When Words Speak Louder Than Actions*

Doron Levit†

Wharton

January 17, 2016

Abstract

This paper studies communication and intervention as mechanisms of conflict resolution. 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 prompt disobedience, since it tempts the agent to challenge the principal *to back her words with actions.* This result provides a novel argument as to why a commitment not to intervene (and therefore, relying solely on communication) can be optimal. In this respect, *words do speak louder than actions.* The analysis sheds new light on the effectiveness of different organizational structures, parenting styles, and corporate governance arrangements.

Keywords: Communication, Delegation, Intervention, Organization, Cheap-Talk, Authority, Obedience.

JEL Classification: D74, D82, D83, G34

*The paper was previously circulated under the title "Governing Through Communication and Intervention".

†The author is from the University of Pennsylvania, Wharton School, Finance Department. I thank Adrian Corum and Michael Lee for excellent research assistance. I also thank Simon Gervais, Vincent Glode, Itay Goldstein, Richard Kilstrom, Andrey Malenko, Christian Opp, Bilge Yilmaz, participants at the Finance Theory Group in Austin 2015, and seminar participants at Interdisciplinary Center in Herzliya, London Business School, London School of Economics, and University of Pennsylvania for helpful comments. Email: dlevit@wharton.upenn.edu.

"Actions speak louder than words, but not nearly as often." Mark Twain

Introduction

In many real world situations, a conflict of interests is resolved if one party is either *persuaded* or *forced* by the other party to change the course of action. For example, most managers would first explain their subordinates why the task has to be performed, especially if it is unpleasant, and overrule them only if their instructions are ignored. Parents would typically speak to their children about their expected behavior before imposing any sanctions. Sophisticated investors (e.g., private equity funds or activist hedge funds), who often demand firms to restructure their balance-sheet and operations, would exercise their control rights only if their ideas are rejected or ignored by the management of the firm.

In all of the situations above, and in many others, communication and intervention are the primary mechanisms of conflict resolution. Communication involves transmission of information, and its effectiveness depends on the credibility of the message. By contrast, intervention involves unilateral enforcement of one's will, and therefore, it requires non-trivial amount of effort, time, and resources. Since communication is less confrontational and costly than intervention, there is a natural "pecking order" where communication precedes intervention, and intervention is only used as a last resort. Naturally, compliance obviates the need for intervention, but the incentives to comply, among other things, depend on the expectations that disobedience would trigger intervention. The effectiveness of communication and intervention is therefore tightly interrelated. Does intervention reinforce compliance? If not, is a commitment not to intervene optimal?

The main result of this paper demonstrates that *intervention can prompt disobedience.* That is, the ability to resolve the conflict through communication is undermined by the possibility of intervention. The key insight is that intervention is counterproductive since it tempts the agent to challenge the principal to *back her words with actions.* In those cases, a commitment not to intervene (and therefore, relying solely on communication) can be optimal. In this respect, *words do speak louder than actions.* The analysis provides a novel explanation for the common wisdom that effective deliberation and compliance can be sustained when the power to make unilateral decisions is restrained. This insight has many implications. For example, a manager can induce more compliance by adopting a hands-off management style. Parents can exert more influence on their children by not keeping them on a tight leash. Sophisticated investors can get their voice heard more effectively if they develop reputation for being non-confrontational and working constructively with the management in companies they invest in, and so on.

To study this topic, I consider a generic principal-agent setup with incomplete contracts and a "top-down" information structure. In the model, the agent (he) has private benefits that create a conflict of interests with the principal (she).[1] The agent chooses from a set of non-contractible actions. Before the agent makes his decision, the principal, who has private information about the action that maximizes her utility, sends the agent a message. This message can be interpreted as instructions, a recommendation, or a nonbinding demand. 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 action to implement. In particular, the agent can disobey or ignore the principal's message. 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 unilaterally chooses the action that will eventually be implemented, even if it is different from the agent's initial choice. If the principal does not intervene, the agent does not incur additional costs and his initial decision is implemented.

In equilibrium, the principal's instructions are informative. The effectiveness of communication can be measured by the probability that the agent voluntarily follows the principal's instructions. 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. In those cases, *intervention reinforces compliance.* This result is intuitive. 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.

[1] While the incentives of the principal and the agent are misaligned, the misalignment is imperfect. For example, the employee has private benefits or costs from performing certain tasks, but he still considers its effect on firm value, perhaps because the firm's well-being affects his future employment.

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. More generally, the likelihood that the agent follows the principal's instructions *increases* with the cost that the principal has to incur in order to intervene. In these cases, *intervention prompts disobedience.*

How can intervention prompt disobedience? Why does the intuition behind the result that intervention reinforces compliance no longer hold? To understand the intuition, note that a standard result in the literature on communication is that with a conflict of interests the principal never fully reveals her private information in equilibrium. In a sense, the principal deliberately conceals information the agent is likely to abuse. The key observation is that 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. Importantly, intervention is an informed decision, and since intervention is costly, the principal intervenes if and only if she believes that the agent's initial decision is detrimental. Therefore, if the principal does not intervene, she effectively "confirms" the decision of the agent to disobey her instructions. This confirmation benefits the agent since he can consume his private benefits without significantly harming the principal's utility. On the other hand, if the principal intervenes, she "corrects" the agent's initial decision. As long as the agent is concerned about the principal's utility (i.e., "doing the right thing"), this correction benefits the agent since it reverses course exactly when the consequences of his actions are detrimental. Either way, the agent can condition on the information that is reflected by the decision of the principal to intervene. This line of reasoning is similar to the winner's curse in common value auctions and pivotal considerations in strategic voting. Altogether, by disobeying the agent forces the principal to make a decision that inevitably reveals information the principal was trying to conceal.[2]

The agent trades off the benefit from "eliciting" additional information from the principal with the direct cost of intervention (i.e., "punishment"). 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,

[2] In the baseline model, the agent cannot change his initial decision after the principal decides whether to intervene. This assumption is relaxed in Section 3.1.

the possibility of intervention creates additional tension between the principal and the agent by providing the agent with opportunities to challenge the principal to *back her words with actions*. Through this channel intervention prompts disobedience.

When intervention prompts disobedience, intervention and communication substitute each other. The principal can benefit from a commitment to limit her ability to intervene in the agent's decision making, depending on the magnitude of the cost that intervention imposes on the agent relative to the cost that is incurred by the principal.[3] In this respect, the analysis provides novel predictions about the circumstances where the benefit from reducing the capacity to intervene is most pronounced. Importantly, different from existing theories (see the literature review below), such commitment is optimal even in situations where the agent is uninformed or there are no hold-up problems.

The mechanism through which the principal commits not to intervene depends on the application. In some situations, the allocation of decision rights can be contracted. For example, the organizational structure of the firm can be decentralized such that the headquarters have a limited authority over its different business units and divisions. Alternatively, the principal can prepare in advance an option to exit her relationship with the agent. All else equal, with the option to exit, the principal has fewer incentives to intervene ex-post. For example, investors can exit their investment by selling their share if they are dissatisfied with management. Similarly, a "busy" principal has a higher alternative cost of intervention. Thus, abstracting from other sources of externalities, an organizational structure where the principal oversees multiple agents can be optimal. Moreover, when the principal interacts repeatedly with the agent (the same agent or a different one in each interaction), she can benefit from developing a reputation of being non-confrontational. Finally, a lower quality of private information reduces the incentives of the principal to intervene, and thereby, benefits the principal by reinforcing compliance (see Section 3.2). Thus, a generalist principal can in fact be more influential than a specialist principal, whose private information is more precise.

To complete the analysis, I extend the baseline model in several dimensions. First, I study the role of intervention as a costly channel of information transmission. I assume that the agent can voluntarily revise his initial decision after observing the principal's decision to intervene, and show two results. First, intervention can be more effective when it operates as

[3] The principal can benefit from the option to intervene even if it prompts disobedience. In this respect, the analysis rationalizes the existence of confrontational but sustainable relationships.

a costly signaling device than when it operates as a corrective action. Second, intervention as a pure costly signaling device (e.g., "burning money") typically prompts disobedience.[4] In the second extension, I study the effect of the quality of the principal's private information on the interaction between communication and intervention. I show that intervention can prompt disobedience even if in applications where the informational advantage of the principal is minor. Moreover, I show that the likelihood of disobedience can increase with the quality of the principal's private information, as higher quality increases the informational benefits from disobedience. In the third extension, I consider a variant of the baseline model in which the agent chooses from a continuum of actions. In this setup, if intervention reinforces compliance then it also increases the amount of information that is revealed in equilibrium, but if intervention prompts disobedience, less information is transmitted. Finally, I extend to analysis to environments in which the agent is also privately informed, and show that with two-sided information asymmetry intervention is more likely to prompt disobedience and reduce the precision of the information that is communicated by the principal.

The paper is organized as follows. The remainder of the section highlights the contribution of the paper to the existing literature. Section 1 presents the setup of the baseline model and discusses in details its mapping to several applications. Section 2 presents the core analysis. Section 3 considers extensions of the baseline model. Section 4 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

The paper is related to the literature on obedience and authority. A number of papers considered interactions where a privately informed principal first communicates with an agent and then, depending on the outcome, unilaterally takes actions which affect the utility of both parties (e.g., Matthews (1989), Shimizu (2008), Marino, Matsusaka, and Zábojník (2010), Van den Steen (2010), and Levit (2013)). While each paper has a different context, they all share the observation that intervention reinforces compliance. By contrast, I show that intervention

[4] Austen-Smith and Banks (2000) and Kartik (2007) study the conditions under which "burning money" improves cheap-talk communication. Different from their analysis, here intervention is not simply burning money, and when it is, burning money occurs after the initial stage of communication and after the principal observes the agent's decision.

can prompt disobedience. To the best of my knowledge, this result is new in the literature. The difference stems from a combination of two assumptions which is absent from the papers above but is important in many applications (see the discussion in Section 1.1). First, intervention in my model does not necessarily terminate the relationship with the agent (although it could), but instead, it allows the principal to unilaterally change the agent's initial decision. Second, there is a common value aspect: the agent believes that the principal's private information is payoff relevant, and hence, would benefit from learning about it. Intervention prompts disobedience because the agent knows the principal will intervene exactly when changing the agent's initial decision is desirable also from the agent's perspective.

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, Bag, and Chakraborty (2014), Alonso and Matouschek (2007), Chakraborty and Yilmaz (2011), Grenadier, Malenko and Malenko (2014), Harris and Raviv (2005, 2008, 2010), and Mylovanov (2008)). 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.

This paper is also related to Aghion and Tirole (1997) and Burkart, Gromb, and Panunzi (1997). In both studies, formal authority can be undesirable because it weakens the agent's incentives to collect information.[5] Generally, with hold-up problems, the agent's incentives to undertake a firm-specific investment decrease with his disutility from intervention. Different from these studies, here there is no hold-up problem of any sort. Therefore, the larger is the direct cost that intervention imposes on the agent, the more effective it is. Instead, intervention is counterproductive since it tempts the agent to challenge the principal to back her words with actions. In this respect, the value of a commitment not to intervene in the agent's decision is derived from different economic forces, and hence, a commitment is expected under different circumstances.

[5] See also Baker, Gibbons, and Murphy (1999), who assume that authority is non-contractible, but can be informally given through commitments enforced by reputation.

1 Setup

Consider a principal-agent environment in which payoffs depend on action $a \in \{L, R\}$ and random variable θ, which has a continuous probability density function f with full support over $\left[\underline{\theta}, \overline{\theta}\right]$. The principal's payoff is given by $v(\theta, a)$, where

$$\Delta(\theta) \equiv v(\theta, R) - v(\theta, L) \tag{1}$$

is a strictly increasing and continuous function, and $\Delta(\underline{\theta}) < 0 < \Delta\left(\overline{\theta}\right)$. The first assumption implies that the relative benefit from action R increases with θ, and the second assumption guarantees that the principal's preferences are not trivial, that is, she prefers action R over L if and only if $\Delta(\theta) > 0$.

The agent's payoff is given by

$$v(\theta, a) + \beta \cdot \mathbf{1}_{\{a=R\}}, \tag{2}$$

where β is a random variable, privately known to the agent, independent of θ, with a continuous probability density function g and full support over $[0, \infty)$.[6] The agent prefers action R over L if and only if $\Delta(\theta) > -\beta$. Thus, when $\Delta(\theta) \in (-\beta, 0)$ the principal and the agent have different preferences over actions. Effectively, β captures the intrinsic conflict of interests between the principal and the agent. One interpretation of this specification is that the agent trades off the principal's utility with his private benefits from action R, where a larger β results in a larger bias toward action R.

Following Grossman and Hart (1986) and Hart and Moore (1990), I assume that contracts are incomplete. In particular, the agent and the principal cannot contract over the decisions themselves or the communication protocol that is used to transmit information. Alternatively, β can be interpreted as the residual conflict of interests between the principal and the agent. A residual conflict of interests is likely to arise even if writing an optimal contract that balances between the agent's incentives and the principal's utility net of the agent's compensation is allowed. Intuitively, eliminating entirely the private benefits of the agent is "too expansive" as it may require the principal to give all the pecuniary benefits from the project to the agent.

[6] The main results continue to hold when β is a common knowledge.

