UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 524

Motivated Agents with Career Concerns:
Signalling Skills and Organizational Involvement.

Oleg Shchetinin

January 2012

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se   info@handels.gu.se

EFMD
EQUIS
ACCREDITED

Motivated Agents with Career Concerns: Signalling Skills and Organizational Involvement.

Oleg Shchetinin[*]

January 23, 2012

Abstract

The paper studies the interaction between two kinds of incentives: career concerns and intrinsic motivation emerging from agent's alignment with organization's objectives or another source of organizational involvement. The information on both skills and involvement can be asymmetric and is updated over time, as in standard career concerns model. It is shown that career concerns is weakened for the involved agent. The agent with low involvement can be more aggressive in career and reputation building at the earlier stages of career, but will be outperformed by the more involved agent in the long-run. The results of the analysis are applied to a number of contexts.

1 Introduction

In many organizations, agents are motivated by alignment with organizational objectives or attaching certain value to their output. Examples are abundant and include bureaucracy, public services provision (health care, education), other organizations with socially valuable output or those having a socially sound mission (non-profits, development institutes and other "mission-driven" organizations).[1] At the same time, in most of these organizations, explicit in-

---

[*]The University of Gothenburg, School of Business, Economics and Law, Department of Economics. Mailing address: Department of Economics, School of Business, Economics and Law at the University of Gothenburg. Vasagatan 1, PO Box 640, SE 40530 Gothenburg, Sweden. Phone: +46 31 7861344. E-mail: oleg.shchetinin@economics.gu.se

[1]Heckman et al. (1996) documented that some bureaucrats do care about the results of their effort. Wilson (1989) writes "bureaucrats have preferences... among them is the desire to do the job". A comprehensive summary of evidence of altruistic behavior in the non-profit sector is provided by Rose-Ackerman (1996). Delfgaauw (2007) summarizes the evidence of non-pecuniary motivation for doctors.

centives are difficult to use,[2] and agents are motivated by implicit incentives, for instance, career concerns (see Dewatripont et al. (1999)). Hence, in such organizations incentives emerge from the interplay between career concerns and intrinsic motivation driven by alignment with organizational objectives. Therefore, arguably, developing a model incorporating both kinds of incentives is highly relevant.

This article extends the career concerns model by Holmström (1999). The new ingredient is agent's organizational involvement taking the form of sharing the organization's (principal's) objective. The agent is, therefore, intrinsically motivated. In the literature, such kind of agents are often referred to as *involved*, *committed*, *reliable* or *mission-driven* agents. I examine incentive structure and agent's performance. The results are applied in different contexts varying from mission-driven organizations to an interpersonal relationship.

First, I characterize the interaction between career concerns and organizational involvement. It is shown that career concerns is weakened for the involved agents. Therefore, the fact that the agent is more aligned with organization has a dual effect on performance. On the one hand, intrinsic motivation improves performance. On the other hand, career concerns is weakened, and, as a result, performance is lowered. I show that one or another effect can dominate, and the dominance is magnified by agent's involvement. As a result, performance reversal can emerge, i.e. the effort of the more involved agent can be lower than the effort of the less involved one.

Second, I examine the impact of information asymmetry. In addition to unknown skills, as in the standard model, the intensity of agent's involvement is not known to the principal; therefore, the agent has an incentive to signal a high value of it. Put differently, agent's reputation is two-dimensional, where one dimension, as standard, is productivity and another dimension is organizational involvement. I focus on the incentive associated to the second dimension of reputation. I show that this incentive can be stronger for the less involved agent; moreover, it can be positive or negative. This, for instance, has implications for the organization design. When information asymmetry creates positive incentive, it is optimal to have a heterogeneous pool of agents, avoiding agents screening. On the other hand, when the incentive is negative, it is optimal to select highly motivated agents, so that they do not need to signal strong involvement by reducing effort.

The article contributes to the theoretical literature on contracting, which incorporates the non-pecuniary motivation of the agent, the so called behavioral incentive theory. Besley and Ghatak (2005) study mission-oriented organizations. Francois (2000) and Delfgaauw and Dur (2008) study implications of workers' intrinsic motivation emerging from involvement or adherence to organization's mission in the public sector. Francois and Vlassopoulos (2008) provide a survey of the literature on motivated agents. Bénabou and Tirole (2006) develop a general framework for modelling pro-social behaviour and ap-

---

[2]More generally, providing explicit incentives in the multitasking setting is complicated and can lead to serious distortions in effort provision - see, e.g., Holmström and Milgrom (1991).

ply it to a number of contexts. The above mentioned studies are based on the assumption that some agents are intrinsically motivated. At the same time, agents' heterogeneity in the non-pecuniary component of preferences is an essential constituent. I follow this approach by assuming that there are agents with different intensity of organizational involvement. The above-mentioned studies, however, are focused on a static setting, whereas my setting is dynamic.

The article is also related to the literature on career concerns. According to Dewatripont et al. (1999), the impact of career concerns is significant in both private and public sector.

The novelty of this article is the introduction of a new dimension of heterogeneity among agents, for instance, the degree of organizational involvement. Some recent papers have extended the career concerns model in a similar fashion. Kőszegi and Li (2008) consider agents differing in responsiveness to explicit incentives, i.e., drive. In their model, the agent has a motive to signal high drive, but it may affect the principal's inference on the agent's ability in the undesired way (shift it downwards). Therefore, the agent has to trade off two motives – to signal high drive and high ability. This has implications for agent's performance and organizational design. Shchetinin (2010) uses the setting close to this paper and studies the interaction between altruistic agent and a principal focusing on matching. Penczynski (2007) assumes that agents are heterogeneous in the degree of career concerns. He found, among other things, that effort can be non-monotone in the degree of uncertainty about the strength of agent's career concerns. An interesting extension is developed by Casas-Arce (2010) who introduces heterogeneous principals.

This paper provides new insights on the interaction between intrinsic motivation and implicit incentives in organizations. While it is typically stressed that intrinsic motivation can be crowded out by the extrinsic incentives[3], I show that the crowding effect can work in the opposite direction – an extrinsic incentive (career concerns) can be weakened in the presence of intrinsic motivation.

The paper proceeds as follows. The model setting is presented in section 2. I analyze the benchmark 2-period case when there is no information asymmetry on agent's involvement in Section 3. Then the main case encompassing information asymmetry on agent's involvement is studied in section 4. The extension to more than two periods, together with other extensions are studied in section 5. Section 6 discusses applications of the model to different contexts. Section 7 concludes.

2 The model setup

Consider the career concerns model following Holmström (1999). I start with two-period version, and will discuss the extension for more than two periods in subsection 5.2.

---

[3]See, e.g., theoretical studies by Bénabou and Tirole (2006), Ellingsen and Johannesson (2008), Shchetinin (2009) and experimental evidence in Falk and Kosfeld (2006), Fehr and Rockenbach (2003).

There is an agent who can in every period be hired by a principal on a competitive market. The agent is characterized by his type (productive skill or talent) θ, as in the standard model.

Additionally, the agent has intrinsic valuation for organization's (principal's) objective, characterized by its intensity β ($0 \leq \beta \leq 1$). In other words, the agent is involved into the organization. This involvement can emerge, for instance, from alignment with organization's objectives (mission), internalization of organization's profit (e.g., in a family enterprize), feeling responsibility for the task or amenability for the output, pure or impure altruism towards the principal, etc.[4]

We start with the case when β is commonly known in section 3 and then proceed to the case when β is agent's private information in section 4.

In every period t the agent exerts effort (or chooses some action) a_t. Output is determined by

$$y_t = \theta + a_t + \varepsilon_t$$

As standard, assume that θ is not known neither by the agent nor by the principals, but at the beginning of period 1 there is common prior belief that θ is normally distributed with mean $\overline{\theta}_1$ and variance σ_1^2: $\theta \sim \mathcal{N}\left(\overline{\theta}_1, \sigma_1^2\right)$.

The output noises ε_t are independent of each other and from θ. Each of them is normally distributed with mean 0 and variance ω^2: $\varepsilon_t \sim \mathcal{N}\left(0, \omega^2\right)$.

Effort is costly, the cost function $C(a)$ assumed to be increasing, convex, and with no cost at zero effort: $C'(a) \geq 0$, $C''(a) \geq 0$, $C(0) = 0$, $C'(0) = 0$.

At the beginning of each period, the principals compete on the labor market by offering salary. After this, the agent chooses one principal whose offer he accepts and exerts effort. The output is then observed by everyone. On the basis of this observation, the principals ("the market") update beliefs on the distribution of agent's type θ or (θ, β) and the next period starts.

The pecuniary components of utilities of the agent and the principal are respectively $u_t = w_t - C(a_t)$ and $v_t = y_t - w_t$.

The total utility of the agent, which encompasses the pro-social component, is given by
$$U_t = u_t + \beta v_t = w_t - C\left(a_t\right) + \beta\left(y_t - w_t\right)$$

The two-period utility is

$$U = U_1 + \delta U_2$$

where $\delta > 0$ is the relative value of the second period utility with respect to the first period. It is possible to have $\delta \geq 1$ as well as $\delta < 1$. The latter is the case, for example, in the model with time discounting when periods 1 and 2 are equally long; the former is the case if, for example, the first period (trial, untenured job) is shorter than the second (permanent, tenured job). Furthermore, a multi-period (with 3 or more periods) interaction can be considered as

---

[4]One can introduce the same type of involvement in the organization's (principal's) preferences. Such extension is considered in subsection 5.1.

a sequence of two-period interactions with the first period corresponding to the first period in the multi-period interaction, and the second period corresponding to "the rest of life" in the multi-period interaction. In this case, the value $\delta \geq 1$ is appropriate. This interpretation is supported by the analysis of the multi-period interaction (see subsection 5.2)

The actors maximize their respective two-period total utilities.

3 The Benchmark Case: No Information Asymmetry about Agent's Involvement.

I proceed backwards in the analysis of the game.

3.1 Period 2

At the beginning of the second period, the principal should update beliefs on the agent's skill, taking into account a signal on it, i.e., the observed period-1 output y_1. Let the Principal have belief a_1^μ on the period-1 effort (in equilibrium it will be equal to the actually exerted effort a_1). Let $\overline{\theta}_2^P$ and $\overline{\theta}_2^A$ denote principal's and agent's beliefs on the mean of agent's skill distribution at the beginning of period 2. The following lemma describes the beliefs update. This lemma is standard and relies on the properties of the (multidimensional) normal distribution.

