c_P) = -\mathbb{E}[\theta|\theta<0]$ for all $c_p \geq -\lambda\underline{\theta}$. Third, if $c_P < -\lambda\underline{\theta}$ then

$$\frac{\partial b^*(c_P)}{\partial c_P} = \frac{f\left(-\frac{c_P}{\lambda}\right)}{F(0) - \lambda F\left(-\frac{c_P}{\lambda}\right)}\left[\frac{c_P}{\lambda} - c_A - b^*(c_P)\right], \tag{25}$$

which implies $\frac{\partial b^*(c_P)}{\partial c_P} > 0 \Leftrightarrow \frac{c_P}{\lambda} - c_A > b^*$. Suppose $c_p < -\lambda\underline{\theta}$ and let $h(c_p) = \frac{c_P}{\lambda} - c_A$. We start by arguing that if there is $c'_P \leq -\lambda\underline{\theta}$ such that $h(c'_P) = b^*(c'_P)$, then $h(c_p) > b^*(c_P)$ for all $c_P > c'_P$. To see why, consider two cases. First, suppose $c'_P = -\lambda\underline{\theta}$. Recall, $c_P \geq -\lambda\underline{\theta} \Rightarrow b^*(c_P) = -\mathbb{E}[\theta|\theta<0]$. Moreover, $h(c_p)$ strictly increases with $c_P$ and $h(-\lambda\underline{\theta}) = -\mathbb{E}[\theta|\theta<0]$. Therefore, $h(c_p) > b^*(c_P)$ for all $c_P > c'_P$. Second, suppose $c'_P < -\lambda\underline{\theta}$. Suppose on the contrary, there is $c''_P > c'_P$ such that $h(c''_P) \leq b^*(c''_P)$. Since $h(c'_P) = b^*(c'_P)$, from continuity, there is $c'''_P \in (c'_P, c''_P]$ and $\varepsilon > 0$ such that $h(c'''_P) = b^*(c'''_P)$ and if $c_P \in (c'''_P - \varepsilon, c'''_P)$ then $h(c_P) > b^*(c_P)$. Based on (25), $\frac{\partial b^*(c_P)}{\partial c_P}|_{c_P = c'''_P} = 0$. Note that $\frac{\partial h(c_P)}{\partial c_P}|_{c_P = c'''_P} = \frac{1}{\lambda} > 0$. Therefore, from continuity, there is $\delta > 0$ such that $h(c_P) < b^*(c_P)$ for all $c_P \in (c'''_P - \delta, c'''_P)$. This contradicts our observation that there is $\varepsilon > 0$ such that $h(c_P) > b^*(c_P)$ for all $c_P \in (c'''_P - \varepsilon, c'''_P)$. We conclude that if there is $c'_P \leq -\lambda\underline{\theta}$ such that $h(c'_P) = b^*(c'_P)$, then $h(c_p) > b^*(c_P)$ for all $c_P > c'_P$.

Next, we argue that if there is $c'_P \leq -\lambda\underline{\theta}$ such that $h(c'_P) = b^*(c'_P)$, then $h(c_p) < b^*(c_P)$ for all $c_P \in [0, c'_P)$. Suppose that, contrary to this statement, there is $c''_P \in [0, c'_P)$ such that $h\left(c''_p\right) \geq b^*(c''_P)$. From continuity, there is $c'''_P \in [c''_P, c'_P)$ such that $h\left(c'''_p\right) = b^*(c'''_P)$. Based on the first step, if $c_P > c'''_P$ then $h(c_p) > b^*(c_P)$. However, since $c'_P > c'''_P$ and $h(c'_P) = b^*(c'_P)$, we get a contradiction. Overall, we conclude that if there is $c'_P \leq -\lambda\underline{\theta}$ such that $h(c'_P) = b^*(c'_P)$, then $h(c_p) > b^*(c_P)$ for all $c_P > c'_P$, and $h(c_p) < b^*(c_P)$ for all $c_P < c'_P$. Based on (25), if there is $c'_P \leq -\lambda\underline{\theta}$ such that $h(c'_P) = b^*(c'_P)$, then $\frac{\partial b^*(c_P)}{\partial c_P} > 0$ for all $c_P \in (c'_P, -\lambda\underline{\theta})$, and $\frac{\partial b^*(c_P)}{\partial c_P} < 0$ for all $c_P < c'_P$.

We proceed in two steps. First, suppose $c_A \geq \mathbb{E}[\theta|\theta<0] - \underline{\theta}$, which implies $h(-\lambda\underline{\theta}) \leq b^*(-\lambda\underline{\theta})$. We argue that $h(c_p) < b^*(c_P)$ for all $c_P \in [0, -\lambda\underline{\theta})$. Suppose on the contrary there is $\hat{c}_p \in [0, -\lambda\underline{\theta})$ such that $h(c_p) \geq b^*(c_P)$. Note that $h(0) < b^*(0)$. From the continuity of $h$ and

$b^*$, there is $c'_P \in (0, \hat{c}_p]$ such that $h(c'_P) = b^*(c'_P)$. Based on our claim above, $h(c_p) > b^*(c_P)$ for all $c_P > c'_P$. Since $c'_P \leq \hat{c}_p < -\lambda\underline{\theta}$, then $h(-\lambda\underline{\theta}) > b^*(-\lambda\underline{\theta})$, a contradiction.

Second, suppose $c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$, which implies $h(-\lambda\underline{\theta}) > b^*(-\lambda\underline{\theta})$. Since $h(0) < b^*(0)$, by the intermediate value theorem, there is $c^*_P \in (0, -\lambda\underline{\theta}]$ such that $h(c^*_P) = b^*(c^*_P)$. Based on our claim above, $c^*_P$ is unique, and $\frac{\partial b^*(c_P)}{\partial c_P} < 0$ for all $c_P < c'_P$, and $\frac{\partial b^*(c_P)}{\partial c_P} > 0$ for all $c_P \in (c'_P, -\lambda\underline{\theta})$, as required. ■

**Proof of Proposition 4.** Recall that if $c_P \geq -\lambda\underline{\theta}$ then $b^* = -\mathbb{E}[\theta|\theta < 0]$. Suppose $c_P < -\lambda\underline{\theta}$. Based on (12),

$$\begin{aligned} \frac{\partial W}{\partial c_P} &= -\frac{\partial b^*}{\partial c_P} g(b^*) \begin{pmatrix} \Pr\left[-\frac{c_P}{\lambda} < \theta < 0\right] \mathbb{E}\left[\theta | -\frac{c_P}{\lambda} < \theta < 0\right] \\ + \Pr\left[\theta < -\frac{c_P}{\lambda}\right] \left((1-\lambda) \mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right] - c_P\right) \end{pmatrix} \\ &\quad -F\left(-\frac{c_P}{\lambda}\right)(1 - G(b^*)). \end{aligned}$$

Therefore, if $\frac{\partial b^*}{\partial c_P} \leq 0$ then $\frac{\partial W}{\partial c_P} < 0$. Note that based on Corollary 1, if $c_A \geq \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$ then $\frac{\partial b^*}{\partial c_P} \leq 0$. Suppose $c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$. Based on (12), $W(c_P) < W_0$ if and only if

$$\lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right] \frac{\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right] + \frac{c_P}{\lambda}}{\Pr[\theta < 0]\,\mathbb{E}[\theta|\theta < 0]} < \frac{G(-\mathbb{E}[\theta|\theta < 0]) - G(b^*)}{1 - G(b^*)}. \tag{26}$$