The model has four stages:

Stage 1: The first stage involves communication between the principal and the agent. While the principal is uninformed about β, she has information about θ that the agent does not have. For simplicity, I assume that the agent is uninformed about θ while the principal privately observes θ. The former assumption is relaxed in Section 3.4, and the latter assumption is relaxed in Section 3.2. Based on her private information, the principal sends the agent message $m \in [\underline{\theta}, \overline{\theta}]$. 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. I denote by $\rho(\theta)$ the principal's communication strategy and by $M \subseteq [\underline{\theta}, \overline{\theta}]$ the set of messages on the equilibrium path.

Stage 2: In the second stage, the agent observes the message from the principal and 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 benefits β.

Stage 3: The key departure of the model from the existing literature is the third stage. In the third stage, the principal observes the agent's decision and then decides whether to intervene. I denote by $e(\theta, a_A) = 1$ the principal's decision to intervene and by $e(\theta, a_A) = 0$ her decision not to intervene. If the principal intervenes, she chooses between reversing the agent's decision and keeping it in place.[7,8] I denote by $a_P(\theta, a_A) \in \{L, R\}$ the principal's decision after intervention. If the principal intervenes and reverses the agent's decision, then the principal incurs a cost $c_P > 0$. Apart from the effort, time and resources that are needed for intervention, parameter c_P also captures the principal's alternative cost of dealing with the task or her aversion for confrontation. I also assume $c_P < -\Delta(\underline{\theta})$. Without this assumption, the intervention stage will have no effect on the game. In addition, if the principal intervenes and reverses the agent decision, the agent incurs a cost $c_A > 0$. This cost can be both pecuniary (e.g., the loss of compensation) and non-pecuniary (e.g., damaged reputation, loss of respect, or embarrassment).[9]

[7] Intuitively, the principal can 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.

[8] The action does not have to be perfectly reversible. In a previous version of the paper, similar results were derived under the assumption that when the principal intervenes, the action is reversed only with some probability.

[9] If the principal intervenes in order to reinforce the agent's initial decision, the principal and the agent do not incur additional costs. Formally, $c_P(a_A, a_P) = c_P \cdot \mathbf{1}_{\{a_A \neq a_P\}} > 0$ and $c_A(a_A, a_P) = c_A \cdot \mathbf{1}_{\{a_A \neq a_P\}} > 0$.

Stage 4: In the final period, the payoffs are realized and distributed to the principal and the agent.

Overall, the principal's utility function is given by

$$u_P(\theta, a_A, a_P, e) = \begin{cases} v(\theta, a_P) - c_P & \textit{if } e = 1 \\ v(\theta, a_A) & \textit{else,} \end{cases} \tag{3}$$

and the agent's utility function is given by

$$u_A(\theta, a_A, a_P, e, \beta) = \begin{cases} v(\theta, a_P) + \beta \cdot \mathbf{1}_{\{a_P = R\}} - c_A & \textit{if } e = 1 \\ v(\theta, a_A) + \beta \cdot \mathbf{1}_{\{a_A = R\}} & \textit{else.} \end{cases} \tag{4}$$

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

1.1 Applications of the model

Before analyzing the model, I discuss its mapping to several applications.

1.1.1 Organizations

The model can be applied to study interactions between managers and their subordinates, owners of small businesses and their employees, firms and labor unions, or CEOs\headquarters and division managers. As an example, consider the interaction between the CEO of a company (principal) and one of its division managers (agent). The company has to decide whether to reinvest the divisional profits in a new divisional project ($a = R$) or to allocate the capital elsewhere ($a = L$). This decision is not contractible since the investment opportunities depend on a variety of macro, industry, and firm-specific factors which cannot be perfectly anticipated. The CEO has superior knowledge of $\Delta(\theta)$, which measures the profitability of investment in the divisional project relative to the alternative use of capital. For example, θ captures the exter-

Intuitively, intervention imposes costs only when there is a conflict. In the baseline model, there is no difference between intervention to reinforce the agent's decision and non-intervention. This assumption only plays a role in Section 3 where the agent is allowed to revise his initial decision.

nalities that the divisional project has on other business units, the attractiveness of investment opportunities in other divisions, the cost of external financing, and market conditions.[10]

While the CEO is interested in maximizing the value of the firm, the division manager is biased toward investment in the divisional project ($\beta > 0$) either because of the private benefits from increasing the assets under his direct control, or his career concerns (his skills will be mostly reflected in the performances of his division). Based on her superior information, the CEO will either demand that the profits are reinvested in the division or allocated elsewhere. If needed, the CEO will step in and force the redistribution of divisional profits and replace the division manager. Intervention of this sort, however, not only requires corporate resources but also the attention of the CEO, which is typically limited ($c_P > 0$). In turn, intervention is likely to harm the division manager's reputation, ego, and compensation ($c_A > 0$).

The optimal internal organization of the firm and the allocation of authority are the key questions in this context. The analysis demonstrates that decentralization is optimal if and only if c_P is large relative to c_A.

1.1.2 Parenting

The interaction between parents (principal) and their children (agent) is another application of the model. For example, a teenager may face a dilemma whether to experiment with the consumption of addictive substances (e.g., drinking alcohol) ($a = R$) or avoid it ($a = L$). The long-term implications of substances abuse on the child's well being, $\Delta(\theta)$, depend on his personality, social status, and health condition.[11] Parents often have a better understanding of the adverse consequences of substance abuse and their child's ability to cope with challenges and avoid addiction. Since children seldom internalize the long-term consequences of their

[10] The CEO may even have expertise related to the division's specific line of business, perhaps because the CEO was the manager of this division in his previous job. A CEO may also be privately informed about the ability of the division manager or on the match between her own skill set and the firm, and thus, about the firm value under her leadership.

[11] If $\Delta(\theta) > 0$ then the model implies that a modest consumption of, e.g., alcohol, is not harmful. The main results go through even if $\Delta(\theta)$ is bounded from above by zero. However, $\Delta(\theta) > 0$ is reasonable in this context for a variety of reasons, other than (perhaps) the immediate pleasure of consumption. First, by letting their child experimenting the parents can mitigate the forbidden fruit element. Second, the child will have fewer incentives to hide it, thereby giving the parents more control on the quantity and quality of the substances he is consuming.

actions, a conflict between parents and their teenager children is not uncommon.[12] Most parents will deal with the problem by first explaining their children the negative consequences of substance abuse and ask them to avoid it. However, if they are ignored, parents can monitor the behavior of their children more closely (e.g., by following their children activity on social networks...), restrict their social interactions after school time, or limit their spending budget. These forms of intervention not only impose social costs on the children (c_A), but also require the parents to invest time and energy in confronting their children (c_P).

The application of the model to parenting highlights that a permissive parenting style can dominate an authoritative style, depending on the vulnerability of the child to sanctions (c_A) and the capacity of the parents to confront him (c_P). For example, that analysis predicts that parents can have a larger influence on their children in families in which both of parents pursue a career (and hence, have less time to confront their children) or the number of children is large (parents' attention per child is limited).

1.1.3 Expert investors

Situations in which investors have substantial expertise and sophistication are also captured by the model. For example, investors in entrepreneurial ventures provide advice and guidance to help entrepreneurs turn their innovative projects into commercial success. The dual role of investors explains the demand for active investors such as venture capitalists. In a typical leveraged buyout, the private equity fund populates the target firm's board of directors with experts from the industry (e.g., Ex-CEOs) and its own general partners. The fund uses its expertise to instruct the CEO how to turnaround the operations of the firm, and if needed, uses the authority of the board to intervene in the CEO's decision making.

Another application is shareholder activism. Activist hedge funds have a market-wide perspective on assets valuation and performances of peer companies that corporate boards often lack. In a typical campaign, the activist engages with the management or the board of directors of the target firm, expressing her dissatisfaction or view of 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.

[12] An additional source of conflict stems from the parents' concerns that a controversial behavior by one child will set a bad example for his siblings.

In all of the corporate finance applications above, investors have control rights as well as expertise that is incremental to the information held by corporate insiders.[13] The analysis of the model suggests that higher expertise should not necessarily be coupled with more control rights, and that sophisticated investors can benefit from committing not to intervene in their portfolio companies, either by creating a reputation for being non-confrontational, by having a large portfolio, or by designing an exit strategy with a low price impact.

1.1.4 Other applications

The model has other plausible applications which for brevity are not discussed here in details: corporate governance (board of directors and the CEO), international relations (diplomacy and soft power vs. military force), regulation (central bank and financial institutions) and education (dean and faculty, teacher and student).

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 decision is not reversed and the principal's payoff is $v(\theta, R)$. If the principal intervenes in order to reverse the agent's decision, her payoff is $v(\theta, L) - c_P$. Therefore, conditional on $a_A = R$, the principal intervenes if and only if $\Delta(\theta) < -c_P$. Suppose the agent chooses action L. Similarly, if the principal intervenes her payoff is $v(\theta, R) - c_P$, and otherwise, her payoff is $v(\theta, L)$. Conditional on $a_A = L$, the principal intervenes if and only if $\Delta(\theta) > c_P$. Overall, the principal intervenes whenever she finds the agent's initial decision detrimental.[14]

Lemma 1 *In any equilibrium, the principal intervenes if and only if $a_A = R$ and $\Delta(\theta) < -c_P$, or $a_A = L$ and $\Delta(\theta) > c_P$.*

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.

[13] As I show in Section 3.4, the main results continue to hold when the agent (the management in the corporate finance application) also has private information about θ.

[14] For this reason, similar results hold if instead the principal chooses the probability that intervention succeeds, λ, at a cost of $c_P(\lambda)$, where $c_P' > 0$ and $c_P'' > 0$.

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

Communication is effective only if in equilibrium the principal reveals information about 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 $\beta_0 \geq 0$ 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.*

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 communication, the agent cannot avoid intervention even if he forgoes his private benefits and chooses action L. Based on Lemma 2, if the equilibrium is non-influential, there is a constant b^{NI} such that the agent chooses action R if and only if $\beta \geq b^{NI}$. Based on his prior beliefs, the agent "follows his bias" and chooses action R with probability one, that is, $b^{NI} \leq 0$.[15]

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 $\Delta(\theta) < -c_P$.*

Consider the properties of an influential equilibrium. 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

[15] In the Appendix I provide sufficient conditions on $\Delta(\theta)$ under which $b^{NI} \leq 0$. The main results continue hold when $b^{NI} > 0$. A non-influential equilibrium with $b^{NI} > 0$ is possible if two conditions are met: the prior puts a large weight on small values of θ such that the agent believes that intervention is less likely when he chooses action L, and c_A is sufficiently large such that the agent prefers forgoing his private benefits in order to avoid the cost of intervention.

message convinces all types $\beta > \beta_R$ ($\beta < \beta_L$) to choose action R (action L) as well. If the equilibrium is influential then it must be $\min_{m\in M} b(m) < \max_{m\in M} b(m)$. Let

$$\begin{aligned} M_R &\equiv \arg\min_{m\in M} b(m) \\ M_L &\equiv \arg\max_{m\in M} b(m). \end{aligned} \tag{5}$$

Since the principal uses her influence to maximize her payoff as given by (3), 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. The principal asks the agent to choose action R if $\Delta(\theta) \geq 0$ and action L if $\Delta(\theta) < 0$.

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

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 $m \in M_R$. The agent follows the instructions and chooses action R if and only if

$$\begin{aligned} \mathbb{E}[v(\theta,R)+\beta|m] \geq\ & \Pr[\Delta(\theta) \leq c_P|m]\,\mathbb{E}[v(\theta,L)\,|\Delta(\theta) \leq c_P, m] + \\ & \Pr[\Delta(\theta) > c_P|m]\,(\mathbb{E}[v(\theta,R)+\beta|\Delta(\theta) > c_P, m] - c_A). \end{aligned} \tag{6}$$

The left hand side of (6) is the agent's expected payoff if he follows instructions. In this case, the principal does not intervene, action R is implemented, and the agent consumes his private benefits. The right hand side of (6) is the agent's expected payoff if he disobeys the principal and chooses action L. According to Lemma 1, if $\Delta(\theta) > c_P$ then the principal intervenes, the agent's decision is reversed, and he incurs an additional cost c_A. Since $\beta \geq 0$ and $m \in M_R \Rightarrow \Delta(\theta) \geq 0$, the inequality in (6) always holds. Therefore, the agent always follows the principal's instructions to choose action R.

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

$\Delta(\theta) < 0$. The agent follows the instructions and chooses action L if and only if

$$\begin{aligned} \mathbb{E}\left[v(\theta, L) | m\right] \geq & \Pr\left[\Delta(\theta) \geq -c_P | m\right] \mathbb{E}\left[v(\theta, R) + \beta | \Delta(\theta) \geq -c_P, m\right] + \\ & \Pr\left[\Delta(\theta) < -c_P | m\right] \left(\mathbb{E}\left[v(\theta, L) | \Delta(\theta) < -c_P, m\right] - c_A\right) \end{aligned} \tag{7}$$

The left hand side of (7) is the agent's expected payoff if he follows instructions. In this case, the principal does not intervene and action L is implemented. The right hand side of (7) is the agent's expected payoff if he disobeys the principal and chooses action R. If $\Delta(\theta) < -c_P$ then the principal intervenes, the agent's decision is reversed, and he incurs an additional cost c_A. If $\Delta(\theta) \geq -c_P$ then the principal does not intervene, the agent's decision is unchanged, and he consumes his private benefits. The next result shows that the agent follows the principal's instructions if and only if $\beta \leq b^*(c_A, c_P)$, as given below.

Proposition 2 *An influential equilibrium always exists. In any influential equilibrium, the principal instructs the agent to choose action R if and only if $\Delta(\theta) \geq 0$. 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*

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

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 $\Delta(\theta) < -c_P$.