Lemma 1. *Suppose that the principals' beliefs on the period-1 effort is a_1^μ, and the actual effort is a_1. Then the principals' and agent's updated beliefs are respectively $\theta \sim \mathcal{N}\left(\overline{\theta}_2^P, \sigma_2^2\right)$ and $\theta \sim \mathcal{N}\left(\overline{\theta}_2^A, \sigma_2^2\right)$, where*

$$\overline{\theta}_2^P = \lambda\left(y_1 - a_1^\mu\right) + (1-\lambda)\overline{\theta}, \quad \overline{\theta}_2^A = \lambda(y_1 - a_1) + (1-\lambda)\overline{\theta}$$

$$\lambda = \frac{\sigma_1^2}{\sigma_1^2 + \omega^2}, \quad \frac{1}{\sigma_2^2} = \frac{1}{\omega^2} + \frac{1}{\sigma_1^2}$$

For the proof of the Lemma see, e.g., Holmström (1999).

Notice that the agent's belief update doesn't depend on the actually exerted effort. Indeed, since any increase in effort results in the equal increase in output, the difference $y_1 - a_1$ doesn't change. On the other hand, principals' beliefs can be manipulated by a deviation in effort.

Consider now effort choice and wage setting in the second period.

The agent's expected utility is $E_2^A\left[w_2 - C\left(a_2\right) + \beta(y_2 - w_2)\right] =$
$E_2^A\left[w_2 - C\left(a_2\right) + \beta\left(\theta + a_2 + \varepsilon_2 - w_2\right)\right] = \beta a_2 - C\left(a_2\right) + (1-\beta)w_2 + \beta\overline{\theta}_2^A.$

The agent's program is then

$$\max_{a_2}\left\{\beta a_2 - C(a_2) + (1-\beta)w_2 + \beta\overline{\theta}_2^A\right\}$$

Its solution is given by

$$C'\left(a_2\right) = \beta \tag{1}$$

We have, therefore, established the following result.

Lemma 2. *Period-2 effort is given by (1). Effort increases with agent's involvement, and doesn't depend on the belief on θ.*

Expected utility obtained in period 2, viewed at the beginning of period 2, is

$$U_2 = w_2 - C(a_2) + \beta\left(\overline{\theta}_2^A + a_2 - w_2\right) \tag{2}$$

Principals' competition should make salary offered to the agent equal to his expected output:

$$w_2 = \overline{\theta}_2^P + a_2 \tag{3}$$

3.2 Period 1

In period 1, the agent takes into account not only the period-1 outcome, but also the impact of his effort on his period-2 utility, since salary w_2 depends on principals' beliefs on agent's talent $\overline{\theta}_2^P$, which is, in turn, affected by the period-1 output. This dynamic impact creates career concerns incentive.

Agent's expected utility is

$$U = U_1 + \delta U_2 = w_1 - C(a_1) + \beta(y_1 - w_1) + \delta\left[(1-\beta)E^A w_2 - C(a_2) + \beta\overline{\theta_2^A} + \beta a_2\right]$$

This leads to the following agent's maximization program in period 1:

$$\max_{a_1}\{\beta a_1 - C(a_1) + \delta(1-\beta)\lambda a_1 + \{const\ on\ a_1\}\}$$

Agent's effort in the first period is then determined by

$$C'(a_1) = \underbrace{(1-\beta)\delta\lambda}_{weakened\ career\ concerns} + \underbrace{\beta}_{intrinsic\ motivation} \tag{4}$$

I will show that for the case of more than 2 periods, agent's effort is determined by a very similar formula (see lemma 8 below).

The following proposition summarizes the main results of the analysis.

Proposition 1. *The period-1 effort a_1 is determined by (4) and is an increasing function of agent's organizational involvement β if and only if $\delta\lambda < 1$. Otherwise, a_1 is a decreasing function of β.*

The proof of proposition 1 is given in the appendix.

The comparison of (4) with the "standard" case when the agent is selfish, in which $C'(a_1) = \delta\lambda$, demonstrates the impact of agent's involvement on effort.

The impact is two-fold. First, organizational involvement weakens career concerns. This is reflected by the multiplier $(1 - \beta)$ to the career concerns intensity $\delta\lambda$. In other words, agent's involvement weakens implicit incentive. Second, involvement gives rise to intrinsic motivation, which increases effort. This is reflected by the term $(+\beta)$ in (4).

Importantly, the two effects of agent's organizational involvement influence effort in the opposite directions. Indeed, the weakening of career concerns lowers effort, whereas intrinsic motivation increases it.

The overall effect is unclear and depends on the strength of career concerns, as stated in proposition 1.

If career concerns is relatively strong ($\delta\lambda > 1$), the effect of weakening of career concerns dominates. Stronger involvement β magnifies this dominance, and, as a consequence, leads to lower effort.

For the case of relatively weak career concerns, the positive impact of intrinsic motivation dominates the weakening of career concerns, and, as before, involvement magnifies the dominance. As a result, stronger involvement leads to greater effort.

It is, therefore, shown that agent's performance is an increasing function of involvement only for relatively weak career concerns, whereas the direction of influence of involvement on effort switches when career concerns is strong. In the latter case, the more involved, and, as a consequence, intrinsically motivated agent, exerts effort lower, compared to the agent with weak involvement. I will refer to this switching of monotonicity as *"performance reversal"*. This is a surprising result and could seem to be only a potential and unrealistic possibility, but I will argue that lower effort of more pro-social agents is observed in many circumstances.

Finally, let us develop the intuition of the mechanism of career concerns weakening. The career concerns incentive is based on the desire of the agent to engage in signal jamming: exerting higher effort would in the future boost the beliefs on agent's skill and, as a result, increase agent's wealth. In equilibrium, however, such potential is taken into account, and, consequently, higher effort is expected. However, the wealth of the agent could increase only at the account of the principal, who would pay a higher wage. The desire of engaging in signal jamming is, therefore, weaker for the more pro-social agent, because he internalizes the potential losses of the principal.

4 The Impact of Information Asymmetry about Agent's Involvement

In this section I consider the setting when both agent's skill and organizational involvement are not known and output conveys a signal on both of them. To motivate the modification of the setting, let us observe that involvement, like skill, influences agent's performance, and because of this the agent is concerned about his reputation on involvement. Finally, the variant of the model with unknown involvement can be more realistic.

4.1 Equilibrium

Let agent's organizational involvement β be his private information. For simplicity, assume that β can take only two values $\beta_H > \beta_L \geq 0$, and let p be the

prior probability that $\beta = \beta_H$:

$$p = Prob(\beta = \beta_H)$$

Period-t output, as before, is determined by $y_t = \theta + a_t + \varepsilon_t$, with the standard assumptions on θ and ε_t: there is common prior beliefs that θ is normally distributed with mean $\overline{\theta}_1$ and variance σ_1^2, each of ε_t is normally distributed with mean 0 and variance ω^2, random variables $\theta, \beta, \varepsilon_1, \varepsilon_2$ are independent.

In this setting, the Agent not only seeks to convince the market in high productivity θ, but also builds reputation on involvement, i.e. seeks to convince the market in high β. Both high skill and good reliability (i.e. high β), as perceived by the market, increase agent's wealth in the future (at $t = 2$ in the 2-period setting), because both lead to higher expected output. Therefore, it is in the agent's interest to produce a signal implying both high θ and high β.

The inference on the two agent's characteristics: skill θ and organizational involvement β should be based on only one observed statistic y_1.[5]

Intuitively, taking into account the "performance reversal" result, obtained for the benchmark case (see proposition 1), a high output y_1 is an evidence of high skill θ and strong involvement β when career concerns is weak. However, when career concerns is strong, high output implies high θ and low β. While in the former case, it is definitely beneficial for the agent to increase output, in the latter it is unclear whether high output y_1 will be interpreted by the market in a way beneficial for the agent. In other words, "performance reversal" can cause "inference switch", when higher output increases the expectation on involvement in the case of relatively weak career concerns, but decreases the expectation when career concerns is relatively strong. This reasoning on the possibility of "inference switch" will be formally justified below in this section.

I turn now to the formal analysis.

Clearly, period-2 performance is still determined by (1): $C'(a_2(\beta_j)) = \beta_j$ for $j = L, H$.

Let $a_1^H(a_1^L)$ denote the principals' beliefs about the period-1 effort of the $\beta_H(\beta_L)$-agent, $a_1(\beta)$ be the actually exerted effort (in equilibrium, $a_1(\beta_j) = a_1^j$), and let y_1 be the observed period-1 output.

Let $f(t)$ denote the probability density function for the random variable $\theta + \varepsilon_1$. Since θ and ε_1 are independent, $\theta + \varepsilon_1 \sim \mathcal{N}\left(\overline{\theta}_1, \sigma_1^2 + \omega^2\right)$.

The beliefs update is described by the following lemma.

Lemma 3. *The principals' updated beliefs at the beginning of period-2 are*

$$\widetilde{p} = Pr\left(\beta = \beta_H \,|\, y_1\right) = \frac{p\gamma_H}{p\gamma_H + (1-p)\gamma_L}$$

$$\text{with prob. } \widetilde{p}\text{: } \theta \sim \mathcal{N}\left(\overline{\theta}_2^H, \sigma_2^2\right)$$

$$\text{with prob. } 1 - \widetilde{p}\text{: } \theta \sim \mathcal{N}\left(\overline{\theta}_2^L, \sigma_2^2\right)$$

---

[5] Even if the interaction lasts for more than two periods, and more than one realization of output y_t is observed, this will only refine the information on β and θ, without making them known with certainty - see section 5.2.

where $\gamma_j = f(y_1 - a_1^j)$, $\overline{\theta}_2^j = \lambda(y_1 - a_1^j) + (1 - \lambda)\overline{\theta}$, $f(t)$ is the probability density function of the $\mathcal{N}\left(\overline{\theta}_1, \sigma_1^2 + \omega^2\right)$ distribution, λ and σ_2 are determined in Lemma 1.

The proof of the lemma is given in the Appendix.