The derivative of the LHS of (26) with respect to $c_P$ is $\frac{\Pr[\theta < -c_P/\lambda]}{\Pr[\theta<0]\mathbb{E}[\theta|\theta<0]} < 0$, and the derivative of the RHS of (26) with respect to $c_P$ is $-g(b^*)\frac{\partial b^*}{\partial c_P}\frac{1-G(-\mathbb{E}[\theta|\theta<0])}{(1-G(b^*))^2}$. Using (25), this derivative can be rewritten as

$$-g(b^*)\frac{f\left(-\frac{c_P}{\lambda}\right)}{F(0) - \lambda F\left(-\frac{c_P}{\lambda}\right)}\left[\frac{c_P}{\lambda} - c_A - b^*(c_P)\right]\frac{1 - G(-\mathbb{E}[\theta|\theta < 0])}{(1 - G(b^*))^2}, \tag{27}$$

and note that it is negative if and only if $c_P > c^*_{\min}$. Also, note that as $c_P \to -\lambda\underline{\theta}$ both the LHS of (26) and its derivative converge to zero, while the RHS of (26) converges to zero and its derivative converges to

$$-f(\underline{\theta})\frac{\mathbb{E}[\theta|\theta < 0] - \underline{\theta} - c_A}{F(0)}\frac{g(-\mathbb{E}[\theta|\theta < 0])}{1 - G(-\mathbb{E}[\theta|\theta < 0])}. \tag{28}$$

Since $c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$ and $f(\underline{\theta}) > 0$, (28) is strictly negative. Therefore, there is $\varepsilon > 0$ such that if $c_P \in (-\lambda\underline{\theta} - \varepsilon, -\lambda\underline{\theta})$ then (26) holds. ■

## A.2 Proofs of Section 3

Definition 3 extends the concept of influential equilibrium to Section 3.

**Definition 3** *An equilibrium is influential if there exist $m_1 \neq m_2 \in M$ and $\beta_0 \in [0, \overline{\beta}]$ such that $\mathbb{E}[\theta|m_1] \neq \mathbb{E}[\theta|m_2]$ and either $a_A(m_1, \beta_0) \neq a_A(m_2, \beta_0)$ or $a_F(m_1, \beta_0) \neq a_F(m_2, \beta_0)$.*

**Proof of Proposition 5.** Lemma 4 in the Online Appendix shows that if the equilibrium is influential according to Definition 3, then it is also influential according to Definition 1. Suppose an influential equilibrium exists. The principal sends message $m \in M_R$ if and only if $\theta > 0$. Suppose $m \in M_R$. If $a_A = R$ then the principal never intervenes and the agent gets $\mathbb{E}[\theta + \beta|m] > 0$. If $a_A = L$ and $e = 0$ then action $L$ is implemented and the agent gets zero. If $a_A = L$ and $e = 1$ action $R$ is implemented whether or not intervention succeeds. Indeed, since $m \in M_R \Rightarrow \theta > 0$, the agent has incentives to reverse his initial decision. However, if intervention succeeds, the agent incurs a cost $c_A > 0$. We conclude, if $m \in M_R$ then the agent chooses $R$ with probability one, and the principal never intervenes.

Suppose $m \in M_L$. If $a_A = L$ then the principal never intervenes and the agent gets zero. Suppose $a_A = R$. As was argued in the main text, the principal intervenes if and only if $\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}$. Note that if the agent chooses action $R$ despite $m \in M_L$, the principal benefits if the agent believes that $\theta$ is as small as possible. The reason is that if the agent believes that $\theta$ is small, he is more likely to choose action $L$ if the principal intervenes and intervention fails. At the same time, if the agent is more likely to choose $L$ upon failed intervention, it means that intervention is more effective, which increases the expected payoff of the principal. Thus, ex-ante, when the principal sends the message, he has (weak) incentives to signal that $\theta$ is low as possible. For this reason, conditional on $\theta < 0$, the only information about $\theta$ that can be revealed in equilibrium is that $\theta$ is negative. Note that the agent keeps his initial decision $a_A = R$ upon failed intervention if and only if $\mathbb{E}\left[\theta + \beta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] \geq 0$. Overall, if $m \in M_L$ and $a_A = R$ then the agent's expected payoff is

$$\begin{aligned}&\Pr\left[\theta > -\frac{c_P}{1-(1-\lambda)\mu_R}|\theta<0\right]\mathbb{E}\left[\theta+\beta|-\frac{c_P}{1-(1-\lambda)\mu_R}<\theta<0\right] \\ &+\Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}|\theta<0\right]\left((1-\lambda)\max\left\{0,\mathbb{E}\left[\theta+\beta|\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]\right\}-\lambda c_A\right).\end{aligned} \tag{29}$$

Therefore, $a_A = L$ if and only if the expression above is negative, which holds if and only if

$$\beta \leq \min\left\{-\mathbb{E}\left[\theta|\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right],\eta(\mu_R)\right\}, \tag{30}$$

or

$$-\mathbb{E}\left[\theta|\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]<\beta\leq\delta(\mu_R). \tag{31}$$

where

$$\eta(\mu_R)=\lambda c_A\frac{\Pr\left[\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr\left[-\frac{c_P}{1-(1-\lambda)\mu_R}<\theta<0\right]}-\mathbb{E}\left[\theta|-\frac{c_P}{1-(1-\lambda)\mu_R}<\theta<0\right], \tag{32}$$

and

$$\begin{aligned}\delta(\mu_R) \;=\;& \lambda c_A\frac{\Pr\left[\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr[\theta<0]-\lambda\Pr\left[\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]} \\ &-\frac{\Pr[\theta<0]\,\mathbb{E}[\theta|\theta<0]-\lambda\Pr\left[\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]\mathbb{E}\left[\theta|\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr[\theta<0]-\lambda\Pr\left[\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]}.\end{aligned} \tag{33}$$

Define

$$h(\mu_R)\equiv\frac{\Pr[\theta<0]}{\lambda}\frac{\mathbb{E}[\theta|\theta<0]-\mathbb{E}\left[\theta|\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr\left[\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right]},$$

and note that $h(\mu_R)$ is increasing in $\mu_R$. It can be verified that if $c_A < h(\mu_R)$ then

$$\eta(\mu_R)<\delta(\mu_R)<-\mathbb{E}\left[\theta|\theta<-\frac{c_P}{1-(1-\lambda)\mu_R}\right],$$

and if $c_A > h(\mu_R)$ then

$$-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] < \delta(\mu_R) < \eta(\mu_R).$$

Based on (30) and (31) there are two cases to consider. First, suppose $c_A \geq h(1)$. Since $h(\cdot)$ is an increasing function, then $c_A \geq h(\mu_R)$ for all $\mu_R \in [0,1]$. Therefore, in equilibrium, it must be $a_A = L$ if and only if $\beta \leq \delta(\mu_R)$. If the agent ignores a recommendation to choose $L$, then $\delta(\mu_R) < \beta$. Since $c_A \geq h(\mu_R)$ implies $-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] \leq \delta(\mu_R)$ then it must be $\mu_R = 1$ in equilibrium. In this case, $b^{**} = b^* = \delta(1) > 0$ and $\mu_R^{**} = 1$.