The threshold $b^*(c_A, c_P)$ measures the effectiveness of communication, where higher $b^*(c_A, c_P)$ implies higher probability that the agent follows the instructions of the principal. Note that $b^*(c_A, c_P)$ increases with c_A. Intuitively, the principal intervenes only if the agent ignores her instructions. In order to avoid the negative consequences of intervention, the agent is more likely to follow the principal's instructions when c_A is higher. The next result shows that the effect of c_P on $b^*(c_A, c_P)$ is more subtle.

Corollary 1 *The threshold* $b^*(c_A, c_P)$ *strictly increases with* c_P *if and only if* $c_p \in \left(c_P^{\min}, -\Delta(\underline{\theta})\right)$ *and* $c_A < \mathbb{E}[\Delta(\theta)|\Delta(\theta) < 0] - \Delta(\underline{\theta})$, *where* $c_P^{\min} \in (0, -\Delta(\underline{\theta}))$ *is the unique solution of*

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

Figure I depicts $b^*(c_A, c_P)$ as a function of c_P when $\Delta(\theta) = \theta$, $\theta \sim U[-1, 1]$, and $c_A = 0.02$. 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 $b^*(c_A, c_P)$ 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, as in those instances the threat of intervention is less credible. Corollary 1 shows that this intuition can be misleading. The next section explains the intuition behind this result.

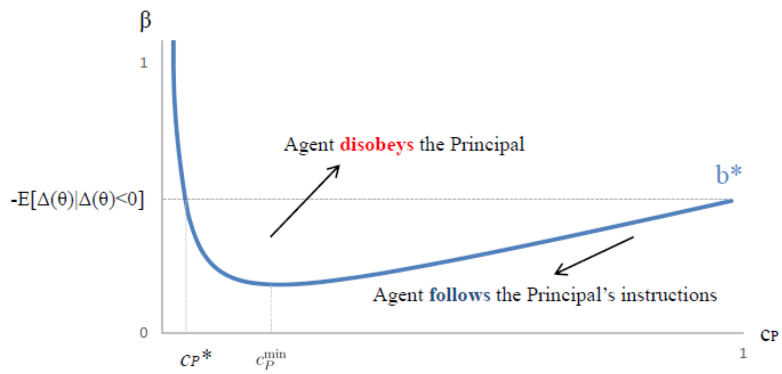

Figure I - The comparative statics of b^* with respect to c_P

Since influential equilibria always exist (and as is standard in the literature), hereafter I assume that the equilibrium in play is always influential. This assumption is also supported by the following result.

Proposition 3 *Every influential equilibrium Pareto dominates every non-influential equilibrium.*

2.1 Does intervention prompt disobedience?

To understand the interaction between communication and intervention, I consider a benchmark in which intervention is either prohibitively costly or entirely ineffective. This is a special case of the baseline model, where $c_P > -\Delta(\underline{\theta})$. According to Proposition 2,

$$\lim_{c_P \to -\Delta(\underline{\theta})} b^*(c_A, c_P) = -\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0]. \tag{10}$$

Therefore, communication is considered less effective with intervention than without it if and only if

$$b^*(c_A, c_P) < -\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0]. \tag{11}$$

That is, if condition (11) holds then intervention prompts disobedience, and otherwise, intervention reinforces compliance. The next result follows immediately from the comparison between $b^*(c_A, c_P)$, as given by (8), and $-\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0]$.

Proposition 4 *Suppose $c_P < -\Delta(\underline{\theta})$. Intervention prompts disobedience if and only if*

$$c_A < \mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0] - \mathbb{E}[\Delta(\theta) | \Delta(\theta) < -c_P]. \tag{12}$$

Figure II illustrates that when $\Delta(\theta) = \theta$ and the distribution of θ is uniform and symmetric around zero, condition (12) becomes $\frac{c_A}{c_P} < \frac{1}{2}$. The cutoff point c_P^* in Figure 1 is the unique value of c_P that satisfies condition (12) with equality.

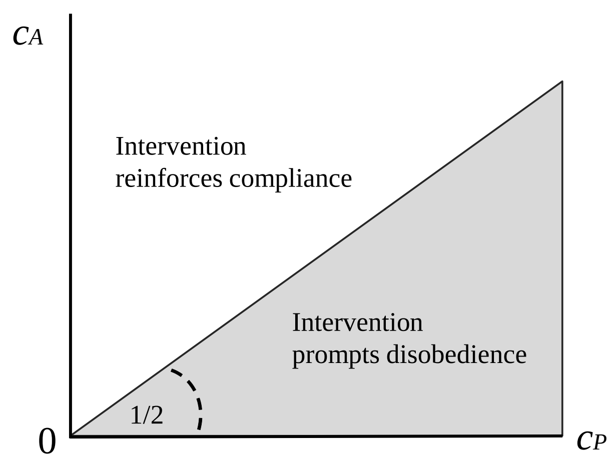

Figure II - The effect of intervention on disobedience

Proposition 4 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 $b^*(c_A, c_P)$ increases with c_A.

The second implication of Proposition 4 is somewhat surprising: if condition (12) holds then intervention prompts disobedience. 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 $\Delta(\theta) < -\beta$. However, in equilibrium, the instructions of the principal do not reveal whether $\Delta(\theta) < -\beta$ or $\Delta(\theta) \in [-\beta, 0)$. The principal has no incentives to do so, because if she did, the agent would have chosen action R when $\Delta(\theta) > -\beta$. Instead, the principal pretends that θ is lower than it really is in order to persuade the agent to choose L even when $\Delta(\theta) > -\beta$. The agent understands the principal's incentives, and hence, the only information that can be inferred from the instructions to choose action L is $\Delta(\theta) < 0$.

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 $\Delta(\theta)$ relative to $-c_P$. In particular, if the principal does not intervene, the agent infers that the principal believes that choosing action R does not justify intervention, that is, $\Delta(\theta) > -c_P$. These are the states in which the agent prefers consuming his private benefits even at the expense of a lower utility for the principal. In this respect, the principal's decision not to intervene "confirms" the agent's initial decision to disobey. On the other hand, if the principal intervenes, the agent infers that the principal believes that choosing action R is detrimental, that is, $\Delta(\theta) < -c_P$. Since the agent is also concerned about the principal's utility, he prefers forgoing his private benefits when he learns that $\Delta(\theta)$ is low. In this respect, intervention in those cases benefits the agent since it "corrects" his initial decision when it is indeed detrimental.

Overall, 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.[16] 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 $\Delta(\theta) < -c_A - \beta$. Note that when deciding whether to intervene, the principal behaves as if she is biased toward action R, where the bias is c_P. Therefore, if $c_P \approx c_A + \beta$ then the principal's "bias" coincides with the preferences of the agent. As can be seen by (9), the minimum of $b^*(c_A, c_P)$ as a function of c_P is obtained when $c_P = c_A + b^*(c_A, c_P)$. While the value of the "correction" effect increases with c_P, the value of the "confirmation" effect decreases with c_P. When $c_P = c_A + b^*(c_A, c_P)$ 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.

To conclude, the possibility of intervention creates additional tension between the principal and the agent. Condition (12) 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 prompts disobedience 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 reinforce compliance. 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 the informational benefit of correction and confirmation.

2.2 Is intervention valuable?

If intervention prompts disobedience, then the principal can benefit from a commitment not to intervene in the agent's decision. Building on Proposition 2, the principal's expected payoff in any influential equilibrium can be written as

$$\begin{aligned} W(c_A, c_P) \;=\; & \mathbb{E}[v(\theta, L)] + \Pr[\Delta(\theta) \geq 0]\,\mathbb{E}[\Delta(\theta) \,|\Delta(\theta) \geq 0] + \\ & \Pr[\beta > b^*(c_A, c_P)]\Pr[\Delta(\theta) < 0]\,\mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < 0] + \\ & \Pr[\beta > b^*(c_A, c_P)]\Pr[\Delta(\theta) < -c_P]\left(-\mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < -c_P] - c_P\right)\mathbf{1}_{\{c_P < -\Delta(\underline{\theta})\}}. \end{aligned} \tag{13}$$

[16] In the Online Appendix I show that the agent prefers a principal with the capacity to intervene over a principal without the capacity to intervene if and only if β is below some threshold that depends on c_A and c_P.

The first line in (13) is the principal's expected utility if her first best is implemented. The second line is the principal's expected disutility from the agent's disobedience. The third line is the expected utility the principal recovers from intervention, conditional on the agent's disobedience.

Expression (13) has several implications. First, since $b^*(c_A, c_P)$ increases with c_A, $W(c_A, c_P)$ increases with c_A as well. That is, the principal is always better off when intervention imposes a higher cost on the agent. Second, the principal gets the highest payoff when $c_P = 0$, in which case she can either enforce or threaten to enforce her optimal strategy. Interestingly,

$$\lim_{c_P \to 0} b^*(c_A, c_P) = \begin{cases} \infty & \textit{if } \; c_A > 0 \\ 0 & \textit{if } \; c_A = 0. \end{cases} \tag{14}$$

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. In most applications, 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 5 *If and only if* $c_A < \mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0] - \Delta(\underline{\theta})$, *there is* $\overline{c}_P \in (c_P^*, -\Delta(\underline{\theta}))$ *such that* $W(c_A, c_P) < W(c_A, -\Delta(\underline{\theta}))$ *for all* $c_P \in (\overline{c}_P, -\Delta(\underline{\theta}))$.

How can the principal be better off without the option to intervene? This is possible only if intervention prompts disobedience. Based on Proposition 4, it is necessary that $c_P > c_P^*$. If $c_P \in (c_P^*, \overline{c}_P]$ then intervention can be preferred by the principal even though it prompts disobedience. 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 $c_P > \overline{c}_P$ then the principal is better off without the option to intervene. These observations are illustrated by the left panel of Figure III, which plots the principal's expected payoff as a function of c_P when $\Delta(\theta) = \theta$, $\theta \sim U[-1, 1]$, 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 prompts disobedience, but is also ineffective on its own. Therefore, the principal prefers effective communication over ineffective intervention.

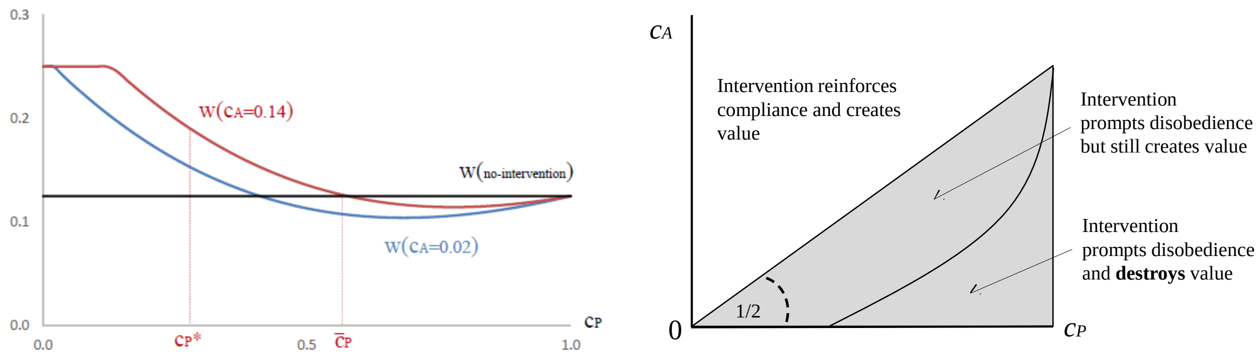

Figure III - The value of intervention

2.3 When is intervention expected?

The probability of intervention in equilibrium is given by

$$\eta(c_A, c_P) \equiv \Pr[\beta \geq b^*(c_A, c_P)] \times \Pr[\Delta(\theta) < -c_P]. \tag{15}$$

Since $b^*(c_A, c_P)$ increases with c_A, $\eta(c_A, c_P)$ always decreases with c_A. Intuitively, with higher c_A the agent has stronger incentives to follow the principals's instructions in order to avoid the costs of intervention. Generally, the probability of intervention has an inverted U-shape as a function of c_P. Figure IV illustrates this point by plotting $\eta(c_A, c_P)$ as a function of c_P when $\Delta(\theta) = \theta$, $\theta \sim U[-1, 1]$ and $\beta \sim U[0, 1]$. Intuitively, if c_P is small then the principal can effectively influence the agent through communication, and intervention serves only as a threat, which results with a low probability of intervention. 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. However, if c_P increases even further, intervention becomes too costly as a corrective tool, and the principal will intervene less often.

The comparative static of η with respect to c_P 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. Unobserved intervention is not necessarily evidence that intervention is ineffective – quite the opposite. With communication, 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.

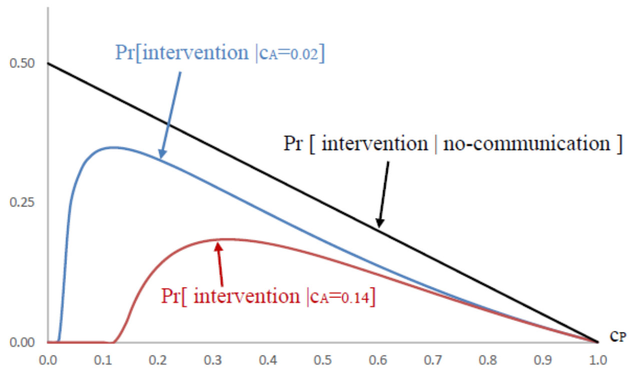

Figure IV - The probability of intervention in equilibrium

3 Extensions

3.1 Intervention as a costly channel of communication

In the baseline model, intervention conveys information about θ that is not revealed through direct communications, but the agent has no opportunity to act based on this information. If the agent could revise his initial decision if the principal did not to intervene or if intervention failed, then intervention would play a dual role: it is a corrective tool as well as a costly channel through which information is transmitted.