Wage w_2 is equal to principals' expectation on agent's output at $t = 2$,

$$w_2 = E^P\left[\theta + a_2(\beta)|\, y_1\right] = \widetilde{p}\left(\overline{\theta}_2^H + a_2^H\right) + (1 - \widetilde{p})\left(\overline{\theta}_2^L + a_2^L\right) =$$
$$= \lambda y_1 + \widetilde{p}(\Delta a_2 - \lambda \Delta a_1) + a_2^L - \lambda a_1^L + (1 - \lambda)\overline{\theta} \quad (5)$$

where $\Delta a_t = a_t^H - a_t^L$ is the difference in performance between agents with high and low involvement in period t.

Agent's expected utility, viewed from $t = 1$ is $U = U_1 + \delta U_2$, and

$$E_1^A U_1 = \beta(E_1\theta + a_1) - C(a_1) + (1 - \beta)w_1$$
$$E_1^A U_2 = \beta(E_1\theta + a_2) - C(a_2) + (1 - \beta)E_1^A w_2$$

Agent's optimization then determines period-1 effort $a_1(\beta)$ by

$$C'(a_1^j) = (1 - \beta_j)\lambda\delta + \beta_j + (1 - \beta_j)\delta\frac{\partial}{\partial a_1}\left(E^A\widetilde{p}\right)(\Delta a_2 - \lambda \Delta a_1) \quad (6)$$

By comparing (6) and (4), one can see that the uncertainty about agent's involvement β creates an additional incentive (or disincentive) for the agent to signal high involvement. Two other incentives - the weakened career concerns and intrinsic motivation are still in place and not altered by the presence of a new dimension of information asymmetry.

The additional incentive to signal strong organizational involvement is represented by

$$(1 - \beta) \cdot \underbrace{\frac{\partial}{\partial a_1}\left(E^A\widetilde{p}\right)}_{reliability\ signalling} \cdot \underbrace{\delta\left(\Delta a_2 - \lambda \Delta a_1\right)}_{the\ value\ of\ being\ perceived\ \beta_H}$$

The incentive is driven by three components: signalling reliability (i.e. strong organizational involvement), the value of being perceived strongly involved and the weakening of the incentive by the involvement. I discuss these three components below.

First, the term

$$k_j = \left.\frac{\partial}{\partial a_1}\left(E^A\widetilde{p}\right)\right|_{a_1 = a_1(\beta_j)}$$

shows the marginal influence of effort on beliefs update, i.e. the marginal increase in $\widetilde{p}$ - the belief that the agent is of β_H-type. The coefficients k_j is analogous to λ, which shows the marginal influence on the belief update on skill:

$$\lambda = \frac{\partial}{\partial a_1}\left(E^A\widetilde{\theta}\right)$$

For brevity, I will refer to k_j's as the intensities of signalling reliability, i.e., shaping belief over organizational involvement. I distinguish k_j's from λ, which is essentially also a measure of belief shaping, but λ is related to the belief on skill, and is a part of the career concerns intensity $\lambda\delta$.

Second, The term $(\Delta a_2 - \lambda\Delta a_1)$ shows the agent's value of being perceived the β_H-type. Being perceived highly pro-social brings about a trade-off. On the one hand, the β_H-agent exerts in period-2 effort, greater by Δa_2, as compared to the β_L-agent, and because of this is offered a salary, higher by this amount. On the other hand, the period-1 effort of the β_H-agent is by Δa_1 greater, as compared to the β_L-agent. Consequently, the skill θ of the β_H-agent is discounted by $\lambda\Delta a_1$ more, compared to the β_L-agent, and the period-2 wage is decreased accordingly. The difference of the two effects results in the agent's total valuation. Notice that this valuation, in principle, can be positive or negative.[6]

Third, similarly to the weakening of career concerns, organizational involvement weakens the incentive to shape the principals' belief about agent's involvement, which is represented by the multiplier $1 - \beta$. Indeed, signalling high involvement leads to a future wage increase. However, since the pro-social agent partially sacrifices the higher wage paid to him by the principal, he is less prone to engage in the signal jamming.

Finally, it is worth to stress that there are clear parallels between career concerns and the incentive to signal involvement. They are the two dimensions of the reputation building incentive and are driven by "belief shaping" (λ and k_j respectively), the value of the shaped beliefs (δ and $\delta(\Delta a_2 - \Delta a_1)$), and weakened by agent's care for the principal's objective ($1 - \beta$ in both cases).

I will now derive some properties of the reliability signalling component k_j and equilibrium performance.

Lemma 4. *The values of k_j are given by*

$$k_H = p(1-p)\frac{\Delta a_1}{\sigma_1^2 + \omega^2} \iint\limits_{\Theta,\mathcal{E}} \frac{f(\theta+\varepsilon)f(\theta+\varepsilon+\Delta a_1)}{[pf(\theta+\varepsilon) + (1-p)f(\theta+\varepsilon+\Delta a_1)]^2}\, d\Omega(\theta,\varepsilon) \quad (7)$$

$$k_L = p(1-p)\frac{\Delta a_1}{\sigma_1^2 + \omega^2} \iint\limits_{\Theta,\mathcal{E}} \frac{f(\theta+\varepsilon)f(\theta+\varepsilon-\Delta a_1)}{[pf(\theta+\varepsilon-\Delta a_1) + (1-p)f(\theta+\varepsilon)]^2}\, d\Omega(\theta,\varepsilon) \quad (8)$$

where $d\Omega(\theta,\varepsilon) = \varphi_\theta(\theta)\varphi_\varepsilon(\varepsilon)d\theta d\varepsilon$; $\varphi_\theta(\theta)$ and $\varphi_\varepsilon(\varepsilon)$ are the density functions of the random variables θ and ε respectively.

The proof of Lemma 4 is given in the Appendix.

The following proposition characterizes the period-1 effort.

---

[6]To justify the possibility of negative valuation in equilibrium, notice that the considered valuation is only a part of the overall benefits of exerting effort. For instance, another component of the benefits emerges from signalling skill θ. The effort decision is based on the maximization of the overall benefits (net of cost) and may lead to choosing effort at such level that one component of the overall benefits is negative.

Proposition 2. *For the case of information asymmetry about involvement, effort exerted in period 1 by the β_j-agent ($j = L, H$) is determined by*

$$C'(a_1^j) = (1 - \beta_j)\lambda\delta + \beta_j + (1 - \beta_j)\delta k_j \left(\Delta a_2 - \lambda \Delta a_1\right) \tag{9}$$

The proposition is a restatement of the first order conditions for the agent's optimization (6).

Based on proposition 2 and lemma 4 one can find the equilibrium of the model by solving the system of four equations: two equations from lemma 4 determine k_j and two equations (9) with $j = L, H$ determine a_1^j.

We will now obtain sharper results on reliability signalling and performance reversal.

Lemma 5. *For the marginal influence on belief on agent's organizational involvement holds:*

- *if $\delta\lambda < 1$, then $a_1^H > a_1^L$ and $k_H, k_L > 0$*

- *if $\delta\lambda > 1$, then $a_1^H < a_1^L$ and $k_H, k_L < 0$*

- *$|k_H| > |k_L|$ for $p < \frac{1}{2}$, $|k_H| < |k_L|$ for $p > \frac{1}{2}$, and $k_H = k_L$ for $p = \frac{1}{2}$.*

The proof of lemma 5 is given in the Appendix.

Let us discuss the implications of the obtained results. First, Lemma 5 confirms the robustness of the "performance reversal" result, obtained for the benchmark case: effort increases with involvement when career concerns is relatively week, and decreases with involvement if career concerns is relatively strong for the case of asymmetric information as well.

Second, the lemma provides formal grounds for the already discussed "inference switch": higher output implies lower involvement when career concerns is relatively strong, and vice versa when career concerns is relatively weak. Indeed, the lemma states that the marginal influence on belief that the agent is highly pro-social k_j is positive if career concerns is relatively low, and negative if it is relatively high. The fact that in the considered setting the sign of k_j switches exactly at $\delta\lambda = 1$ is related to the "performance reversal" result.

Finally, the lemma characterizes the "reliability signalling". It turns out that, depending on the composition of the population of the agents, either more or less involved agent is more aggressive in shaping principal's belief on the involvement. When the ex-ante share of highly involved agents is small ($p < 1/2$), the more involved agent is more aggressive, i.e., more deeply engaged into signal jamming. In other words, the highly involved agent makes a stronger attempt to be distinguished from the majority. However, when the share of highly involved agents is relatively large ($p > 1/2$), the agent with low involvement is more deeply involved in signal jamming, i.e. makes a harder attempt to be indistinguishable from the majority.
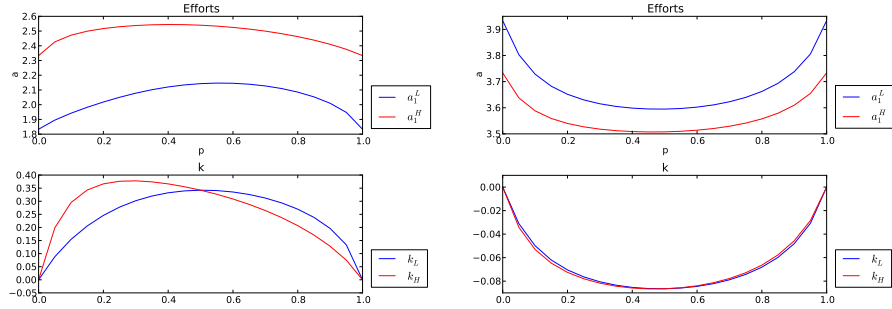
Figure 1: Efforts as function of p. Left panel: $\delta = 0.5$, Right panel: $\delta = 1.2$. Other parameters: $\beta_H = 0.4$, $\beta_L = 0.1$, $C(a) = 0.3\frac{a^2}{2}$, $\theta \sim N(2, 0.25)$

4.2 The impact of information asymmetry

Figure 4.2 depicts effort a_1^j and the marginal influence on the beliefs k_j as a functions of probability p for the equilibrium levels of effort. On the left panel $\delta < 1$, on the right panel $\delta > 1$. The values of p close to 0 or 1 correspond to the case of little uncertainty (and little information asymmetry) about agent's involvement, whereas for p close to $1/2$ the information asymmetry is the greatest.