Second, suppose $c_A < h(1)$. Suppose on the contrary that in equilibrium $\mu_R^{**}$ satisfies $c_A \geq h(\mu_R^{**})$. In that case, it must be $a_A = L$ if and only if $\beta \leq \delta(\mu_R^{**})$, and similar to the argument in the previous case, $\mu_R^{**} = 1$. Since $c_A < h(1)$ it must be a contradiction. Therefore, in equilibrium $\mu_R^{**} < 1$ satisfies $c_A < h(\mu_R^{**})$. In this case, $a_A = L$ if and only if $\beta \leq \eta(\mu_R^{**})$. Therefore, $b^{**} = \eta(\mu_R^{**})$. Note that $c_A < h(\mu_R^{**})$ implies $\eta(\mu_R^{**}) < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}}\right]$, and hence, $\mu_R^{**}$ must solve $\mu_R^{**} = \phi(\mu_R^{**}, \eta(\mu_R^{**}))$. It is left to show that the equation $y = \phi(y, \eta(y))$ has a solution in $(0,1)$ that satisfies $c_A < h(y)$. Since $c_A < h(1)$ then $\eta(1) < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right]$, and hence, $\phi(1, \eta(1)) < 1$. Let $\underline{y} = \max\{0, h^{-1}(c_A)\}$, and note that $0 \leq \underline{y} < 1$. If $\underline{y} = h^{-1}(c_A)$ then $\eta(\underline{y}) = -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\underline{y}}\right]$, and hence, $\phi(\underline{y}, \eta(\underline{y})) = 1$. If $\underline{y} = 0$ then $\eta(\underline{y}) < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\underline{y}}\right]$, and hence, $\phi(\underline{y}, \eta(\underline{y})) > 0$. Either way, there is $\mu_R^{**} \in (\underline{y}, 1)$ such that $\phi(\mu_R^{**}, \eta(\mu_R^{**})) = \mu_R^{**}$, where $c_A < h(\mu_R^{**})$. Letting $\hat{c}_A = h(1)$ concludes the proof. ■

**Proof of Corollary 2.** I prove that there is $c_A^* \in (0, h(1)]$ such that $c_A \leq c_A^* \Rightarrow \eta(\mu_R^{**}(c_A), c_A) < \delta(1, c_A)$. Based on (32) and (33), $\eta(\mu_R, c_A)$ and $\delta(\mu_R, c_A)$ are linear in $c_A$, $\frac{\partial \eta(\mu_R, c_A)}{\partial c_A} > \frac{\partial \delta(\mu_R, c_A)}{\partial c_A} > \frac{\partial \delta(1, c_A)}{\partial c_A}$ for all $c_A$ and $\mu_R$, and $\eta(\mu_R, 0) < \delta(\mu_R, 0)$ for all $\mu_R$. Therefore, $\eta(1,0) < \delta(1,0)$. Moreover, note that $\eta(\mu_R, 0)$ increases with $\mu_R$. Therefore, if $\mu_R < 1$ then $\eta(\mu_R, 0) < \eta(1,0) < \delta(1,0)$. Since $h(1) > 0$, then $\mu_R^{**}(0) < 1$, and hence, $\eta(\mu_R^{**}(0), 0) < \delta(1,0)$. From continuity, there is $c_A^* \in (0, h(1)]$ such that, if $c_A \leq c_A^*$ then $\eta(\mu_R^{**}(c_A), c_A) < \delta(1, c_A)$, as required. Note that for any $c_A \leq h(1)$, $b^{**} = \eta(\mu_R^{**}(c_A), c_A)$ and $b^* = \delta(1, c_A)$. ■

**Proof of Corollary 3.** Based on Proposition 5, if $c_A \geq \hat{c}_A$ then $\mu_R^{**} = 1$ and the agent never reverses his initial decision. If $c_A < \hat{c}_A$ then $\mu_R^{**} = \phi\left(\mu_R^{**}, \eta\left(\mu_R^{**}, c_A\right)\right)$. An application of the implicit function theorem implies

$$\frac{\partial \mu_R^{**}}{\partial c_A} = \frac{\phi\left(\mu_R^{**}, \eta\left(\mu_R^{**}, c_A\right)\right) \frac{g\left(\eta\left(\mu_R^{**}, c_A\right)\right)}{1-G\left(\eta\left(\mu_R^{**}, c_A\right)\right)}}{1 - \frac{\phi\left(\mu_R^{**}, \eta\left(\mu_R^{**}, c_A\right)\right)}{\partial \mu_R}} \frac{\partial \eta\left(\mu_R^{**}, c_A\right)}{\partial c_A}. \tag{34}$$

We focus on stable equilibria, that is, we focus on solutions of $\mu_R^{**} = \phi\left(\mu_R^{**}, \eta\left(\mu_R^{**}, c_A\right)\right)$ where the function $\phi\left(\mu_R, \eta\left(\mu_R, c_A\right)\right)$ intersects with the 45 degrees line from above. Based on the proof of Proposition 5, such an intersection point always exists. Therefore, $1 - \frac{\phi\left(\mu_R^{**}, \eta\left(\mu_R^{**}, c_A\right)\right)}{\partial \mu_R} > 0$. Moreover, note that $\frac{\partial \eta\left(\mu_R^{**}, c_A\right)}{\partial c_A} > 0$. Therefore, $\frac{\partial \mu_R^{**}}{\partial c_A} > 0$. ■

# Online Appendix for "Governing Through Communication and Intervention"

Doron Levit

The Online Appendix has two sections. The first section contains supplemental results, and the second section contains two extensions to the baseline model: continuum of actions and verifiable information.

## A Supplemental results

**Proposition 6** *Every influential equilibrium Pareto dominates every non-influential equilibrium.*

**Proof.** According to Proposition 1, the principal's expected payoffs is

$$W^{Non-Influential} = \Pr\left[\theta > -\frac{c_P}{\lambda}\right] \mathbb{E}\left[\theta|\theta > -\frac{c_P}{\lambda}\right] + \Pr\left[\theta < -\frac{c_P}{\lambda}\right]\left(-c_P + (1-\lambda)\,\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right]\right), \tag{35}$$

and the agent's expected payoff (conditional on $\beta$) is given by

$$U_A^{Non-Influential}(\beta) = \mathbb{E}[\theta+\beta] - \lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right]\left(\mathbb{E}\left[\theta+\beta|\theta < -\frac{c_P}{\lambda}\right] + c_A\right) \tag{36}$$

According to Proposition 2, the principal's expected payoff, $W^{Influential}$, is given by (12), and the agent's expected payoff (conditional on $\beta$) is given by

$$U_A^{Influential}(\beta) = \begin{cases} \Pr[\theta > 0]\,\mathbb{E}[\theta+\beta|\theta > 0] & \textit{if } \beta \leq b^* \\ \mathbb{E}[\theta+\beta] - \lambda \Pr\left[\theta < -\frac{c_P}{\lambda}\right]\left(\mathbb{E}\left[\theta+\beta|\theta < -\frac{c_P}{\lambda}\right] + c_A\right) & \textit{else.} \end{cases} \tag{37}$$

A direct comparison shows that $U_A^{Influential}((\beta)) \geq U_A^{Non-Influential}(\beta)$ for all $\beta$, and $W^{Influential} \geq W^{Non-Influential}$ ■

**Proposition 7** *A non-influential equilibrium always exists. In any non-influential equilibrium there are* $b_S^{NI}$ *and* $\mu_R^{NI} \in (0,1)$ *such that following hold:*

(*i*) *The agent chooses action* $L$ *if and only if* $\beta \leq b_S^{NI}$.

