Governing Through Communication and Intervention

Doron Levit[*]

Wharton

April 8, 2015

Abstract

This paper studies communication and intervention as mechanisms of governance. I develop a model in which a privately informed principal can overrule the decisions of the agent (intervention) if the agent disobeys the principal's instructions (communication). The main result shows that intervention can be counterproductive, as it *undermines* the principal's ability to resolve the conflict with the agent through communication. I show that 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". This result provides a novel argument as to why a commitment not to intervene can be beneficial to the principal, echoing the common wisdom that the capacity to make unilateral decisions can discourage effective deliberation, cooperation and compliance. The analysis sheds new light on the effectiveness of different governance arrangements and provides novel predictions about expected patterns of intervention.

Keywords: Governance, Communication, Intervention, Organization, Cheap-Talk, Authority, Obedience

JEL Classification: D72, D74, D82, D83, G28, 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. By intervening, the principal can either overrule the agent and perform the task on her own, force the agent to repeat the work, monitor the agent closely, or find a replacement. Either way, intervention requires a non-trivial amount of effort, time and resources from the principal, and therefore, it is often used as a last resort. Instead, the principal may prefer communicating with the agent, resolving the tension by discussion and persuasion. Unlike intervention, communication is effective only if the agent is carrying out the principal's demand at his own free will. 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 instructions, the principal has the option to intervene.

By nature, communication and intervention are interrelated. Is communication more effective with intervention than without it? Can the principal benefit from a commitment not to intervene in the agent's decision? The main result of this paper demonstrates that intervention can be counterproductive, as it *undermines* the principal's ability to resolve the tension with the agent through communication. In other words, intervention prompts disobedience, and a commitment not to intervene can be beneficial to the principal. By characterizing the conditions under which a commitment not to intervene is optimal, the paper can explain observed patterns of intervention and shed new light on the effectiveness of different governance arrangements in a variety of applications: organizations (manager and employees, CEO and division manager), corporate governance (corporate board and CEO, venture capital fund and entrepreneur, activist investor and the board of directors), regulation (central bank and financial intermediaries), 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.[1] The principal is privately informed about the action that maximizes her utility. 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 con-

[1] In the Online Appendix, I consider a variant of the model in which the agent chooses from a continuum of actions, and show that the main results continue to hold.

flict 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 instructions, a recommendation, or a nonbinding demand from 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 can disobey and ignore the principal's instructions. After observing the agent's decision, the principal decides whether to intervene. Intervention imposes a cost on the principal, as well as 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 decision. If the principal does not intervene, the agent does not incur additional costs and his initial decision is unaffected.

In equilibrium, the principal's instructions are informative. The effectiveness of communication can be measured by the probability that the agent uses this information and voluntarily follows the instructions of the principal in equilibrium.[2] As one might expect, if the cost that intervention imposes on the agent is *large* relative to the cost that is incurred by the principal, communication is *more effective* with intervention than without it. Intuitively, when the cost of intervention is small, the agent understands that the principal is likely to intervene if he does not follow her instructions. Since intervention imposes a large cost on the agent, the agent prefers following the principal's instructions 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.

Perhaps surprisingly, the main result of the paper shows that communication is *less effective* with intervention than without it, if the cost that intervention imposes on the agent is *small* relative to the cost that is incurred by the principal. Under this condition, intervention *harms* communication. Since the agent is less likely to follow the principal's instructions when the principal has the option to intervene, intervention and communication substitute each other. In this case, the principal can benefit from a commitment not to intervene in the agent's decisions. Interestingly, this result holds even though there are no hold-up problems in the model, the agent is uninformed, and intervention imposes a strictly positive direct cost on the agent. In this respect, the analysis provides a novel explanation of why a commitment not to intervene

[2] With a continuum of actions, more effective communication also implies that more information is revealed by the principal in equilibrium.

in the agent's decisions is valuable.

How can intervention harm communication? Why does the intuition behind the result that intervention reinforces communication no longer hold? To understand the intuition, note that a standard result in the literature on communication is that because of the conflict of interests the principal does not fully reveal her private information in equilibrium. In a sense, the principal deliberately conceals information the agent is likely to abuse. By ignoring the principal's instructions, the agent can elicit additional information from the principal. To see how, note that if the agent ignores the principal's instructions, the principal must 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. Therefore, if the principal does not intervene, the agent infers that the principal believes that the initial decision did 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 instructions. On the other hand, if the principal intervenes, the agent infers that the principal believes that the initial decision was detrimental. 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. When calculating the benefit from ignoring the principal's instructions, the agent "conditions" on events in which the principal intervenes and on events in which she does not.[3] This line of reasoning is similar to the winner's curse in common value auctions and pivotal considerations in models with strategic voting. Altogether, by ignoring the principal's instructions the agent forces the principal to make a decision that inevitably reveals information she was trying to conceal.

The agent trades off the benefit from eliciting additional information from the principal 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 deciding whether to intervene. This bias, which is exploited by the agent, increases the agent's benefit from the "confirmation" and "correction" effects. In other words, the possibility of intervention creates additional tension between the principal and the agent by providing the

[3] While in the baseline model the agent cannot change his initial decision after the principal decides whether to intervene, in Section 3 I analyze a variant of the model in which the agent can voluntarily revise his initial decision upon failed intervention.

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 effective deliberation and cooperation can be sustained when the power to make unilateral decisions is restrained. When intervention harms communication, the principal can benefit from a commitment to reduce her capacity to intervene. This observation has important implications. For example, it can rationalize the benefits of flat organizations, hands-off management, lax regulation and various corporate governance arrangements (e.g., "friendly" boards). Importantly, the analysis suggests that these governance arrangements can be optimal even in situations where the principal has information the agent does not have.

There are several empirical implications that follow from the analysis. First, the analysis highlights the importance of the relative magnitude of the cost that intervention imposes on the agent to the cost that is incurred by the principal. The characterization of the conditions under which intervention harms communication provides novel predictions about the circumstances where the benefit from reducing the capacity to intervene is most pronounced.[4] Second, the analysis suggests that the observation of rare instances of intervention is not necessarily evidence that intervention is ineffective – quite the opposite. Indeed, with communication, the principal does not need to intervene if the threat of intervention is sufficient to convince the agent to follow her instructions. Generally, there is a non-monotonic relationship between the cost of intervention and the likelihood that the option to intervene is exercised.[5] Intervention occurs only if it is not credible enough to deter the agent from ignoring the principal's instructions, but it is sufficiently profitable as a corrective tool. Unless this non-monotonicity is explicitly addressed, factors that have an effect on the cost of intervention may seem unrelated to the empirical frequency of instances of intervention.

Intervention may fail for various reasons. For example, successful intervention requires confrontation, arbitration or cooperation with a third party. If intervention fails, the agent may act on the new information that is conveyed by the principal's (failed) attempt to intervene. That is, intervention plays a dual role: it is a corrective tool as well as a channel through

[4]Under some conditions, the principal can be better off with the option to intervene even if it harms communication. In this respect, the analysis also identifies cases where the relationship between the principal and the agent is confrontational but sustainable.

[5]Without communication, the threat of intervention does not affect the agent's decision, and the likelihood of intervention decreases with the cost of intervention.

which information is transmitted to the agent. To capture this possibility, I extend the model by allowing the agent to revise his initial decision if the principal intervenes but intervention fails. With the option to revise the initial decision, the agent has even weaker incentives to follow the principal's instructions. The reason is that if the principal intervenes and fails, the agent learns that his initial decision was detrimental. Unlike the baseline model, here the agent can himself correct the decision if needed. At the extreme, if intervention is doomed to failure and only serves as a costly signaling device by the principal ("burning money"), intervention always harms communication.[6] 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 paper proceeds as follows. The remainder of the section highlights the contribution of the paper relative to the existing literature. Section 1 presents the setup of the baseline model, 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 regulation and corporate governance. 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.

Related literature

This paper is related to Aghion and Tirole (1997) who study the optimal allocation of authority within the organization, and distinguish between the concepts of formal and real authority. Formal authority can be undesirable because it weakens the agent's incentives to collect information. Different from their study, here there is no hold-up problem, communication is strategic, and intervention is costly. The right to intervene can be harmful since it weakens the principal's ability to influence the agent through communication. Therefore, the value of a

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

commitment not to intervene in the agent's decision is derived from different economic forces.[7]

The paper is also related to the literature on obedience and interpersonal authority. Van den Steen (2010) shows that the firm is a mechanism to give the manager interpersonal authority (effective communication) over employees when there is open disagreement. Shifting asset ownership from the agent to the principal lowers the outside option of the agent and raises that of the principal, making it more costly for the agent to get fired and easier for the principal to commit to firing a disobeying agent. Marino et al. (2010) study how the allocation of control and the external labor market characteristics affect disobedience, and show that the agent is given more authority when replacement is costly or the agent does not fear dismissal. In both studies, intervention reinforces communication and increases obedience. By contrast, I show that intervention can harm communication and decrease obedience. To the best of my knowledge, this result is new in the literature. The difference in the results stems from the combination of two assumptions which I believe are first order in many applications. First, intervention in my model does not necessarily terminate the relationship with the agent (although it could). Instead, intervention means that the principal can directly and unilaterally change the agent's action. Second, the agent believes that the principal's private information is payoff relevant, and hence, he benefits from learning about it.