The bottom panel of figure 4.2 illustrates the marginal influence on the beliefs k_j as function of p for the equilibrium levels of effort. On the one hand, when p is close to 0 or 1, the values of k_j are close to zero. This follows from Lemma 4, and is very intuitive: since there is little uncertainty about agent's type (the involvement dimension), it is hard to shift beliefs about it. On the other hand, the values of k_j are maximal for p close to $1/2$, i.e. when the uncertainty about agent's type is the greatest, and, therefore, the period-1 output has a big impact on the belief update. This is implied by Lemma 4. Indeed, according to the lemma, $k_j = p(1-p)I_j(p) \cdot const$, where $I_j(p)$ are the integrals in (7) and (7). The expression $p(1-p)$ is maximal at $p = 1/2$, which is adjusted by the presence of the multiplier $I_j(p)$.

The total incentive to signal high organizational involvement is determined by the intensity of reliability signalling k_j, multiplied by the marginal impact of being perceived highly involved on the future benefits. First, the relative importance of the future δ. Second, the value of being perceived highly pro-social $\Delta a_2 - \lambda \Delta a_1$. These two factors are identical for all agents. Finally, the weakening of the incentive, represented by $(1 - \beta_j)$, which plays essentially the same role as weakening of career concerns: the incentive is more deflated for the more pro-social agent.

To sum up, effort is determined by the three forces: intrinsic motivation and two components, associated with reputation building (shaping belief on type):

signalling high skill and high involvement. The information asymmetry about agent's involvement influences only the last component. The following lemma characterizes this influence.

Lemma 6. *The effort component motivated by reliability signalling is stronger for the β_H-agent if $p > \hat{p}$, otherwise it is stronger for the β_L-agent. The value $\hat{p}$ satisfies $\hat{p} < \frac{1}{2}$.*

The proof of the Lemma is given in the Appendix.

Table 1 summarizes the structure of incentive and resulting performance.

Table 1: Incentive structure for altruistic agent in the case of career concerns.

	Case	
Incentive	$\delta\lambda < 1$	$\delta\lambda > 1$
Signalling θ (career concerns)	$L > H$	$L > H$
Signaling $\beta = \beta_H$ (reliability signalling)	$k_j > 0$, $L, H > 0$ $L < H$ for $p < \hat{p} < 0.5$ $L > H$ for $p > \hat{p}$	$k_j < 0$, $L, H < 0$ $|L| < |H|$ for $p < \hat{p} < 0.5$ $|L| > |H|$ for $p > \hat{p}$
Intrinsic motivation	$L < H$	$L < H$
Performance	$a_L < a_H$	$a_L > a_H$ (performance reversal)

The letters L, H in the table correspond to effort component determined by incentives in the first column.

Let us now be back to the discussion of the sign of the expression $\Delta a_2 - \lambda\Delta a_1$ representing the value of being perceived the β_H-agent.

First, notice that if $\delta\lambda > 1$, then $\Delta a_1 < 0$, and, given, that $\Delta a_2 > 0$, this leads to $\Delta a_2 - \lambda\Delta a_1 > 0$. Second, for the case of $\delta\lambda < 1$, one can guarantee that $\Delta a_2 - \lambda\Delta a_1 > 0$ when marginal cost $C'(a)$ is a convex function. However, it is as well possible to have $\Delta a_2 - \lambda\Delta a_1 < 0$ when $C'(a)$ is concave.[7]

5 Extensions

5.1 Caring Principal

It can be reasonable to assume that not only the agent cares about principal's objectives, but also the principal cares about the agent. For example, one can think of a family enterprise, where both the principal and the agent care about each other.[8] Nevertheless, the setting with principal's pro-sociality is plausible only in some specific contexts. Having in mind the family enterprise, I will stick here to interpreting β as "altruism" instead of "involvement", as it

---

[7]This is confirmed by numerical example: with $C'(a) = 0.3\sqrt{a}$, $p = 0.2$, $\theta \sim \mathcal{N}(2, 0.25)$, $\lambda \geq 0.5$, $\delta = 0.5$, $\beta_H = 0.4$, $\beta_L = 0.1$, one can solve the problem numerically and verify that $\Delta a_2 - \lambda\Delta a_1 < 0$.

[8]Another possibility is to consider a relationship between two friends.

better captures the essence of the relationship. I will focus on the impact of the principal's altruism and its interaction with agent's career concerns in this part of the paper.

Let α denote the degree of principal's altruism, and β, as before, the degree of agent's altruism. Then the utilities of the principal and the agent are given by

$$V_t = y_t - w_t + \alpha(w_t - C(a_t))$$
$$U_t = w_t - C(a_t) + \beta(y_t - w_t)$$

Clearly, the agent's effort in period 2 and the belief update are not affected by the fact that the principal is altruistic towards the agent, and the results of the previous analysis stated in lemmas 1 and 2 continue to hold. Principal's expected utility, as viewed at the beginning of the period 2, is given by $V_2 = \overline{\theta_2^P} + a_2 - (1-\alpha)w_2 - \alpha C(a_2)$, and wage is determined by the principals' competition, which drive them to zero utility:

$$w_2 = \overline{\theta_2^P} + a_2 + \frac{\alpha}{1 - \alpha}\left(\overline{\theta_2^P} + a_2 - C(a_2)\right) \tag{10}$$

This leads to the following agent's maximization program in period 1:

$$\max_{a_1}\left\{\beta a_1 - C(a_1) + \delta(1 - \beta)\frac{1}{1 - \alpha}\lambda a_1 + \{const\ on\ a_1\}\right\}$$

Agent's effort in the first period is then determined by

$$C'(a_1) = (1 - \beta)\frac{1}{1 - \alpha}\delta\lambda + \beta \tag{11}$$

The obtained result implies that first, effort increases with principal's altruism. The key for this effect is the principals' competition increasing the period-2 wage. Indeed, principals partially internalize the agent's benefit from a wage increase, and therefore are more inclined to increasing payment to the agent. This reinforces agent's benefits from signal jamming, i.e. manipulating principal's beliefs through exerting higher effort in the first period. Effectively, the competition among altruistic principals[9] strengthens career concerns, and the multiplier $\frac{1}{1-\alpha}\delta\lambda$ can be considered as a measure of career concerns incentive (instead of "standard" $\delta\lambda$).

Finally, let us discuss the consequences of limited competition between altruistic principals. The limits to the competition can be imposed by different circumstances, for example, moving from one principal to another (changing jobs) can be costly for the agent. Clearly, in this case the impact of principals' altruism will be reduced, as worker's second period benefits from signal jamming in the first period are less sensitive to principals' beliefs shaping.

---

[9]It is clear that if principals internalize agent's benefits due to some another reason, the consequences will be the same.

5.2 More than 2 periods

In this part of the paper I consider the same interaction as before which lasts for $T > 2$ periods. I start with the benchmark case when agent's organizational involvement is known, and then turn to the main case when the agent builds reputation on both skill and involvement.

Let δ denote the time-discounting factor, so that the discounted value of the utility stream $u_t, u_{t+1}, \cdots, u_T$ evaluated at period t is $U = u_t + \delta u_{t+1} + \cdots + \delta^{T-t} u_T$. We will now assume that $0 \le \delta \le 1$.

Period-t output noise is ε_t. All random variables $\varepsilon_t \sim \mathcal{N}(0, \omega^2)$ are assumed to be i.i.d. and independent from θ.

5.2.1 The Benchmark Case

Assume that at the beginning of period t the beliefs on the distribution of agent's skill θ is given by $\theta \sim \mathcal{N}(\overline{\theta}_t, \sigma_t^2)$.[10] The Principal also has belief a_t^{μ} about the period-t output.

After observing the period-t output y_t, the beliefs on the distribution of θ is updated. Period-1 belief is commonly known: $\theta \sim \mathcal{N}(\overline{\theta}_1, \sigma_1^2)$, i.e., commonly known are the parameters of the distribution θ_1 and σ_1. The following lemma describes the update.

Lemma 7. *Suppose that at the beginning of period-t the beliefs on the distribution of θ is given by $\theta \sim \mathcal{N}(\overline{\theta}_t, \sigma_t^2)$, and the period-$t$ effort is a_t. Then period-$t+1$ belief is given by*

$$\theta \sim \mathcal{N}(\overline{\theta}_{t+1}, \sigma_{t+1}^2)$$

where $\overline{\theta}_{t+1} = \lambda_t(y_t - a_t^{\mu}) + (1 - \lambda_t)\overline{\theta}_t, \quad \lambda_t = \frac{\sigma_t^2}{\sigma_t^2 + \omega^2}, \quad \frac{1}{\sigma_{t+1}^2} = \frac{1}{\omega^2} + \frac{1}{\sigma_t^2}$

The belief update in period $t + s$, based on period t belief and performance in periods $t, t+1, \ldots t+s-1$ $(t, s \ge 1)$ is given by

$$\overline{\theta}_{t+s} = \sum_{q=0}^{s-1} \Lambda_{t+q, t+s}(y_{t+q} - a_{t+q}^{\mu}) + \prod_{q=0}^{s-1}(1 - \lambda_{t+q}) \cdot \overline{\theta}_t$$

where $\Lambda_{t,t+1} = \lambda_t, \Lambda_{t,t+r} = (1 - \lambda_{t+r-1})(1 - \lambda_{t+r-2}) \cdot \ldots \cdot (1 - \lambda_{t+1})\lambda_t$ for $r \ge 2$,

$$\frac{1}{\lambda_t} = t + \frac{\omega^2}{\sigma_1^2}, \quad \frac{1}{\sigma_{t+1}^2} = \frac{t}{\omega^2} + \frac{1}{\sigma_1^2}$$

The Proof of the Lemma 7 is given in the appendix.

The Lemma not only provides the retrospective view on the determinants of the current beliefs, but also can be interpreted in the forward-looking sense, as it allows to track the future influences of the currently exerted effort. For

---

[10]We will see that if the initial belief is represented by the normal distribution, then all the subsequent updates will also be represented by some normal distribution, so our assumption does not reduce generality.

instance, the coefficients $\Lambda_{t,t+r}$ represent the marginal impact of the period-t effort on the period-$t+r$ belief on the agent's skill:

$$\Lambda_{t,t+r} = \frac{\partial \overline{\theta}^j_{t+r}}{\partial a_t} \tag{12}$$

One can see that the coefficients $\Lambda_{t,t+r}$ are a generalization of the coefficient λ, introduced in the 2-period setting. Indeed, $\lambda = \Lambda_{12}$.

We can now characterize the agent's effort in period t.

