Harmful monitoring

Marco A. Haan* Bart Los† Yohanes E. Riyanto‡

January 19, 2007

Abstract

We show that there may be circumstances in which a principal prefers not to observe the project choice of an agent that acts on her behalf. The ability of the agent is private information. Projects differ with respect to the amount of risk. If the principal can observe the project choice of the agent, the latter will use that choice as a signal. In the separating equilibrium, an agent with high ability then chooses a project that is too risky. If more difficult projects require more effort, there are two opposite effects. The shirking effect implies that the agent chooses a project that is too safe. The signaling effect implies that he chooses a project that is too risky. The net effect is ambiguous. We also discuss the implications of our model for promotion policies.

JEL Classification Codes: D82, L23, M52.

Keywords: Principal-Agent, Project Choice, Career Concerns, Signalling.

*Department of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands. Phone: +31-50-3637327. Fax: +31-50-3637227. e-mail: m.a.haan@rug.nl.

†Department of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands. Phone: +31-50-3637317. Fax: +31-50-3637227. e-mail: b.los@rug.nl.

‡Department of Economics, FASS, National University of Singapore, AS2 1Arts Link, Singapore 117570. e-mail: ecsrye@nus.edu.sg, Phone:+65-8746939, Fax: +65-7752646. The authors thank participants at EARIE 2005, Pim Heijnen, Allard van der Made, Sander Onderstal, Bastiaan Overvest and especially Linda Toolsema for useful comments.

1 Introduction

A wealth of literature studies the problem that a principal faces if she wants an agent to exert effort or choose a project on her behalf. Often, this problem is studied using the canonical principal-agent model. The optimal solution suggested by that literature has the principal offering an explicit incentive scheme that balances the trade-off between incentives and risk (see e.g. Laffont and Martimort, 2001).

Yet, the career concerns literature shows that agents may be willing to act in the principal's interest, even in the absence of explicit incentives (see e.g. Holmstrom, 1982 or 1999, Gibbons and Murphy, 1992, or Dewatripont, Jewitt and Tirole, 1999). Models in this literature typically assume that the agent's ability is unknown to both the principal and the agent. Consider a simple model with two periods. Suppose that the agent's output in period 1 is a function of his ability, his effort, and some noise. In the second period, the agent's wage will depend on his perceived ability which in turn depends on realized output. This gives the agent an incentive to supply effort in the first period, which is in the interest of the principal.

This suggests that career concerns of an agent are good news for the principal. It induces the agent to work harder than he otherwise would, in an attempt to impress the marketplace about his ability. Of course, it does not imply that career concerns yield a first-best outcome. Holmstrom (1982) shows that in a multiperiod setting, from a welfare point of view, the agent's equilibrium effort is excessive in early periods and insufficient in

later stages of his career. Both Hirshleifer and Thakor (1992) and Zwiebel (1995) provide models in which the agent's project choice is either too conservative or too safe.

In this paper, we provide a model in which career concerns, and the agent's desire to impress the marketplace, may actually *hurt* the principal. In a nutshell, the argument is as follows. Consider a case in which the agent knows his own ability, but the principal does not. The agent has to implement a project on behalf of the principal. Perceptions of his ability are important, as they are in the standard career concerns model. The agent then has an incentive to try to impress the principal (or, more generally: the job market) by undertaking difficult projects, and thereby signalling his ability. This is especially true if the outcome of the project is hard to observe in the short term, or if the outcome is only a noisy function of ability. By choosing to undertake difficult projects, the agent gives the impression that he has high ability. In fact, the agent may even be inclined to undertake a project that is too difficult for him. Obviously, this is not in the best interest of the principal. In such a situation, the principal is better off if she cannot observe the difficulty of the project that the agent implements, so he is not able to impress her by his choice of project.

In our benchmark model, the interests of principal and agent are perfectly aligned in the sense that both have the same long-run objectives. For example, both a venture capitalist and an entrepreneur have an interest in implementing a profitable project. Also, both an economist and an economics department have an interest in having high-quality publications. The principal wants the agent to implement the project that gives her the

highest expected long-term pay-off. If the project choice is unobservable, the agent himself also wants to implement that project, as it best serves his long-run prospects on the job market. If the project choice is observable, however, the agent has an incentive to signal his ability by choosing a more difficult project.

We extend the model to allow for the agent's disutility of effort. In that case, the principal faces a trade-off. If she remains uninformed about the agent's project choice, the agent will not choose a project that is too difficult. But he may then choose a project that is too easy, as easy projects require less effort. We thus have a trade-off between signalling and shirking.

Our model applies to situations in which workers have a high discretion to implement their own projects. Knowledge workers are a good example. For instance, consider the case of a young economist who contemplates a suitable outlet for his work. First suppose that his department will only be able to observe his actual publication record. The choice of journal will then be a trade-off between the probability of acceptance and the quality of the journal. The interests of the economist and his department are perfectly aligned. But now consider the case in which the department can also observe where this economist submits his paper, and has to base career decisions on that observation. In that case, he will be inclined to send his paper to a better journal than he otherwise would, in an attempt to signal that his quality is higher than it really is. Similarly, a scientist that applies for an NSF-grant, will be evaluated on the basis of his project, rather than on whether the project will actually be successful. Such a scientist then has an incentive to propose projects that

are too ambitious.

As another example, consider a firm that consists of a number of departments. The manager of each department proposes and implements a risky project that has a high probability of failure. The probability that a project is successful, depends on the unobservable quality of a manager. The remuneration and career opportunities of a manager depend on his perceived ability. When the executive board can only observe whether or not a project has been implemented and the interests of the firm as a whole and each individual manager are perfectly aligned, then the manager will choose the project that maximizes expected profits. But when the executive board can also observe the nature of the project that is implemented, a manager will be inclined to implement a riskier project than he otherwise would, in an attempt to signal that his ability is high. Also in this case, even though the incentives of the individual and the group are perfectly aligned, the desire to signal leads to a decision that is more risky than the first-best solution.

The remainder of the paper is organized as follows. We first discuss some related literature in section 2. Section 3 describes the basic set-up of our benchmark model, in which the interests of principal and agent are perfectly aligned. In section 4, we solve that model. We show that the principal always prefers not to be able to monitor the agent's choice of project — or at least, to commit not to let such an observation influence the agent's career. In section 5, we introduce a conflict of interest between principal and agent. Now, the principal prefers to monitor the agent's project choice when the conflict of interest is sufficiently large. Section 6 considers an extension in which the principal

can commit to her probability of monitoring. In that case, she can often find the perfect balance between signalling and shirking, and always induce the agent to implement the first-best project. Finally, section 7 concludes.

2 Related Literature

In the introduction, we already noted that most of the existing literature predicts that agents will be too conservative or safe in the presence of career concerns. Zwiebel (1995) shows that career concerns may induce an agent to refrain from undertaking innovations that are risky. Suppose that the principal uses some relative performance evaluation. Implementing non-standard, risky innovations then implies that the agent will be evaluated less accurately than his peers who undertake standard innovations. Hence, in this sense career concerns lead to excessive conservatism in the selection of projects.

In a corporate finance framework, Hirshleifer and Thakor (1992) demonstrate that a manager may deliberately distort investment policy in favor of relatively safe projects in an attempt to build his reputation. In their model, the manager's ability and his project choices are private information. The manager's ability can only be inferred from observed successes or failures. As failures are more readily observable than successes, the manager has an incentive to choose projects that are relatively safe and will be more likely to succeed.