In this section I focus on the second role of intervention. I reinterpret the baseline model by assuming that while the principal has to incur a cost c_P in order to reverse the agent's initial decision (in which case, the agent incurs a cost c_A as well), the principal can always reinforce the agent's initial decision without incurring additional costs to her or to the agent. Thus, the principal has three options: not intervening in the agent's decision, intervening in order to reinforce the agent's decision, and intervening in order to reverse the agent's decision. Different from the baseline model, here I assume that any attempt of the principal to reverse the agent's decision is doomed to fail.[17] Nevertheless, by trying to reverse the decision of the agent, the

[17] In the Appendix I show that if the only difference from the baseline model is that the agent can revise his initial decision upon non-intervention, but intervention itself cannot fail, then in equilibrium the agent never reverses his initial decision following non-intervention, and therefore, the set of equilibria does not change

principal incurs a cost c_P and imposes a cost c_A on the agent. Therefore, intervention to reverse the agent's decision is non-binding.[18] The key assumption is that the agent can revise his initial decision at no additional costs after observing the decision of the principal not to intervene, or if the principal intervenes but intervention failed. I denote the agent's final decision by $a_F \in \{L, R\}$.

Proposition 6 *Suppose the agent is allowed to revise his decision if the principal did not intervene or if intervention failed. An influential equilibrium always exists.*[19,20,21] *In any influential equilibrium there are $b^{**} > 0$ and $\mu^{**} \in (0,1)$ such that*

$$b^{**} = b^* \left(c_A, c_P/\mu^{**}\right) \tag{16}$$

and

$$\mu^{**} = \frac{G\left(-\mathbb{E}\left[\Delta\left(\theta\right)\left|\Delta\left(\theta\right) < -c_P/\mu^{**}\right.\right]\right) - G\left(b^{**}\right)}{1 - G\left(b^{**}\right)}. \tag{17}$$

In equilibrium, the principal instructs the agent to choose action R if and only if $\Delta(\theta) \geq 0$. If the principal instructs the agent to choose action R, the agent chooses action R with probability one and the principal reinforces his decision. 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 reinforces his decision. If the agent ignores these instructions, the principal intervenes to reverse the agent's decision if $\Delta(\theta) < -c_P/\mu^{**}$, and does not intervene otherwise. The agent revises his initial decision from R to L if and only if the principal's attempt to reverse the agent's decision failed and $\beta \in (b^{**}, -\mathbb{E}[\Delta(\theta) | \Delta(\theta) < -c_P/\mu^{**}]]$.

Proposition 6 is similar to Proposition 2, but has the following key difference: Intervention is only successful with probability μ^{**}. Therefore, the cost of intervention per unit of success is c_P/μ^{**} instead of c_P. Notice that even though intervention is non-binding, it is still effective since by intervening the principal communicates information that convinces the agent to

relative to the analysis in Section 2. Intuitively, non-intervention provides confirmation to the agent that his disobedience is not detrimental, and hence, the agent will not revise his initial decision.

[18] In a previous version of the paper similar results were derived under the assumption that an intervention to reverse the agent's decision is successful with probability smaller than one.

[19] Here, 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 revising the initial decision. In the Online Appendix I show these equilibria do not exist, and therefore, all influential equilibria satisfy Definition 1.

[20] The analysis of non-influential equilibria is given in the Online Appendix.

[21] We restrict attention to equilibria that survive the Grossman and Perry (1986) criterion.

voluntarily revise his initial decision with a conditional probability μ^{**}. In equilibrium, μ^{**} reflects the principal's beliefs about the circumstances under which the agent revises his initial decision, which in turn, depends on the agent's private benefits as well as on his own beliefs about the circumstances under which the principal intervenes.

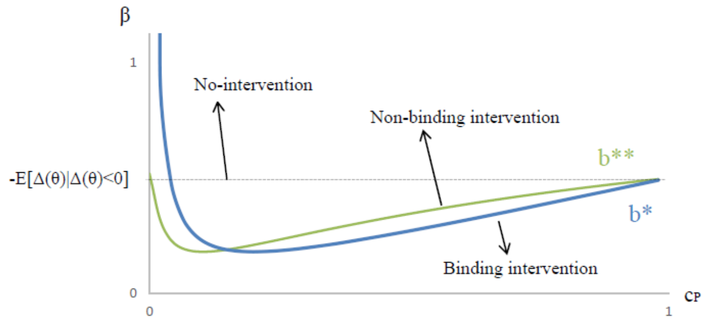

Figure V - Binding vs. Non-binding intervention

Proposition 6 has several implications. First, intervention can result with more compliance when it is non-binding than when it is binding. That is, $b^{**}(c_A, c_P) > b^{*}(c_A, c_P)$ is feasible. For example, this is the case whenever $c_P \geq c_P^{\min}$ as given by Corollary 1. This observation is illustrated in Figure V which plots b^* (blue curve) and b^{**} (green curve) as a function of c_P, when $\Delta(\theta) = \theta$, $\theta \sim U[-1, 1]$, $c_A = 0.02$, and $\beta \sim U[0, 1]$. Intuitively, with non-binding intervention, the cost per unit of success is higher, which in turn reduces the confirmation benefit from disobedience, and thereby, increases compliance.[22] Second, if $c_P = 0$ then $b^{**} = -\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0]$. That is, the benefit from the non-binding intervention exists only if non-binding intervention is costly to the principal. Indeed, if $c_P = 0$ and on the contrary $\mu^{**} > 0$, the principal always intervenes if the agent ignores his instructions, and hence, her non-binding intervention does not convey additional information. Without information, the agent cannot be persuaded to change his mind, and hence, it must be $\mu^{**} = 0$. Finally, non-binding intervention can be counterproductive, that is, $b^{**}(c_A, c_P) < -\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0]$. For example, in Figure 5, b^{**} is for the most part below the horizontal doted line. In fact, non-binding intervention always prompts disobedience when $c_A = 0$. In this case, non-binding intervention has no element of punishing the agent and it is literally a "burning money" operation. The signaling role of intervention, on its own, only harms communication. Since burning money provides

[22] This observation also implies that the principal might prefer non-binding intervention over binding intervention, as a way to commit to not to intervene in the agent's decision.

a stronger signal that action L is optimal, the agent is better off waiting until the principal burns money to demonstrate that indeed action L is optimal. In other words, the resolution of a conflict is possible only after some confrontation occurs. The paradox is that the anticipation for confrontation makes it harder for information to be transmitted through less costly channels of communication.

3.2 Partially informed principal

A key observation from the baseline model is that intervention prompts disobedience because the decision of the principal to intervene is an informed decision. In this section, I demonstrate that a lower quality of the principal's private information can actually *strengthen* the key results. To illustrate this point, suppose the principal observes an imperfect signal of θ, denoted by s, with the following properties:

$$s = \begin{cases} \theta & \text{with probability } \gamma \\ \varepsilon & \text{with probability } 1-\gamma, \end{cases} \tag{18}$$

where random variable ε is independent of θ but has the same distribution as θ. The principal does not know whether $s = \theta$ or $s = \varepsilon$. Under this specification, the principal's signal is precise with probability γ and a pure noise otherwise. Parameter $\gamma \in [0, 1]$ captures the quality of the principal's private information. The baseline model is a special case where $\gamma = 1$. For simplicity, I also assume $\mathbb{E}[\Delta(\theta)] = 0$.

In the Appendix, I show that similar to Proposition 2 and Proposition 4, the agent follows the principal's instructions to choose action L if and only if $\beta \leq \gamma b^*(c_A/\gamma, c_P/\gamma)$, and intervention prompts disobedience if and only if

$$c_A/\gamma < \mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0] - \mathbb{E}[\Delta(\theta) | \Delta(\theta) \leq -c_P/\gamma]. \tag{19}$$

As γ decreases, the instructions of the principal become a weaker signal of θ and the agent's informational benefits from intervention decrease. Therefore, the punishment from intervention has a larger impact on the agent's incentives. This effect, which is reflected by the scaling of c_A by γ in (19), extends the region in which intervention reinforces compliance. On the other hand, the principal has a larger cost of intervention per unit of information, and therefore, she is less likely to intervene. This effect, which is reflected by the scaling of c_P by γ in (19), extends the region in which intervention prompts disobedience. When the distribution of $\Delta(\theta)$ is uniform, these affects are cancelled out, and condition (19) is equivalent to (12) for any $\gamma \in (0,1)$. In these cases, the circumstances under which intervention prompts disobedience are invariant to γ.

It is interesting to note that without intervention, the agent's compliance increases with γ, the quality of the principal's private information. Indeed, with more information the agent has more reasons to follow the principal's instructions. However, with intervention, the agent's compliance can *decrease* with the quality of the principal's private information.[23] Intuitively, the informational benefits of the agent from disobedience, correction and confirmation, are insignificant when the principal's private information is imprecise. Therefore, the agent has fewer incentives to challenge the principal, and overall, he is more likely to follow her instructions. We conclude, high quality of private information can prompt disobedience.

3.3 Continuum of actions and endogenous precision

The main results in this paper can be extended to a setup in which the action space is a continuum. With more than two actions, the analysis sheds light not only on the effect of intervention on disobedience, but also on its effect on the quality of communication, measured by the precision of the principal's message.

For illustration, consider the leading example of Crawford and Sobel (1982), which has been used extensively in the literature on communication and delegation. In particular, suppose $a \in [\underline{\theta}, \overline{\theta}]$, $v(\theta, a) = -(\theta - a)^2$, and the agent's payoff is $v(\theta + \beta, a)$ where $\beta > 0$ is a common knowledge. In addition, suppose that if the agent chooses a_A and the principal intervenes by choosing 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. The leading example of Crawford and Sobel is a special case of this model where $c_P = \infty$ and $c_A = 0$.

[23] In the Appendix I provide sufficient conditions and an example under which $\frac{\partial}{\partial\gamma}[\gamma b^*(c_A/\gamma, c_P/\gamma)] < 0$.

Proposition 7 *Let $\Lambda(\beta, c_P, c_A)$ be the set of equilibria of the game induced by parameter values (β, c_P, c_A). Then*

$$(\rho^*, a_A^*, a_P^*) \in \Lambda(\beta, c_P, c_A) \Leftrightarrow (\rho^*, a_A^*, a_P^*) \in \Lambda\left(\beta\frac{1+c_P}{c_A/c_P + c_P}, \infty, 0\right), \tag{20}$$

where $a_P^*(\theta, a_A) = a_A + \frac{\theta - a_A}{1+c_P}$.[24]

Proposition 7 shows that for a given parameter values (β, c_P, c_A), the set of communication strategies that arise in a game with intervention, ρ^* and a_A^*, is identical to the set of communication strategies that arise in a game without intervention, where the agent's bias is $\beta\frac{1+c_P}{c_A/c_P+c_P}$ instead of β. This observation has several implications. Recall from Crawford and Sobel (1982) that under the most informative equilibrium the quality of communication improves when the principal and the agent have closer preferences, that is, β decreases. Moreover, the agent's compliance decreases with β, in the sense that his actions in equilibrium are further away from the principal's ideal points as inferred from her messages. Therefore, Proposition 7 implies intervention prompts disobedience if and only if

$$\beta\frac{1+c_P}{c_A/c_P + c_P} > \beta \Leftrightarrow c_P > c_A. \tag{21}$$

Consistent with Proposition 4, intervention is counterproductive if and only if c_P is large relative to c_A. However, here, the *quality of communication* also changes. In particular, when intervention reinforces compliance, more information can be revealed by the principal in equilibrium, and when intervention is counterproductive, the information transmitted by the principal becomes nosier.[25]

3.4 Informed agent

A key assumption in the model is that principal has private information. By assuming that the agent is uninformed, the baseline model stacks the cards against the result that a commitment not to intervene in the agent's decision can be optimal. However, the main conclusions do not

[24] Notice that without intervention, $c_P = \infty$, and hence, trivially, $a_P^*(\theta, a_A) = a_A$.

[25] A commitment not to intervene in the agent's decision can be optimal in this setup as well. For example, this is the case if $\theta \sim U[0, 20]$, $c_A = 2$, $\beta = 0.5$, and $c_P = 10$.

change if in addition the agent has private information about θ.[26] For example, consider the setup in Section 3.3 where $\theta = \theta_P + \theta_A$, the principal is privately informed about θ_P, the agent is privately informed about θ_A, and θ_P and θ_A are independent.

Proposition 8 *Let $\Lambda_{linear}(\beta, c_P, c_A)$ be the set of equilibria of the game induced by parameter values (β, c_P, c_A).*[27] *Then*

$$(\rho^*, a_A^*, a_P^*) \in \Lambda_{linear}(\beta, c_P, c_A) \Leftrightarrow (\rho^*, a_A^*, a_P^*) \in \Lambda\left(\beta\frac{1+c_P}{c_P}, \infty, 0\right) \tag{22}$$

where $a_P^*(\theta_P, a_A) = a_A + \frac{h(\theta_P)}{1+c_P}$ and $h(\theta_P)$ is a function that is independent of a_A.