(*ii*) *If the agent chooses action* $L$ *then the principal intervenes if and only if* $\theta > c_P$, *and upon failed intervention the agent voluntarily implements action* $R$ *with probability one.*

(*iii*) *If the agent chooses action* $R$ *then the principal intervenes if and only if* $\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{NI}}$, *and upon failed intervention the agent voluntarily implements action* $L$ *if and only if* $b_S^{NI} \leq \beta \leq -\mathbb{E}\left[\theta | \theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{NI}}\right]$ *where* $\mu_R^{NI}$ *solves* $\mu_R^{NI} = \phi\left(\mu_R^{NI}, b_S^{NI}\right)$.

**Proof.** I denote by $\mu_a$ the probability that the agent chooses action $R$ when intervention fails, conditional on his initial decision to choose action $a \in \{L, R\}$. First note that based on (3), if intervention fails, the agent prefers action $R$ over action $L$ if and only if

$$\beta \geq -\mathbb{E}\left[\theta | e\left(\theta, a_A\right) = 1\right]. \tag{38}$$

Suppose $a_A = L$. The principal gets $\lambda\theta + (1-\lambda)\mu_L\theta - c_P$ if she intervenes, and zero otherwise. Therefore, the principal intervenes if and only if $\theta > \frac{c_P}{\lambda+(1-\lambda)\mu_L}$. Since $\mathbb{E}\left[\theta + \beta | \theta > \frac{c_P}{\lambda+(1-\lambda)\mu_L}\right] > 0$ for all $\beta > 0$ and $\mu_L \in [0,1]$, based on (38), if $e = 1$ and $\chi = 0$ then the agent reverses his decision from $L$ to $R$ with probability one. Therefore, in any equilibrium it must be $\mu_L = 1$. Therefore, if $a_A = L$ then the agent's expected utility can be rewritten as

$$\Pr\left[\theta > c_P\right]\left(\mathbb{E}\left[\theta + \beta | \theta > c_P\right] - \lambda c_A\right). \tag{39}$$

Suppose $a_A = R$. The principal gets $(1-\lambda)\mu_R\theta - c_P$ if she intervenes, and $\theta$ otherwise. Therefore, the principal intervenes if and only if $\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}$. Based on (38), if $e = 1$ and $\chi = 0$ then the agent keeps his initial decision in place if and only if $\mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] \geq$

0. The agent's expected utility is:

$$\Pr\left[\theta \geq -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right]\mathbb{E}\left[\theta+\beta|\theta \geq -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right] \tag{40}$$
$$+\Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right]\left(-\lambda c_A + (1-\lambda)\max\left\{0, \mathbb{E}\left[\theta+\beta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right]\right\}\right).$$

Comparing (40) and (39), the agent chooses $a_A = L$ if and only if,

$$\beta \leq \min\left\{-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right], \eta\,(\mu_R)\right\} \tag{41}$$

or

$$-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right] < \beta \leq \delta\,(\mu_R)\,, \tag{42}$$

where

$$\eta\,(\mu_R) = \lambda c_A \frac{\Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\mu_R}\right] - \Pr\left[\theta > c_P\right]}{\Pr\left[-\frac{c_P}{1-(1-\lambda)\mu_R} < \theta < c_P\right]} - \mathbb{E}\left[\theta| - \frac{c_P}{1-(1-\lambda)\,\mu_R} < \theta < c_P\right]$$

$$\delta\,(\mu_R) = \lambda c_A \frac{\Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\mu_R}\right] - \Pr\left[\theta > c_P\right]}{\Pr\left[\theta < c_P\right] - \lambda\Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}$$
$$+\frac{\lambda\Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\mu_R}\right]\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] - \Pr\left[\theta < c_P\right]\mathbb{E}\left[\theta|\theta < c_P\right]}{\Pr\left[\theta < c_P\right] - \lambda\Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}.$$

Consider the following condition,

$$\lambda c_A \frac{\Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\mu_R}\right] - \Pr\left[\theta > c_P\right]}{\Pr\left[\theta < c_P\right]} > \mathbb{E}\left[\theta|\theta < c_P\right] - \mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right]. \tag{43}$$

It can be verified that if (43) holds then

$$-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right] < \eta\,(\mu_R) < \delta\,(\mu_R)\,, \tag{44}$$

and otherwise,

$$\delta\left(\mu_R\right) < \eta\left(\mu_R\right) < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right]. \tag{45}$$

Define

$$\begin{aligned} l\left(x\right) &= \Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\,x}\right] - \Pr\left[\theta > c_P\right] \\ h\left(x\right) &\equiv \frac{\Pr\left[\theta < c_P\right]}{\lambda}\frac{\mathbb{E}\left[\theta|\theta < c_P\right] - \mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)x}\right]}{l\left(x\right)}. \end{aligned}$$

Condition (43) holds if and only if

$$l\left(\mu_R\right) > 0 \text{ and } c_A > h\left(\mu_R\right).$$

Note that $l\left(x\right)$ decreases in $x$ and $h\left(x\right)$ increases in $x$, and hence, (43) holds if and only if

$$\mu_R < \overline{\mu} \equiv \min\left\{l^{-1}\left(0\right), h^{-1}\left(c_A\right)\right\}.$$

Note that $\overline{\mu}$ can be greater than one or smaller than zero. Based on (41) and (42):

- If $\mu_R \leq \overline{\mu}$ then (43) holds the agent chooses $a_A = L$ if and only if $\beta \leq \delta\left(\mu_R\right)$. Note that $-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] \leq \delta\left(\mu_R\right)$ implies $\mu_R = 1$. Therefore, it has to be $\overline{\mu} \geq 1$.
- If $\mu_R > \overline{\mu}$ then (43) is violated and the agent chooses $a_A = L$ if and only if $\beta \leq \eta\left(\mu_R\right)$. Since $\eta\left(\mu_R\right) < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right]$ then $\mu_R$ must also solve $\mu_R = \phi\left(\mu_R, \eta\left(\mu_R\right)\right)$. Therefore, it has to be $\overline{\mu} < 1$. Suppose $\overline{\mu} < 1$. Then, $\eta\left(1\right) < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right]$, and hence, $\phi\left(1, \eta\left(1\right)\right) < 1$. If $\mu_R = \overline{\mu} \geq 0$ then $\eta\left(\overline{\mu}\right) = -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\overline{\mu}}\right]$, and hence, $\phi\left(\overline{\mu}, \eta\left(\overline{\mu}\right)\right) = 1$. If $\mu_R = 0 > \overline{\mu}$ then $\eta\left(0\right) < -\mathbb{E}\left[\theta|\theta < -c_P\right]$, and hence, $\phi\left(0, \eta\left(0\right)\right) > 0$. Either way, if $\overline{\mu} < 1$ then $\mu_R = \phi\left(\mu_R, \eta\left(\mu_R\right)\right)$ has a solution in $\left[\max\left\{0, \overline{\mu}\right\}, 1\right]$.

We conclude, there are two cases:

1. If $\overline{\mu} \geq 1$ then the agent chooses $a_A = L$ if and only if $\beta \leq \delta\left(1\right)$, where $\mu_R^{NI} = 1$. Note that $b_S^{NI} \equiv \delta\left(1\right) > 0$.