Kanodia, Bushman and Dickhaut (1989) give another reason why a manager may be too conservative. Suppose that a manager, who initially has invested in a new production

equipment, learns that he has made the wrong choice, and that there is equipment available that could do the job at lower cost. The manager may then still be inclined not to switch to the new technology, out of fear that the market will find out that he has made the wrong decision and will therefore perceive him as a low ability manager. In Holmstrom and Ricart i Costa (1986), managers will not undertake projects as often as they should, and are too conservative in that sense. The authors show this in a model of career concerns, where the agent can also be given explicit incentives.

All the models discussed above have one thing in common: career concerns induce decisions by agents that are suboptimal from the point of view of the manager. That is also the case in our model. Yet, in the papers discussed above, the principal is still better off than he would be in the absence of managerial career concerns. In our model, the mere presence of career concerns may actually make the principal strictly worse off.

Landers, Rebitzer and Taylor (1996) consider a law firm consisting of a number of attorneys who work for two periods. In the first period they exert observable effort (the number of working hours), in the second period some attorneys will be selected as partners. There are two types of attorney: the low type dislikes working more than the high type. Partners share revenues, which depend on working hours spend. With complete information attorneys will choose their first-best working hours. With incomplete information, there is a separating equilibrium in which the high type will work inefficiently long hours, for much the same reason as we have in our model: in order to convince the principal of being a high type. Still, there are crucial differences with our set-up. Choosing not to observe

effort will not make principals better off in Landers, Rebitzer and Taylor: the attorney's signalling actually benefits them, as longer hours imply higher output.

Prat (2005) also considers a framework in which a principal may be better off if she is not able to observe the agent's actions. Yet, his set-up is entirely different from ours. In his paper, the agent's ability is unknown to both agent and principal, as is standard in the career concerns literature. The outcome of the agent's action depends on the state of the world. The agent receives a signal about this state of the world. With unobservable actions, the principal can only observe the outcome, but with observable actions, the principal can observe both the action and the outcome. Hence, the agent will disregard the signal.

In our model, we consider whether the principal wants to observe the project choice of the agent. Thus, in a sense we study whether the principal wants to allow 'communication' between the agent and herself to take place. In this respect, our paper is related to Friebel and Raith (2004). They show that when the principal allows the agent to communicate, he will try to convince her that he is better suited for the supervisory role than the incumbent supervisor. As a result the incumbent may refrain from developing the agent's skills and expertise and also may deliberately recruit agents with low abilities. Consequently, it may be better for the principal not to allow the agent to directly communicate with her.

3 The Benchmark Model

In this section, we introduce our benchmark model in which long-run interests of principal and agent are perfectly aligned, and the cost of effort for the agent is independent of his

choice of project. Our set-up is as follows. An agent has the authority to choose some project to implement on behalf of a principal. A continuum of projects is available. Some projects are more risky than others, in the following sense. Riskier projects have a lower probability of success but, if successful, they also yield a higher return. The probability of success of any project will also depend on the ability of the agent. The higher the ability of the agent, the higher the probability of success of any given project. For that reason, we may also refer to 'risky' projects as 'difficult' projects.

The easiest way to capture these assumptions is as follows. The continuum of available projects is indexed by x, with $x \in [0, \bar{x}]$. Projects are identical in terms of their required effort. If project x is successful, it will yield the principal a payoff x. If it fails, it will yield payoff 0. The probability that it is successful, is $\theta - x$, with θ the ability of the agent. Thus, a higher ability agent has a higher probability of success in implementing any project. Also, for given θ, projects that have a higher payoff when successful, also have a lower probability of success, and hence are riskier. The expected return on project x is denoted as $R(x)$, so we have

$$R(x) = (\theta - x)\, x. \tag{1}$$

For simplicity, we assume that the ability of the agent is either high or low: θ_H or θ_L, respectively. The true ability is private information. A priori, the probability of a high type is ρ. The choice of project x thus serves as a signal of the agent's true ability. We denote the posterior belief of facing a high type as μ. We assume that $\bar{x} < \theta_L < \theta_H < 1$,

which guarantees that probabilities are always strictly between 0 and 1.

The payoff of the agent is based on two components. The first is the expected return of the project. There may be several reasons for this. For instance, the agent may have some intrinsic satisfaction of obtaining a higher expected return. Also, in the long run, the outcome of the project is observable and the agent's career will depend on it. The second component is the short run benefit drawn from the perception that others have of his ability. A better perception of that ability may for example imply better job opportunities, either inside or outside the current relationship. When internal job opportunities are important, the agent will primarily care about the principal's perception. When external job opportunities are important, he will primarily care about the perception of outsiders. For our purposes, this is immaterial. Therefore, in the remainder, we will refer to the party that forms the relevant perceptions as the receiver in this game. As an illustration, consider the venture capital example from the introduction. If the venture capitalist has a higher perception of the entrepreneur's ability, then the entrepreneur has a better chance of obtaining financing in the first place. At the same time, the entrepreneur also has an incentive to choose a project with a high expected return: if he does obtain financing, then such a project will also give him higher expected monetary rewards.

Expected payoffs for the agent of choosing project x can be denoted as:

$$B_A = \alpha (\theta - x) x + f(\mu), \tag{2}$$

where the parameter α measures the relative importance of the expected return of the

project, μ is the receiver's posterior belief regarding the agent's type, and f is an increasing function: the agent is better off the higher the receiver's posterior of him being a high type. Expected payoffs for the principal are simply

$$B_P = (\theta - x)\, x \tag{3}$$

In our venture capital example, this would suggest that the venture capitalist obtains all revenues from the project. This, however, is immaterial. The only thing we need for our analysis is that the venture capitalist's return is proportional to the return on the project.

Our model can be summarized by the time line depicted in Figure 1. In the next section, we solve the model using backward induction.

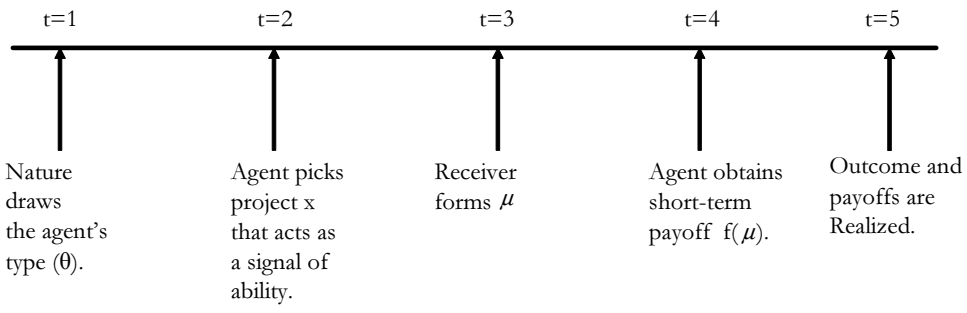

Figure 1: The Time Line

4 Solving the benchmark model

In this section, we solve the model we set out above. In section 4.1 we consider the case in which the project choice is unobservable. Section 4.2 solves the model when the project choice is observable. Section 4.3 compares the two cases.