Similar to Proposition 7, Proposition 8 shows that for a given parameter values (β, c_P, c_A), the set of communication strategies that arise in a game with intervention is identical to the set of communication strategies that arise in a game without intervention, where the agent's bias is $\beta\frac{1+c_P}{c_P}$ instead of β. However, notice that $\beta\frac{1+c_P}{c_P} > \beta$ for all $c_P < \infty$. That is, intervention always prompts disobedience and reduces the quality of communication, even if c_A is arbitrarily large.

To understand the intuition behind Proposition 8, first notice that in sharp contrast to the analysis in Section 3.3, c_A has no effect on the equilibrium. If the agent is uninformed, larger c_A weakens his incentives to choose actions that are distant from the principal's ideal point. However, when the agent has private information about θ, choosing a distant action does not increase the intensity of intervention by the principal. In equilibrium, the principal infers from the agent's decision the value of θ_A, and hence, the principal updates her beliefs about her own ideal point accordingly. That is, the principal rationally attributes distant actions as strong signals about θ_A. Nevertheless, since the agent is biased, the principal intervenes in order to undo the distortion in the agent's initial decision, taking into account the cost of intervention. This distortion, which is given by $\frac{h(\theta_P)}{1+c_P}$, is independent of the agent's initial decision, a_A. Therefore, the intensity of intervention, measured by $|a_P - a_A|$, is beyond the agent's control. For this reason, c_A does not factor into the agent's consideration when choosing a_A. This

[26] In the baseline model, the agent has private information about his private benefits β, but given the agent decision, β is payoff irrelevant from the principal's point of view.

[27] A linear equilibrium means that the agent's decision is linear in θ_A and $E[\theta_P|m]$. Notice that without intervention (i.e., $c_P = \infty$) all equilibria are linear.

observation implies that when the agent is privately informed, the disciplinary punishment that intervention imposes on the agent is mitigated. Nevertheless, disobedience still provides the agent with informational benefits, and hence, intervention always prompts disobedience.

Finally, notice that if $c_A > 0$ then $\frac{1+c_P}{c_P} > \frac{1+c_P}{c_A/c_P+c_P}$. That is, with private information, the agent is less likely to comply with the principal's instructions, and consequently, communication becomes nosier. This result does not hold when the principal does not have the option to intervene. In those cases, the amount of information that is communicated by the principal is invariant to the agent's private information (Harris and Raviv (2005)).

4 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 intervention reinforces compliance if and only if the cost that intervention imposes on the agent is large relative to the cost that is incurred by the principal. Somewhat surprisingly, intervention can prompt disobedience since it tempts the agent to challenge the principal to back her words with actions. In this respect, intervention is counterproductive, and a commitment not to intervene can be optimal. The analysis provides a novel argument as to why a commitment not to intervene can be beneficial to the principal, echoing the common wisdom that effective deliberation and compliance can be sustained when the power to make unilateral decisions is restrained.

The insights derived in this paper can be applied to various contexts: organizations, parenting, corporate finance, regulation, politics, and diplomacy. 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] Burkart, Mike, Denis Gromb, and Fausto Panunzi. 1997. "Large shareholders, monitoring, and the value of the firm." T*he Quarterly Journal of Economics* 112 (3), 693–728.

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

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

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

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

[13] 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.

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

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

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

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

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

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

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

[23] Mylovanov, Tymofiy. 2008. "Veto-based delegation," J. Econ. Theory 138 297–307.

[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

A.1 Proofs of Section 2

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 $a_A^*(m, \beta) : [\underline{\theta}, \overline{\theta}] \times [0, \infty) \to \{L, R\}$, 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, $\rho^*(\theta) \in \arg\max_{m \in M} \mathbb{E}[u_P(\theta, a_A^*(m, \beta), a_P^*(\theta, a_A), e^*(\theta, a_A))]$, where the expectations are taken with respect to β.*

(*ii*) *For any $m \in M$, $a_A^*(m, \beta) \in \arg\max_{\hat{a} \in \{L,R\}} \mathbb{E}[u_A(\theta, \hat{a}, a_P^*(\theta, \hat{a}), e^*(\theta, \hat{a}), \beta) | \rho^*(\theta) = m]$, where the expectations are taken with respect to θ.*

(*iii*) *For any θ and $a_A \in \{L, R\}$, $e^*(\theta, a_A) \in \arg\max_{\hat{e} \in \{0,1\}} u_P(\theta, a_A, a_P^*(\theta, a_A), \hat{e})$.*

(*iv*) *For any θ and a_A, if $e^*(\theta, a_A) = 1$ then $a_P^*(\theta, a_A) \in \arg\max_{\hat{a} \in \{L,R\}} u_P(\theta, a_A, \hat{a}, 1)$.*

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

Proof of Lemma 2. Suppose $m_0 \in M$. Based on Lemma 1, $e = 1$ if and only if $a_A = R$ and $\Delta(\theta) < -c_P$, or $a_A = L$ and $\Delta(\theta) > c_P$. Therefore, if $a_A = R$, the agent's expected utility is

$$\begin{aligned} &\Pr[\Delta(\theta) \geq -c_P | m_0] \, \mathbb{E}[v(\theta, R) + \beta | \Delta(\theta) \geq -c_P, m_0] \\ &+ \Pr[\Delta(\theta) < -c_P | m_0] \, [\mathbb{E}[v(\theta, L) | \Delta(\theta) < -c_P, m_0] - c_A], \end{aligned} \tag{23}$$

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

$$\begin{aligned} &\Pr[\Delta(\theta) \geq c_P | m_0] \, [\mathbb{E}[v(\theta, R) + \beta | \Delta(\theta) \geq c_P, m_0] - c_A] \\ &+ \Pr[\Delta(\theta) < c_P | m_0] \, \mathbb{E}[v(\theta, L) | \Delta(\theta) < c_P, m_0], \end{aligned} \tag{24}$$

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

$$b(m_0) \equiv c_A \frac{\Pr[\Delta(\theta) < -c_P | m_0] - \Pr[\Delta(\theta) \geq c_P | m_0]}{\Pr[-c_P \leq \Delta(\theta) < c_P | m_0]} - \mathbb{E}[\Delta(\theta) | -c_P \leq \Delta(\theta) < c_P, m_0] \tag{25}$$

as required. ■

Proposition 1 [extended] *A non-influential equilibrium always exists. In any non-influential equilibrium the agent chooses action R if and only if $\beta \geq b^{NI}$* where

$$b^{NI} = c_A \times \frac{\Pr\left[\Delta\left(\theta\right) < -c_P\right] - \Pr\left[\Delta\left(\theta\right) > c_P\right]}{\Pr\left[-c_P < \Delta\left(\theta\right) < c_P\right]} - \mathbb{E}\left[\Delta\left(\theta\right) \mid -c_P < \Delta\left(\theta\right) < c_P\right]. \tag{26}$$

Proof. Given message m_0, the agent chooses action R if and only if $\beta \geq b\left(m_0\right)$, where $b\left(m_0\right)$ is given by (25). When the equilibrium is non-influential, no message is informative about θ. Therefore, $b\left(m_0\right) = b^{NI}$ for any m_0, where b^{NI} is given by (26). If random variable $\Delta\left(\theta\right)$ is symmetric and $\mathbb{E}\left[\Delta\left(\theta\right)\right] \geq 0$, then $b^{NI} \leq 0$, and hence, the agent always chooses action R in a non-influential equilibrium ■

Proof of Lemma 3. According to Lemma 1, $e\left(\theta, a_A\right) = 1$ if and only if $a_A = R$ and $\Delta\left(\theta\right) < -c_P$, or $a_A = L$ and $\Delta\left(\theta\right) > c_P$. Based on Lemma 2, $\Pr\left[a_A\left(m, \beta\right) = R\right] = 1 - G\left(b\left(m\right)\right)$, where $b\left(m\right)$ is given by (25). Thus, if the principal sends message m, her expected utility conditional on θ is:

$$\mathbb{E}\left[u_P|\theta, m\right] = v\left(\theta, L\right) + \begin{cases} \left(1 - G\left(b\left(m\right)\right)\right)\left(-c_P\right) & \text{if } \Delta\left(\theta\right) < -c_P \\ \left(1 - G\left(b\left(m\right)\right)\right)\Delta\left(\theta\right) & \text{if } -c_P \leq \Delta\left(\theta\right) < c_P \\ \left(1 - G\left(b\left(m\right)\right)\right)\Delta\left(\theta\right) + G\left(b\left(m\right)\right)\left(\Delta\left(\theta\right) - c_P\right) & \text{if } c_P \leq \Delta\left(\theta\right) \end{cases} \tag{27}$$

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

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 $\beta \leq b\left(m_0\right)$, where $b\left(m_0\right)$ is given by (25). According to Definition 1, $\min_{m \in M} b\left(m\right) < \max_{m \in M} b\left(m\right)$, and both M_R and M_L are not empty.

Suppose $m_0 \in M_R$. According to Lemma 3, $\Delta\left(\theta\right) \geq 0$. Therefore, (25) can be rewritten as

$$b\left(m_0\right) = -c_A \frac{\Pr\left[\Delta\left(\theta\right) \geq c_P|m_0\right]}{\Pr\left[0 \leq \Delta\left(\theta\right) < c_P|m_0\right]} - \mathbb{E}\left[\Delta\left(\theta\right)|0 \leq \Delta\left(\theta\right) < c_P, m_0\right] \tag{28}$$

which is always negative. Since $\beta > 0$ with probability one, the agent follows the principal's instructions and chooses action R with probability one.

Suppose $m_0 \in M_L$. According to Lemma 3, $\Delta(\theta) < 0$. Therefore, (25) can be rewritten as

$$b(m_0) = c_A \frac{\Pr[\Delta(\theta) < -c_P | m_0]}{\Pr[-c_P \le \Delta(\theta) < 0 | m_0]} - \mathbb{E}[\Delta(\theta) | -c_P \le \Delta(\theta) < 0, m_0]. \tag{29}$$

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 $\Delta(\theta) < 0$, some algebra and integration over all $m_0 \in M_L$ show that $b(m_0) = b^*(c_A, c_P)$. Therefore, in any influential equilibrium, the agent follows the principal's instructions and chooses action L if and only if $\beta \le b^*(c_A, c_P)$. 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 $\Delta(\theta) \ge 0$ and message $m_L \ne m_R$ otherwise. As was proved above, the agent always follows the principal's instructions if he observes message m_R. Since $m = m_L$ if and only if $\Delta(\theta) < 0$, (25) evaluated at m_L can be rewritten as $b^*(c_A, c_P)$. Thus, the agent follows the principal's instructions to implement action L if and only if $\beta \le b^*(c_A, c_P)$. 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

$$\Pr[\Delta(\theta) < -c_P] \, \mathbb{E}[\Delta(\theta) | \Delta(\theta) < -c_P] > \Pr[\Delta(\theta) < 0] \, \mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0],$$

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

Proof of Proposition 3. According to Proposition 1, the principal's expected payoff in any non-influential equilibrium is

$$\mathbb{E}\left[u_P^{NI}\right] = \mathbb{E}[v(\theta, R)] - \Pr[\Delta(\theta) < -c_P] \left(\mathbb{E}[\Delta(\theta) | \Delta(\theta) < -c_P] + c_P\right) \mathbf{1}_{\{c_P < -\Delta(\underline{\theta})\}}, \tag{30}$$

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

$$\mathbb{E}\left[u_A^{NI}(\beta)\right] = \mathbb{E}[v(\theta, R) + \beta] - \Pr[\Delta(\theta) < -c_P] \left(\mathbb{E}[\Delta(\theta) + \beta | \Delta(\theta) < -c_P] + c_A\right) \mathbf{1}_{\{c_P < -\Delta(\underline{\theta})\}}.$$

According to Proposition 2, the principal's expected payoff in any influential equilibrium, $\mathbb{E}\left[u_P^I\right]$, is given by (13). The agent's expected payoff (conditional on β) is given by

$$\mathbb{E}\left[u_A^I(\beta)\right] = \begin{cases} \mathbb{E}[v(\theta, L)] + \Pr[\Delta(\theta) > 0] \, \mathbb{E}[\Delta(\theta) + \beta | \Delta(\theta) > 0] & \textit{if } \beta \le b^*(c_A, c_P) \\ \mathbb{E}\left[u_A^{NI}(\beta)\right] & \textit{else,} \end{cases} \tag{31}$$

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 $\mathbb{E}\left[u_P^I\right] \geq \mathbb{E}\left[u_P^{NI}\right]$. ■

Proof of Corollary 1. Consider several properties of $b^*(c_P)$ as a function of c_P, as given by (8): $b^*(0) > 0$, $b^*(c_P) = -\mathbb{E}[\Delta(\theta)|\Delta(\theta) < 0]$ for all $c_p \geq -\Delta(\underline{\theta})$, and if $c_P \in [0, -\Delta(\underline{\theta}))$, then

$$\frac{\partial b^*(c_P)}{\partial c_P} = [b^*(c_P) - h(c_p)] \frac{f(\Delta^{-1}(-c_P))}{F(\Delta^{-1}(0)) - F(\Delta^{-1}(-c_P))} \frac{\partial \Delta^{-1}(-c_P)}{\partial c_P}, \tag{32}$$

where $h(c_p) \equiv c_P - c_A$. Since $\Delta(\cdot)$ is an increasing function, $\frac{\partial \Delta^{-1}(-c_P)}{\partial c_P} < 0$ and (32) implies $\frac{\partial b^*(c_P)}{\partial c_P} > 0 \Leftrightarrow h(c_p) > b^*(c_P)$.