Lemma 8. *The equilibrium period-t effort is given by*

$$C'(a_t) = \beta + (1-\beta)\Gamma_t$$

where

$$\Gamma_t = \sum_{s=1}^{T-t} \delta^s \Lambda_{t,t+s}$$

The lemma is a generalization of the two-period result stated in proposition 1. The coefficient Γ_t shows the intensity of the career concerns incentive in the first period of t-period interaction and can be thought as an aggregated value of the future discounted career concerns, and in this sense is a generalization of the career concerns intensity in the 2-period case $\delta\lambda$. It is shown in the proof of lemma 9 that $\{\Gamma_t\}$ is a decreasing sequence, so that the relative importance of the future declines, when fewer periods if left.

The lemma shows that performance of a motivated agent is driven, as for the 2-period case, by two motives: intrinsic motivation and weakened career concerns.

Finally, we generalize the "performance reversal" result.

Lemma 9. *Let $\beta_L < \beta_H$ be any two values.*

If in period τ $a_\tau(\beta_H) < a_\tau(\beta_L)$, then in all previous periods $t < \tau$ also holds $a_\tau(\beta_H) < a_\tau(\beta_L)$.

If in period τ $a_\tau(\beta_H) > a_\tau(\beta_L)$, then in all consequent periods $t > \tau$ also holds $a_\tau(\beta_H) > a_\tau(\beta_L)$.

As the lemma claims, the agents with low involvement can exert higher effort at the beginning of the career, i.e., when career concerns are high. In this case, effort is a decreasing function of involvement, which manifests the "performance reversal". However, the direction of monotonicity switches, as time goes and career concerns declines: effort becomes an increasing function of agent's intrinsic motivation at least at the end of career. It follows from the Lemma that monotonicity can switch from "reversed" to "normal" direction only once.

Figure 5.2.1 shows the sets of parameters (δ, μ) ($\mu = \frac{\omega^2}{\sigma^2}$) for which the period-1 effort in the T-period interaction (for $T = 3, \ldots, 9$) is a decreasing function of agent's involvement., i.e. there is "performance reversal". For each T the set lies below the corresponding line. For example, the shaded area shows the set of parameters for the 9-period interaction.
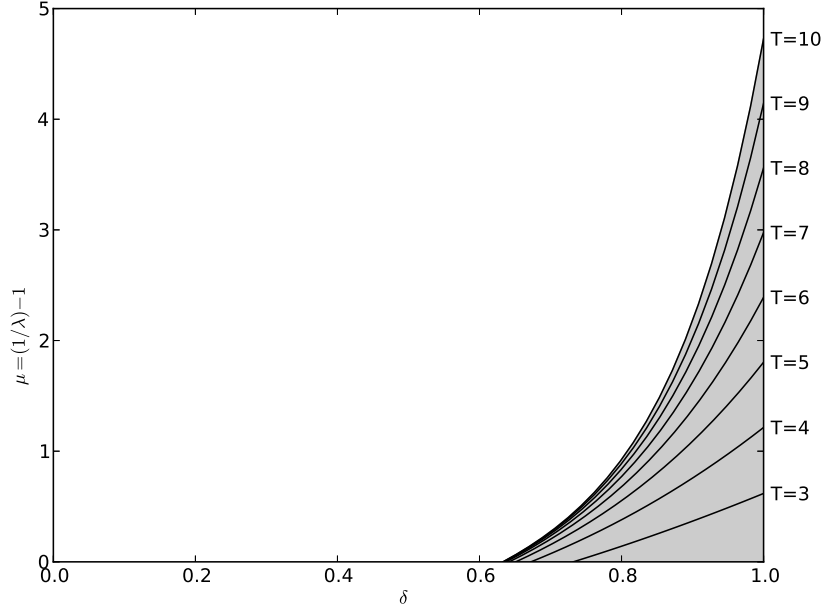
16

Figure 2: The set of parameters (δ, μ) for which "performance reversal" appears in the first period of the T-period interaction.

5.2.2 The Main Case

We will now extend the analysis to the setting where agent's organizational involvement β is not known to the principal. As previously in the article, I assume that agent's involvement can take only two values $\beta_L < \beta_H$.

We shall start with the beliefs update. There are two important differences with the 2-period case, when the update is based on the ex-ante information and one observation (period-1 output). First, after period 1, the variables (θ, β) in the ex-ante distribution are correlated. Second, while the mechanism of the influence of the current effort on the expected value of skill is the same, as in the benchmark case, the mechanism of the influence on the belief over agent's involvement is more complex. This is illustrated by the diagram on Figure 5.2.2. On the diagram, the arrows show the direction of the influence, and the expressions on the arrows are the marginal impacts of the change in a_t. One can see, for example, that a_t influences directly on next period belief on θ and β through the already familiar mechanisms. However, a_t also affects, for example, belief p_{t+2} indirectly through θ_{t+1}^j and p_{t+1}. Such influence is absent in the 2-period setting. Clearly, a_t indirectly affects all the future p_{t+s} and $\overline{\theta}_{t+s}$.
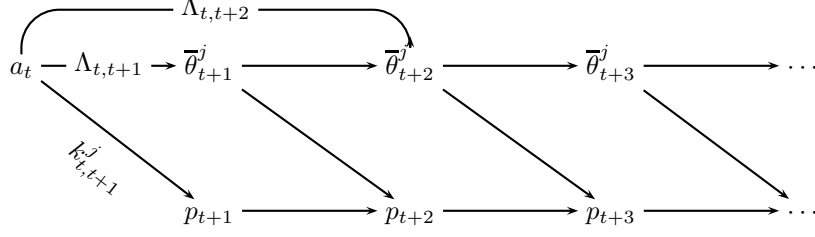
Figure 3: Marginal effects of a_t on beliefs.

The following Lemma formally describes the belief update.

Lemma 10. *Let beliefs at the begining of period-t are given by:*

$$\text{with prob. } p_t\text{: } \beta = \beta_H, \theta \sim \mathcal{N}\left(\overline{\theta}_t^H, \sigma_t^2\right)$$

$$\text{with prob. } (1 - p_t)\text{: } \beta = \beta_L, \theta \sim \mathcal{N}\left(\overline{\theta}_t^L, \sigma_t^2\right)$$

The beliefs on period-t effort are a_t^j for β_j-agent, and the observed output is y_t. Then the period-$(t+1)$ beliefs have the same structure, as period-t beliefs with $\overline{\theta}_{t+1}^j = \lambda_t(y_t - a_t^j) + (1 - \lambda_t)\overline{\theta}_t^j$, σ_{t+1} and λ_t determined in Lemma 7,

$$p_{t+1} = \frac{p_t \gamma_t^H}{p_t \gamma_t^H + (1 - p_t)\gamma_t^L}$$

where $\gamma_t^j = f_t^j(y_t - a_t^j)$, and $f_t^j(x)$ is the p.d.f. of the $\mathcal{N}\left(\overline{\theta}_t^j, \sigma_t^2 + \omega^2\right)$ distribution.

The proof of the Lemma is given in the appendix.

We shall now characterize the long-range (for period $t + s$) impact of the current (period t) effort in order to determine the total marginal benefit from exerting effort.

Denote by

$$k_{t,t+s}^j = \left.\frac{\partial}{\partial a_t} E_t p_{t+s}\right|_{a_t = a_t(\beta_j)} \tag{13}$$

the expected (from the period-t perspective) marginal impact of period-t effort on the period-$(t + s)$ belief that the agent is of β_j-type. The coefficients k_H and k_L determined in Lemma 4 are particular cases of $k_{t,t+s}^j$ for $s = 1$. The parallels between coefficients k_j and λ hold for the coefficients $k_{t,t+s}^j$ and $\Lambda_{t,t+s}$: they represent the marginal impact on the belief on two dimensions of the agent's type of the period-t effort, as follows from (12) and (13).

The following lemma provides the formulas for $k_{t,t+s}^j$.

Lemma 11. *The marginal impacts $k_{t,t+s}^j$ are given by the following formulas:*
for $s = 1$

$$k_{t,t+1}^H = p_t(1-p_t)\frac{\Delta a_t + \Delta\overline{\theta}_t}{\sigma_t^2 + \omega^2} \iint\limits_{\Theta,\mathcal{E}} \frac{f_t^L(\theta + \varepsilon + \Delta a_t)f_t^H(\theta + \varepsilon)}{\left[p_t f_t^H(\theta + \varepsilon) + (1-p_t)f_t^L(\theta + \varepsilon + \Delta a_t)\right]^2} \, d\Omega(\theta, \varepsilon)$$

$$k_{t,t+1}^L = p_t(1-p_t)\frac{\Delta a_t + \Delta\overline{\theta}_t}{\sigma_t^2 + \omega^2} \iint\limits_{\Theta,\mathcal{E}} \frac{f_t^L(\theta + \varepsilon)f_t^H(\theta + \varepsilon) + \Delta a_t}{\left[p_t f_t^L(\theta + \varepsilon) + (1-p_t)f_t^H(\theta + \varepsilon - \Delta a_t)\right]^2} \, d\Omega(\theta, \varepsilon)$$

for $s > 1$: $k_{t,t+s+1}^j = \left(k^\theta\right)_{t,t+s+1}^j + \left(k^p\right)_{t,t+s+1}^j$
where

$$\left(k^\theta\right)_{t,t+s+1}^j = \int\cdots\int\limits_{\mathcal{E}_t,\ldots,\mathcal{E}_{t+s},\Theta} p_{t+s}(1-p_{t+s})\Lambda_{t,t+s}\frac{-\Delta a_{t+s} - \Delta\overline{\theta}_{t+s}}{\sigma_{t+s}^2 + \omega^2} \times$$

$$\times \frac{\gamma_{t+s}^L\gamma_{t+s}^H}{\left[p_{t+s}\gamma_{t+s}^H + (1-p_{t+s})\gamma_{t+s}^L\right]^2} \, d\Omega\left(\varepsilon_t,\ldots,\varepsilon_{t+s},\theta\right)$$

$$\left(k^p\right)_{t,t+s+1}^j = \int\cdots\int\limits_{\mathcal{E}_t,\ldots,\mathcal{E}_{t+s},\Theta} \frac{\partial p_{t+s}}{\partial a_t} \frac{\gamma_{t+s}^L\gamma_{t+s}^H}{\left[p_{t+s}\gamma_{t+s}^H + (1-p_{t+s})\gamma_{t+s}^L\right]^2} \, d\Omega\left(\varepsilon_t,\ldots,\varepsilon_{t+s},\theta\right)$$

and $\gamma_{t+s}^j = f_{t+s}^j(y_{t+s} - a_{t+s}^j)$ are functions of $\varepsilon_t,\ldots,\varepsilon_{t+s},\theta$ (f_{t+s}^j is p.d.f. for the $\mathcal{N}\left(\overline{\theta}_{t+s}^j, \sigma_{t+s}^2 + \omega^2\right)$ distribution, where $\overline{\theta}_{t+s}^j$ depends on $\varepsilon_t,\ldots,\varepsilon_{t+s-1}$, $y_{t+s}^j = \theta + a_{t+s}^j + \varepsilon_{t+s}$).