4.1 Unobservable Project Choice

Suppose that the agent's choice of project is unobservable. In that case, the agent's choice cannot influence the receiver's perception of his ability, hence $\mu = \rho$. The agent then chooses

$$x_i = \arg\max_x \alpha\left(\theta_i - x\right) x + f\left(\rho\right),$$

which is maximized by setting

$$x_i = \theta_i/2,\ i \in \{L, H\}. \tag{4}$$

Note that the project chosen by the agent is also the one that is preferred by the principal, in the sense that it maximizes B_P. In the remainder of the paper, we will refer to the first-best project of type i as x_i^{fa}, with $i \in \{L, H\}$. Thus, in this model, we have $x_i^{fa} = \theta_i/2$. Similarly, we will refer to the first-best of the principal facing a type i agent as x_i^{fp}. Hence, we here have $x_i^{f_p} = \theta_i/2$. In this case, we thus have $x_i^{fa} = x_i^{fp}$. This, however, is no longer the case once we introduce a conflict of interest, as we do in section 5.

4.2 Observable Project Choice

If the principal can observe the agent's choice of project, we have a signalling model. We first define our equilibrium concept:

Definition 1 *A sequential equilibrium of the game described above consists of a strategy x_L^* for the low type, a strategy x_H^* for the high type, and a system of beliefs $\mu(x)$ such that the following conditions hold:*

1. *Optimality for the agent; For $i = L, H$,*

$$x_i^* \in \arg\max_x \alpha\left(\theta_i - x\right) x + f\left(\mu(x)\right) \tag{5}$$

2. *Bayes' consistency of beliefs;*

$$\mu(x) = \begin{cases} 1 & \text{if } \; x = x_H^* \neq x_L^* \\ \rho & \text{if } \; x = x_H^* = x_L^* \\ 0 & \text{if } \; x = x_L^* \neq x_H^* \\ \in [0,1] & \text{otherwise} \end{cases} \tag{6}$$

As an equilibrium refinement, we use Cho and Kreps' (1987) Intuitive Criterion, as is standard in these models. This refinement is based on the following observation. Suppose that the receiver observes some out-of-equilibrium message x. Then it makes no sense to believe that this message comes from a type i if, regardless of the out-of-equilibrium beliefs of the receiver, that type is always strictly worse off playing x rather than playing his equilibrium strategy x_i^*. Formally:

Definition 2 *A sequential equilibrium surviving the Intuitive Criterion is a sequential equilibrium in which, when observing an out-of-equilibrium action x, the receiver attaches zero probability to the event that she faces type i if* $\alpha(\theta_i - x)x + f(1) < \alpha(\theta_i - x_i^*)x_i^* + f(\mu(x_i^*))$.

We introduce some additional notation. The ability difference of the high type and the low type is denoted $\Delta_\theta \equiv \theta_H - \theta_L$. We write f_L for the perception payoffs of an agent that is known to be a low type, and f_H for those of an agent known to be a high type: $f_L \equiv f(0)$, $f_H \equiv f(1)$. The difference between the two is $\Delta_f \equiv f_H - f_L$.

Throughout the analysis, we restrict attention to the case where it is necessary for the high type to distort his project choice in order to convince the principal of his type. That is, we assume that $(x_L^*, x_H^*) = (x_L^{fa}, x_H^{fa})$ is *not* a sequential equilibrium. For simplicity, we will also assume that the parameters of the problem are such that both types will always choose an internal solution, thus $x_L^*, x_H^* \in (0, 1)$. Later, we will show that these two assumptions imply:

Assumption 1 $\frac{1}{2}\sqrt{\alpha}\Delta_\theta \leq \sqrt{\Delta_f} \leq \frac{1}{2}\sqrt{\alpha}\,(2 - \theta_L)$.

We take advantage of the arbitrariness of out-of-equilibrium beliefs to assume that an out-of-equilibrium message is interpreted as being sent by a low type: $\mu(x) = 0$ for $x \neq x_H^*$. We can then show:

Theorem 1 *The unique separating equilibrium surviving the Intuitive Criterion has*

$$(x_L^*, x_H^*) = \left(x_L^{fa}, \frac{\theta_L}{2} + \sqrt{\frac{\Delta_f}{\alpha}} \right) \tag{7}$$

Proof. With the usual arguments, the low type will implement his first-best project in any separating equilibrium, so $x_L^* = x_L^{fa} = \theta_L/2$. Also, we need that x_H^* is such that type L is just not willing to mimic the high type and being perceived as having high ability. Thus

$$\alpha\left(\theta_L/2\right)^2 + f_L \geq \alpha\left(\theta_L - x_H^*\right)x_H^* + f_H,$$

or

$$x_H^* \geq \frac{\theta_L}{2} + \frac{1}{\sqrt{\alpha}}\sqrt{\Delta_f}. \tag{8}$$

If the high type chooses to defect, his best possible defection is x_H^{fa}. Incentive compatibility for the high type thus requires

$$\alpha\left(\theta_H - x_H^*\right)x_H^* + f_H \geq \alpha\left(\theta_H/2\right)^2 + f_L,$$

or

$$x_H^* \leq \frac{\theta_H}{2} + \frac{1}{\sqrt{\alpha}}\sqrt{\Delta_f}.$$

With $\theta_H > \theta_L$, this upper bound on x_H^* is always higher than the lower bound given by (8). Therefore, a separating equilibrium always exists. With the usual arguments, the unique equilibrium surviving Cho and Kreps (1987) Intuitive Criterion has (8) binding for the high ability type, which establishes the result. ■

As noted, to have signalling in equilibrium, we need that $(x_L^*, x_H^*) = (x_L^{fa}, x_H^{fa})$ is not a sequential equilibrium. In other words; we assume that when a type H would set x_H^{fa} in equilibrium, a type L would have an incentive to mimic that strategy. Hence we need $x_H^* > x_H^{fa}$. It is easy to see that this requires $\sqrt{\Delta_f} > \frac{1}{2}\sqrt{\alpha}\Delta\theta$, which is true due to

Assumption 1. Also note that, to assure that $x_H^* < 1$, we need $\sqrt{\Delta_f} < \frac{1}{2}\sqrt{\alpha}\,(2-\theta_L)$, which is also satisfied due to Assumption 1.

4.3 Comparing the two regimes

From the analyses above, we have that in the equilibrium of our benchmark model, the high type chooses a project that is riskier than his first-best choice. This is not in the best interest of the principal. The project choice of the low type is unaffected. Both in the case where his project choice is observable, and in the case where it is not, he chooses to set his first-best project. We therefore have:

Theorem 2 *In the benchmark model, the a priori expected payoffs of the principal are strictly higher when the project choice is unobservable.*

Proof. With unobservability, expected payoffs equal $(1-\rho)\,\theta_L^2/4 + \rho\theta_H^2/4$. With observability, they equal $(1-\rho)\,(\theta_H - x_L^*)\,x_L^* + \rho\,(\theta_H - x_H^*)\,x_H^*$. With $x_L^* = \theta_L/2$ and $x_H^* > \theta_H/2 = \arg\max_x (\theta_H - x)\,x$, we have the result. ■

Hence, in this case, the principal would like to commit not to be able to observe the project choice of the agent. In doing so, she would avoid costly signalling by the high type, that ultimately is also costly for herself. There is a detrimental *signalling effect*, as we depict in Figure 2.