Suppose there is $c'_P \in [0, -\Delta(\underline{\theta})]$ 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 (32), 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 $c'_P = -\Delta(\underline{\theta})$. Recall, $c_P \geq -\Delta(\underline{\theta}) \Rightarrow b^*(c_P) = -\mathbb{E}[\Delta(\theta)|\Delta(\theta) < 0]$. Since $h'(c_p) > 0$ and $h(c'_P) = b^*(c'_P)$, then $h(c_p) > b^*(c_P)$ for all $c_P > c'_P$ as required. Second, suppose $c'_P \in [0, -\Delta(\underline{\theta}))$ 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 $\frac{\partial h(c_P)}{\partial c_P}|_{c_P=c'_P} = 1 > 0$. 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 (32), $h(c'''_P) = b^*(c'''_P)$ implies $\frac{\partial b^*(c_P)}{\partial c_P}|_{c_P=c'''_P} = 0 < 1 = \frac{\partial h(c_P)}{\partial c_P}|_{c_P=c'''_P}$. 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 $c'_P \in [0, -\Delta(\underline{\theta})]$ such that $h(c'_P) = b^*(c'_P)$, then c'_P is unique. Note that $h(0) < b^*(0)$. Moreover, $h(-\Delta(\underline{\theta})) > b^*(-\Delta(\underline{\theta}))$ if and only if $c_A < \mathbb{E}[\Delta(\theta)|\Delta(\theta) < 0] - \Delta(\underline{\theta})$. Therefore, if $c_A < \mathbb{E}[\Delta(\theta)|\Delta(\theta) < 0] - \Delta(\underline{\theta})$, then from continuity there is a unique $c_P^{\min} \in (0, -\Delta(\underline{\theta}))$ 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 $c_A \geq \mathbb{E}[\Delta(\theta)|\Delta(\theta) < 0] - \Delta(\underline{\theta})$ and on the contrary there is c'_P such that $h(c'_P) = b^*(c'_P)$, then it has to be that $h(-\Delta(\underline{\theta})) > b^*(-\Delta(\underline{\theta}))$, 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 5. First note that $W(c_A, c_P)$ is given by (13). Therefore, with the help of some algebra, $W(c_A, c_P) < W(c_A, -\Delta(\underline{\theta}))$ if and only if

$$1 - \frac{1 - G(b_\infty^*)}{1 - G(b^*(c_P))} > \frac{\int_{\underline{\theta}}^{\Delta^{-1}(-c_P)} (\Delta(\theta) + c_P)\, dF(\theta)}{\int_{\underline{\theta}}^{\Delta^{-1}(0)} \Delta(\theta)\, dF(\theta)}, \tag{33}$$

where $b_\infty^* \equiv -\mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < 0]$. Note that the right hand side of (33) is strictly positive, and based on Corollary 1, its proof, and the properties of $b^*(c_P)$, the left hand side is positive if and only if $c_A < \mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < 0] - \Delta(\underline{\theta})$ and $c_P > c_P^*$, where c_P^* is the unique solution of (12). Therefore, (33) holds only if $c_A < \mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < 0] - \Delta(\underline{\theta})$ and $c_P > c_P^*$. Suppose that is the case. Next, note that both sides of (33) converge to zero as $c_P \to -\Delta(\underline{\theta})$. The derivatives with respect to c_P of the right and left hand side, respectively, are given by

$$\begin{aligned} RHS(c_P) &= \frac{F(\Delta^{-1}(-c_P))}{\int_{\underline{\theta}}^{\Delta^{-1}(0)} \Delta(\theta)\, dF(\theta)} \\ LHS(c_P) &= -\frac{1 - G(b_\infty^*)}{(1 - G(b^*(c_P)))^2} g(b^*(c_P)) \frac{\partial b^*(c_P)}{\partial c_P}, \end{aligned}$$

respectively, where $\frac{\partial b^*(c_P)}{\partial c_P}$ is given by (32). Note that $RHS(c_P) < 0$, $\lim_{c_P \to -\Delta(\underline{\theta})} RHS(c_P) = 0$, and

$$\lim_{c_P \to -\Delta(\underline{\theta})} LHS(c_P) = \left[\begin{array}{c} f(\underline{\theta}) \frac{g(b_\infty^*)}{1 - G(b_\infty^*)} \frac{\partial \Delta^{-1}(-c_P)}{\partial c_P}\Big|_{c_P = -\Delta(\underline{\theta})} \\ \times \frac{\mathbb{E}[\Delta(\theta)|\Delta(\theta) \geq 0] - \Delta(\underline{\theta}) - c_A}{F(\Delta^{-1}(0))} \end{array} \right] < 0$$

From continuity, there is $\overline{c}'_P \in (0, -\Delta(\underline{\theta}))$ such that if $c_P \in (\overline{c}_P, -\Delta(\underline{\theta}))$ then $LHS(c_P) < RHS(c_P) < 0$, and hence, there is $\overline{c}_P \in (\overline{c}'_P, -\Delta(\underline{\theta}))$ such that (33) holds, as required. ■

A.2 Proofs of Section 3

A.2.1 Proofs of Section 3.1

The next definition extends the concept of influential equilibrium to Section 3.1.

Definition 3 *An equilibrium is influential if there exist $m_1 \neq m_2 \in M$ and $\beta_0 \geq 0$ 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 $a_F(m_1, \beta_0) \neq a_F(m_2, \beta_0)$.*

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 $m \in M$ and $\beta \geq 0$, $a_A(m,\beta) = a_A^* \in \{L, R\}$. Moreover, there are $m_1, m_2 \in M$ and β_0 such that $a_F(m_1,\beta_0) \neq a_F(m_2,\beta_0)$. Without the loss of generality, suppose that

$$\Pr[a_F = R | a_A = a_A^*, e = 1, m_1] < \Pr[a_F = R | a_A = a_A^*, e = 1, m_2], \tag{34}$$

where $e = 1$ represents cases where the principal intervenes in order to reverse the agent's initial decision. Suppose $a_A^* = R$ ($a_A^* = L$). If $e = 1$ then the principal strictly prefers message m_1 over m_2 (m_2 over m_1). Since by assumption no message can affect the agent's initial decision, it must be $m_2 \notin M$ ($m_1 \notin M$), yielding a contradiction. ■

Proof of Proposition 6. Suppose an influential equilibrium exists. As in Lemma 3, the principal sends message $m \in M_R$ if and only if $\Delta(\theta) \geq 0$. Therefore, if $m \in M_R$ then $a_A = R$ with probability one and the principal reinforces that decision. Suppose $m \in M_L$. Therefore, $\Delta(\theta) < 0$. If $a_A = L$ then the principal reinforces that decision and the agent's utility is $\mathbb{E}[v(\theta,L)|m]$. Suppose $a_A = R$. The principal never has strict incentives to reinforce that decision. Thus, if the principal intervenes, it is only to reverse the agent's decision to L. Let μ_1 be the probability the agent chooses action L if the principal intervenes to reverse the agent's initial decision. Similarly, let μ_0 be the probability the agent chooses action L if the principal does not intervene. If the principal intervenes, her payoff is $(1-\mu_1)v(\theta,R) + \mu_1 v(\theta,L) - c_P$, and if she does not, her payoff is $(1-\mu_0)v(\theta,R) + \mu_0 v(\theta,L)$. Therefore, the principal intervenes if and only if,

$$(\mu_1 - \mu_0)\Delta(\theta) < -c_P \tag{35}$$

Notice that if $\mu_1 - \mu_0 \leq 0$ then the principal never intervenes. Intuitively, if upon non-intervention the probability the agent chooses L is larger than upon intervention, the principal prefers not intervening and saving the cost c_P. Therefore, intervention must be off the equilibrium path. Since in this equilibrium the agent learns nothing from principal's decision not to intervene, he never revises his decision, that is, $\mu_0 = 0$. Since $\mu_1 - \mu_0 \leq 0$, then $\mu_1 = 0$ as well. Below we show that these equilibria do not survive the Grossman and Perry (1986) criterion.

Consider an equilibrium where $\mu_1 - \mu_0 > 0$. Based on 35, the principal intervenes if and only if $\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}$. Therefore, if $m \in M_L$ then $a_A = L$ if and only if,

$$\begin{aligned} 0 \;\; \geq \;\; & \Pr\left[\Delta(\theta) \geq -\frac{c_P}{\mu_1-\mu_0}|m\right]\max\left\{0, \mathbb{E}\left[\Delta(\theta)+\beta|\Delta(\theta) \geq -\frac{c_P}{\mu_1-\mu_0}, m\right]\right\} \\ & + \Pr\left[\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}|m\right]\left[\max\left\{0, \mathbb{E}\left[\Delta(\theta)+\beta|\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}, m\right]\right\} - c_A\right] \end{aligned}$$

Using some algebra, the above condition holds if and only if

$$\begin{aligned} -\mathbb{E}\left[\Delta(\theta)|\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}, m\right] \;\; < \;\; & \beta \leq c_A \Pr\left[\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}|m\right] - \mathbb{E}\left[\Delta(\theta)|m\right] \\ & \textit{or} \\ \beta \;\; \leq \;\; & \min\left\{\begin{array}{c} -\mathbb{E}\left[\Delta(\theta)|\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}, m\right], \\ c_A\frac{\Pr\left[\Delta(\theta)<-\frac{c_P}{\mu_1-\mu_0}|m\right]}{\Pr\left[\Delta(\theta)\geq-\frac{c_P}{\mu_1-\mu_0}|m\right]} - \mathbb{E}\left[\Delta(\theta)|\Delta(\theta) \geq -\frac{c_P}{\mu_1-\mu_0}, m\right] \end{array}\right\}, \end{aligned}$$

which is equivalent to

$$\beta \leq \begin{cases} c_A \Pr\left[\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}|m\right] - \mathbb{E}\left[\Delta(\theta)|m\right] & \textit{if} \;\; \frac{\mathbb{E}[\Delta(\theta)|m]-\mathbb{E}\left[\Delta(\theta)|\Delta(\theta)<-\frac{c_P}{\mu_1-\mu_0},m\right]}{\Pr\left[\Delta(\theta)<-\frac{c_P}{\mu_1-\mu_0}|m\right]} < c_A \\ c_A\frac{\Pr\left[\Delta(\theta)<-\frac{c_P}{\mu_1-\mu_0}|m\right]}{\Pr\left[\Delta(\theta)\geq-\frac{c_P}{\mu_1-\mu_0}|m\right]} - \mathbb{E}\left[\Delta(\theta)|\Delta(\theta) \geq -\frac{c_P}{\mu_1-\mu_0}, m\right] & \textit{else.} \end{cases} \tag{36}$$

Notice that in the first case, where c_A is large, the threshold below which $a_A = L$ is larger than $-\mathbb{E}\left[\Delta(\theta)|\Delta(\theta) < -\frac{c_P}{\mu_1-\mu_0}, m\right]$ and $-\mathbb{E}\left[\Delta(\theta)|\Delta(\theta) \geq -\frac{c_P}{\mu_1-\mu_0}, m\right]$. This means that the agent never revises his initial decision if he chooses $a_A = R$, and therefore, the principal never has incentives to intervene. Formally, $\mu_0 = \mu_1 = 1$, which contradicts, $\mu_1 - \mu_0 > 0$. We conclude that the first case cannot be an equilibrium. In the second case, where c_A is small, note that the threshold below which $a_A = L$ is larger than $-\mathbb{E}\left[\Delta(\theta)|\Delta(\theta) \geq -\frac{c_P}{\mu_1-\mu_0}, m\right]$, which means that the agent never revises his initial decision to choose R upon non-intervention. That is, in this equilibrium it must be $\mu_0 = 0$. Notice, that the principal, as in the baseline model, benefits if the agent believes that $\Delta(\theta)$ is as small as possible: if the agent believes that $\Delta(\theta)$ is small, he is more likely to choose action L if the principal intervenes and intervention fails. At the same time, if the agent is more likely to choose L upon failed intervention, it means that intervention is more effective, which increases the expected payoff of the principal. Therefore, the only information revealed by the principal if $m \in M_L$ is that $\Delta(\theta)$ is negative. We conclude, if $m \in M_L$ then $a_A = L$ if and only if, $\beta \leq b^*(c_A, c_P/\mu)$ where μ must be a solution of

$$1-\mu = \frac{1-G\left(-\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < -c_P/\mu\right]\right)}{1-G\left(b^*\left(c_A, c_P/\mu\right)\right)}$$

Notice that if μ solves the above equation, then $b^*(c_A, c_P/\mu) \leq -\mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < -c_P/\mu]$, and hence, $\frac{\mathbb{E}[\Delta(\theta)|\Delta(\theta)<0]-\mathbb{E}[\Delta(\theta)|\Delta(\theta)<-c_P/\mu]}{\frac{\Pr[\Delta(\theta)<-c_P/\mu]}{\Pr[\Delta(\theta)<0]}} \geq c_A$ as required. To show that a solution in $(0,1)$ always exists, note that $RHS(\mu)$ is a continuous function where $\lim_{\mu\to 0} RHS(\mu) < 1$ and $\lim_{\mu\to 1} RHS(\mu) > 0$. Therefore, by the intermediate value theorem, a solution in $(0,1)$ always exists.