2. If $\overline{\mu} < 1$ then the agent chooses $a_A = L$ if and only if $\beta \leq \eta(\mu_R^*)$, where $1 > \mu_R^{NI} > \max\{0, \overline{\mu}\}$ and $\mu_R^{NI}$ solves $\mu_R = \phi(\mu_R, \eta(\mu_R))$. Note that $b_S^{NI} = \eta(\mu_R^{NI})$.

Since $f$ is symmetric and $\mathbb{E}[\theta] \geq 0$, then $\Pr\left[\theta \leq -\frac{c_P}{1-(1-\lambda)\mu_R}\right] - \Pr[\theta > c_P] \leq 0$ for all $\mu_R \in [0,1]$. Therefore, $\overline{\mu} = -\infty$, (43) is violated and only the second case applies, that is, $b_S^{NI} = \eta(\mu_R^{NI})$. Note that $\eta(\mu_R^{NI})$ can be positive or negative. Moreover, since $\overline{\mu} = -\infty$ then $\mu_R^{NI} < 1$. ■

**Lemma 4** *If the equilibrium is influential according to Definition 3, then it is also influential according to Definition 1.*

**Proof.** Suppose on the contrary there is an equilibrium that is influential according to Definition 3, but it is not influential according to Definition 1. Therefore, for any $m_1, m_2 \in M$ and $\beta \in [0, \overline{\beta}]$, $a_A(\beta_0, m_1) = a_A(\beta_0, m_2)$. Moreover, there are $m_1, m_2 \in M$ and $\beta_0$ such that $a_F(\beta_0, m_1) \neq a_F(\beta_0, m_2)$. Similar to the arguments in the proof of Proposition 7, $\mu_L = 1$. That is, if $a_A = L$, $e = 1$ and $\chi = 0$, then the agent chooses $a_F = R$ with probability one. Therefore, it has to be that $a_A(\beta_0, m) = R$ for all $m \in M$. Suppose $a_A = R$, the principal intervenes, but intervention failed. Since $a_F(\beta_0, m_1) \neq a_F(\beta_0, m_2)$, similar to the logic behind Lemma 2, suppose that, without the loss of generality,

$$\begin{aligned} & \Pr[a_F = R | a_A = R, e = 1, \chi = 0, m_1] \\ < \; & \Pr[a_F = R | a_A = R, e = 1, \chi = 0, m_2]. \end{aligned} \tag{46}$$

Note that if $a_A = R$ and $e = 1$ then it has to be that $\theta < 0$. Therefore, the principal strictly prefers $m_1$ over $m_2$. This contradicts the assumption that $m_2 \in M$. ■

**Lemma 5** *Suppose $\lambda = 0$. An equilibrium in which intervention is off the equilibrium path does not survive the Grossman and Perry (1986) criterion.*

**Proof.** Consider an equilibrium in which intervention is off the equilibrium path. In this equilibrium, the agent always follows the recommendation to choose action $R$. The agent follows the recommendation to choose action $L$ if and only if $\beta \leq -\mathbb{E}[\theta|\theta<0]$. Suppose $\theta<0$ and the principal recommends the agent to choose action $L$ but the agent decides on $R$. Let

$$\hat{\mu}_R = \frac{1-G\left(-\mathbb{E}\left[\theta|\theta<-\frac{c_P}{1-\hat{\mu}_R}\right]\right)}{1-G\left(-\mathbb{E}\left[\theta|\theta<0\right]\right)}, \tag{47}$$

and note that a solution in the unit interval always exists. Consider the following deviation: the principal intervenes if and only if $\theta<-\frac{c_P}{1-\hat{\mu}_R}$. If the agent expects that upon deviation the principal intervenes if and only if $\theta<-\frac{c_P}{1-\hat{\mu}_R}$, the agent has incentives to revise the decision from $R$ to $L$ if and only if $\beta \leq -\mathbb{E}\left[\theta|\theta<-\frac{c_P}{1-\hat{\mu}_R}\right]$. Given this behavior, the principal has incentives to deviate and intervene if and only if $\theta<-\frac{c_P}{1-\hat{\mu}_R}$. Indeed, since $\hat{\mu}_R$ solves (47), if the principal deviates and intervenes, he expects the agent to revise his decision with probability $1-\hat{\mu}_R$. Therefore, the benefit from intervention is $\hat{\mu}_R\theta-c_A$. If the principal does not intervene, then her payoff is $\theta$. Therefore, the principal intervenes if and only if $\theta<-\frac{c_P}{1-\hat{\mu}_R}$. The existence of this deviation violates Grossman and Perry (1986) criterion. ■

**Proposition 8** *Suppose the agent is allowed to reverse his initial decision upon non-intervention. Then, all the equilibria in Proposition 5 continue to exist.*

**Proof.** Consider an equilibrium as described by Proposition 5. Suppose $m \in M_R$. In any equilibrium described by Proposition 5, the agent chooses action $R$ for sure. Since $m \in M_R \Rightarrow \theta > 0$, based on (3), the agent has incentives to maintain his original decision if the principal does not intervene and regardless of his inference from this event. The principal has rational expectations, and hence, her incentives to intervene are the same as in Proposition 5. Therefore, the agent has the same incentives as in Proposition 5 to choose action $R$.

Suppose $m \in M_L$. Based on Proposition 5 and its proof, the agent chooses $a_A = R$ if and only if $\beta > b^{**}$ where:

- if $c_A \geq h(1)$ then $b^{**} = b^*$ and $\mu_R^{**} = 1$.
- if $c_A < h(1)$ then $b^{**} = \eta(\mu_R^{**})$ and $\mu_R^{**} < 1$.

Furthermore, if the agent chooses $a_A = L$, the principal never intervenes, and if the agent chooses $a_A = R$, the principal intervenes if and only if $-\frac{c_P}{1-(1-\lambda)\mu_R^{**}} < \theta$. Last, note that in both cases,

$$b^{**} > -\mathbb{E}\left[\theta | -\frac{c_P}{1-(1-\lambda)\,\mu_R^{**}} < \theta < 0\right]. \tag{48}$$

Suppose $\beta > b^{**}$. We argue that even with the option to reverse the initial decision upon non-intervention, the agent has incentives to choose action $R$. If the agent chooses $a_A = L$, the principal never intervenes. The agent does not infer new information from the decision not to intervene, and hence, if he chooses $L$, his expected payoff is zero. If the agent chooses $a_A = R$, and the principal does not intervene, the agent reverses his initial decision from $R$ to $L$ if and only if

$$\beta \leq -\mathbb{E}\left[\theta | -\frac{c_P}{1-(1-\lambda)\,\mu_R^{**}} < \theta < 0\right]. \tag{49}$$

However, note that $\beta > b^{**}$ and (48) imply that (49) never holds. Therefore, the agent has no incentives to reverse his decision to $L$ upon non-intervention. It follows, the agents's expected payoff from choosing $R$ is given by (29). By construction, $\beta > b^{**}$ implies that (29) is non negative, and hence, the agent indeed has incentives to choose action $R$.