Of course, the principal may also try to avoid signaling by trying to make sure that Assumption (1) is not satisfied. One way to do so is to make Δ_f small. Note that Δ_f can be interpreted as the difference between the short-run benefits obtained by an agent

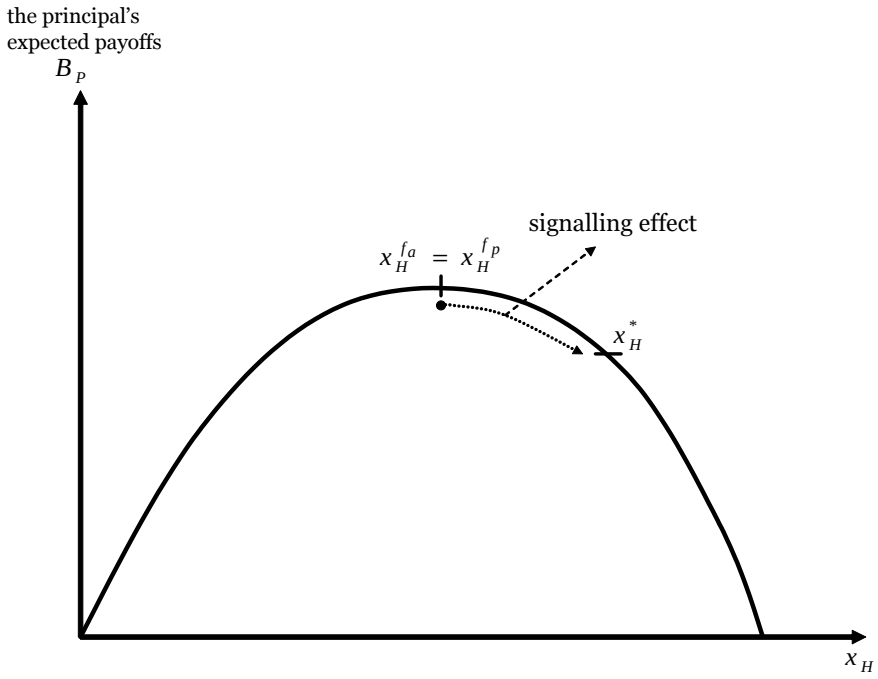

Figure 2: The Benchmark Case

that is known to be of high ability, and the short-run benefits obtained by an agent that is known to be of low ability. Interestingly, such an interpretation implies that distortive signalling can be avoided by having a policy of promoting personnel only on the basis of some objective measure of output, rather than on the basis of managerial discretion, granted of course that this is feasible. With such a policy, the high ability type would not have an incentive to try to impress the principal, and hence would not engage in costly signalling.

5 Introducing a conflict of interest

5.1 The Model

We now introduce a conflict of interest between the principal and the agent. In our benchmark model with unobservable actions, the interests of interests and principal are perfectly aligned, as we argued in the introduction. Suppose now that the agent has to exert some effort in implementing a project and, more importantly, that that effort differs among projects. We will assume that more difficult projects not only require more ability, but also more effort. To capture this, we assume that project x requires effort γx^2. For simplicity, this effort is independent of the quality of the agent. This does not affect our qualitative results.[1] The agent's payoff now is

$$B_A = \alpha\left(\theta_i - x\right)x - \gamma x^2 + f(\mu),\ i \in \{L, H\}, \tag{9}$$

while the principal's payoff function is unaffected.

5.2 Solving the Model

First consider the case where project choice is unobservable. The agent then simply sets x_i to maximize (9) and will therefore choose

$$x_i^{fa} = \frac{\theta_i}{2}\frac{\alpha}{\alpha+\gamma},\ i \in \{L, H\}. \tag{10}$$

Note that $\alpha/(\alpha+\gamma)$ is strictly between 0 and 1. Hence, if x_i^{fa} is well-defined in the benchmark model, then it also is in this model. The principal, however, still wants to

[1] If more difficult projects require less effort of a high type, then the high type needs less of a distortion to his project choice to convince the principal that he really is a low type. Ceteris paribus, this makes it more attractive for the principal to observe his agent's project choice.

maximize (3), and hence prefers to have

$$x_i^{fp} = \frac{\theta_i}{2},\ i \in \{L, H\}. \tag{11}$$

We now have a conflict of interest between principal and agent, even in the case where project choice is unobservable. Similar to the standard principal-agent model, the principal wants the agent to exert more effort than the latter's first-best choice. We will refer to the difference between x_i^{fa} and x_i^{fp} as the *shirking effect.*

Consider the case that project choice is observable. A sequential equilibrium now requires

$$x_i^* \in \arg\max_x \alpha\left(\theta_i - x\right)x - \gamma x^2 + f\left(\mu(x)\right),$$

while (6) is unaffected. This yields the following:

Theorem 3 *The unique separating equilibrium surviving the Intuitive Criterion has*

$$(x_L^*, x_H^*) = \left(x_L^{fa}, \frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f}.\right) \tag{12}$$

Proof. Again, obviously $x_L^* = x_L^{fa}$. Incentive compatibility for the low type requires

$$\alpha\left(\theta_L - x_L^*\right)x_L^* - \gamma\left(x_L^*\right)^2 + f_L \geq \alpha\left(\theta_L - x_H^*\right)x_H^* - \gamma\left(x_H^*\right)^2 + f_H.$$

or

$$x_H^* \geq \frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f}. \tag{13}$$

Incentive compatibility for the high type requires

$$\alpha\left(\theta_H - x_H^*\right)x_H^* - \gamma\left(x_H^*\right)^2 + f_H \geq \alpha\left(\theta_H - x_H^{fa}\right)x_H^{fa} - \gamma\left(x_H^{fa}\right)^2 + f_L,$$

or

$$x_H^* \leq \frac{\theta_H}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f}.$$

With $\theta_H > \theta_L$, this upper bound on x_H^* is always higher than the lower bound given by (13). Therefore, a separating equilibrium always exists. With the usual arguments, the unique equilibrium surviving Cho and Kreps (1987) Intuitive Criterion has (13) binding for the high ability type, which establishes the result. ■

Compared to the benchmark model, the absolute deviation relative to the first-best of the high type is smaller. Choosing a project different from the first-best is costlier than in the benchmark model, as it also implies an additional cost of effort. Hence, the high type does not have to distort as much as before to discourage the low type from mimicking his choice.

As in the benchmark model, we need $x_H^{fa} < x_H^* < 1$. Fortunately, the same parameter restrictions are sufficient to guarantee that in this model:

Lemma 1 *In the model with effort, assumption 1 is sufficient to have* $x_H^{fa} < x_H^* < 1$.

Proof. The inequality $x_H^{fa} < x_H^*$ requires

$$\frac{\theta_H}{2}\frac{\alpha}{\alpha+\gamma} < \frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f}$$

or

$$\sqrt{\Delta_f} > \frac{\alpha\Delta_\theta}{2\sqrt{\alpha+\gamma}},$$

The inequality $x_H^* < 1$ requires

$$\frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f} < 1$$

or

$$\sqrt{\Delta_f} < \frac{\frac{1}{2}\alpha\left(2-\theta_L\right)+\gamma}{\sqrt{\alpha+\gamma}}$$

Combining these inequalities, we need

$$\sqrt{\Delta_f} \in \left(\frac{\alpha\Delta_\theta}{2\sqrt{\alpha+\gamma}}, \frac{\frac{1}{2}\alpha\left(2-\theta_L\right)+\gamma}{\sqrt{\alpha+\gamma}}\right) \tag{14}$$

Note that

$$\frac{\partial}{\partial\gamma}\left(\frac{\alpha\Delta_\theta}{2\sqrt{\alpha+\gamma}}\right) = \frac{-\alpha\Delta_\theta}{4\left(\alpha+\gamma\right)^{\frac{3}{2}}} < 0$$

and

$$\frac{\partial}{\partial\gamma}\left(\frac{\frac{1}{2}\alpha\left(2-\theta_L\right)+\gamma}{\sqrt{\alpha+\gamma}}\right) = \frac{2\alpha+2\gamma+\alpha\theta_L}{4\left(\alpha+\gamma\right)^{3/2}} > 0.$$