The proof of lemma 11 is given in the appendix.

One can see that lemma 4 is a particular case of lemma 11 with $f^L = f^H = f$ and $\Delta\overline{\theta}_1 = 0$.

The formulas for the one period ahead update ($k_{t,t+1}^j$) show the direct impact of the period-t effort on p_{t+1} (see figure 5.2.2). For more than one period ahead update, the impact of current effort is indirect and goes through the impact on beliefs in the previous period. The corresponding marginal effects are represented by the terms k^p and k^θ in the lemma.

Finally, we will characterize agent's effort.

Proposition 3. *The equilibrium effort in period t is determined by*

$$C'(a_t) = \beta + (1-\beta)\Gamma_t + (1-\beta)\sum_{s=1}^{T-t}\delta^s k_{t,t+s}\left(\Delta a_{t+s} - \Lambda_{t,t+s}\Delta a_t\right)$$

The proof of the proposition 3 is given in the appendix.

It is clear that proposition 2, characterizing equilibrium for the 2-period setting, is a particular case of proposition 3. The incentive structure for the more than 2 periods interaction is similar to this for the 2-period case. Agent's performance is driven by intrinsic motivation, weakened career concerns and belief (on degree of alignment with organizational objectives) shaping. See section 4 for the detailed discussion.

5.3 Valuing output vs valuing objective

In the previous analysis the agent was assumed to be pro-social in the sense that he internalizes the organizational objective, represented by the net principal's wealth $(y - w)$. It could, however, be the case that the agent attaches different values to output y and payment from the principal w, so that utility of the pro-social agent takes the form

$$U = w - C(a) + \beta_y y - \beta_w w$$

In the previous analysis, $\beta_y = \beta_w = \beta$. It could, however, be that $\beta_w \ll \beta_y$, and in the extreme case $\beta_w = 0$, i.e., the agent values only output.

The assumption of $\beta_y = \beta_w$ or β_w comparable with β_y is suitable to model the agency relationship in, e.g., family enterprise, charity organizations, or the case of courting. Generally, the model with $\beta_y = \beta_w$ seem to be suitable to model relationship where both the agent and the principal are individuals rather than organizations or a society. The model with $\beta_w = 0$ is better suitable to model the agency relationship when the principal is the organization or a society. Examples are bureaucracy, especially at the state or regional level (bureaucrats at the local level can internalize society's wealth and have $\beta_w > 0$), public education and healthcare, different kinds of mission driven organizations.

In this part of the paper I will discuss how the main findings of the previous analysis when $\beta_y = \beta_w = \beta$ change if agent's involvement is related only to the organization's output, rather than to wealth. One can immediately see that in the considered case the agent is still intrinsically motivated, but his career concerns will not be weakened. Therefore, agent's involvement will not alter the engagement into signal jamming with respect to his productivity.

Consider the two-period setting. Let agent's per-period utility be given by $U = w - C(a) + \beta y$. The period-2 effort is given by $C'(a_2) = \beta$, as before. Notice that all belief updates are not affected by the different assumption on the structure of agent's involvement. Agent's optimization for period-1 leads to

$$C'(a_1) = \beta + \delta \lambda$$

We can immediately see that the weakening of career concerns is absent now, and agent's performance is determined by intrinsic motivation and career concerns, which are identical to career concerns for selfish agent. An important consequence of changing the assumption on the agent's involvement is that now there is no "performance reversal": effort is an increasing function of agent's involvement β, independently of the strength of career concerns.

In case of information asymmetry about agent's involvement β, the performance is determined by intrinsic motivation, career concerns, and signalling high involvement:

$$C'(a_1) = \beta + \delta \lambda + \delta k_j \left(\Delta a_2 - \lambda \Delta a_1 \right)$$

This formula can be obtained by straightforward adoption of the proof of proposition 2.

The "performance reversal" and "inference switch" do not emerge in this case too, and performance is always an increasing function of agent's involvement, independently of the strength of career concerns. However, the "reliability signalling" differences between the agents with strong and weak involvement are still there, and one can show that the result of lemmas 5 and 6 on the comparison of "reliability signalling" intensities k^j hold.

Lemma 12. *For the setting with $\beta_w = 0$:*

- $a_1^H > a_1^L > 0$ *and* $k^j > 0$

- $k^L > k^H$ *for* $p > 1/2$, *and* $k^H > k^L$ *otherwise*

- *the effort component motivated by reliability signalling is stronger for the β_H-agent if $p > \hat{p} = 1/2$; otherwise it is stronger for the β_L-agent.*

The proof is obtained by straightforward adaption of the proofs of lemmas 5 and 6.

The incentive structure and the resulting performance for the setting with $\beta_w = 0$ is summarized in Table 2

Table 2: Incentive structure for altruistic agent in the case of career concerns.

Incentive	
Signaling θ	$L = H$
Signaling $\beta = \beta_H$	$k_j > 0$, $L, H > 0$ $L < H$ for $p < \hat{p} = 0.5$ $L > H$ for $p > \hat{p}$
Intrinsic motivation	$L < H$
Performance	$a_L < a_H$

The letters L, H in the table correspond to effort component determined by incentives in the first column.

Let us now discuss how the difference in assumptions ($\beta_w = 0$, instead of $\beta_w = \beta_y$) alters the results. First, there is no more weakening effect for highly pro-social agent. As a result, the strength of career concerns incentives is now the same for all agents, and coincides with the strength for the selfish agent. Second, the incentive to shape belief on agent's involvement is also altered, but the comparison between β_L and β_H-agent remains qualitatively the same for both settings. Finally, intrinsic motivation does not depend on the structure of the agent's involvement. The above discussed alterations in the incentive structure exclude now the possibility of performance reversal even for string career concerns.

6 Applications and Discussion

In this part of the paper, I discuss implications of the obtained results. As it was discussed in Section 2, the parameter β can be interpreted in different ways. It

can be seen as the degree of alignment with organization's (principal's) mission or objectives. It can also be interpreted as the degree of agent's amenability for the results of the effort, irrespectively of the organization's objectives, or another kind of involvement or reliability of the agent. It can as well be thought of as agent's (unconditional) altruism with respect to the principal. In any case, the agent with greater value of β has stronger intrinsic motivation, and are often referred to as "motivated agents".

6.1 Performance and motivation over career path in mission-driven organizations

The preceding analysis shows that in mission-driven organizations (and in other organizations involving motivated agents), the less involved agents (β_L-agents) can outperform the deeply involved and highly intrinsically motivated agents. More precisely, according to lemma 5 such "performance reversal" can emerge only at the early stages of career, when career concerns are high.[11] The "performance reversal" is caused by the greater weakening of career concerns for the deeply involved agents. The effect will, however, decline over time, as there will be less uncertainty about the agent's productivity; indeed, career concerns decrease, and so does the weakening effect. Therefore, at the later stages of career, the more involved agent will (on average) outperform the ones with low involvement.[12]

The findings suggest that the less involved agents can advance faster in the career progression at the early stages, compared to the deeply involved agents, but this is only a short-term effect. Put differently, the more selfish agents exhibit more "careerist" attitude towards organization. An interesting question is whether such peculiarity of career patterns in mission-driven organizations (and other organizations involving motivated agents) is optimal for the organizations. I leave this question open for future investigation.

For the case when the "performance reversal" does not emerge even at the early stages of career, the more deeply involved agent outperforms the self-serving one and will advance faster in the career.

Whether "performance reversal" emerges or not, the difference in effort between agents with high and low involvement increases over time, perhaps, starting with negative values, if performance reversal is there at the beginning.

In sum, the performance of the agent with stronger involvement, and as a consequence, stronger intrinsic motivation, is steadier over time, whereas the more self-serving agent's performance will decrease over time relatively fast. At the same time, the more self-serving agent can be more successful in career development at the early stages of career.

---

[11] And for the case of agents caring about organization's objectives (and not only about output).

[12] To be more precise, the more involved agent will exert higher effort. It will lead to higher output, holding other things equal, i.e., provided that productivity θ and output noise ε are the same.

Let us now discuss the sources of agent's motivation and their dynamics over the career path. First, intrinsic motivation does not change over time, and it is high for the deeply involved agents, and low for the less involved ones. This difference is the main determinant of agent's performance in the long-run. In the short-run, i.e., at the beginning of career, reliability signalling incentives are strong and alter effort to a large degree.

Second, the incentive for shaping beliefs on the agent's involvement. The results summarized in Tables 1 and 2 imply that the less involved agent will be more aggressive in signalling the deep involvement (alignment with organization's objectives) than the agent whose objectives are more aligned with the organization. This is the case when the uncertainty about agent's involvement is high (p close to $1/2$), and when the a-priori expectation that the agent is deeply involved is high. However, over time the belief on the agent's involvement will approach its true value, i.e. $p \to 0$ for the less involved agent and $p \to 1$ for the more involved agent, and the incentive based on reliability signalling will weaken, although it can still be more powerful for the less involved agent.

The dynamics of these two components of incentives does not depend on the structure of agent's social preferences, i.e. whether he cares about organizational objective or only on output ($\beta_w = \beta_y$ or $\beta_w = 0$, see discussion in subsection 5.3).

Third, the career concerns incentive, or the incentive to shape beliefs on skills (productivity). The incentive decreases over time for all agents. However, if career concerns is high enough at the beginning of career, the incentive is stronger for the less involved agent. This finding is, however, sensitive, to the model setting. If the organization's mission gives rise to agent's concern about organization's objective, and not only about output, then the career concerns incentive is stronger for the more self-serving agents. Conversely, if the organization's mission gives rise to agent's concern only about output, the career concerns incentive is the same for all agents.

To sum up, the more steady performance over the career path of the more involved agent in mission driven organizations is attributed to the dynamic stability of the intrinsic motivation. The career building incentive, i.e., the two dimensions of reliability signalling, can be stronger for the less involved agents, but these incentives decrease over time, which determines a more steady decrease in performance of the less socially involved agent.