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), Alonso and Matouschek (2007), Chakraborty and Yilmaz (2011), Grenadier et al. (2014) and Harris and Raviv (2005, 2008, 2010)). In these models, delegation is beneficial since it avoids the distortion of the agent's private information. Thus, the principal never delegates decision rights to the biased agent if the agent has no private information. A commitment not to intervene in the agent's decision can be viewed as a form of delegation. However, here the agent has no payoff relevant private information that the principal does not have, and hence, the motive for "delegation" is fundamentally different. Moreover, unlike previous studies, the principal cannot perform the task on her own without incurring the cost of intervention. Instead, the alternative to "delegation" in this model is a combination of communication and the option to intervene.

[7] See also Baker et al. (1999), who unlike Aghion and Tirole (1997) assume that authority is non contractible, but can be informally given through commitments enforced by reputation.

Matthews (1989) studies a model of strategic communication in which the sender (principal) has the option to veto the receiver's (agent's) decision and enforce an exogenously given status quo. In his model, the agent's payoff is independent of the state of the world, which is the principal's private information. By contrast, here the agent trades off the principal's utility with his private benefits, and hence, the agent's utility and desired actions depend on the true state. The aforementioned "confirmation" and "correction" effects arise only when the principal and the agent have a common value and the agent benefits from eliciting additional information from the principal. Related, Levit and Malenko (2011) study whether intervention by a third party can improve the aggregation of information in a nonbinding voting setup. In their model, the player who intervenes is different from the sender (the voters) and he is uninformed. Here, the main results rely on the assumption that the informed player who communicates is also the one who decides whether to intervene.[8]

1 Setup

Consider a principal-agent model in which the payoffs to the principal and the agent depend on random variable θ and action $a \in \{L, R\}$. In the Online Appendix, I show that similar results hold when the agent chooses from a continuum of actions. Random variable θ 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 $\mathbb{E}\left[\theta\right] \geq 0$. Similar results hold if f is asymmetric. If action $a \in \{L, R\}$ is implemented then the principal's payoff is given by

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

and the agent's payoff is given by $v\left(\theta + \beta, a\right)$. Random variable β is privately known to the agent, independent of θ, and has a continuous probability density function g with full

[8] Levit (2013) and Shimizu (2008) study models of strategic communication in which the outside option is exiting the relationship. 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 the receiver's decisions. In Shimizu (2008), exit imposes a large and exogenous disutility on the receiver. Both studies, conclude that exit can reinforce communication. Different from intervention, exit does not change the receiver's initial decision once the decision is made. This distinction between exit and intervention is crucial: intervention harms communication only because the agent knows that the principal will intervene and change his initial decision exactly in those states where changing this decision is desirable from the agent's perspective.

support over $[0, \overline{\beta}]$.[9] 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 the agent have different preferences. Effectively, β captures the intrinsic conflict of interests between the principal and the agent. The agent faces a trade off between the principal's utility and his private benefits from action R, where a larger beta results in a larger 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, β can be interpreted as the residual conflict of interests between the principal and the agent.[10]

The model has four stages. The first stage involves communication between the principal and the agent. I assume that the principal has information about θ that the agent does not have. For simplicity, the principal privately observes θ while the agent is uninformed about θ.[11] Based on her private information, the principal sends the agent message $m \in \left[\underline{\theta}, \overline{\theta}\right]$. The principal's information about θ is non-verifiable and the content of m does not affect the agent's or the principal's payoff directly. These assumptions capture the informal nature of communication. In the Online Appendix, I show that similar results hold if the information about θ 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 β.

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

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

[10] To fix ideas, the model can be given the following interpretation. If $a = R$ investment is undertaken and the project's return is θ. 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 β. Other applications of the model are discussed in Section 4.

[11] 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 for 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.

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 its success is independent of θ and β. If the principal intervenes, whether or 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 alternative cost of dealing with the task or her aversion for confrontation. I assume that $c_P < -\lambda\underline{\theta}$. As I show below, without this assumption, the intervention stage has no effect on the game. 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, the agent incurs a cost $c_A > 0$ if and only if the principal intervenes and intervention succeeds. This cost can be both pecuniary (e.g., the loss of compensation) and non-pecuniary (e.g., damaged reputation, loss of respect or embarrassment).[12]

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

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

and the agent's utility is given by

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

The principal and the agent are risk-neutral and their preferences, up to θ and β, are common knowledge.

[12] 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.

2 Analysis

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

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 θ. Since intervention is costly, the principal intervenes only if she intends to reverse the agent's decision. In this case, the principal's expected payoff is $(1-\lambda)\,\theta - c_P$. It follows, conditional on $a_A = R$, the principal intervenes 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. Conditional on $a_A = L$, the principal intervenes if and only if $\theta > \frac{c_P}{\lambda}$. Overall, the principal intervenes whenever she finds the agent's initial decision detrimental.[13] 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 benefits β are 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 $\beta \le b(m)$.*

Communication is effective only if in equilibrium the principal reveals information about θ 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 θ conditional on observing message m and unconditional on the principal's decision to intervene.*

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

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 about θ and trigger different decisions by the agent. By contrast, if the equilibrium is non-influential, the 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 forgoes 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 β.

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 $\theta < -\frac{c_P}{\lambda}$.*[14]

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 β_R (type β_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 that $\min_{m\in M} b(m) < \max_{m\in M} b(m)$. 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 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 instructions to choose R, and messages in M_L can be interpreted as instructions to choose L. Based on (2), the principal asks the agent to choose action R if $\theta \geq 0$ and action L if $\theta < 0$.

[14] 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 small values of θ, 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 benefits and choose action L in order to avoid the cost of intervention.

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 $\theta \geq 0$.*

An influential equilibrium exists only if the agent finds it in his best interests to follow the principal's instructions. 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 instructions 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

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

On the other hand, if the agent ignores the principal's instructions and chooses action L, the principal will intervene whenever $\theta > \frac{c_P}{\lambda}$. The agent's expected payoff 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 instructions 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 instructions and chooses action L, the principal will not intervene. The agent's expected payoff is zero. However, if the agent ignores the principal's instructions and chooses action R, the principal intervenes whenever $\theta < -\frac{c_P}{\lambda}$. In this case, the agent's expected payoff is

$$\underbrace{\Pr\left[\theta \geq -\frac{c_P}{\lambda} | m\right] \mathbb{E}\left[\theta + \beta | \theta \geq -\frac{c_P}{\lambda}, m\right]}_{\text{Agent's payoff if the principal does not intervene}} \tag{7}$$
$$+ \underbrace{\Pr\left[\theta < -\frac{c_P}{\lambda} | m\right] \left((1-\lambda)\,\mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{\lambda}, m\right] - \lambda c_A\right)}_{\text{Agent's payoff if the principal intervenes}}.$$

The agent follows the principal's instructions 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 instructs the agent to choose action R if and only if $\theta \geq 0$. If the principal instructs the agent to choose action R, the agent chooses action R with probability one and the principal never intervenes. If the principal instructs the agent to choose action L, the agent chooses action L if and only if $\beta \leq b^*(c_A, c_P)$, where*

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

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

According to Proposition 2, the agent follows the principal's instructions if and only if $\beta \leq b^*$. Therefore, b^* measures the effectiveness of communication, where higher b^* implies more effective communication in equilibrium. The comparative statics of b^* follow directly from (8). First, note that b^* increases with c_A. Intuitively, the principal intervenes only if the agent ignores her instructions. In order to avoid the cost of intervention, the agent is more likely to follow the principal's instructions when c_A is higher. The comparative static of b^* with respect to c_P is more subtle.

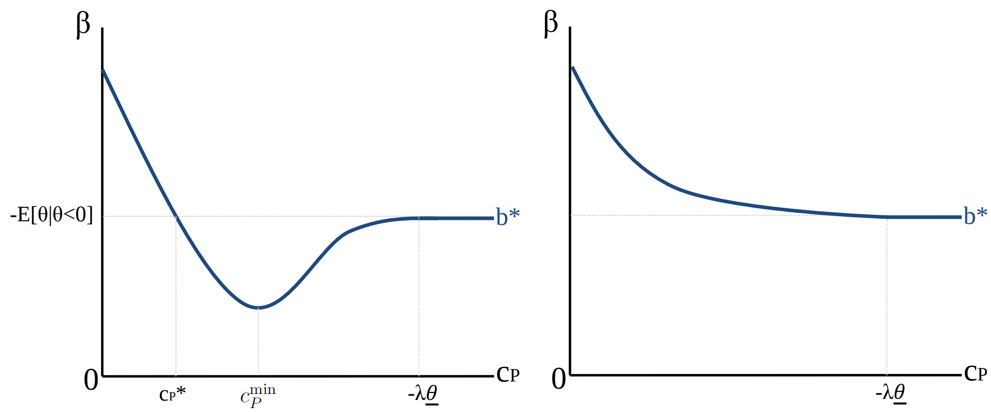
Corollary 1

(i) If $c_A \geq \mathbb{E}[\theta | \theta < 0] - \underline{\theta}$, then $b^*(c_A, c_P)$ is non-increasing with c_P.

(ii) If $c_A < \mathbb{E}[\theta | \theta < 0] - \underline{\theta}$, then $b^*(c_A, c_P)$ 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

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

Figure 1 depicts b^* as a function of c_P. At any point above the blue curve the agent ignores the principal's instructions and chooses action R with probability one. At any point below the blue curve the agent follows the principal's instructions. 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 instructions 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.



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

Figure 1 - The comparative statics of b^* with respect to c_P

In the Appendix, I show that influential equilibria Pareto dominate non-influential equilibria. Since influential equilibria always exist, and as is standard in the literature, hereafter I assume that the equilibrium in play is always influential.