Suppose $\beta \leq b^{**}$. Similar to the previous case, if the agent chooses $a_A = L$, the principal never intervenes, and the agent's expected payoff is zero. By construction, $\beta < b^{**}$ implies that (29) is negative. As before if (49) is violated, the agent has no incentives to reverse his decision from $R$ to $L$ upon non-intervention. Therefore, his expected payoff is (29), and since it is negative, he is better off choosing action $L$. Suppose (49) holds. Then, the agent will reverse his decision form $R$ to $L$ upon non-intervention. The agents expected payoff from choosing $R$ is

$$\Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R^{**}} | \theta < 0\right]\left((1-\lambda)\max\left\{0, \mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R^{**}}\right]\right\} - \lambda c_A\right).$$

Note that (49) implies $\mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}}\right] \leq 0$. Therefore, upon failed intervention, the agent always reverses his decision, and his expected payoff from choosing $R$ is $-\lambda c_A \Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}} | \theta < 0\right.$ which is always negative. Therefore, the agent has strict incentives to choose action $L$.

We conclude, that even with the option to reverse the decision upon non-intervention, the agent's initial decision is incentive compatible. Given this strategy, it follows directly from the proof of Proposition 5, the principal's communication and intervention polices are incentives compatible as well. Therefore, all the equilibria in Proposition 5 continue to exist as required.

■

**Proposition 9** *With voluntary revision, intervention harms communication if and only if*

$$c_A < \frac{1}{\lambda}\left(\mathbb{E}\left[\theta|\theta<0\right] - \mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R^{**}}\right]\right). \tag{50}$$

**Proof.** Based on the proof of Proposition 5, if $c_A \geq h(1)$ then $b^{**} = \delta(1, c_A) > 0$. Note that $\delta(1, h(1)) = -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right]$. Since $c_A > h(1) \Rightarrow \delta(1, c_A) > \delta(1, h(1))$, and since $-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right] > -\mathbb{E}\left[\theta|\theta<0\right]$, then $c_A \geq h(1)$ implies $\delta(1, c_A) > -\mathbb{E}\left[\theta|\theta<0\right]$, and intervention enhances communication. If $c_A < h(1)$ then $b^{**} = \eta\left(\mu_R^{**}(c_A), c_A\right)$ and

$$\eta\left(\mu_R^{**}(c_A), c_A\right) > -\mathbb{E}\left[\theta|\theta<0\right] \Leftrightarrow c_A > \frac{\mathbb{E}\left[\theta|\theta<0\right] - \mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\lambda},$$

and note that

$$\frac{\mathbb{E}\left[\theta|\theta<0\right] - \mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\lambda} < \frac{\Pr\left[\theta<0\right]}{\Pr\left[\theta < -\frac{c_P}{\lambda}\right]}\,\frac{\mathbb{E}\left[\theta|\theta<0\right] - \mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right]}{\lambda} = h(1).$$

Combined, intervention enhances communication if and only if (50) holds. ■

# B Extensions

## B.1 Continuum of actions

Consider a variant of the baseline model in which the action space is a continuum. Specifically, suppose $a \in \mathbb{R}$ and let

$$v(\theta, a) = -(\theta - a)^2. \tag{51}$$

For simplicity, I assume that the agent's bias $\beta$ is a common knowledge and strictly positive. I make the following assumptions about intervention. First, intervention is always successful, that is, $\lambda = 1$. Second, if the agent chooses $a_A$ and the principal intervenes and chooses $a_P$, the principal incurs an additional cost of $c_P (a_P - a_A)^2$ and the agent incurs an additional cost of $c_A (a_P - a_A)^2$, where $c_P \geq 0$ and $c_A \geq 0$. These functional forms capture the idea that as $|a_P - a_A|$ increases, both the principal and the agent incur larger costs due to intervention. As in the baseline model, I denote by $a_A(m)$ the agent's action strategy and by $\rho(\theta)$ the principal's messaging strategy. I also denote by $\Delta(a_A, \theta)$ the difference between $a_P$ and $a_A$ when principal intervenes, as a function of $a_A$ and $\theta$.

**Proposition 10** *Let $\Lambda(\beta, c_P, c_A)$ the set of equilibria of the game. In any equilibrium, $\Delta^*(a_A, \theta) = \frac{\theta - a_A}{1 + c_P}$. Moreover,*

$$\left(a_A^*(m), \rho^*(\theta), \Delta^*(a_A, \theta)\right) \in \Lambda(\beta, c_P, c_A)$$

*if and only if*

$$\left(a_A^*(m), \rho^*(\theta), \Delta^*(a_A, \theta)\right) \in \Lambda\left(\beta \times \frac{c_P + c_P^2}{c_A + c_P^2}, \infty, c_A\right).$$

**Proof.** Given the agent's decision $a_A$ and the observation of $\theta$, regardless of the message that the principal sent the agent, the principal solves

$$\begin{aligned} \Delta(a_A, \theta) &\in \arg\max_{\Delta} \left\{-(\theta - (a_A + \Delta))^2 - c_P \Delta^2\right\} \\ \Rightarrow \ \Delta(a_A, \theta) &= \frac{\theta - a_A}{1 + c_P}. \end{aligned}$$

Thus, if the agent chooses action $a_A$, the principal's utility conditional on $\theta$ is

$$\begin{aligned} u_P &= -(\theta - (a_A + \Delta(a_A, \theta)))^2 - c_P \Delta(a_A, \theta)^2 \\ &= -\frac{c_P}{1 + c_P}(\theta - a_A)^2. \end{aligned}$$

The agent expects the principal to follow intervention policy $\Delta(a_A, \theta)$, and therefore, given message $m$, he solves

$$\begin{aligned} a_A^* &\in \arg\max_{a_A} \mathbb{E}\left[-(\theta+\beta-(a_A+\Delta(a_A,\theta)))^2 - c_A\Delta(a_A,\theta)^2 \,|m\right] \\ &\Rightarrow a_A^* = \mathbb{E}[\theta|m] + \beta\frac{c_P+c_P^2}{c_A+c_P^2}. \end{aligned}$$

It follows, at the communication stage, the principal behaves as if her preferences are represented by the utility function $-(\theta-a_A)^2$, and the agent behaves as if $c_P = \infty$ and his preferences are represented by the utility function $-\left(\theta+\beta\frac{c_P+c_P^2}{c_A+c_P^2}-a_A\right)^2$. ■

Proposition 10 implies that the quality of communication between the principal and the agent in equilibrium is equivalent to the quality of communication when intervention is not possible and the agent's bias is $\beta\frac{c_P+c_P^2}{c_A+c_P^2}$ instead of $\beta$. Note that Crawford and Sobel's (1982) setup with a quadratic loss function is a special case of this model when $c_P = \infty$. Therefore, intervention harms communication if and only if

$$\frac{c_P+c_P^2}{c_A+c_P^2} > 1 \Leftrightarrow c_P > c_A.$$

## B.2 Verifiable information

Consider a variant of the baseline model in which $\theta$ is verifiable. I argue that intervention can harm the principal's ability to affect the agent's decision through disclosure in this setup as well. When information is verifiable, $\rho(\theta) \in \{\theta, \phi\}$ where $\rho = \phi$ is interpreted as the principal's decision not to disclose information, and $\rho(\theta) = \theta$ is the principal's decision to disclose the exact value of $\theta$. To keep the analysis simple, I assume $\lambda = 1$ and $c_A > 0$.

Suppose the principal discloses $\theta$. If $\theta \leq -c_P$ then the principal intervenes if and only if the agent chooses $R$. Since $c_A > 0$, the agent will avoid intervention and choose $L$. If $\theta > -c_P$ the principal intervenes if and only if the agent chooses $L$ and $\theta > c_A$. However, according to

(3), the agent will choose $L$ if and only if $\theta < -\beta$. To conclude,

$$\Pr\left[a_A = R | \rho(\theta) = \theta\right] = \begin{cases} 0 & \textit{if } \theta \leq -c_P \\ \Pr\left[\theta \geq -\beta\right] & \textit{if } \theta > -c_P, \end{cases} \tag{52}$$

and note that if $\rho(\theta) = \theta$ then the principal never intervenes. The next result characterizes the equilibria of the game with verifiable information.