Hence, the lower bound of (14) is decreasing in γ, while the upper bound is increasing in γ. This implies that if the condition is satisfied in the benchmark case, where effectively $\gamma = 0$, then it is also satisfied for any $\gamma > 0$. Hence, if assumption 1 is satisfied, then we also have $x_H^{fa} < x_H^* < 1$ in our model with effort. ■

5.3 Comparing the two regimes

As was the case in our benchmark model, a low type chooses the same project in both regimes, so we only have to consider the choice of a high type. With a high type, the first-best of the principle still is $x_H^{fp} = \theta_H/2$. Since her pay-off function is quadratic, the

preference of the principal boils down to whether x_H^* or x_H^{fa} is closer to x_H^{fp}. For ease of exposition, we assume that the principal can choose whether or not to observe the actions of the agent. We will refer to that decision as whether or not to *monitor* the agent's project choice. We thus explicitly assume that the principal is the receiver. In terms of the time line of figure 1, we add a stage at $t = 0$. For simplicity, we assume that monitoring is costless. Provided that true monitoring costs are not too high, this assumption does not affect our qualitative results.

For the purposes of this section, it is convenient to define an upper bound on Δ_θ. Note that from assumption 1, we always have $\Delta_\theta \leq 2\sqrt{\Delta_f/\alpha}$. For any $\gamma \geq 0$, this immediately implies that $\Delta_\theta \leq \left(2\sqrt{\alpha+\gamma}\sqrt{\Delta_f}\right)/\alpha$. We will refer to the latter value as $\Delta_\theta^{\max}$.

We can now establish the following result:

Theorem 4 (Undersignalling) *If $\Delta_\theta > \Delta_\theta^{\max} - \gamma\theta_H/\alpha$, the principal chooses to monitor. If she does, however, the agent will still choose a project that is too safe from the principal's perspective.*

Proof. For the result to hold, we need $x_H^* < x_H^{fp}$, hence

$$\frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f} < \theta_H/2,$$

which, using the definition of $\Delta_\theta^{\max}$, implies the condition stated in the theorem. ■

Essentially the result is driven by a trade-off between two effects. The first one is the *shirking effect*. Given that we have a conflict of interest, there will be a divergence

between the principal's first best project choice x_H^{fp} and the high-type agent's first best project choice x_H^{fa}. The shirking effect represents the extent of this divergence. The second effect is the *signalling effect*. In a separating equilibrium, a high type will have an incentive to separate himself from the low type by choosing a riskier project (a higher x_H^*) than he otherwise would. The incentive to signal pushes the agent's project choice back towards the principal's first best choice. Hence, in this case signalling will be beneficial for the principal. The result is depicted in Figure 3. In this case, the signalling effect is not strong enough to fully compensate for the shirking effect. That is why we refer to this case as one of undersignalling.

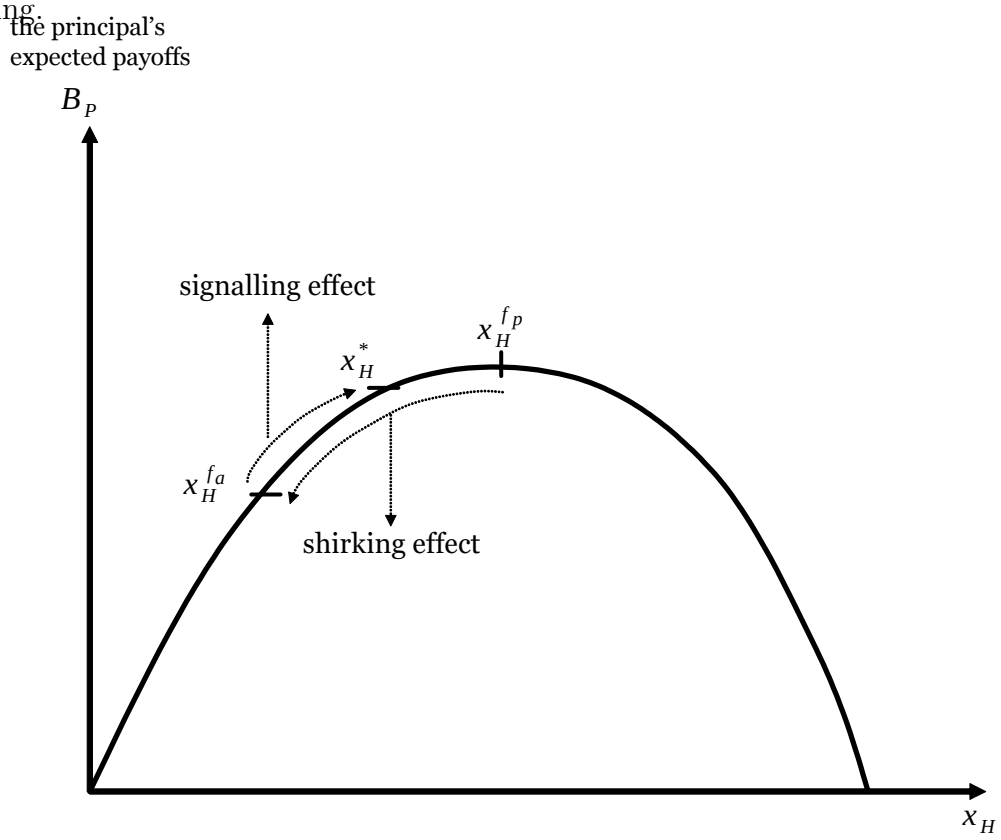

Figure 3: Undersignalling ($\Delta_\theta > \Delta_\theta^{\max} - \gamma\theta_H/\alpha$)

However, there may also be cases in which the signalling effect more than compensates for the shirking effect. But also in that case, observing the project choice may still make the principal better off, as long as the project choice with observability is still closer to the principal's first best than the agent's first-best project is. We refer to this case as one of oversignalling, and depict it in figure 4.

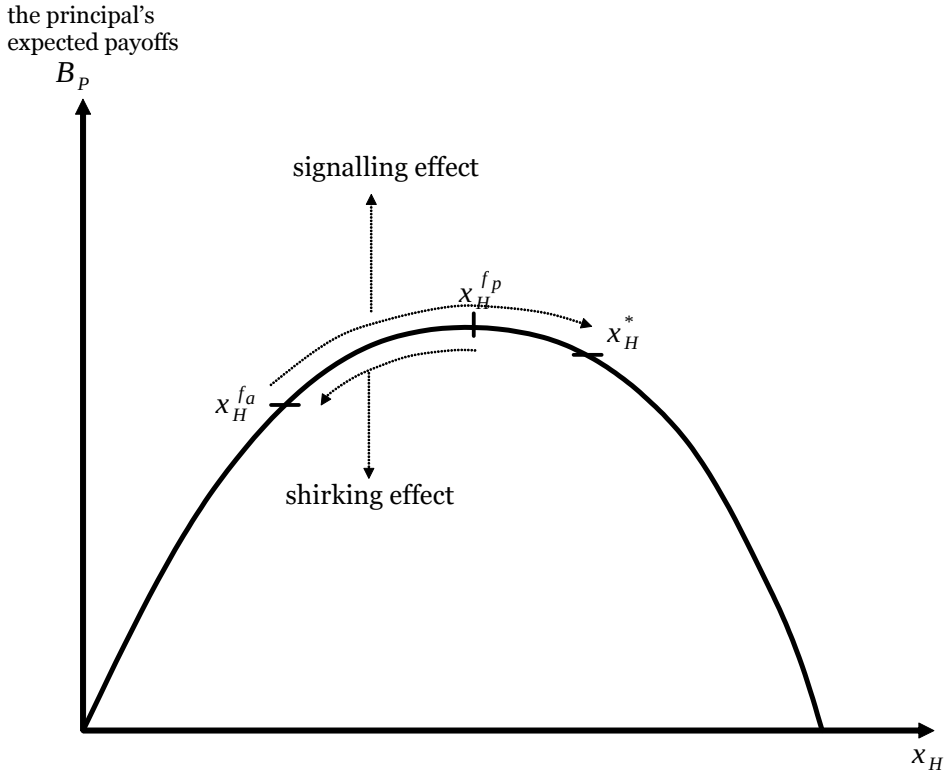

Figure 4: Oversignalling ($\Delta_\theta^{\max} - 2\gamma\theta_H/\alpha < \Delta_\theta < \Delta_\theta^{\max} - \gamma\theta_H/\alpha$)