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 instructions 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, $\lim_{c_P/\lambda \to -\underline{\theta}} b^*(c_A, c_P) = -\mathbb{E}[\theta|\theta < 0]$. Thus, the same threshold emerges when intervention is prohibitively costly or entirely ineffective. Communication is considered less effective with intervention than without it if and only if $b^* < -\mathbb{E}[\theta|\theta < 0]$. If $b^* < (>) - \mathbb{E}[\theta|\theta < 0]$ the agent is strictly less (more) likely to follow the principal's instructions, and hence, intervention harms (enhances) communication. 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 large relative to the cost that is incurred by the principal, the agent is more likely to follow the principal's instructions when the principal can intervene. 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 θ is uniform and symmetric around zero, condition (11) becomes $\frac{c_A}{c_P} < \frac{1}{2\lambda}$. The cutoff point c_P^* in the left panel of Figure 1 is the unique value of c_P that satisfies condition (11) with equality.

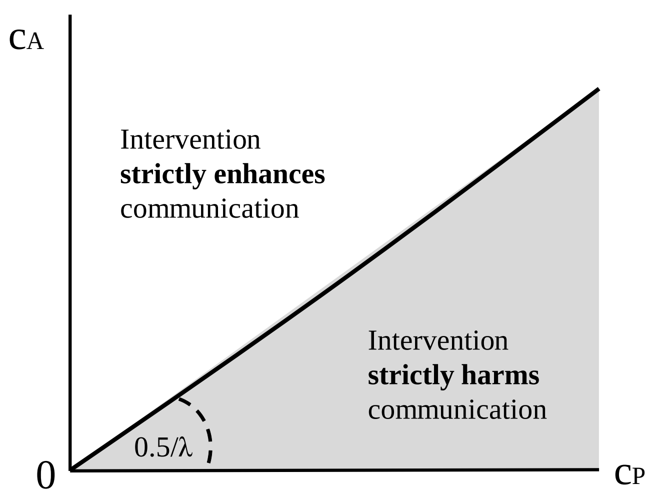


Figure 2 - The effect of intervention on communication

The second implication of Proposition 3 is somewhat surprising: if condition (11) holds, then communication is strictly *less effective* with the option to intervene. How can intervention harm communication? To understand the intuition behind this result, note that the agent is willing to forgo her private benefits and choose action L if he learns that $\theta < -\beta$. However, in equilibrium, the instructions of the principal to choose action L do not reveal whether $\theta < -\beta$ or $\theta \in [-\beta, 0)$. The principal has no incentives to do so, because if she did the agent would have chosen action R when $\theta > -\beta$. Instead, the principal pretends that θ is lower than it really is in order to persuade the agent to choose L even when $\theta > -\beta$. The agent understands the principal's incentives, and hence, the only information that can be inferred from the instructions to choose action L is $\theta < 0$.

Intervention allows the agent to elicit information from the principal that is otherwise not revealed by her instructions. If the agent ignores the principal's instructions, the principal has to decide whether to intervene. Intervention is an informed decision. In equilibrium, the principal intervenes only if she is convinced that the implementation of action R is sufficiently detrimental to justify incurring the costs of intervention. Therefore, the principal's decision to intervene reveals the value of θ relative to $-\frac{c_P}{\lambda}$. In particular, if the principal does not intervene, her decision reveals that $\theta > -\frac{c_P}{\lambda}$. The agent infers that the principal believes that choosing action R did not justify intervention. These are the states in which the agent prefers consuming his private benefits even at the expense of a lower utility for the principal. Indeed, by revealed preferences, the "cost" to the principal is small. In this respect, the principal's decision not to intervene "confirms" the agent's initial decision. On the other hand, if the principal intervenes, her decision reveals that $\theta < -\frac{c_P}{\lambda}$. The agent infers that the principal believes that choosing action R was detrimental. Since the agent is also concerned about the principal's utility, he prefers forgoing his private benefits when he learns that θ is low. In this respect, intervention in those cases benefits the agent since it "corrects" his initial decision when it is indeed detrimental.

We see that by ignoring the principal's instructions, the agent effectively "forces" the principal to make an informed decision which inevitably reveals information about θ she was trying to conceal. The benefit from eliciting this additional information is reflected in the aforementioned "confirmation" and "correction" effects. Against these informational benefits, the agent suffers the direct cost that is imposed by intervention, c_A. Combined, the agent benefits from the principal's intervention if and only if $\theta < -c_A - \beta$. Note that when deciding whether to intervene, the principal behaves as if she is biased toward action R, where the bias is $\frac{c_P}{\lambda}$. Therefore, if $\frac{c_P}{\lambda} \approx c_A + \beta$, the principal's "bias" coincides with the preferences of the agent. As can be seen by (9), the minimum of b^* as a function of c_P is obtained when $\frac{c_P}{\lambda} = c_A + b^*(c_A, c_P)$. While the value of the "correction" effect increases with c_P, the value of the "confirmation" effect decreases with c_P. When $\frac{c_P}{\lambda} = c_A + b^*(c_A, c_P)$ 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 instructions is the lowest. This also explains the intuition behind Corollary 1 and the left panel of Figure 1.

To conclude, the possibility of intervention creates additional tension between the principal and the agent. 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 instructions, then intervention would necessarily enhance communication. Intuitively, with commitment, the principal's decision to intervene does not depend on θ. Therefore, ignoring the instructions of the principal imposes a direct cost on the agent without providing him with the informational 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

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

The first term in (12) is the principal's expected utility from her ability to convince the agent to choose action R when $\theta \geq 0$. The second term is the principal's expected utility from her

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

attempt to convince the agent to choose action L when $\theta < 0$. The second term in the brackets captures the events in which the principal intervenes when her instructions are 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. Notice that this feature differentiates the current paper from models in which the tension between the principal and the agent stems from a hold-up problem. In these models (e.g., Aghion and Tirole (1997)), the agent's incentives to undertake a firm-specific investment decrease with his disutility from intervention (the analog of c_A). Therefore, in those models, the principal's relative benefit from intervention is likely to decrease with c_A.

Second, the principal gets the highest payoff when $c_P = 0$, in which case she can either enforce or threaten to enforce her optimal strategy. Interestingly, if $\lambda = 1$ then

$$\lim_{c_P \to 0} b^*(c_A, c_P) = \begin{cases} \infty & \textit{if } c_A > 0 \\ 0 & \textit{if } c_A = 0. \end{cases} \tag{13}$$

That is, if $c_A > 0$ the principal obtains her first best without ever intervening in equilibrium. By contrast, if $c_A = 0$ the principal obtains her first best because she has to intervene with probability one. Technologically, 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* $\bar{c}_P \in (c_P^*, -\lambda\underline{\theta})$ *such that* $W(c_A, c_P) < \lim_{c_P \to -\lambda\underline{\theta}} W(c_A, c_P)$ *for all* $c_P \in (\bar{c}_P, -\lambda\underline{\theta})$*.*

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 $c_P \in (c_P^*, \bar{c}_P]$ 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 expensive, and hence, less desirable as a substitute for communication. If $c_P > \bar{c}_P$ the principal is better off without the option to intervene. These observations 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, intervention not only harms communication, but is also ineffective on its own. Therefore, the principal prefers effective communication over ineffective intervention.

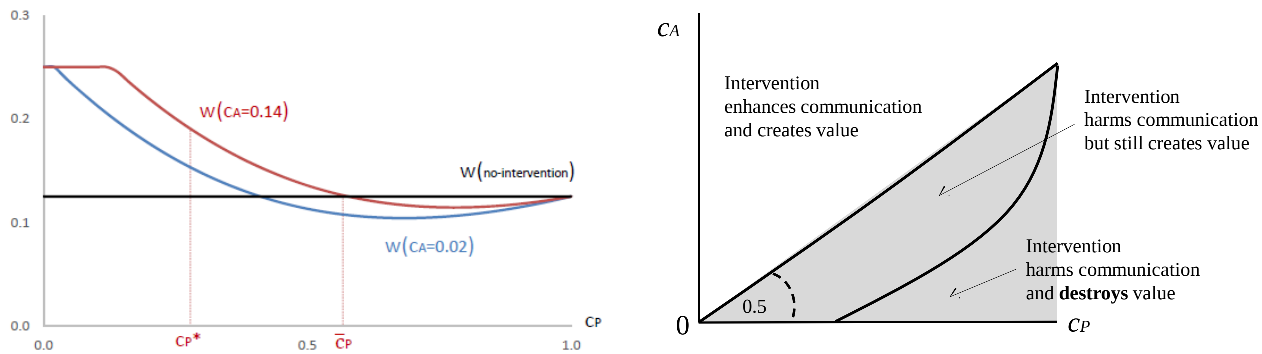


Figure 3 - The value of intervention

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. Trivially, 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 instructions. All else equal, with the option to exit, the principal has fewer 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 instructions to choose action L and the benefit from changing the agent's decision from R to L justifies incurring the cost of intervention. Therefore, the probability of intervention in equilibrium is given by

$$\Pr\left[\beta \geq b^*\left(c_A, c_P\right)\right] \times \Pr\left[\theta < -c_P/\lambda\right]. \tag{14}$$

Expression (14) 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 instructions in order to avoid the costs that are associated with intervention. 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, as can be inferred from Proposition 1, always decreases with c_P when the equilibrium is non-influential. Second, the principal does not need to intervene if the agent follows her instructions. Therefore, if b^* decreases with c_P, the agent is less likely to follow the principal's instructions, and higher c_P can result in a higher probability of intervention. Generally, the probability of intervention increases with c_P if and only if

$$\frac{1-G\left(b^*\right)}{g\left(b^*\right)} < -\frac{\partial b^*}{\partial c_P}\frac{\lambda F\left(-c_P/\lambda\right)}{f\left(-c_P/\lambda\right)}. \tag{15}$$

Condition (15) demonstrates that this comparative static can be reversed once communication is considered, that is, when the equilibrium is influential. 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 less credible, and the agent is more likely to ignore the principal's instructions. 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 in a lower probability of intervention.