6.2 Organization's composition and performance

It is commonly accepted that organizations should seek employees sharing their objectives (having their interests aligned with those of the organization) in order to increase the organization's output. The results of the previous analysis can be applied to study the optimal composition of the pool of employees in the organization. We will now address the question of whether the organization should always screen its potential employees and select those with views aligned

with the views of the organization. [13]

Proposition 2 and lemma 5 imply that if career concerns are weak enough and the signalling of strong alignment with the organization's objectives is beneficial ($\Delta a_2 - \Delta a_1 > 0$), then the greater heterogeneity of the agents (p close to $1/2$) creates a positive incentive to signal strong alignment and drives the effort of all agents upwards. Would the organization screen the agents and hire those with strong alignment (making p close to 1), the incentive to signal strong alignment will be almost absent, and the overall effort of all workers can be lower, compared to the case of the unscreened pool of workers. Therefore, strategically avoiding agents' selection can be beneficial for the organization.

On the contrary, for the case of strong enough career concerns, it is always in the interest of the organization to have a pool of workers strongly aligned with the organization's objectives. Importantly, these workers should know that they are "in a good company," as the incentive to signal strong alignment drives effort downward, i.e. the incentive becomes a counter-incentive.

6.3 Charities and anti-socially behaving agents

Charities usually employ generous and considerate people. However, recently there were a number of scandals engaging some employees of charities who were convicted in antisocial behavior. This seems to be puzzling, especially when taking into account the fact that there is usually a sufficient supply of people willing to work for charity.

The analysis in the paper, however, sheds some light on the causes of the penetration of "bad seeds" into charities, or, more generally, mission-oriented organizations. Indeed, before being employed in a charity, there is usually a candidate selection process, during which those having non-social goals for being employed in a charity may engage into a much stronger signal jamming than socially oriented candidates. As a result, these "bad seeds" can be selected to be employed, as they can be perceived as more productive and even as sharing the organization's objective to a larger degree. Note that the analysis suggests that such adverse outcome is likely to emerge under strong career concerns. Therefore, reducing the career concerns may help charities to avoid attracting people with anti-social intentions.

6.4 Courting

Let us consider a rather common situation when a gentleman (the agent) strives to build a long-term relationship or, perhaps, to marry a lady. In this case, the "productivity" θ is the attractiveness of a gentleman for a lady, or the value of the relationship with the gentleman. It can include a pecuniary component (e.g., future income). The lady does not know the exact value of θ, but should rather get an idea of what it is (i.e. form a belief).

The period-2 agent's benefit (w_2) comes from the establishment of the relationship and can be interpreted as love, care and attention given by the lady

---

[13]We will assume for now that the screening is costless.

to the gentleman. It is reasonable to expect that a gentleman receives greater benefits if he is perceived by a lady as more caring about her (having a greater β), because the lady is very likely to reciprocate this care. If the two gentlemen differ in their degree of altruism towards the lady, it is often the case that the less altruistic gentleman uses much more energy (exerts higher effort a_1) when competing for the establishment of the future relationship. Clearly, in this case "career concerns" are high, and the less caring agent engages in a stronger signal jamming. The analysis in the article point to one of the mechanisms in this social interaction[14].

7   Conclusion

This article incorporates pro-social preferences into the career concerns model. This extension is relevant, since many organizations in which career concerns is crucial in motivating agents are mission-driven; therefore, many employees in these organizations have some degree of alignment with organization's objectives, or feel involvement into organization's activity. The main purpose of the paper is to study the incentive structure and to characterize the performance patterns in organizations where career concerns is the main source of extrinsic incentives, and organization's mission is socially sound.

The paper demonstrates that when the degree of alignment with organization's objectives is not known by the organization, the reputation becomes two-dimensional: skills, as in the standard model, and the degree of alignment. The additional component of reputation contributes to the incentive provision. The paper studies the structure of incentives and performance of agents, paying particular attention to the impact of the pro-social component in the agent's preferences, i.e., their alignment with organization's objectives.

The analysis shows that involvement weakens career concerns (i.e. building reputation over skills) so that the less involved agent has stronger career concerns. Moreover, information asymmetry about the degree of agent's involvement creates an incentive to shape beliefs on it. This incentive is similar in its nature to the "standard" career concerns incentive, as it is based on the desire to build reputation and leads to the agent's engagement in signal jamming. However, to distinguish it from the "standard" career concerns, in this paper it is referred to as "reliability signalling".

The analysis shows that the two incentives related to reputation building can be stronger for the less involved agent. The less involved agent can, therefore, be more aggressive in signal jamming; consequently, can be more successful in career building at the early stages of career. This is, however, only short-term effect; over time, the performance of the less involved agent will decrease faster, whereas the performance of the deeply involved agent is more steady (but is also decreasing, as reputation building incentives decrease). At the later stages of career, the more involved agent will outperform the less involved one, primarily

---

[14]Other motives of the agents engaged in the interaction can also play an important role in explaining their behaviour. For instance, the role of emotions should not be underestimated

because intrinsic motivation plays the more important role in influencing agents' effort.

The paper adds a new perspective to the analysis of incentive provision in mission-driven organizations. Taking into account the pro-social component of (some) agents in such organizations influences the interpretation of the outcomes (agent's output), wage setting and the optimal composition of the agents' pool in the organization.

8 Appendix

Proof of Proposition 1

The first part of proposition is the straightforward consequence of the first order condition of the agent's optimization problem.

To prove the second part, notice that the right-hand side of (4) is a linear function of β with the coefficient $(1 - \delta\lambda)$. It is an increasing function if this coefficient is positive and a decreasing function when the coefficient is negative. The left-hand side of (4) is an increasing function of a_1.

Proof of Lemma 3

Proof. The claim of the lemma follows from a two-stage belief update. First, the belief on the distribution of β is updated by applying the standard Bayesian rule. Second, conditional on β taking values β_H (β_L), the belief on distribution of θ is updated by applying Lemma 1. □

Proof of Lemma 4

Proof. The derivatives are computed by the "brute force". From Lemma 3, $\widetilde{p} = \frac{p\gamma_H}{p\gamma_H + (1-p)\gamma_L}$, then

$$E^A\widetilde{p} = \iint\limits_{\Theta,\mathcal{E}} \widetilde{p}\big|_{y_1=\theta+a_1+\varepsilon_1} \, d\Omega(\theta,\varepsilon)$$

where $d\Omega(\theta,\varepsilon) = \varphi_\theta(\theta)\varphi_\varepsilon(\varepsilon)\,d\theta d\varepsilon$.

Let $g_j(a_1;\theta,\varepsilon) = f(a_1 + \theta + \varepsilon - a_1^j)$, and to simplify notation denote by $g_j' = \frac{\partial}{\partial a_1} g_j(a_1;\theta,\varepsilon)$ (a_1^j are constants).

Then

$$\frac{\partial}{\partial a_1}\left(E^A\widetilde{p}\right) = \iint\limits_{\Theta,\mathcal{E}} \frac{pg_H'(pg_H + (1-p)g_L) - pg_H(pg_H' + (1-p)g_L')}{(pg_H + (1-p)g_L)^2} \, d\Omega(\theta,\varepsilon)$$

After rearrangements this gives

$$\frac{\partial}{\partial a_1}\left(E^A\widetilde{p}\right) = p(1-p) \iint\limits_{\Theta,\mathcal{E}} \frac{g_H'g_L - g_Hg_L'}{(pg_H + (1-p)g_L)^2} \, d\Omega(\theta,\varepsilon)$$

26

To obtain k_H, substitute $a_1 = a_1^H$ (in equilibrium, the actually exerted effort a_1 by the β_H-agent will be equal to the belief on it, i.e., a_1^H) in the arguments of functions g_j. This gives $g_H\left(a_1^H\right) = f(\theta+\varepsilon)$, $g_H'\left(a_1^H\right) = f'(\theta+\varepsilon)$, $g_L\left(a_1^H\right) = f(\theta+\varepsilon+\Delta a_1)$, $g_L'\left(a_1^H\right) = f'(\theta+\varepsilon+\Delta a_1)$ and leads to

$$k_H = p(1-p) \iint\limits_{\Theta,\mathcal{E}} \frac{f(\theta+\varepsilon)f(\theta+\varepsilon+\Delta a_1)}{\left[pf(\theta+\varepsilon) + (1-p)f(\theta+\varepsilon+\Delta a_1)\right]^2} L(\theta,\varepsilon,\Delta a_1)\, d\Omega(\theta,\varepsilon)$$

where

$$L(\theta,\varepsilon,\Delta a_1) = \left[\frac{f'(\theta+\varepsilon)}{f(\theta+\varepsilon)} - \frac{f'(\theta+\varepsilon+\Delta a_1)}{f(\theta+\varepsilon+\Delta a_1)}\right]$$

Since function f is the density function for the normally distributed random variable $\theta+\varepsilon$ with mean $\overline{\theta}$ and variance $\sigma_1^2 + \omega^2$, for any x holds $\frac{f'(x)}{f(x)} = -\frac{x-\overline{\theta}}{\sigma_1^2+\omega^2}$. Therefore,

$$L(\theta,\varepsilon,\Delta a_1) = \frac{\Delta a_1}{\sigma_1^2 + \omega^2}$$

and L doesn't depend on θ and ε. This leads to the required expression for k_H.

The expression for k_L is obtained in the same way. Substitute $a_1 = a_1^L$ in the arguments of functions g_j, then $g_H\left(a_1^L\right) = f(\theta+\varepsilon-\Delta a_1)$, $g_H'\left(a_1^L\right) = f'(\theta+\varepsilon-\Delta a_1)$, $g_L\left(a_1^L\right) = f(\theta+\varepsilon)$, $g_L'\left(a_1^L\right) = f'(\theta+\varepsilon)$. The expression for k_L is obtained after similar substitution and rearrangements. $\square$

Proof of Lemma 5

Proof. Notice that $k_j > 0$ $(j = L, H)$ if and only if $\Delta a_1 > 0$, as follows immediately from Lemma 4.

Proceed now to prove the first two claims of the Lemma, i.e. that $\Delta a_1 > 0$ if and only if $\delta\lambda < 1$. To do so, we start with the case of p close to 0 or 1, and then generalize the result to all values of p.