Theorem 5 (Oversignalling) *If* $\Delta_\theta \in (\Delta_\theta^{\max} - 2\gamma\theta_H/\alpha, \Delta_\theta^{\max} - \gamma\theta_H/\alpha)$*, the principal chooses to monitor. If she does the agent will choose a project that is too risky from the principal's perspective.*

Proof. For the result to hold, two conditions need to be satisfied. First, we need

$x_H^* > x_H^{fp}$: the project choice with observability is riskier than the principal's first-best. Second, we need $x_H^* - x_H^{fp} < x_H^{fp} - x_H^{fa}$: the project choice with observability is closer to the principal's first best than the project choice without observability. From the previous theorem, we have that the first condition is satisfied if $\Delta_\theta < \Delta_\theta^{\max} - \gamma\theta_H/\alpha$. The second condition implies

$$\frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f} - \frac{\theta_H}{2} < \frac{\theta_H}{2} - \frac{\theta_H}{2}\frac{\alpha}{\alpha+\gamma}.$$

or $\Delta_\theta > \Delta_\theta^{\max} - 2\gamma\theta_H/\alpha$. Combining the two establishes the result. ■

Finally, there may be cases in which the signalling effect is so strong that the net effect is to make the principal worse off if he monitors. We depict this case in figure 5. Note that we also had this case in the previous section, where the shirking effect was zero.

Theorem 6 (Excessive signalling) *If $\Delta_\theta < \Delta_\theta^{\max} - 2\gamma\theta_H/\alpha$, the principal chooses not to monitor. The agent chooses a project that is too safe from the principal's perspective.*

Proof. For the result to hold, two conditions need to be satisfied. First, we need $x_H^* > x_H^{fp}$: the project choice with observability is riskier than the principal's first-best. Second, we need $x_H^* - x_H^{fp} < x_H^{fp} - x_H^{fa}$: the project choice with observability is closer to the principal's first best than the project choice without observability. From the previous theorem, we have that the first condition is satisfied if $\Delta_\theta < \Delta_\theta^{\max} - \gamma\theta_H/\alpha$. The second condition implies

$$\frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{\Delta_f} - \frac{\theta_H}{2} < \frac{\theta_H}{2} - \frac{\theta_H}{2}\frac{\alpha}{\alpha+\gamma}.$$

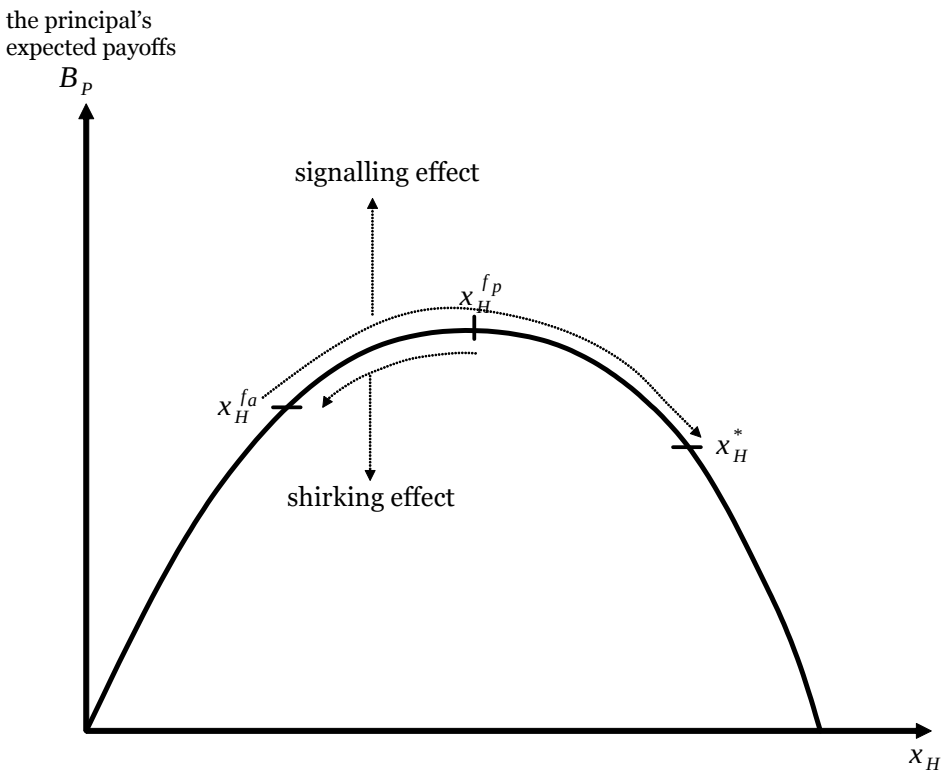

Figure 5: Excessive signalling ($\Delta_\theta < \Delta_\theta^{\max} - 2\gamma\theta_H/\alpha$)

or $\Delta_\theta > \Delta_\theta^{\max} - 2\gamma\theta_H/\alpha$. Combining the two establishes the result. ■

Hence the principal chooses not to monitor the agent's project choice. With observability, she would have excessive signalling. Just like we had in the previous case, the agent's project choice is too risky from her point of view. However, in this case, the signalling effect is so excessive that it is better for the principal to avoid it altogether and settle for the agent's first-best choice.

The following figure illustrates the possible outcomes under observability. For given values of the other parameters, we have that for low Δ_θ, there is excessive signalling, for intermediate Δ_θ, there is oversignalling, and for high Δ_θ, there is undersignalling. The

intuition for this result is as follows. The shirking effect is independent of the ability difference between the two types. Yet, the closer the two types are in terms of ability, the more the high type has to signal in order to truly differentiate himself from the low type, and the more likely that the principal is worse off.