Condition (15) has important implications for empirical work as it suggests that the probability of observed intervention is generally non-monotonic with respect to the cost of intervention. In particular, unobserved intervention is not necessarily 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 instructions of the principal. Figure 4 illustrates this point by plotting the probability of intervention in equilibrium as a function of c_P when $\lambda = 1$, $\theta \sim U\left[-1, 1\right]$ and $\beta \sim U\left[0, 1\right]$.

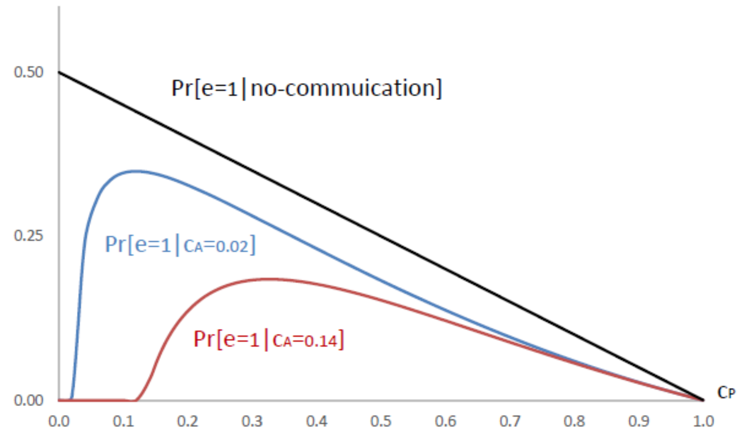


Figure 4 - The probability of intervention in equilibrium

3 Learning from (failed) intervention attempts

The analysis in the previous section suggests that intervention conveys information about θ that is not revealed through direct communications. The agent may use this additional information if he has the opportunity. Such opportunity exists when $\lambda < 1$ and the principal's attempt to intervene fails. So far, however, it was assumed that the agent cannot revise his initial decision. Without this assumption, intervention plays a dual role: it is a corrective tool as well as a channel through which information is transmitted to the agent. Naturally, the interaction between communication and intervention can change when the agent is allowed to revise his initial decision. In this section, I consider the following variant of the baseline model: I assume that if the principal intervenes but intervention fails ($e = 1$ and $\chi = 0$), the agent can 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. Intuitively, the principal can always intervene if she has not done so already. In the Online Appendix, I show that all the influential equilibria that exist under this assumption continue to exist even if it is relaxed.

Consider an influential equilibrium of the modified setup.[16] The analysis of non-influential

[16] When the agent has the option to revise his initial decision, an equilibrium can be influential even if the agent ignores the messages from the principal when making his initial decision, but responds to them when deciding whether to revise the initial decision. In the Online Appendix, I show these equilibria do not exist, and therefore, that all influential equilibria must satisfy Definition 1.

equilibria is given in the Online Appendix. Similar to Section 2, the principal asks the agent to choose action R if and only if $\theta \geq 0$, the agent always follows the instructions to choose action R and the principal never intervenes if the agent follows her instructions. The main departure from the analysis in Section 2 arises when the agent ignores the principal's instructions to choose action L. Let μ_R be the probability that $a_F = R$ in those circumstances. If the principal intervenes, she gets $(1-\lambda)\,\mu_R\theta - c_P$, and otherwise she gets θ. The principal intervenes if and only if

$$\theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}. \tag{16}$$

Lower μ_R implies a higher probability that the agent's decision is revised even if intervention fails. Effectively, intervention is more likely to succeed and the principal has stronger incentives to intervene. In equilibrium, μ_R must reflect the principal's beliefs about the circumstances under which the agent revises his initial decision. The incentives of the agent to revise his initial decision depend on his private benefits. In particular, since the principal intervenes whenever (16) holds, the agent revises his decision from R to L if and only if

$$\beta \leq -\mathbb{E}\left[\theta | \theta < -\frac{c_P}{1-(1-\lambda)\,\mu_R}\right]. \tag{17}$$

Moreover, since the agent's initial decision also depends on β, μ_R must reflect the information about β in that decision as well.

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

- *(i) The principal instructs the agent to choose action R if and only if $\theta \geq 0$. If the principal instructs the agent to choose action R, the agent chooses action R with probability one and the principal never intervenes. If the principal instructs the agent to choose action L, the agent chooses action L if and only if $\beta \leq b^{**}$. If the agent follows the instructions to choose action L, the principal never intervenes. If the agent ignores the instructions 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*

$$b^{**} = c_A \times \frac{\lambda \Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}}\right]}{\Pr\left[-\frac{c_P}{1-(1-\lambda)\mu_R^{**}} < \theta < 0\right]} - \mathbb{E}\left[\theta | -\frac{c_P}{1-(1-\lambda)\,\mu_R^{**}} < \theta < 0\right] \tag{18}$$

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

$$\phi\left(x, y\right) \equiv \Pr\left[\beta > -\mathbb{E}\left[\theta | \theta < -\frac{c_P}{1-(1-\lambda)\,x}\right] | \beta > y\right]. \tag{19}$$

According to Proposition 5, if $c_A \geq \hat{c}_A$, then the equilibrium coincides with the equilibrium in Proposition 2. Intuitively, when c_A is large intervention is costly and ignoring the instructions of the principal to choose action L is risky. The agent is willing to take this risk only if his private benefits from action R are large. Since the principal intervenes only if her instructions are ignored, the additional information in the principal's decision to intervene is not sufficient to convince the agent to revise the initial decision and voluntarily forgo his (large) private benefits. In other words, in this range the agent's option to revise his initial decision is "out of the money". By contrast, when $c_A < \hat{c}_A$ the nature of the equilibrium changes: the agent revises his initial decision with a positive probability and the threshold below which the agent follows the instructions to choose action L, as reflected by (18), is generally different from the one in Proposition 2.

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

When $c_A \leq c_A^*$ the agent is willing to take the risk and ignore the instructions of the principal to choose action L even if his private benefits are relatively small. Since intervention provides a stronger signal that action L is optimal, the agent revises his initial decision. In other words, the agent's option to correct his initial decision is "in the money", and therefore, he has fewer incentives to follow the principal's instructions before observing her attempt to intervene. For this reason, communication in this range is less effective. Notice that when $c_A \in (c_A^*, \hat{c}_A]$ it is possible to construct examples in which $b^{**} > b^*$. Intuitively, if $\mu_R^{**} < 1$ then intervention is more effective from the principal's point of view and she is more likely to intervene. When c_A is not too small ($c_A \in (c_A^*, \hat{c}_A]$), this effect can increase the incentives of the agent to follow the instructions of the principal in order to avoid the cost of intervention. In those cases, communication is more effective.

In the baseline model, the principal never intervenes when $c_P \geq -\lambda\underline{\theta}$, even though technically she could. In this range, intervention is ineffective as a corrective tool, but it can still be used by the principal as a channel to transmit additional information about θ. If this information can convince the agent to revise his initial decision ($\mu_R^{**} < 1$), ex-post, the principal may nevertheless choose to intervene. To illustrate this point, suppose $\lambda = 0$ and $\underline{\theta} = -\infty$. As can be seen by Proposition 5 part (ii.b), $\lambda = 0$ implies that

$$b^{**} = -\mathbb{E}\left[\theta | -\frac{c_P}{1-\mu_R^{**}} < \theta < 0\right], \tag{20}$$

$\mu_R^{**} < 1$, and the principal intervenes if and only if $\theta < -\frac{c_P}{1-\mu_R^{**}}$.[17] Note that (20) is smaller than $-\mathbb{E}\left[\theta|\theta < 0\right]$, 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 of the principal to "burn money".[18] 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.

3.1 When is voluntary revision expected?

The model has implications for the likelihood that the agent voluntarily revises his initial decision in equilibrium. Conditional on failed intervention, the probability of voluntarily revision is given by $1 - \mu_R^{**}$. 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 $b^{**} < \beta$ and (17) holds.

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

[18] Austen-Smith and Banks (2000) and Kartik (2007) study the conditions under which burning money improves communication. Different from their analysis, here, burning money takes place after the initial stage of communication and after the principal observes the agent's decision.

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 instead of choosing action R, triggering costly intervention, and then reversing the initial decision to L if intervention fails.[19]

4 Applications

In this section I discuss several applications of the model.

4.1 Regulation

The tension between a regulator (e.g., the central bank) and a regulatee (e.g., financial intermediaries) 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 macroeconomic indicators, the stability of the financial system as a whole, and future policy. The central bank can use this information to evaluate the consequences of excessive leverage or risk-taking by financial institutions, and propose policies that limit the contribution or exposure of various financial intermediaries to systemic risk. If these intermediaries refuse to comply with the proposed policies, the central bank can intervene by forcing changes on their balance sheets (both on the asset side and on the liability side) or trigger changes in senior management. Intervention is costly since it may require raising taxes or using the policy makers' political capital. The analysis highlights that financial intermediaries are less likely to comply with the proposed policies, even if non-compliance imposes sanctions, if they anticipate that the central bank would save them from a financial meltdown or crisis. Therefore, there are circumstances

[19] 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 Sotheby's board of directors. His effort was hampered 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 the 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.)

where the regulator can be more effective if its capacity to intervene is limited. In this respect, lax regulation can be optimal.