**Proposition 11** *Let $\Upsilon^* \equiv \{\theta : \rho(\theta) = \phi\}$ and $\varphi^* = \Pr\left[a_A = R | \rho = \phi\right]$. In any equilibrium the principal intervenes with a zero probability. Moreover:*

*(i) For any $\Theta \subseteq \left[0, \overline{\theta}\right]$ there is an equilibrium in which $\Upsilon^* = \Theta$, $\varphi^* = 1$ and $a_A = L$ if and only if $\theta < \max\{-c_P, -\beta\}$.*

*(ii) An equilibrium with $\varphi^* = 0$ exists if and only if $\overline{\beta} \leq \overline{b}_{\text{ver}}$ where*

$$\overline{b}_{\text{ver}} \equiv \max_{\Upsilon : \left[\max\{-c_P, -\overline{\beta}\}, 0\right] \subseteq \Upsilon \subseteq \left[\underline{\theta}, 0\right]} \left\{ c_A \frac{\Pr\left[\theta < -c_P | \theta \in \Upsilon\right]}{1 - \Pr\left[\theta < -c_P | \theta \in \Upsilon\right]} - \mathbb{E}\left[\theta | \theta \geq -c_P, \theta \in \Upsilon\right] \right\}. \tag{53}$$

*In this equilibrium, $\left[\max\left\{-c_P, -\overline{\beta}\right\}, 0\right] \subseteq \Upsilon^* \subseteq \left[\underline{\theta}, 0\right]$ and $a_A = L$ if and only if $\theta < 0$.*

*(iii) No other equilibrium exists.*

**Proof.** First, consider an equilibrium with $\varphi^* = 1$. If $\theta < 0$ the principal strictly prefers disclosing $\theta$ and thereby reducing the probability that $R$ is chosen from one to $\Pr\left[a_A = R | \rho(\theta) = \theta\right] < 1$ as given by (52). If $\theta \geq 0$ the principal is indifferent with respect to her disclosure policy, since in both cases the agent chooses $R$ with probability one. Moreover, the principal has no incentives to intervene if $\theta > 0$ and $a_A = R$. Therefore, for any $\Upsilon \subseteq \left[0, \overline{\theta}\right]$, if $\rho = \phi$ then the agent infers that $\theta > 0$ for sure, and according to (3), he strictly prefers choosing $R$. Overall, in this equilibrium, $a_A = L$ if and only if $\theta < \max\{-c_P, -\beta\}$, and the principal never intervenes.

Second, consider an equilibrium with $\varphi^* \in (0, 1)$. If $\theta < \max\left\{-c_P, -\overline{\beta}\right\}$ and $\rho = \theta$ then the principal expects the agent to choose $L$ with probability one. Since $\varphi^* > 0$, the principal strictly prefers disclosing $\theta$, thereby saving on the cost of intervention when the agent chooses

$R$. If $\theta > 0$ and $\rho = \theta$ then the principal expects the agent to choose $R$ with probability one. Since $\varphi^* < 1$, the principal strictly prefers disclosing $\theta$ in this range. Combined, it is necessary that $\Upsilon^* \subseteq \left[\max\left\{-c_P, -\overline{\beta}\right\}, 0\right]$. Since $\left[\max\left\{-c_P, -\overline{\beta}\right\}, 0\right] \subseteq [-c_P, 0]$ the agent knows that upon non-disclosure $\theta \in [-c_P, 0]$, and hence, the principal will not intervene. It follows, the agent will choose $R$ upon non-disclosure if and only if $-\beta \leq \mathbb{E}\left[\theta | \theta \in \Upsilon^*\right]$. Therefore, it must be $\varphi^* = \Pr\left[-\beta \leq \mathbb{E}\left[\theta | \theta \in \Upsilon^*\right]\right]$. Let $\hat{\theta} \in \Upsilon^*$ be such that $\hat{\theta} < \mathbb{E}\left[\theta | \theta \in \Upsilon^*\right]$. If $\rho = \hat{\theta}$ then the agent will choose $R$ if and only if $-\beta \leq \hat{\theta}$. Therefore, by disclosing $\hat{\theta}$ the principal strictly increases the probability that the agent chooses $L$ from $1 - \varphi^*$ to $\Pr[-\beta > \hat{\theta}]$. Since $\hat{\theta} \in \Upsilon^* \Rightarrow \hat{\theta} < 0$, the principal has strict incentives to deviate and disclose $\hat{\theta}$. By this logic, if $\varphi^* \in (0, 1)$ then $\Upsilon^* \in \{\varnothing, \{0\}\}$. In both cases, $a_A = L$ if and only if $\theta < \max\{-c_P, -\beta\}$, which is a special case of part (i).[22]

Last, consider an equilibrium with $\varphi^* = 0$. Since $\varphi^* = 0$, the principal has strict incentives disclose $\theta$ when $\theta > 0$. Moreover, if $\theta \in \left[\max\left\{-c_P, -\overline{\beta}\right\}, 0\right]$ the principal has strict incentives to conceal $\theta$, since if she discloses $\theta$, there is a strictly positive probability that the agent chooses $a = R$. If $\theta \in \left[\underline{\theta}, \max\left\{-c_P, -\overline{\beta}\right\}\right]$ then the principal is indifferent between disclosing and concealing $\theta$, as in both cases the agent chooses $L$ for sure. Therefore, it is necessary that $\left[\max\left\{-c_P, -\overline{\beta}\right\}, 0\right] \subseteq \Upsilon^* \subseteq [\underline{\theta}, 0]$. If $\rho = \phi$ the agent infers $\theta \in \Upsilon^*$. Since $\Upsilon^* \subseteq [\underline{\theta}, 0]$ the agent expects that if he chooses $L$ the principal never intervenes and his payoff will be zero. Instead, if the agent chooses $R$, his expected utility is

$$\Pr\left[\theta \geq -c_P | \theta \in \Upsilon^*\right] \mathbb{E}\left[\theta + \beta | \theta \geq -c_P, \theta \in \Upsilon^*\right] - c_A \Pr\left[\theta < -c_P | \theta \in \Upsilon^*\right].$$

Therefore, if $\rho = \phi$ the agent chooses $L$ if and only if

$$\beta \leq c_A \frac{\Pr\left[\theta < -c_P | \theta \in \Upsilon^*\right]}{1 - \Pr\left[\theta < -c_P | \theta \in \Upsilon^*\right]} - \mathbb{E}\left[\theta | \theta \geq -c_P, \theta \in \Upsilon^*\right].$$

Note that $\varphi^* = 0$ requires $\overline{\beta}$ being smaller than the RHS of the above condition. Therefore, an equilibrium with $\varphi^* = 0$ exists if and only if $\overline{\beta} \leq \overline{b}_{\text{ver}}$. If $\overline{\beta} \leq \overline{b}_{\text{ver}}$ then the agent effectively

[22] If $\Upsilon^* \in \{\varnothing, \{0\}\}$ then $\Pr\left[\theta \in \Upsilon^*\right] = 0$, and hence, $\varphi^*$ can take any value without changing the outcome of the equilibrium.