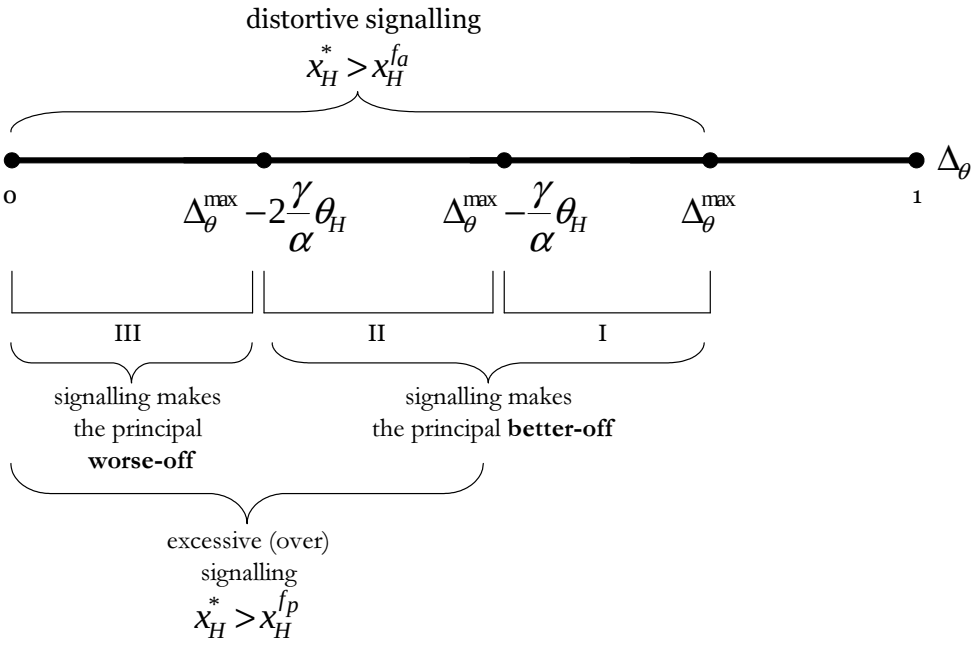

Figure 6: Summary of Results

Of course, it would also be interesting to see how a change in the extent of the conflict of interest would affect the result. We now address that issue.

Theorem 7 *For given $\Delta_\theta \leq 2\sqrt{\Delta_f/\alpha}$, there exist threshold values $\bar{\gamma}_1(\Delta_\theta) < \bar{\gamma}_2(\Delta_\theta) < \bar{\gamma}_3(\Delta_\theta)$ such that the following holds:*

1. *With $\gamma < \bar{\gamma}_1(\Delta_\theta)$, there is excessive signalling,*

2. *with* $\bar{\gamma}_1(\Delta_\theta) < \gamma < \bar{\gamma}_2(\Delta_\theta)$*, there is oversignalling,*

3. *with* $\gamma > \bar{\gamma}_2(\Delta_\theta)$*, there is undersignalling.*

Proof. Consider the cut-off point between excessive signalling and oversignalling, given by $\Delta_\theta^{\max} - 2\gamma\theta_H/\alpha = \left(2\sqrt{\alpha+\gamma}\sqrt{\Delta_f}\right)/\alpha - 2\gamma\theta_H/\alpha$. Denote this cut-off as $C_{eo}(\gamma)$.With $\gamma = 0$, the expression becomes $2\sqrt{\Delta_f/\alpha}$, so we have excessive signalling for all admissible values of Δ_θ. Note that

$$\frac{\partial^2 C_{eo}}{\partial \gamma^2} = -\frac{1}{2}\frac{\sqrt{\Delta_f}}{\alpha\left(\alpha+\gamma\right)^{\frac{3}{2}}} < 0.$$

This implies that C_{eo} is a strictly concave function. Also note

$$\lim_{\gamma\to\infty}\frac{\partial C_{eo}}{\partial \gamma} = -\infty.$$

Consider some $\Delta'_\theta < C_{eo}(0) = 2\sqrt{\Delta_f/\alpha}$. The two properties derived above imply that there is some γ' such that $C_{eo}(\gamma') = \Delta'_\theta$. Concavity implies that for all $\gamma'' \in (0,\gamma')$, we have

$$C_{eo}(\gamma'') > \frac{\gamma''}{\gamma'}C_{eo}(0) + \left(\frac{\gamma'-\gamma''}{\gamma'}\right)C_{eo}(\gamma') > C_{eo}(\gamma').$$

Next, concavity of C_{eo} and the fact that $C_{eo}(\gamma') < C_{eo}(0)$ together imply that C_{eo} is decreasing in γ'. Concavity then implies that for all $\gamma''' \in (\gamma',\infty)$, we have that $C_{eo}(\gamma''') < C_{eo}(\gamma')$. Hence, γ' is the unique threshold such that the principal faces excessive signaling with $\gamma < \gamma'$, and faces either oversignaling or undersignaling with $\gamma > \gamma'$.

Next, consider the cut-off point between over- and undersignalling, given by $\Delta_\theta^{\max} - \gamma\theta_H/\alpha = \left(2\sqrt{\alpha+\gamma}\sqrt{\Delta_f}\right)/\alpha - \gamma\theta_H/\alpha$. Denote this cut-off as $C_{ou}(\gamma)$. With the exact same

arguments as above, we can show that, for given Δ'_θ, there is a unique threshold $\hat{\gamma}$ such that the principal faces oversignaling or excessive signaling with $\gamma < \hat{\gamma}$, and faces undersignaling with $\gamma > \hat{\gamma}$. Since we have $C_{eo}(\gamma) < C_{ou}(\gamma)$ for all $\gamma > 0$, the result is established. ■

Thus for low enough γ, and hence if the conflict of interest is not too strong, we always have excessive signaling, just as we had in our benchmark model. For intermediate values of γ, we have a case of oversignalling. If the conflict of interest is strong enough, we have undersignalling. This also implies that for low enough γ, the principal chooses not to monitor. Yet, once γ is high enough, and the conflict of interest is sufficiently high, the principal does want to monitor.

The intuition for this result is as follows. First, the signalling effect becomes weaker as γ increases. When the conflict of interest increases, it is less difficult for the high ability type to differentiate from the low ability type: when mimicking the high type, the low type now has to incur not only the lower payoff associated with a more risky project, but also the higher effort. Second, an increase in γ implies that the shirking effect is stronger. For the same project, the agent now has to exert more effort, regardless of his type. Both effects lead to the agent choosing safer projects. Hence, excessive signalling becomes less of an issue.

6 Extension: probabilistic monitoring

So far, we have assumed that the principal faces the binary choice whether to monitor or not to monitor. Yet, there may be circumstances in which she can commit to monitor with

some probability. In this section we analyze this scenario.

Suppose that *a priori,* the principal can commit to monitor with some probability m. The principal will choose m to maximize expected payoffs. The agent's payoff now is

$$B_A = \alpha\left(\theta_i - x\right)x - \gamma x^2 + \left(1-m\right)f(\rho) + mf(\mu),\ i \in \{L, H\}. \tag{15}$$

This can be seen as follows. With probability $1-m$, there is no monitoring, and the principal's posterior belief regarding the agent's quality equals the prior belief ρ. With probability m, there is monitoring, and the principal can update her belief. Thus, with $m=1$ we are in the observability case, and with $m=0$, we are in the unobservability case. The first-best outcomes of agent and principal are still given by (10) and (11) respectively. Consider the subgame that starts after the principal has chosen m. A sequential equilibrium of that subgame requires

$$x_i^* \in \arg\max_x \alpha\left(\theta_i - x\right)x - \gamma x^2 + mf(\mu(x)),$$

while (6) is unaffected. This yields the following:

Lemma 2 *The unique separating equilibrium surviving the Intuitive Criterion has* $x_L^* = x_L^{fa}$ *and*

$$x_H^* = \begin{cases} x_H^{fa} & \text{if } m < \frac{1}{4\Delta_f}\frac{\alpha^2\Delta_\theta^2}{\alpha+\gamma} \\ \frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{m\Delta_f} & \text{otherwise.} \end{cases}. \tag{16}$$