4.2 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 complement 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. General partners can use their expertise to advise founders on 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, general partners are board members and hold various voting and liquidation rights. If a founder is not focused on profit maximization, the general partners can exercise their power to liquidate the start-up. In this respect, the analysis in this paper suggests that venture capital funds can benefit from holding fewer control rights even if the entrepreneur is inexperienced, an observation which is consistent with Kaplan and Stromberg (2004).

4.3 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 on a variety of issues, for example, mergers and acquisitions. 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 in the labor market, which can be proxied by tenure or age. The analysis therefore sheds light on the optimal board composition, suggesting that boards with a low capacity of intervention can be optimal. This result is consistent with Adams, Hermalin, and Weisbach (2010), who observe that corporate boards are often too "friendly" to their CEOs.

4.4 Shareholder activism

Activist investors have a market-wide perspective on assets valuation that corporate boards often lack. In a typical campaign, the activist buys a sizeable stake in a public company and then engages the management or the board of directors, expressing her dissatisfaction or view how the company should be managed. Occasionally, if the company refuses to comply with the activist's demand, the activist ends up litigating or launching a proxy fight in order to gain board seats, and thereby, forcing her 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 the ownership of other institutional investors, the support of proxy advisory firms, and other corporate governance measures of the company. A proxy fight can damage incumbent directors' public image and result in their resignations.

The analysis highlights the factors that might affect the success of a campaign, and suggests that the threat of intervention can be counter-productive. Therefore, the analysis can explain why some activist hedge funds choose to be adversarial (e.g., Pershing Square), while others build reputations for working constructively with management (e.g., ValueAct). Other than putting their reputational capital at risk, activists can commit not to intervene by filing schedule 13-G (instead of schedule 13-D) or by targeting companies for which coordinating with other investors is harder (e.g., dispersed ownership).

Shareholder activism and board structure

The model provides an interesting link between shareholder activism and board structure. Intervention is always optimal ex-post, but 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 board does not intervene even after the CEO ignores its demand. In those instances, an activist may launch a campaign to replace board members with more aggressive and demanding directors, effectively dissolving 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 ex-ante, shareholder activism can nevertheless be undesired.

5 Concluding remarks

In this paper I consider a principal-agent environment in which a privately informed principal can communicate with the agent and subsequently intervene if the agent ignores her instructions. 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. 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 effective deliberation and cooperation, and it further 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 insights derived in this paper can be applied to various contexts: organizations, corporate governance, regulation, politics, diplomacy and households. In some of these applications, the key parameters of the model, the cost that intervention imposes on the agent (c_A) and the cost that is incurred by the principal when she intervenes (c_P), might not only be a function of the environment but also the outcome of an incomplete contract that is agreed upon at the outset. The development of these applications is left for future research.

References

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

[2] Adams, Renee B., Benjamin E. Hermalin, and Michael S. Weisbach. 2010. "The Role of Boards of Directors in Corporate Governance: A Conceptual Framework and Survey." *Journal of Economic Literature*, 48(1): 58-107.

[3] Agastya, Murali, Parimal K. Bag, and Indranil Chakraborty. 2014. "Communication and authority with a partially informed expert." The *RAND Journal of Economics*, 45: 176–197

[4] Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105: 1-29

[5] Alonso, Ricardo, and Niko Matouschek. 2007. "Relational Delegation." The *RAND Journal of Economics*, 38 (4): 1070-1089.

[6] Austen-Smith, David, and Jeffrey S. Banks. 2000. "Cheap Talk and Burned Money." *Journal of Economic Theory*, 91(1),1-16.

[7] Baker, George, Robert Gibbons, and Kevin J. Murphy. 1999. "Informal Authority in Organizations," 15 *Journal of Law, Economics, and Organization* 56–73.

[8] Chakraborty, Archishman, and Bilge Yilmaz. 2011. "Authority, Consensus and Governance." Working paper.

[9] Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50, 1431-1451.

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

[11] Grenadier, Steven R., Andrey Malenko, and Nadya Malenko. 2014. "Timing Decisions in Organizations: Communication and Authority in a Dynamic Environment." Working paper.

[12] Grossman, Sanford J., and Oliver Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94, 691-719.

[13] Grossman, Sanford J., and Motty Perry. 1986. "Perfect Sequential Equilibrium." *Journal of Economic Theory* 39(1), 97-119.

[14] Harris, Milton, and Artur Raviv. 2005. "Allocation of Decision-Making Authority." *Review of Finance* 9, 353–383.

[15] Harris, Milton, and Artur Raviv. 2008. "A Theory of Board Control and Size." *Review of Financial Studies*, 21, 1797-1832.

[16] Harris, Milton, and Artur Raviv. 2010. "Control of corporate decisions: Shareholders vs. management." *Review of Financial Studies* 23:4115–47

[17] Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98, 1119-1158.

[18] Kaplan, Steven N., and Per Strömberg. 2003. "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts." *Review of Economic Studies* 70, 281–315.

[19] Kartik, Navin. 2007. "A Note On Cheap Talk and Burned Money." *Journal of Economic Theory*, 136(1), 749-758

[20] Levit, Doron. 2013. "Soft Shareholder Activism." Working paper.

[21] Levit, Doron, and Nadya Malenko. 2011. "Nonbinding Voting for Shareholder Propoals." *Journal of Finance*, 66(5), 1579-1614.

[22] Marino, Anthony, John Matsusaka, and Jan Zábojník. 2010. "Disobedience and Authority," 26 *Journal of Law, Economics and Organization* 427–59.

[23] Matthews, Steven A. 1989. “Veto Threats: Rhetoric in a Bargaining Game.” *Quarterly Journal of Economics*, 104(2): 347-69.

[24] Shimizu, Takashi. 2008. “Cheap Talk with the Exit Option: a Model of Exit and Voice, Working paper.

[25] Van den Steen, Eric. 2010. “Interpersonal authority in a theory of the firm”. *American Economic Review* 100 (1), 466–490.

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

(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 θ is consistent with Bayes' rule.*

(iii) *For any θ 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 χ.*

(iv) *For any θ 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 χ.*

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

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 $\theta > \frac{c_P}{\lambda}$. Therefore, if $a_A = R$, the agent's expected utility is

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

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

$$\Pr\left[\theta > \frac{c_P}{\lambda}|m_0\right] \lambda \left(\mathbb{E}\left[\theta + \beta|\theta > \frac{c_P}{\lambda}, m_0\right] - c_A\right). \tag{22}$$

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

$$\begin{aligned} b(m_0) \equiv\ & c_A \frac{\Pr\left[\theta < -\frac{c_P}{\lambda}|m_0\right] - \Pr\left[\theta > \frac{c_P}{\lambda}|m_0\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}|m_0\right]} \\ & - \frac{\frac{1-\lambda}{\lambda}\mathbb{E}\left[\theta|m_0\right] + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}|m_0\right] \mathbb{E}\left[\theta| -\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}, m_0\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}|m_0\right]} \end{aligned} \tag{23}$$

as required. ■

Proof of Proposition 1. Given message m_0, the agent chooses action R if and only if $\beta \geq b(m_0)$, where $b(m_0)$ is given by (23). When the equilibrium is non-influential, no message is informative about θ. Therefore, $b(m_0) = b^{NI}$ for any m_0, where

$$\begin{aligned} b^{NI} =\ & c_A \times \frac{\Pr\left[\theta < -\frac{c_P}{\lambda}\right] - \Pr\left[\theta > \frac{c_P}{\lambda}\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right]} \\ & - \frac{\frac{1-\lambda}{\lambda}\mathbb{E}\left[\theta\right] + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right] \mathbb{E}\left[\theta| -\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right]}{\frac{1-\lambda}{\lambda} + \Pr\left[-\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right]}. \end{aligned} \tag{24}$$

Since f is symmetric and $\mathbb{E}[\theta] \geq 0$, $\mathbb{E}\left[\theta| -\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda}\right] \geq 0$ 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, $\Pr[a_A(m, \beta) = R] = 1 - G(b(m))$, where $b(m)$ is given by (23). Thus, if the principal sends message m, then her expected utility is:

$$\mathbb{E}\left[u_P(\theta, a_A, a_P, e, \chi)|\theta\right] = \begin{cases} (1 - G(b(m)))((1-\lambda)\theta - c_P) & \text{if } \theta < -\frac{c_P}{\lambda} \\ (1 - G(b(m)))\theta & \text{if } -\frac{c_P}{\lambda} < \theta < \frac{c_P}{\lambda} \\ (1 - G(b(m)))\theta + G(b(m))(\lambda\theta - c_P) & \text{if } \frac{c_P}{\lambda} < \theta \end{cases} \tag{25}$$

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 (23). 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 instructions and chooses action R with probability one. Suppose $m_0 \in M_L$. According to Lemma 3, $\theta < 0$. Therefore, (23) can be rewritten as

$$\begin{aligned} b(m_0) &= c_A \times \frac{\lambda \Pr\left[\theta < -\frac{c_P}{\lambda} | m_0\right]}{1 - \lambda \Pr\left[\theta < -\frac{c_P}{\lambda} | m_0\right]} \\ &\quad + \frac{\lambda \Pr\left[\theta < -\frac{c_P}{\lambda} | m_0\right] \mathbb{E}\left[\theta | \theta < -\frac{c_P}{\lambda}, m_0\right] - \mathbb{E}\left[\theta | m_0\right]}{1 - \lambda \Pr\left[\theta < -\frac{c_P}{\lambda} | m_0\right]}. \end{aligned} \tag{26}$$

Since $b(m_0) = \max_{m \in M} b(m)$ for all $m_0 \in M_L$, $b(m_0)$ is invariant to $m_0 \in M_L$. Since $m_0 \in M_L$ if and only if $\theta < 0$, some algebra and an integration over all $m_0 \in M_L$ show that $b(m_0) = b^*$, where b^* is given by (8). Therefore, in any influential equilibrium, the agent follows the principal's instructions 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 instructions if he observes message m_R. Since $m = m_L$ if and only if $\theta < 0$, (23) evaluated at m_L can be rewritten as b^*. Thus, the agent follows the principal's instructions 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 (Lemma 3). Finally, note that

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

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: $b^*(0) > 0$, $b^*(c_P) = -\mathbb{E}[\theta|\theta < 0]$ for all $c_p \geq -\lambda\underline{\theta}$, and if $c_P \in [0, -\lambda\underline{\theta})$, then

$$\frac{\partial b^*(c_P)}{\partial c_P} = \frac{f\left(-\frac{c_P}{\lambda}\right)}{F(0) - \lambda F\left(-\frac{c_P}{\lambda}\right)}\left[h(c_p) - b^*(c_P)\right], \tag{27}$$

where $h(c_p) \equiv \frac{1}{\lambda}c_P - c_A$. Note that (27) implies $\frac{\partial b^*(c_P)}{\partial c_P} > 0 \Leftrightarrow h(c_p) > b^*(c_P)$.