chooses $a_A = R$ if and only if $\theta > 0$, and the principal never intervenes. This argument proves part (ii). Part (iii) and the claim that in any equilibrium the principal intervenes with a zero probability, follow by noting that all cases where $\varphi^* \in [0,1]$ have been covered by the proof. ■

When $\varphi^* = 0$ the principal can conceal enough information to convince the agent to choose action $L$ whenever $\theta < 0$. In this respect, the agent is following the principal's demand, and the principal's first best is obtained in equilibrium. By contrast, when $\varphi^* = 1$ (and $c_P > 0$), the principal's expected payoff is strictly less than her first best. In this respect, equilibria with $\varphi^* = 0$ ($\varphi^* = 1$) are the analog of influential (non-influential) equilibria in the baseline model. Proposition 11 shows that the existence of an equilibrium with $\varphi^* = 0$ depends on how $\overline{\beta}$ compares with $\overline{b}_{\text{ver}}$. The next result gives an example where an equilibrium with $\varphi^* = 0$ exists without intervention, but it does not exist with intervention. In this respect, intervention harms the principal's ability to influence the agent through communication, even with verifiable information. The intuition behind the result is similar to the one in the baseline model, and this can be seen by the similarity of expression (53) to expression (8).

**Proposition 12** *Suppose $c_P \in (0, -\underline{\theta})$ and $-\mathbb{E}[\theta| - c_P \leq \theta < 0] < \overline{\beta} \leq -\mathbb{E}[\theta|\theta < 0]$. There is $\bar{c}_A > 0$ such that if $c_A \in (0, \bar{c}_A)$ then the principal's first best is obtained in equilibrium without intervention, but it is not obtained in equilibrium with intervention.*

**Proof.** I start by arguing that $\overline{\beta} \leq \overline{b}_{\text{ver}}(c_A = 0, c_P)$ if and only if $\overline{\beta} \leq -\mathbb{E}[\theta| - c_P \leq \theta < 0]$. Consider three cases. First, suppose $\overline{\beta} \leq -\mathbb{E}[\theta| - c_P \leq \theta < 0]$. Let $\Upsilon = [-c_P, 0]$ and note that $\overline{\beta} \leq -\mathbb{E}[\theta| - c_P \leq \theta < 0] = -\mathbb{E}[\theta|\theta \geq -c_P, \theta \in \Upsilon]$. Since $-\mathbb{E}[\theta|\theta \geq -c_P, \theta \in \Upsilon] \leq \overline{b}_{\text{ver}}$, we have $\overline{\beta} \leq \overline{b}_{\text{ver}}(c_A = 0, c_P)$ as required. Second, suppose $\overline{\beta} > -\mathbb{E}[\theta| - c_P \leq \theta < 0]$ and $\overline{\beta} \geq c_P$. Then $[\max\{-c_P, -\overline{\beta}\}, 0] \subseteq \Upsilon$ implies $[-c_P, 0] \subseteq \Upsilon$. Therefore, $\mathbb{E}[\theta|\theta \geq -c_P, \theta \in \Upsilon]$ is invariant to $\Upsilon$ and is equal to $\mathbb{E}[\theta| - c_P \leq \theta < 0]$. Therefore, $\overline{b}_{\text{ver}} = -\mathbb{E}[\theta| - c_P \leq \theta < 0]$. Since $\overline{\beta} > -\mathbb{E}[\theta| - c_P \leq \theta < 0]$ then $\overline{b}_{\text{ver}}(c_A = 0, c_P) < \overline{\beta}$, as required. Third, suppose $c_P > \overline{\beta} > -\mathbb{E}[\theta| - c_P \leq \theta < 0]$. Relative to $\Upsilon' = [-c_P, 0]$, any $[-\overline{\beta}, 0] \subseteq \Upsilon \subseteq [\underline{\theta}, 0]$ such that $[-c_P, 0] \backslash \Upsilon \neq \varnothing$ is missing from its pool $\theta \in [-c_P, -\overline{\beta}]$. Since $-\overline{\beta} < \mathbb{E}[\theta| - c_P \leq \theta < 0]$, then $\mathbb{E}[\theta| - c_P \leq \theta < 0] \leq \mathbb{E}[\theta|\theta \geq -c_P, \theta \in \Upsilon]$. This implies $-\overline{\beta} < \mathbb{E}[\theta|\theta \geq -c_P, \theta \in \Upsilon]$ for all $[\max\{-c_P, -\overline{\beta}\}, 0] \subseteq \Upsilon \subseteq [\underline{\theta}, 0]$. Therefore, $\overline{b}_{\text{ver}}(c_A = 0, c_P) < \overline{\beta}$ as required.

A special case of Proposition 11 is $c_P \geq -\underline{\theta}$, that is, intervention is not allowed. Based on Proposition 11, without intervention, an equilibrium with $\varphi^* = 0$ exists if and only if $\overline{\beta} \leq \overline{b}_{\text{ver}}(c_A, -\underline{\theta})$. Note that $\overline{b}_{\text{ver}}(c_A, -\underline{\theta}) = \overline{b}_{\text{ver}}(0, -\underline{\theta})$ for any $c_A$. According to the argument above, without intervention, an equilibrium with $\varphi^* = 0$ exists if and only if $\overline{\beta} \leq -\mathbb{E}[\theta|\theta < 0]$. Next, suppose $c_P \in (0, -\underline{\theta})$ and $-\mathbb{E}[\theta| - c_P \leq \theta < 0] < \overline{\beta} \leq -\mathbb{E}[\theta|\theta < 0]$. Note that for all $c_A$ and $c_P$ we have $\overline{b}_{\text{ver}}(c_A, c_P) \leq H(c_A, c_P)$ where

$$\begin{aligned} H(c_A, c_P) \equiv\; & c_A \frac{\Pr\left[\theta < -c_P | \theta \in [\underline{\theta}, -c_P] \cup [\max\{-c_P, -\overline{\beta}\}, 0]\right]}{1 - \Pr\left[\theta < -c_P | \theta \in [\underline{\theta}, -c_P] \cup [\max\{-c_P, -\overline{\beta}\}, 0]\right]} \\ & + \max_{\Upsilon : [\max\{-c_P, -\overline{\beta}\}, 0] \subseteq \Upsilon \subseteq [\underline{\theta}, 0]} \{-\mathbb{E}[\theta | \theta \geq -c_P, \theta \in \Upsilon]\} \end{aligned}$$

and $H(c_A, c_P)$ is continuous and increasing in $c_A$. Moreover, $\lim_{c_A \to 0} H(c_A, c_P) = \overline{b}_{\text{ver}}(0, c_P)$. Therefore, for any $\varepsilon \in \left(0, \overline{\beta} + \mathbb{E}[\theta| - c_P \leq \theta < 0]\right)$ there is $\bar{c}_A > 0$ such that if $c_A \in (0, \bar{c}_A)$ then

$$\overline{b}_{\text{ver}}(c_A, c_P) \leq H(c_A, c_P) < -\mathbb{E}[\theta| - c_P \leq \theta < 0] + \varepsilon$$

Therefore, if $c_A \in (0, \bar{c}_A)$ then $\overline{b}_{\text{ver}}(c_A, c_P) < \overline{\beta}$, and according to Proposition 11, an equilibrium with $\varphi^* = 0$ exists without intervention, but not with intervention. ■