Finally note that in any equilibrium where $\mu^{**} \in (0,1)$ intervention and non-intervention are on the equilibrium path, and hence, the Grossman and Perry (1986) criterion is trivially satisfied. We argue that an equilibrium where $\mu^{**} = 0$ (and hence, intervention is off the equilibrium path) does not survive the Grossman and Perry (1986) criterion. To see why, 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 $\beta \leq -\mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < 0]$. Suppose $\Delta(\theta) < 0$ and the principal recommends on action L but the agent decides on R. Let $\hat{\mu}$ be the solution of

$$1-\mu = \frac{1-G\left(-\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < -c_P/\mu\right]\right)}{1-G\left(-\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < 0\right]\right)}, \tag{37}$$

Note that a solution in $(0,1)$ always exists as $\lim_{\mu\to 0} RHS(\mu) < 1$ and $\lim_{\mu\to 1} RHS(\mu) > 0$. Consider the following deviation: the principal intervenes if and only if $\Delta(\theta) < -c_P/\hat{\mu}$. If the agent expects the principal to behave this way, the agent has incentives to revise the decision from R to L upon intervention if and only if $\beta \leq -\mathbb{E}[\Delta(\theta)\,|\Delta(\theta) < -c_P/\hat{\mu}]$. Given this behavior, the principal has incentives to deviate and intervene if and only if $\Delta(\theta) < -c_P/\hat{\mu}$. Indeed, since $\hat{\mu}$ solves (37), if the principal deviates and intervenes, she expects the agent to revise his decision to L with probability $\hat{\mu}$. Therefore, the net benefit from intervention is $-c_P - \Delta(\theta)/\hat{\mu}$. The existence of this deviation violates the Grossman and Perry (1986) criterion. ■

A.2.2 Proofs of Section 3.2

Proposition 9 *Consider the setup in Section 3.2. Then,*

(*i*) *Proposition 2 holds with the exception that if the principal instructs the agent to choose* *action L, the agent chooses action L if and only if $\beta \leq \gamma b^* (c_A/\gamma, c_P/\gamma)$.*

(ii) *Proposition 4 holds with the exception that (12) is replaced by (19).*

(iii) *If $c_A = 0$, $\Delta(\theta) = \theta$, and $f(-c_P/\gamma)(c_P/\gamma) > F(-c_P/\gamma)$ then $\frac{\partial}{\partial\gamma}[\gamma b^* (c_A/\gamma, c_P/\gamma)] < 0$.*

Proof. Consider part (i). Based on (18) and since $\mathbb{E}[\Delta(\theta)] = 0$, $\mathbb{E}[\Delta(\theta)|s] = \gamma\Delta(s)$. Based on (3), conditional on s, the principal prefers action L over action R if and only if $\Delta(s) \leq 0$, and notice that s has the same distribution as θ. If $a_A = R$ then the principal intervenes if and only if $\Delta(s) \leq -c_P/\gamma$. Similar to (7) and the derivation of Proposition 2, the agent follows the instructions of the principal to choose action L if and only if,

$$\begin{aligned}
\mathbb{E}[v(\theta, L)|\Delta(s) \leq 0] &\geq \Pr[\Delta(s) \geq -c_P/\gamma|\Delta(s) \leq 0]\,\mathbb{E}[v(\theta, R) + \beta| - c_P/\gamma \leq \Delta(s) < 0] \\
&\quad + \Pr[\Delta(s) < -c_P/\gamma|\Delta(s) \leq 0]\,(\mathbb{E}[v(\theta, L)|\Delta(s) < -c_P/\gamma] - c_A) \Leftrightarrow \\
\beta &\leq c_A \frac{\Pr[\Delta(s) < -c_P/\gamma]}{\Pr[-c_P/\gamma \leq \Delta(s) < 0]} - \mathbb{E}[\Delta(\theta)| - c_P/\gamma \leq \Delta(s) < 0].
\end{aligned}$$

Since

$$\mathbb{E}[\Delta(\theta)| - c_P/\gamma \leq \Delta(s) < 0] = \gamma\mathbb{E}[\Delta(\theta)| - c_P/\gamma \leq \Delta(\theta) < 0]$$

and s has the same distribution as θ, the agent follows the instructions of the principal to choose action L if and only if,

$$\begin{aligned}
\beta &\leq c_A \frac{\Pr[\Delta(\theta) < -c_P/\gamma]}{\Pr[-c_P/\gamma \leq \Delta(\theta) < 0]} - \gamma\mathbb{E}[\Delta(\theta)| - c_P/\gamma \leq \Delta(\theta) < 0] \\
&= \gamma b^* (c_A/\gamma, c_P/\gamma),
\end{aligned}$$

as required.

Consider part (ii). Without intervention, the agent follows the principal instructions if and only if $\beta \leq -\mathbb{E}[\Delta(\theta)|\Delta(s) < 0]$. Notice that

$$-\mathbb{E}[\Delta(\theta)|\Delta(s) < 0] = -\gamma\mathbb{E}[\Delta(\theta)|\Delta(\theta) < 0].$$

Therefore,

$$\begin{array}{rcl}
\gamma b^*\left(c_A/\gamma, c_P/\gamma\right) & < & -\mathbb{E}\left[\Delta\left(\theta\right)|\Delta\left(s\right)<0\right] \Leftrightarrow \\
c_A \dfrac{\Pr\left[\Delta\left(\theta\right)<-c_P/\gamma\right]}{\Pr\left[-c_P/\gamma \leq \Delta\left(\theta\right)<0\right]} - \gamma\mathbb{E}\left[\Delta\left(\theta\right)| - c_P/\gamma \leq \Delta\left(\theta\right)<0\right] & < & -\gamma\mathbb{E}\left[\Delta\left(\theta\right)|\Delta\left(\theta\right)<0\right] \Leftrightarrow \\
c_A/\gamma & < & \mathbb{E}\left[\Delta\left(\theta\right)|\Delta\left(\theta\right)<0\right] \\
& & -\mathbb{E}\left[\Delta\left(\theta\right)|\Delta\left(\theta\right)<-c_P/\gamma\right],
\end{array}$$

as required.

Consider part (iii). Suppose $c_A = 0$ and $\Delta\left(\theta\right) = \theta$. Then,

$$\gamma b^*\left(c_A/\gamma, c_P/\gamma\right) = -\gamma\mathbb{E}\left[\theta| - c_P/\gamma \leq \theta < 0\right].$$

Therefore,

$$\begin{array}{rcl}
\dfrac{\partial}{\partial\gamma}\left[\gamma b^*\left(c_A/\gamma, c_P/\gamma\right)\right] & = & -\dfrac{\int_{-c_P/\gamma}^{0}\theta f\left(\theta\right)d\theta}{F\left(-c_P/\gamma\right)} \\
& & -\dfrac{\left(c_P/\gamma\right)^2 f\left(-c_P/\gamma\right)F\left(-c_P/\gamma\right) - \left[\int_{-c_P/\gamma}^{0}\theta f\left(\theta\right)d\theta\right] f\left(-c_P/\gamma\right)\left(c_P/\gamma\right)}{F\left(-c_P/\gamma\right)^2},
\end{array}$$

and

$$\frac{\partial}{\partial\gamma}\left[\gamma b^*\left(c_A/\gamma, c_P/\gamma\right)\right] < 0 \Leftrightarrow \frac{F\left(-c_P/\gamma\right) - f\left(-c_P/\gamma\right)\left(c_P/\gamma\right)}{F\left(-c_P/\gamma\right)f\left(-c_P/\gamma\right)\left(c_P/\gamma\right)} < \frac{c_P/\gamma}{-\int_{-c_P/\gamma}^{0}\theta f\left(\theta\right)d\theta}.$$

Therefore, if $f\left(-c_P/\gamma\right)\left(c_P/\gamma\right) > F\left(-c_P/\gamma\right)$ then $\frac{\partial}{\partial\gamma}\left[\gamma b^*\left(c_A/\gamma, c_P/\gamma\right)\right] < 0$. ■

A.2.3 Proofs of Section 3.3

Proof of Proposition 7. Conditional on θ and the agent's decision, a_A, and regardless of the message sent by the principal to the agent, the principal solves

$$\begin{array}{rcl}
a_P & \in & \arg\max_x \left\{-\left(\theta - x\right)^2 - c_P\left(x - a_A\right)^2\right\} \\
& \Rightarrow & a_P\left(a_A, \theta\right) = a_A + \dfrac{\theta - a_A}{1 + c_P}.
\end{array}$$

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

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

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

$$\begin{aligned} a_A^* &\in \arg\max_{a_A} \mathbb{E}\left[-(\theta + \beta - a_P(a_A, \theta))^2 - c_A(a_P(a_A, \theta) - a_A)^2 |m\right] \\ &\Rightarrow a_A^* = \mathbb{E}[\theta|m] + \beta \frac{1 + c_P}{c_A/c_P + c_P}. \end{aligned}$$

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 $c_P = \infty$, $c_A = 0$ and his preferences are represented by the utility function $-\left(\theta + \beta \frac{1+c_P}{c_A/c_P+c_P} - a_A\right)^2$. ■

A.2.4 Proofs of Section 3.4

Proof of Proposition 8. Consider the game with two-sided information asymmetry and no intervention. As in Harris and Raviv (2005), in any equilibrium, if the agent observes θ_A and the principal sends message m, the agent chooses $a_A(\theta_A, m) = \theta_A + \mathbb{E}[\theta_P|m]$. Therefore, the principal's expected utility from sending message m is

$$-\mathbb{E}\left[(\theta_P + \theta_A - a_A(\theta_A, m))^2 |\theta_P, m\right] = -\mathbb{E}\left[(\theta_P - \mathbb{E}[\theta_P|m])^2 |\theta_P, m\right],$$

which is independent of θ_A, and hence, perfectly predictable by the principal. Intuitively, the unknown private information of the agent is always canceled by the agent's (optimal) choice of a_A. We are back to the standard Crawford and Sobel (1982), where $a_A(\theta_A, m) = \theta_A + \hat{a}_A(m)$ where $\hat{a}_A(m)$ is given by the Crawford and Sobel's model when only the principal has private information.

Consider the model with intervention. Suppose that in equilibrium the agent follows a linear equilibrium strategy

$$a_A(\theta_A, m) = \alpha_1\theta_A + \alpha_2 E[\theta_P|m] + B.$$

Conditional on θ_P, message m, and the agent's decision a_A, the principal solves

$$\max_x \left\{ -\left(\theta_P + H\left(a_A, m\right) - x\right)^2 - c_P \left(x - a_A\right)^2 \right\}$$

where

$$H\left(a_A, m\right) = \frac{1}{\alpha_1} a_A - \frac{\alpha_2}{\alpha_1} E\left[\theta_P | m\right] - \frac{B}{\alpha_1}.$$

Therefore, principal chooses

$$a_P\left(a_A, \theta_P, m\right) = a_A + \frac{\theta_P + H\left(a_A, m\right) - a_A}{1 + c_P}.$$

The agent expects the principal to follow intervention policy $a_P\left(a_A, \theta_P, m\right)$, and therefore, given message m and the observation of θ_A, he solves

$$\max_y \mathbb{E}\left[-\left(\theta_P + \theta_A + \beta - a_P\left(y, \theta_P, m\right)\right)^2 - c_A \left(a_P\left(y, \theta_P, m\right) - y\right)^2 | m\right] \Rightarrow$$

$$a_A = \frac{1}{1 + \frac{1-\alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}} \theta_A + \frac{1 + \frac{\alpha_2 - \alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}}{1 + \frac{1-\alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}} E\left[\theta_P | m\right] + \frac{\beta + \frac{1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P} B}{1 + \frac{1-\alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}}.$$

Matching coefficients implies

$$\alpha_1 = \frac{1}{1 + \frac{1-\alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}} \Leftrightarrow \alpha_1 \in \left\{ 1, \frac{c_A - c_P}{c_A + c_P^2} \right\}.$$

Notice that $\alpha_1 = \frac{c_A - c_P}{c_A + c_P^2}$ cannot be an equilibrium as the matching of

$$\frac{\beta + \frac{1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P} B}{1 + \frac{1-\alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}} = B$$

requires $\beta = 0$. If $\alpha_1 = 1$ then

$$\alpha_2 = \frac{1 + \frac{\alpha_2 - \alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}}{1 + \frac{1-\alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}} \Rightarrow \alpha_2 = 1$$

$$B = \frac{\beta + \frac{1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P} B}{1 + \frac{1-\alpha_1}{\alpha_1} \frac{1 + c_A \frac{1-\alpha_1}{c_P \alpha_1 + 1}}{1 + c_P}} \Rightarrow \beta \frac{1 + c_P}{c_P}$$

This implies

$$a_A(\theta_A, m) = \theta_A + E[\theta_P | m] + \beta \frac{1 + c_P}{c_P},$$

and hence,

$$a_P(a_A, \theta_P, m) = a_A + \frac{\theta_P - E[\theta_P | m] - \beta \frac{1+c_P}{c_P}}{1 + c_P},$$

and note that the second term is independent of the actual action taken by the agent. Thus, if the agent chooses action $a_A(\theta_A, m)$, the principal's utility conditional on θ_P and m is

$$\begin{aligned} u_P(\theta_P, \theta_A, a_A) &= E\left[-\left(\theta_P + \theta_A - a_P(a_A, \theta_P, m)\right)^2 - c_P\left(a_P(a_A, \theta_P, m) - a_A\right)^2 | m, \theta_P\right] \\ &= E\left[\begin{array}{cc} -\left(\theta_P + \theta_A - a_A - \frac{\theta_P - E[\theta_P|m] - \beta\frac{1+c_P}{c_P}}{1+c_P}\right)^2 & \\ -c_P\left(\frac{\theta_P - E[\theta_P|m] - \beta\frac{1+c_P}{c_P}}{1+c_P}\right)^2 & | m, \theta_P \end{array}\right] \end{aligned}$$

Using the fact that in equilibrium $a_A = a_A(\theta_A, m)$ we get

$$\begin{aligned} u_P(\theta_P, \theta_A, a_A(\theta_A, m)) &= -\frac{c_P}{1+c_P} E\left[\left(\theta_P - E[\theta_P|m] - \beta\frac{1+c_P}{c_P}\right)^2 | m, \theta_P\right] \\ &= -\frac{c_P}{1+c_P} E\left[\left(\theta_P + \theta_A - a_A(\theta_A, m)\right)^2 | m, \theta_P\right] \end{aligned}$$

which is perfectly predictable by the principal. It follows that at the communication stage, the principal behaves as if her preferences are represented by the utility function $-(\theta_P + \theta_A - a_A)^2$, and the agent behaves as if $c_P = \infty$, $c_A = 0$ and his preferences are represented by the utility function $-\left(\theta_P + \theta_A + \beta\frac{1+c_P}{c_P} - a_A\right)^2$. ■

Online Appendix for "Can Words Speak Louder Than Actions?"