Suppose there is $c'_P \in [0, -\lambda\underline{\theta}]$ such that $h(c'_P) = b^*(c'_P)$. I argue $c_P \in [0, c'_P) \Rightarrow h(c_P) < b^*(c_P)$ and $c_P > c'_P \Rightarrow h(c_P) > b^*(c_P)$. If true, based on (27), c'_P is the unique minimum of $b^*(c_P)$. To prove this argument, I show that $c_P > c'_P \Rightarrow h(c_P) > b^*(c_P)$. Consider two cases. First, suppose $c'_P = -\lambda\underline{\theta}$. Recall, $c_P \geq -\lambda\underline{\theta} \Rightarrow b^*(c_P) = -\mathbb{E}[\theta|\theta < 0]$. Since $h'(c_p) = \frac{1}{\lambda} > 0$ and $h(-\lambda\underline{\theta}) = -\mathbb{E}[\theta|\theta < 0]$, $h(c_p) > b^*(c_P)$ for all $c_P > c'_P$ as required. Second, suppose $c'_P \in [0, -\lambda\underline{\theta})$ and on the contrary there is $c''_P > c'_P$ such that $h(c''_P) \leq b^*(c''_P)$. Note that both $b^*(c_P)$ and $h(c_P)$ are continuous, where $\frac{\partial b^*(c_P)}{\partial c_P}|_{c_P = c'_P} = 0$ and $\frac{\partial h(c_P)}{\partial c_P}|_{c_P = c'_P} = \frac{1}{\lambda} > 0$. Therefore, there is $\varepsilon > 0$ such that $c_P \in (c'_P, c'_P + \varepsilon) \Rightarrow h(c_P) > b^*(c_P)$. Moreover, since $h(c''_P) \leq b^*(c''_P)$ there is $c'''_P \in (c'_P, c''_P]$ such that $h(c'''_P) = b^*(c'''_P)$ and $h(c_P) > b^*(c_P)$ for all $c_P \in (c'_P, c'''_P)$. However, based on (27), $h(c'''_P) = b^*(c'''_P)$ implies $\frac{\partial b^*(c_P)}{\partial c_P}|_{c_P = c'''_P} = 0 < \frac{1}{\lambda} = \frac{\partial h(c_P)}{\partial c_P}|_{c_P = c'''_P}$. From continuity, there is $\delta > 0$ such that $h(c_P) < b^*(c_P)$ for all $c_P \in (c'''_P - \delta, c'''_P)$, yielding a contradiction. This proves $c_P > c'_P \Rightarrow h(c_P) > b^*(c_P)$. The proof of $c_P \in [0, c'_P) \Rightarrow h(c_P) < b^*(c_P)$ follows a similar set of arguments, and for brevity, it is omitted.

The previous argument shows that if there is $c'_P \in [0, -\lambda\underline{\theta}]$ such that $h(c'_P) = b^*(c'_P)$, then c'_P is unique. Note that $h(0) < b^*(0)$. Moreover, $h(-\lambda\underline{\theta}) > b^*(-\lambda\underline{\theta})$ if and only if $c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$. Therefore, if $c_A < \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$, then from continuity there is a unique $c_P^{\min} \in (0, -\lambda\underline{\theta})$ such that $h\left(c_P^{\min}\right) = b^*\left(c_P^{\min}\right)$. Based on the argument above, $c_P^{\min}$ is the unique minimum of $b^*(c_P)$. If $c_A \geq \mathbb{E}[\theta|\theta < 0] - \underline{\theta}$ and on the contrary there is c'_P such that $h(c'_P) = b^*(c'_P)$, then it has to be that $h(-\lambda\underline{\theta}) > b^*(-\lambda\underline{\theta})$, which yields a contradiction. $h(0) < b^*(0)$ implies $h(c_P) < b^*(c_P)$ for all c_P, and hence, $\frac{\partial b^*(c_P)}{\partial c_P} < 0$ as required. ■

Proof of Proposition 4. Based on (12), (8) and some algebra, $W(c_A, c_P) < W(c_A, -\lambda\underline{\theta})$ if and only if

$$\frac{G\left(-\mathbb{E}\left[\theta|\theta<0\right]\right)-G\left(b^*\left(c_P\right)\right)}{1-G\left(b^*\left(c_P\right)\right)}>\frac{\int_{\underline{\theta}}^{-\frac{c_P}{\lambda}}\left(\lambda\theta+c_P\right)dF\left(\theta\right)}{\int_{\underline{\theta}}^{0}\theta dF\left(\theta\right)}. \tag{28}$$

Note that the right hand side of (28) is strictly positive. Based on Corollary 1, its proof and the properties of $b^*\left(c_P\right)$, if $c_A<\mathbb{E}\left[\theta|\theta<0\right]-\underline{\theta}$ or $c_P\leq c_P^*$ (where c_P^* is the unique solution of (11)), then $b^*\left(c_P\right)\geq-\mathbb{E}\left[\theta|\theta<0\right]$, and hence, (28) never holds. Therefore, it is necessary that $c_A>\mathbb{E}\left[\theta|\theta<0\right]-\underline{\theta}$ and $c_P>c_P^*$.

Next, note that both sides of (28) converge to zero as $c_P\rightarrow-\lambda\underline{\theta}$. The derivatives with respect to c_P of the right and left hand side, respectively, are given by

$$\begin{aligned}
RHS\left(c_P\right) &= F\left(-c_P/\lambda\right)/\int_{\underline{\theta}}^{0}\theta dF\left(\theta\right)\\
LHS\left(c_P\right) &= -\frac{\partial b^*\left(c_P\right)}{\partial c_P}g\left(b^*\left(c_P\right)\right)\frac{1-G\left(-\mathbb{E}\left[\theta|\theta<0\right]\right)}{\left[1-G\left(b^*\left(c_P\right)\right)\right]^2},
\end{aligned}$$

where $\frac{\partial b^*(c_P)}{\partial c_P}$ is given by (27). Note that $RHS\left(c_P\right)<0$, $\lim_{c_P\rightarrow-\lambda\underline{\theta}}RHS\left(c_P\right)=0$, $LHS\left(c_P\right)<0$, and

$$\lim_{c_P\rightarrow-\lambda\underline{\theta}}LHS\left(c_P\right)=f\left(\underline{\theta}\right)\frac{c_A-\mathbb{E}\left[\theta|\theta<0\right]+\underline{\theta}}{F\left(0\right)}\frac{g\left(-\mathbb{E}\left[\theta|\theta<0\right]\right)}{1-G\left(-\mathbb{E}\left[\theta|\theta<0\right]\right)}<0$$

From continuity, there is $\bar{c}_P'\in\left(0,-\lambda\underline{\theta}\right)$ such that if $c_P\in\left(\bar{c}_P,-\lambda\underline{\theta}\right)$ then $LHS\left(c_P\right)<RHS\left(c_P\right)<0$, and hence, there is $\bar{c}_P\in\left(\bar{c}_P',-\lambda\underline{\theta}\right)$ such that (28) holds. ■

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

Proof. According to Proposition 1, the principal's expected payoff in any non-influential equilibrium is

$$\mathbb{E}\left[u_P^{NI}\right]=\Pr\left[\theta>-\frac{c_P}{\lambda}\right]\mathbb{E}\left[\theta|\theta>-\frac{c_P}{\lambda}\right]+\Pr\left[\theta<-\frac{c_P}{\lambda}\right]\left(\left(1-\lambda\right)\mathbb{E}\left[\theta|\theta<-\frac{c_P}{\lambda}\right]-c_P\right), \tag{29}$$

and the agent's expected payoff (conditional on β) is given by

$$\mathbb{E}\left[u_A^{NI}\left(\beta\right)\right]=\mathbb{E}\left[\theta+\beta\right]-\Pr\left[\theta<-\frac{c_P}{\lambda}\right]\left(\lambda\mathbb{E}\left[\theta+\beta|\theta<-\frac{c_P}{\lambda}\right]+c_A\right). \tag{30}$$

According to Proposition 2, the principal's expected payoff in any influential equilibrium, $\mathbb{E}\left[u_P^I\right]$, is given by (12). The agent's expected payoff (conditional on β) is given by

$$\mathbb{E}\left[u_A^I(\beta)\right] = \begin{cases} \Pr[\theta > 0]\,\mathbb{E}[\theta + \beta | \theta > 0] & \text{if } \beta \leq b^* \\ \mathbb{E}\left[u_A^{NI}(\beta)\right] & \text{else,} \end{cases} \tag{31}$$

where b^* is given by (8). A direct comparison shows that $\mathbb{E}\left[u_A^I(\beta)\right] \geq \mathbb{E}\left[u_A^{NI}(\beta)\right]$ for all β, and $\mathbb{E}\left[u_P^I\right] \geq \mathbb{E}\left[u_P^{NI}\right]$. ∎

A.2 Proofs of Section 3

The next definition 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 $\beta_0 \in \left[0, \overline{\beta}\right]$ 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. As in Lemma 3, the principal sends message $m \in M_R$ if and only if $\theta > 0$. Consider two cases.