Proof. Obviously $x_L^* = x_L^{fa}$. Again, the natural candidate for x_H^* in a separating equilibrium is the $\hat{x}$ for which the incentive compatibility constraint for the low type is just

binding. The constraint is

$$\alpha\left(\theta_L - x_L^*\right)x_L^* - \gamma\left(x_L^*\right)^2 + mf_L \geq \alpha\left(\theta_L - \hat{x}\right)x - \gamma\left(\hat{x}\right)^2 + mf_H,$$

hence

$$\hat{x} = \frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{m\Delta_f}. \tag{17}$$

For this to be the separating equilibrium we need that $\hat{x} > x_H^{fa}$; otherwise the high type would simply set x_H^{fa}, without the low type having an incentive to mimic that strategy. This implies that to have $x_H^* = \hat{x}$, we require $m \geq \frac{1}{4\Delta_f}\frac{\alpha^2\Delta_\theta^2}{\alpha+\gamma}$. This establishes the result. ■

We now have:

Theorem 8 *The optimal choice for the principal is to set*

$$m^* = \begin{cases} \frac{(\alpha\Delta_\theta+\gamma\theta_H)^2}{4(\alpha+\gamma)\Delta_f} & \text{if } \Delta_\theta > \Delta_\theta^{\max} - \gamma\theta_H/\alpha, \\ 1 & \text{otherwise.} \end{cases}$$

Proof. The best the principal can do is to achieve his first-best. From (16) and (11), we have that $x_H^* = x_H^{fp}$ if

$$\frac{\theta_L}{2}\frac{\alpha}{\alpha+\gamma} + \frac{1}{\sqrt{\alpha+\gamma}}\sqrt{m\Delta_f} = \frac{\theta_H}{2}$$

or

$$m^* = \frac{\left(\alpha\Delta_\theta + \gamma\theta_H\right)^2}{4\left(\alpha+\gamma\right)\Delta_f}.$$

For this to be the optimal m, we require, first, that $m^* \in [0,1]$ and, second, that m^* is such that the high type indeed chooses a distortive signal ($x_H^* > x_H^{fa}$). For the first condition, note that $m^* > 0$. It is easy to verify that $m^* = 1$ if $\Delta_\theta = \Delta_\theta^{\max} - \gamma\theta_H/\alpha$. With m^* strictly

increasing in Δ_θ, this implies that the first condition is satisfied if $\Delta_\theta < \Delta_\theta^{\max} - \gamma\theta_H/\alpha$. For the second condition, we require, using lemma 2,

$$\frac{(\alpha\Delta_\theta + \gamma\theta_H)^2}{4(\alpha+\gamma)\Delta_f} > \frac{1}{4\Delta_f}\frac{\alpha^2\Delta_\theta^2}{\alpha+\gamma},$$

which immediately simplifies to $(\alpha\Delta_\theta + \gamma\theta_H)^2 > \alpha^2\Delta_\theta^2$, and is satisfied for any $\gamma > 0$. This establishes the result. ■

The intuition for this result can be seen from the figures that we used earlier. If there is oversignalling (so we are either in the case depicted in Figure 4, or in that depicted in Figure 5), the principal can achieve her first-best by lowering the probability of monitoring. Yet, in the case of undersignalling (so we are in the case depicted in Figure 3), it is not possible to reach the first-best, as the probability of monitoring cannot chosen to be higher than 1.

We also have:

Theorem 9 *Suppose $m^* < 1$. We then have $\partial m^*/\partial\gamma > 0$ and $\partial m^*/\partial\Delta_\theta > 0$.*

Proof. Follows directly from Theorem 8. ■

Thus, as the conflict of interest increases, the principal wants to monitor more intensively; shirking now becomes more of a problem, while signalling is less of an issue. Also, the principal wants to monitor more intensively as the difference in ability between the two types of agent increases. If that difference is high, there will be less signalling, so observing the agent's choice will be less of a problem. As a result, the principal wants to monitor more intensively.

7 Conclusion

In this paper, we have shown that an agent's career concerns may hurt the principal. If the agent's project choice is observable, then he has an incentive to try to impress the job market by choosing a project that is more difficult and more risky than he can actually handle. This hurts his principal, who would prefer the agent to pick a project that is more suitable for his capabilities. We have also shown that if more difficult projects require more effort, there is a trade-off between the signaling effect explained above, and a shirking effect. The principal will be more eager to observe the agent's project choice if shirking is more of an issue. If quality differences between 'good' and 'bad' agents are larger, the principal will also be more eager to observe the agent's project choice. In that case, a 'good' agent does not have that much of a need to distort his project choice, so the downside to the principal of observing the project choice of the agent, is also smaller.

One implication of our results is that it may be better to base remuneration and promotion on strict output criteria rather than on the discretion of the principal, provided that this is feasible. With strict rules, the principal's impression of her agent does not play a role in the remuneration and promotion decision. Thus, effectively, a signalling attempt can be prevented. Of course, the principal only wants to do this in cases where she prefers not to observe the agent's project choice.

Interestingly, our paper may also shed new light on the overconfidence or self-serving bias (see e.g. Camerer, 1997 or Babcock and Loewenstein, 1997). A wealth of psychological

evidence suggests that a vast majority of subjects overrate their own skills. In our model the same phenomenon occurs: agents choose projects that are more difficult than what they can actually handle. Yet, in our model, this is not due to a systematic overestimate of one's own abilities, but rather to a deliberate strategic attempt to distort the impression of others. In a related paper (Haan et al., 2006), we show that the behavior of contestants in the TV game show *The Weakest Link* follows the same pattern. Contestants who have to bet money on whether they will be able to answer some general knowledge question, make riskier choices if the perception of their skill by others is more important.

References

Babcock, L., and G. Loewenstein (1997): "Explaining Bargaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives*, 11, 109–126.

Camerer, C. F. (1997): "Progress in Behavioral Game Theory," *Journal of Economic Perspectives*, 11(4), 167–188.

Cho, I.-K., and D. M. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2), 179–221.

Dewatripont, M., I. Jewitt, and J. Tirole (1975): "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, 66(1), 183–198.

Friebel, G., and M. Raith (2004): "Abuse of Authority and Hierarchical Communication," *RAND Journal of Economics*, 35, 224–244.

Gibbons, R., and K. J. Murphy (1992): "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 100(3), 468–505.

Haan, M., B. Los, and Y. Riyanto (2006): "Signaling strength? An analysis of economic decision making in The Weakest Link," mimeo, University of Groningen.

Hirshleifer, D., and A. V. Thakor (1992): "Managerial Conservatism, Project Choice and Debt," *Review of Financial Studies*, 5, 437–470.

Holmstrom, B. (1982): "Managerial Incentives – A Dynamic Perspective," in *Essays in Economics and Management in Honors of Lars Wahlbeck*. Swedish School of Economics, Helsinki, Finland.

——— (1999): "Managerial Incentive Problems - A Dynamic Perspective," *Review of Economic Studies*, 66, 169–182.

Holmstrom, B., and J. Ricart i Costa (1986): "Managerial Incentives and Capital Management," *Quarterly Journal of Economics*, 101, 835–860.

Kanodia, C., R. Bushman, and J. Dickhaut (1989): "Escalation Errors and the Sunk Cost Effect: An Explanation Based on Reputation and Information Asymmetries," *Journal of Accounting Research*, 27, 59–77.

Landers, R. M., J. B. Rebitzer, and L. J. Taylor (1996): "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms," *American Economic Review*, 86(3), 329–348.

Prat, A. (2005): "The Wrong Kind of Transparency," *American Economic Review*, 95(3), 862–877.

Zwiebel, J. (1995): "Corporate Conservatism and Relative Compensation," *Journal of Political Economy*, 103, 1–25.