Doron Levit

Proposition 10 *The agent prefers a principal with the capacity to intervene over a principal without the capacity to intervene if and only if*

$$\beta < \max\left\{-c_A - \mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < -c_P\right], -\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < -c_P\right]\right\}. \tag{38}$$

Proof. The agent's relative expected payoff from disobedience (when asked to choose action L) is

$$\mathbb{E}\left[\Delta(\theta) + \beta|\Delta(\theta) < 0\right] \tag{39}$$

if the principal does not have the capacity to intervene and

$$\Pr\left[\Delta(\theta) \geq -c_P|\Delta(\theta) < 0\right]\mathbb{E}\left[\Delta(\theta) + \beta|\Delta(\theta) \geq -c_P, \Delta(\theta) < 0\right] - \Pr\left[\Delta(\theta) < -c_P|\Delta(\theta) < 0\right]c_A \tag{40}$$

otherwise. The former is greater than the latter if and only if

$$\beta > -c_A - \mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < -c_P\right]. \tag{41}$$

Notice that either

$$-c_A - \mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < -c_P\right] < -\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < 0\right] < b^*(c_A, c_P) \tag{42}$$

or

$$b^*(c_A, c_P) < -\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < 0\right] < -c_A - \mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < -c_P\right]. \tag{43}$$

Consider the first case. Based on Proposition 2, if $\beta \leq -\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < 0\right]$ then, with or without intervention, the agent follows instructions of the principal, and hence, he is indifferent. If $-\mathbb{E}\left[\Delta(\theta)\,|\Delta(\theta) < 0\right] < \beta < b^*(c_A, c_P)$ then the agent follows instructions if and only if the principal has the capacity to intervene. This necessarily means that the agent is worse off with intervention. Indeed, the agent can always secure an expected payoff of $\mathbb{E}\left[v(\theta, L)\,|\Delta(\theta) < 0\right]$ if he chooses to follow instructions. Since the principal never intervenes if her instructions are followed, this payoff is invariant to the capacity of the principal to intervene. Therefore, by revealed preferences, the agent is worse off when the principal has the capacity to intervene.

If $\beta > b^*(c_A, c_P)$ then the agent disobeys the principal whether or not she has the capacity to intervene Since $\beta > -c_A - \mathbb{E}[\Delta(\theta) | \Delta(\theta) < -c_P]$, the agent is worse off with intervention. We conclude, if condition (42) holds, the agent prefers a principal with the capacity to intervene over a principal without the capacity to intervene if and only if $\beta < -\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0]$.

Consider the second case. Based on Proposition 2, if $\beta \leq b^*(c_A, c_P)$ then, with or without intervention, the agent follows instructions of the principal, and hence, he is indifferent. Suppose $b^*(c_A, c_P) < \beta < -\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0]$. The agent follows instructions if and only if the principal does not have the capacity to intervene. This necessarily means that the agent is better off with intervention. Since the agent can always secure an expected payoff of $\mathbb{E}[v(\theta, L) | \Delta(\theta) < 0]$ if he chooses to follow instructions, with or without intervention, by revealed preferences, the fact that he chooses to disobey implies that he is getting a higher payoff than when he is following the instructions. If $-\mathbb{E}[\Delta(\theta) | \Delta(\theta) < 0] < \beta$ then the agent disobeys the principal whether or not she has the capacity to intervene. Based on the analysis above, the agent suffers from intervention if and only if (41) holds. We conclude, if condition (43) holds, the agent prefers a principal with the capacity to intervene over a principal without the capacity to intervene if and only if $\beta < -c_A - \mathbb{E}[\Delta(\theta) | \Delta(\theta) < -c_P]$. The combination of the two cases proves the statement in the proposition, as required. ∎

Proposition 11 *Suppose the agent is allowed to reverse his decision if the principal did not intervene. Then, the agent never revises his initial decision in equilibrium. The set of equilibria is identical to the one in Section 2.*

Proof. As in the baseline model, the agent always follows the demand of the principal to implement action R. Suppose the principal demands the agent to choose action L. If the agent chooses L then the principal can reinforce it and the action is never reversed. Suppose $a_A = R$ and the principal believes that conditional on her decision not to intervene, the agent chooses $a_F = L$ with probability μ. Then, the principal intervenes if and only if

$$v(\theta, L) - c_P > \mu v(\theta, L) + (1-\mu) v(\theta, R) \Leftrightarrow \Delta(\theta) \leq -\frac{c_P}{1-\mu}.$$

Therefore, if the principal does not intervene, the agent chooses $a_F = R$ if and only if

$$\beta \geq -\mathbb{E}\left[\Delta(\theta) | \Delta(\theta) \geq -\frac{c_P}{1-\mu}, m\right],$$

where m is the message from the principal. Therefore, $a_A = L$ if and only if

$$\begin{aligned} 0 &\geq \Pr\left[\Delta(\theta) \geq -\frac{c_P}{1-\mu}|m\right] \max\left\{\mathbb{E}\left[\Delta(\theta) + \beta|\Delta(\theta) \geq -\frac{c_P}{1-\mu}, m\right], 0\right\} - c_A \Pr\left[\Delta(\theta) < -\frac{c_P}{1-\mu}|m\right] \\ &\Leftrightarrow \beta \leq c_A \frac{\Pr\left[\Delta(\theta) < -\frac{c_P}{1-\mu}|m\right]}{\Pr\left[\Delta(\theta) \geq -\frac{c_P}{1-\mu}|m\right]} - \mathbb{E}\left[\Delta(\theta)|\Delta(\theta) \geq -\frac{c_P}{1-\mu}, m\right]. \end{aligned}$$

It follows that if $a_A = R$ then $\beta > -\mathbb{E}\left[\Delta(\theta)|\Delta(\theta) \geq -\frac{c_P}{1-\mu}, m\right]$ for sure, and hence, $\mu = 0$ in any equilibrium. This concludes the proof. ■

Proposition 12 *Suppose the agent is allowed to reverse his decision if the principal did not intervene or intervention failed. A non-influential equilibrium always exists. In non-influential equilibrium there are $b^{**} > 0$ and $\mu^{**} \in (0,1)$ such that the following hold. The agent chooses action L if and only if $\beta \leq b^{**}(c_A, c_P)$ where*

$$b^{**}(c_A, c_P) = b^*(c_A, c_P/\mu^{**}). \tag{44}$$

If the agent chooses action L, the principal reinforces that decision if and only if $\Delta(\theta) < 0$, and the agent reverses his initial decision if and only if the principal does no reinforces his decision. If the agent chooses action R, the principal reinforces that decision if and only if $\Delta(\theta) \geq 0$. If $\Delta(\theta) < 0$ then the principal intervenes to reverse the agent's decision if $\Delta(\theta) < -c_P/\mu^{**}$, and does not intervene if $-c_P/\mu^{**} \leq \Delta(\theta) < 0$. The agent revises his initial decision from R to L, if and only if the principal attempted to reverse the agent's decision and $b^*(c_A, c_P/\mu^{**}) \leq \beta \leq -\mathbb{E}[\Delta(\theta)|\Delta(\theta) < -c_P/\mu^{**}]$, where $\mu^{**} \in (0,1)$ is as given by Proposition 6.

Proof. The principal reinforces the agent's initial decision if and only if $a_A = R$ and $\Delta(\theta) \geq 0$, or $a_A = L$ and $\Delta(\theta) < 0$. Suppose $a_A = R$ and $\Delta(\theta) < 0$. The principal intervenes if and only if,

$$(\mu_1 - \mu_0)\Delta(\theta) < -c_P$$

where μ_e has the same interpretation as in Proposition 6. Notice that if $\mu_1 - \mu_0 \leq 0$ then the principal never intervenes. Consider an equilibrium where $\mu_1 - \mu_0 > 0$. The principal intervenes if and only if $\Delta(\theta) < -\frac{c_P}{\mu_1 - \mu_0}$. Note that since the $a_A = R$ and the principal did not reinforce that decision, the agent infers $\Delta(\theta) < 0$. The agent's expected payoff from choosing

$a_A = R$ is

$$\begin{aligned}
&\mathbb{E}\left[v\left(\theta, L\right)\right] + \Pr\left[\Delta\left(\theta\right) \geq 0\right]\mathbb{E}\left[\Delta\left(\theta\right) + \beta | \Delta\left(\theta\right) \geq 0\right] \\
&+ \Pr\left[-\frac{c_P}{\mu_1 - \mu_0} \leq \Delta\left(\theta\right) < 0\right]\max\left\{0, \mathbb{E}\left[\Delta\left(\theta\right) + \beta | -\frac{c_P}{\mu_1 - \mu_0} \leq \Delta\left(\theta\right) < 0\right]\right\} \\
&+ \Pr\left[\Delta\left(\theta\right) < -\frac{c_P}{\mu_1 - \mu_0}\right]\left[\max\left\{0, \mathbb{E}\left[\Delta\left(\theta\right) + \beta | \Delta\left(\theta\right) < -\frac{c_P}{\mu_1 - \mu_0}\right]\right\} - c_A\right]
\end{aligned}$$

Suppose $a_A = L$ and $\Delta\left(\theta\right) \geq 0$. The principal does not reinforce the agent's decision, and hence, the agent infers that $\Delta\left(\theta\right) \geq 0$ and he will revise his decision once he is given the opportunity. Therefore, the principal never intervenes, and the agent reverses his decision. The agent's expected payoff from choosing $a_A = L$ is

$$\mathbb{E}\left[v\left(\theta, L\right)\right] + \Pr\left[\Delta\left(\theta\right) \geq 0\right]\mathbb{E}\left[\Delta\left(\theta\right) + \beta | \Delta\left(\theta\right) \geq 0\right].$$

Comparing the two terms, the agent chooses $a_A = L$ if and only if

$$\begin{aligned}
0 \ \geq \ &\Pr\left[-\frac{c_P}{\mu_1 - \mu_0} \leq \Delta\left(\theta\right) < 0\right]\max\left\{0, \mathbb{E}\left[\Delta\left(\theta\right) + \beta | -\frac{c_P}{\mu_1 - \mu_0} \leq \Delta\left(\theta\right) < 0\right]\right\} \\
&+ \Pr\left[\Delta\left(\theta\right) < -\frac{c_P}{\mu_1 - \mu_0}\right]\left[\max\left\{0, \mathbb{E}\left[\Delta\left(\theta\right) + \beta | \Delta\left(\theta\right) < -\frac{c_P}{\mu_1 - \mu_0}\right]\right\} - c_A\right]
\end{aligned}$$

This term is similar to the one in Proposition 6, and hence, similar to that proof, we conclude that the agent chooses action L if and only if $\beta \leq b^{**}\left(c_A, c_P\right)$ where $b^{**}\left(c_A, c_P\right) = b^{*}\left(c_A, c_P/\mu^{**}\right)$ and μ^{**} is given as in Proposition 6. ■

Example of Proposition 9 part (iii). Suppose $f\left(\theta\right) = \frac{2}{\pi r^2}\sqrt{r^2 - x^2}$ (a Wigner semicircle distribution with parameter $r > 0$). This distribution is symmetric around zero, and hence, $F\left(0\right) = \frac{1}{2}$. Since $f\left(-c_P/\gamma\right)\left(c_P/\gamma\right) + F\left(-c_P/\gamma\right) < F\left(0\right)$, and $F\left(0\right) = \frac{1}{2}$, if $f\left(-c_P/\gamma\right)\left(c_P/\gamma\right) > \frac{1}{4}$ then it must be $F\left(-c_P/\gamma\right) < \frac{1}{4}$, and hence, $f\left(-c_P/\gamma\right)\left(c_P/\gamma\right) > F\left(-c_P/\gamma\right)$. Suppose $c_P/\gamma = \frac{r}{\sqrt{2}}$. Then

$$f\left(-c_P/\gamma\right)\left(c_P/\gamma\right) > \frac{1}{4} \Leftrightarrow \frac{2}{\pi r^2}\sqrt{r^2 - \left(-\frac{r}{\sqrt{2}}\right)^2}\frac{r}{\sqrt{2}} > \frac{1}{4} \Leftrightarrow \pi < 4,$$

which always holds. ■