First, suppose $m \in M_R$. If $a_A = R$ then $e = 0$ 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 $e = 1$, then 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 $c_A > 0$. Therefore, if $m \in M_R$, then $a_A = R$ with probability one and $e = 0$.

Second, suppose $m \in M_L$. If $a_A = L$, then $e = 0$ and the agent gets zero. If $a_A = R$, then, as was argued in the main text, $e = 1 \Leftrightarrow \theta < -\frac{c_P}{1-(1-\lambda)\mu_R}$. If $e = 1$ and $\chi = 0$, then $a_F = R \Leftrightarrow \mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{1-(1-\lambda)\mu_R}, m\right] \geq 0$. Overall, if $m \in M_L$ and $a_A = R$, then the agent's expected payoff is:[20]

$$\Pr\left[\theta > -\frac{c_P}{1-(1-\lambda)\mu_R}|\theta<0\right]\mathbb{E}\left[\theta+\beta| -\frac{c_P}{1-(1-\lambda)\mu_R}<\theta<0\right] \tag{32}$$
$$+\Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}|\theta<0\right]\left((1-\lambda)\max\left\{0,\mathbb{E}\left[\theta+\beta|\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]\right\}-\lambda c_A\right).$$

Therefore, $a_A = L$ if and only if (32) is negative, which holds if and only if

$$\beta \leq \min\left\{-\mathbb{E}\left[\theta|\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right],\eta(\mu_R)\right\} \tag{33}$$

or

$$-\mathbb{E}\left[\theta|\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right] < \beta \leq \delta(\mu_R), \tag{34}$$

where

$$\eta(\mu_R) = \lambda c_A \frac{\Pr\left[\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr\left[-\frac{c_P}{1-(1-\lambda)\mu_R}<\theta<0\right]} - \mathbb{E}\left[\theta| -\frac{c_P}{1-(1-\lambda)\mu_R}<\theta<0\right] \tag{35}$$

and

$$\begin{aligned}\delta(\mu_R) &= \lambda c_A \frac{\Pr\left[\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr[\theta<0]-\lambda\Pr\left[\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]} \\ &\quad - \frac{\Pr[\theta<0]\,\mathbb{E}[\theta|\theta<0]-\lambda\Pr\left[\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]\mathbb{E}\left[\theta|\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr[\theta<0]-\lambda\Pr\left[\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}.\end{aligned} \tag{36}$$

Define

$$h(\mu_R) \equiv \frac{\Pr[\theta<0]}{\lambda}\frac{\mathbb{E}[\theta|\theta<0]-\mathbb{E}\left[\theta|\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}{\Pr\left[\theta< -\frac{c_P}{1-(1-\lambda)\mu_R}\right]}. \tag{37}$$

[20] Note that if $m \in M_L$, then the only information about θ that can be revealed in equilibrium is that θ is negative. Indeed, the principal benefits if the agent believes that θ is as small as possible: if the agent believes that θ 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.

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 (33) and (34), there are two cases to consider. First, suppose $c_A \geq h(1)$. Since $h(\cdot)$ is an increasing function (can be verified), then $c_A \geq h(\mu_R)$ for all $\mu_R \in [0,1]$. Therefore, in equilibrium, it must be $a_A = L \Leftrightarrow \beta \leq \delta(\mu_R)$. If the agent ignores the instructions to choose L, then $\delta(\mu_R) < \beta$. Since $c_A \geq h(\mu_R)$ implies $-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] \leq \delta(\mu_R)$, it must be that $\mu_R = 1$ 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 μ_R^{**} satisfies $c_A \geq h(\mu_R^{**})$. In that case, it must be $a_A = L \Leftrightarrow \beta \leq \delta(\mu_R^{**})$, and similar to the argument in the previous case, $\mu_R^{**} = 1$. Since $c_A < h(1)$, it is a contradiction. Therefore, in equilibrium $\mu_R^{**} < 1$ satisfies $c_A < h(\mu_R^{**})$. In this case, $a_A = L \Leftrightarrow \beta \leq \eta(\mu_R^{**})$. 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, μ_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 $c_A < h(y)$. $c_A < h(1)$ implies that $\eta(1) < -\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right]$, and hence, $\phi(1, \eta(1)) < 1$. Let $\underline{y} \equiv \max\{0, h^{-1}(c_A)\}$, 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 $\hat{c}_A \equiv h(1)$ concludes the proof. ■

Proof of Corollary 2. Let $\eta(\mu_R, c_A)$ and $\delta(\mu_R, c_A)$ be defined by (35) and (36), respectively. Note that for any $c_A \leq h(1)$, $b^{**} = \eta(\mu_R^{**}(c_A), c_A)$ and $b^* = \delta(1, c_A)$. I prove that there is $c_A^* \in (0, h(1)]$ such that $c_A \leq c_A^* \Rightarrow \eta(\mu_R^{**}(c_A), c_A) < \delta(1, c_A)$. Note that $\eta(\mu_R, c_A)$ 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 μ_R, and $\eta(\mu_R, 0) < \delta(\mu_R, 0)$ for all μ_R. Therefore, $\eta(1,0) < \delta(1,0)$. Moreover, note that $\eta(\mu_R, 0)$ increases with

μ_R. Therefore, if $\mu_R < 1$ then $\eta(\mu_R, 0) < \eta(1, 0) < \delta(1, 0)$. Since $h(1) > 0$, $\mu_R^{**}(0) < 1$, 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. ■

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 $\mu_R^{**} = \phi(\mu_R^{**}, \eta(\mu_R^{**}, c_A))$. An application of the implicit function theorem implies

$$\frac{\partial \mu_R^{**}}{\partial c_A} = \frac{\phi(\mu_R^{**}, \eta(\mu_R^{**}, c_A)) \frac{g(\eta(\mu_R^{**}, c_A))}{1 - G(\eta(\mu_R^{**}, c_A))}}{1 - \frac{\phi(\mu_R^{**}, \eta(\mu_R^{**}, c_A))}{\partial \mu_R}} \frac{\partial \eta(\mu_R^{**}, c_A)}{\partial c_A}. \tag{38}$$

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

Online Appendix for "Governing Through Communication and Intervention"

Doron Levit

The Online Appendix has two sections. The first section contains supplemental results for Section 3. The second section contains two extensions to the baseline model: continuum of actions and verifiable information.

A Supplemental results for Section 3

Proposition 7 *A non-influential equilibrium always exists. In any non-influential equilibrium there are b_{NI}^{**} and $\mu_{NI}^{**} \in (0,1)$ such that following hold: the agent chooses action L if and only if $\beta \leq b_{NI}^{**}$. If the agent chooses action L then the principal intervenes if and only if $\theta > c_P$, and upon failed intervention the agent revises his initial decision with probability one. If the agent chooses action R then the principal intervenes if and only if $\theta < -\frac{c_P}{1-(1-\lambda)\mu_{NI}^{**}}$, and upon failed intervention the agent revises his initial decision with probability $1-\mu_{NI}^{**}$, where μ_{NI}^{**} solves $\mu_{NI}^{**} = \phi\left(\mu_{NI}^{**}, b_{NI}^{**}\right)$ and ϕ is given by (19).*

Proof. Let μ_{a_A} be the probability that $a_F = R$ if $a_A \in \{L, R\}$ and intervention fails. Suppose $a_A = L$. The principal gets $\lambda\theta + (1-\lambda)\mu_L\theta - c_P$ if she intervenes, and zero otherwise. Therefore, $e = 1 \Leftrightarrow \theta > \frac{c_P}{\lambda+(1-\lambda)\mu_L}$. Since $\mathbb{E}\left[\theta + \beta | \theta > \frac{c_P}{\lambda+(1-\lambda)\mu_L}\right] > 0$ for all $\beta > 0$ and $\mu_L \in [0,1]$, then in any equilibrium it must be $\mu_L = 1$. The agent's expected utility from $a_A = L$ is

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

Suppose $a_A = R$. The principal gets $(1-\lambda)\mu_R\theta - c_P$ if she intervenes, and θ otherwise. Therefore, $e = 1 \Leftrightarrow \theta < -\frac{c_P}{1-(1-\lambda)\mu_R}$. If $e = 1$ and $\chi = 0$, then $a_F = R \Leftrightarrow \mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{1-(1-\lambda)\mu_R}\right] \geq$

0. The agent's expected utility from $a_A = R$ 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

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

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

Consider the following condition,

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

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,

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

Define

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

Condition (43) holds if and only if

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

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

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

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

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

We conclude, there are two cases:

1. If $\overline{\mu} \ge 1$, then the agent chooses $a_A = L$ if and only if $\beta \le \delta'(1)$, where $\mu_{NI}^{**} = 1$. Note that $b_{NI}^{**} \equiv \delta'(1) > 0$.

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

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

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 $\beta \in \left[0, \overline{\beta}\right]$, $a_A\left(m_1, \beta\right) = a_A\left(m_2, \beta\right)$. Moreover, there are $m_1, m_2 \in M$ and β_0 such that $a_F\left(m_1, \beta_0\right) \neq a_F\left(m_2, \beta_0\right)$. Similar to the arguments in the proof of Proposition 7, $\mu_L = 1$. Since $a_A\left(m, \beta\right)$ is invariant to m but $a_F\left(m_1, \beta_0\right) \neq a_F\left(m_2, \beta_0\right)$, it has to be $a_A\left(\beta, m\right) = R$ for all $\beta \in \left[0, \overline{\beta}\right]$ and $m \in M$. Suppose that, without the loss of generality

$$\Pr\left[a_F = R | a_A = R, e = 1, \chi = 0, m_1\right] < \Pr\left[a_F = R | a_A = R, e = 1, \chi = 0, m_2\right].$$

Note that if $a_A = R$ and $e = 1$ then $\theta < 0$. Therefore, the principal strictly prefers m_1 over m_2, a contradiction to $m_2 \in M$. ■

Lemma 5 *Suppose $\lambda = 0$ and $c_P < -\underline{\theta}\frac{G(-\underline{\theta})-G(-\mathbb{E}[\theta|\theta<0])}{1-G(-\mathbb{E}[\theta|\theta<0])}$. An influential equilibrium in which intervention is off the equilibrium path does not survive the Grossman and Perry (1986) criterion. Moreover, an influential equilibrium that survives the Grossman and Perry (1986) criterion always exists, and it satisfies Proposition 5 part (ii.b).*

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

$$\hat{\mu}_R = \frac{1 - G\left(-\mathbb{E}\left[\theta | \theta < -\frac{c_P}{1-\hat{\mu}_R}\right]\right)}{1 - G\left(-\mathbb{E}\left[\theta | \theta < 0\right]\right)}, \tag{46}$$

and note that since $c_P < -\underline{\theta}\frac{G(-\underline{\theta})-G(-\mathbb{E}[\theta|\theta<0])}{1-G(-\mathbb{E}[\theta|\theta<0])}$ then a solution such that $\underline{\theta} < -\frac{c_P}{1-\hat{\mu}_R}$ 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 (46), if the principal deviates and intervenes, she 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 θ. Therefore, the principal intervenes if and only if $\theta < -\frac{c_P}{1-\hat{\mu}_R}$. The existence of this deviation violates the Grossman and Perry (1986) criterion.

Next, note that any equilibrium that is described by Proposition 5, both $e = 1$ and $e = 0$ are on the equilibrium path, and hence, the Grossman and Perry (1986) criterion is trivially satisfied. Based on the proof of Proposition 5, if $\lambda = 0$ then $c_A < h(\mu_R)$ for any $\mu_R \in [0, 1)$. Clearly, if $\mu_R^{**} = 1$ then the $\lambda = 0$ implies that the principal never intervenes. Therefore, if there is an influential equilibrium in which $e = 1$ is on the equilibrium path, then b^{**} and $\mu_R^{**} < 1$ must satisfy part (ii.b). Since $c_P < -\underline{\theta}\frac{G(-\underline{\theta})-G(-\mathbb{E}[\theta|\theta<0])}{1-G(-\mathbb{E}[\theta|\theta<0])}$ then

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

has a solution where $\underline{\theta} < -\frac{c_P}{1-\mu_R}$, μ_R^{**} is given by this solution and b^{**} is given by (20). By construction, an influential equilibrium as in Proposition 5 part (ii.b) indeed exists. ■

Proposition 8 *All the equilibria in Proposition 5 continue to exist when the agent is allowed to revise his initial decision when $e = 0$.*

Proof. Consider an equilibrium as described by Proposition 5. Suppose $m \in M_R$. In any equilibrium described by Proposition 5, $a_A = R$ for sure. Since $m \in M_R \Rightarrow \theta > 0$, based on (3), the agent has incentives to maintain his original decision even if the principal does not intervene, and as in Proposition 5, $e = 0$ for sure. Suppose $m \in M_L$. Based on Proposition 5, $a_A = R \Leftrightarrow \beta > b^{**}$, and $e = 1$ if and only if $a_A = R$ and $\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}}$. If $a_A = L$ then

$e = 0$ for sure. The agent does not infer new information from $e = 0$ and his expected payoff is therefore zero. If $a_A = R$ and $e = 0$ then the agent infers $-\frac{c_P}{1-(1-\lambda)\mu_R^{**}} < \theta < 0$, and he revises his decision to L if and only if

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

Suppose $\beta > b^{**}$. From the proof of Proposition 5, it can be verified that the right hand side of (48) is strictly smaller than b^{**}. Therefore, $\beta > b^{**}$ implies that (48) never holds. That is, the agent has no incentives to revise his decision to L if $e = 0$. The agents's expected payoff from choosing R is given by (32). By construction, $\beta > b^{**}$ implies that (32) is non-negative and the agent is better off choosing action R as prescribed by Proposition 5.

Suppose $\beta \leq b^{**}$. If (48) is violated then the agent has no incentives to revise his decision to L if $e = 0$, and his expected payoff from choosing R is given by (32). By construction, $\beta < b^{**}$ implies that (32) is negative. Therefore, the agent is better off choosing action L as prescribed by Proposition 5. If (48) holds then the agent revises his decision to L if $e = 0$. Note that (48) implies $\mathbb{E}\left[\theta + \beta | \theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}}\right] \leq 0$. Therefore, if $e = 1$ and $\chi = 0$ then the agent always revises his decision to L. The agent's expected payoff from choosing R is $-\lambda c_A \Pr\left[\theta < -\frac{c_P}{1-(1-\lambda)\mu_R^{**}} | \theta < 0\right] < 0$. Therefore, the agent is better off choosing action L as prescribed by Proposition 5.

To conclude, the agent's initial decision as prescribed by Proposition 5 does not change if he has option to revise it when $e = 0$. Therefore, the principal's communication and intervention strategies, as prescribed Proposition 5, are also incentive compatible, and all the equilibria in Proposition 5 continue to exist as required. ■

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

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

where $\mu_R^{**}(c_A, c_P)$ is given by Proposition 5 part (ii.b).

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 $-\mathbb{E}\left[\theta|\theta < -\frac{c_P}{\lambda}\right] > -\mathbb{E}[\theta|\theta < 0]$, $c_A \geq h(1)$ implies $\delta(1, c_A) > -\mathbb{E}[\theta|\theta < 0]$, and intervention enhances communication. If $c_A < h(1)$, then $b^{**} = \eta(\mu_R^{**}(c_A), c_A)$ and

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

and note that

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

Combined, intervention enhances communication if and only if (49) 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

$$v(\theta, a) = -(\theta - a)^2. \tag{50}$$

For simplicity, I assume that the agent's bias β 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 θ.

Proposition 10 *Let $\Lambda(\beta, c_P, c_A)$ be 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 θ, regardless of the message that the principal sent the agent, the principal solves

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

Thus, if the agent chooses action a_A, the principal's utility conditional on θ is

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

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

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

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 β. 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 θ 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 θ. To keep the analysis simple, I assume $\lambda=1$ and $c_A>0$.

Suppose the principal discloses θ. 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,

$$\Pr\left[a_A=R|\rho(\theta)=\theta\right]=\begin{cases}0 & \textit{if } \theta\leq -c_P\\ \Pr\left[\theta\geq-\beta\right] & \textit{if } \theta>-c_P,\end{cases}\tag{51}$$

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 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* $\overline{\beta} \leq \bar{b}_{\text{ver}}$, *where*

$$\bar{b}_{\text{ver}} \equiv \max_{\Upsilon:[\max\{-c_P,-\overline{\beta}\},0]\subseteq\Upsilon\subseteq[\underline{\theta},0]} \left\{ c_A \frac{\Pr\left[\theta < -c_P|\theta \in \Upsilon\right]}{1 - \Pr\left[\theta < -c_P|\theta \in \Upsilon\right]} - \mathbb{E}\left[\theta|\theta \geq -c_P, \theta \in \Upsilon\right] \right\}. \tag{52}$$

In this equilibrium, $\left[\max\left\{-c_P, -\overline{\beta}\right\}, 0\right] \subseteq \Upsilon^* \subseteq [\underline{\theta}, 0]$ *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 θ 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 (51). 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 θ, thereby saving 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 θ 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, that 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).[21]

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

Last, consider an equilibrium with $\varphi^* = 0$. Since $\varphi^* = 0$, the principal has strict incentives to disclose θ when $\theta > 0$. Moreover, if $\theta \in \left[\max\left\{-c_P, -\overline{\beta}\right\}, 0\right]$, the principal has strict incentives to conceal θ, since if she discloses θ, there is a strictly positive probability that the agent chooses $a = R$. If $\theta \in \left[\underline{\theta}, \max\left\{-c_P, -\overline{\beta}\right\}\right]$, the principal is indifferent between disclosing and concealing θ, 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 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 between expression (52) and 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 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 Υ 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 $\overline{\beta} > -\mathbb{E}[\theta| - c_P \leq \theta < 0]$, $\overline{b}_{\text{ver}}(c_A = 0, c_P) < \overline{\beta}$, 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 $-\overline{\beta} < \mathbb{E}[\theta| - c_P \leq \theta < 0]$, $\mathbb{E}[\theta| - c_P \leq \theta < 0] \leq \mathbb{E}[\theta|\theta \geq -c_P, \theta \in \Upsilon]$. 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

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

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. ■