According to proposition 2, the effort is determined by

$$C'(a_1^j) = \lambda\delta + (1-\lambda\delta)\beta_j + (1-\beta_j)\delta k_j(\Delta a_2 - \lambda\Delta a_1) \tag{14}$$

One can see from lemma 4 that $k_j \to 0$ as $p \to 0$ or $p \to 1$ $(j = L, H)$. Therefore, for p close to 0 or 1, the last term in (14) $(1-\beta_j)\delta k_j(\Delta a_2 - \Delta a_1) \to 0$, and then $a_1^H > a_1^L$, i.e., $\Delta a_1 > 0$ when $\lambda\delta < 1$, and $a_1^H < a_1^L$ when $\lambda\delta > 1$.

To generalize the result to all values of p, we prove that Δa_1 doesn't change its sign when p varies. Consider again (14) and subtract the formula for a_1^L from the one for a_1^H. This leads to

$$C'(a_1^H) - C'(a_1^L) + \delta\left(\lambda\Delta a_1 - \Delta a_2\right)\left[(1-\beta_L)(k_H - k_L) - \Delta\beta k_H\right] = \Delta\beta(1-\lambda\delta) \tag{15}$$

The right-hand side of (15) remains constant when p varies.

Consider the case of $\delta\lambda < 1$. Then the right-hand side of (15) is positive. The difference $\Delta a_1 = a_1^H - a_1^L$ is a continuous function of p, and it is already proved that $\Delta a_1 > 0$ for p close to 0 or 1. Assume that for some $p = p_0$ the difference Δa_1 switches sign, then it should be equal to zero for $p = p_0$. One can see from lemma 4 that if $\Delta a_1 = 0$, then $k_H = k_L = 0$. Then, the left-hand side of (15) is equal to zero for $p = p_0$, while the right-hand side is positive, which is contradictory. Therefore, the difference Δa_1 cannot switch sign for $\lambda\delta < 1$ and remains positive for all values of $p \in (0, 1)$.

The same argument applies to the case of $\lambda\delta > 1$.

The first two claims of the Lemma are now proved and we proceed to the the last claim on the comparison of the values of $|k_j|$.

To prove the second part, notice that the expressions for k_j in lemma 4 can be rewritten as (one should make a change of variable $\widetilde{\theta} = \theta - \Delta a_1$ in the expression for k_L and then denote the new variable by θ)

$$k_H = p(1-p)\frac{\Delta a_1}{\sigma_1^2 + \omega^2} \iint\limits_{\Theta,\mathcal{E}} \frac{G(\theta,\varepsilon)}{L^2(\theta,\varepsilon)}\varphi_\theta(\theta)\varphi_\varepsilon(\varepsilon)\,d\theta d\varepsilon$$

$$k_L = p(1-p)\frac{\Delta a_1}{\sigma_1^2 + \omega^2} \iint\limits_{\Theta,\mathcal{E}} \frac{G(\theta,\varepsilon)}{L^2(\theta,\varepsilon)}\varphi_\theta(\theta + \Delta a_1)\varphi_\varepsilon(\varepsilon)\,d\theta d\varepsilon$$

where

$$G(\theta,\varepsilon) = f(\theta + \varepsilon)f(\theta + \varepsilon + \Delta a_1)$$
$$L(\theta,\varepsilon) = pf(\theta + \varepsilon) + (1-p)f(\theta + \varepsilon + \Delta a_1)$$

Then, the difference $k_H - k_L$ is

$$k_H - k_L = p(1-p)\frac{\Delta a_1}{\sigma_1^2 + \omega^2} \iint\limits_{\Theta,\mathcal{E}} \frac{G(\theta,\varepsilon)}{L^2(\theta,\varepsilon)}K(\theta,\varepsilon)\varphi_\varepsilon(\varepsilon)\,d\theta d\varepsilon \qquad (16)$$

where

$$K(\theta,\varepsilon) = \varphi_\theta(\theta) - \varphi_\theta(\theta + \Delta a_1)$$

Clearly, the sign of the difference $k_H - k_L$ in (16) is determined by the sign of the double integral, which, in turn, is determined by the sign of the integral over the variable θ in the following expression

$$\iint\limits_{\Theta,\mathcal{E}} \frac{G(\theta,\varepsilon)}{L^2(\theta,\varepsilon)}K(\theta,\varepsilon)\varphi_\varepsilon(\varepsilon)\,d\theta d\varepsilon = \int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{\infty} \frac{G(\theta,\varepsilon)}{L^2(\theta,\varepsilon)}K(\theta,\varepsilon)\,d\theta\right]\varphi_\varepsilon(\varepsilon)\,d\varepsilon$$

We will now concentrate on the one-variable integral $\int\limits_{-\infty}^{\infty} \frac{G(\theta,\varepsilon)}{L^2(\theta,\varepsilon)}K(\theta,\varepsilon)\,d\theta$.

To simplify notation, let us drop the variable ε from the arguments of the functions G, L, K (they are still considered as functions of two variables, (θ, ε)).
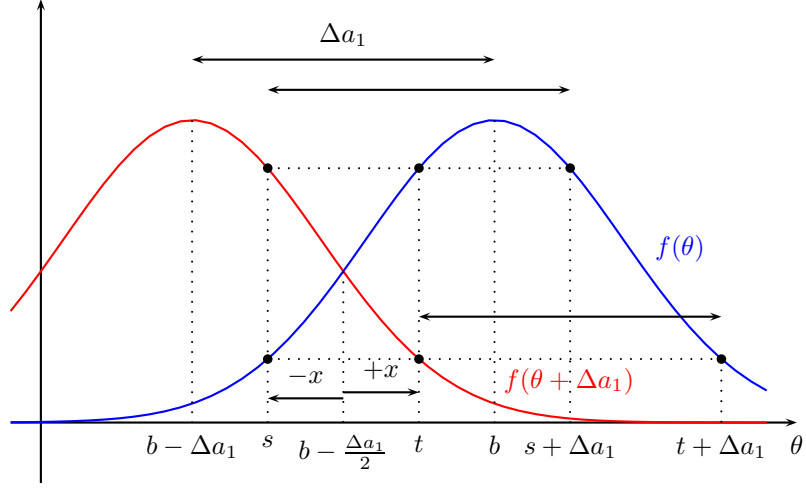
Figure 4: Assessing the integral in (16)

Consider the case of $\Delta a_1 > 0$.

Clearly, $G(\theta) > 0$, $L(\theta) > 0$ for all θ, and $K(\theta) \geq 0$ for $\theta \geq \overline{\theta} - \frac{\Delta a_1}{2}$, $K(\theta) \leq 0$ otherwise.

The idea of assessing this integral is illustrated by Figure 8.

Consider $s = \overline{\theta} - \frac{\Delta a_1}{2} - x$ and $t = \overline{\theta} - \frac{\Delta a_1}{2} + x$ for all $x \geq 0$. Notice that

$$f(t) = f(s + \Delta a_1), \quad f(t + \Delta a_1) = f(s)$$
$$f(t) \geq f(t + \Delta a_1), \quad f(s) \geq f(s + \Delta a_1)$$
$$f'(t) = -f'(s + \Delta a_1), \quad f'(t + \Delta a_1) = -f'(s)$$

Therefore,
$$G(s) = G(t), K(s) = -K(t)$$

Hence, the integrand's value at points t and s have absolute value $G(t)K(t)$, weighted by $\frac{1}{L^2(t)}$ and $\frac{1}{L^2(s)}$ respectively, and have different signs: $K(t) \geq 0$, $K(s) \leq 0$. The sign of the integral is therefore determined by the comparison of the weights. One can see that $L(t) \geq L(s)$ for $p \geq 1/2$, therefore the negative side of the integrand, corresponding to $\theta < \overline{\theta} - \frac{\Delta a_1}{2}$, i.e. $\theta = s$ "outweighs" its positive part, corresponding to $\theta > \overline{\theta} - \frac{\Delta a_1}{2}$, i.e. $\theta = t$, and therefore $k_H - k_L < 0$. Similarly, one can obtain that $k_H - k_L > 0$ for $p > 1/2$.

For the formal proof, consider the integral in (16) and split it into two integrals: $I = I_1 + I_2$, where

29

$$I_1 = \int\limits_{-\infty}^{\overline{\theta} - \frac{\Delta a_1}{2}} \frac{G(\theta)}{L^2(\theta)} K(\theta)\, d\theta, \quad I_2 = \int\limits_{\overline{\theta} - \frac{\Delta a_1}{2}}^{\infty} \frac{G(\theta)}{L^2(\theta)} K(\theta)\, d\theta$$

Make a change of variable in both integrals: $s = \left(\overline{\theta} - \frac{\Delta a_1}{2}\right) - \theta$ in I_1, and $t = \theta - \left(\overline{\theta} - \frac{\Delta a_1}{2}\right)$ in I_2. Then both internal integrals will have limits 0 and ∞. By using the symmetry of f, $f(\overline{\theta} + x) = f(\overline{\theta} - x)$ for all x, rewrite

$$f\left(\overline{\theta} - \frac{\Delta a_1}{2} - s\right) = f\left(\overline{\theta} + \frac{\Delta a_1}{2} + s\right), \; f\left(\overline{\theta} + \frac{\Delta a_1}{2} - s\right) = f\left(\overline{\theta} - \frac{\Delta a_1}{2} + s\right)$$

in the integral I_1, and then rename the integration variable to t. After that we obtain

$$I_1 + I_2 = \int\limits_{0}^{\infty} G(t)K(t)\left[\frac{1}{L_2^2(t)} - \frac{1}{L_1^2(t)}\right] dt$$

where

$$L_1(t) = pf\left(t + \overline{\theta} + \frac{\Delta a_1}{2}\right) + (1-p)f\left(t + \overline{\theta} - \frac{\Delta a_1}{2}\right)$$

$$L_2(t) = pf\left(t + \overline{\theta} - \frac{\Delta a_1}{2}\right) + (1-p)f\left(t + \overline{\theta} + \frac{\Delta a_1}{2}\right)$$

Since for all $t > 0$ holds $f\left(\overline{\theta} - \frac{\Delta a_1}{2} + t\right) \geq f\left(\overline{\theta} + \frac{\Delta a_1}{2} + t\right)$, then $L_2(t) \geq L_1(t)$ for $p \geq 1/2$, and therefore $I_1 + I_2 \leq 0$, which means that $k_H \leq k_L$. For $p \leq 1/2$ one obtains $L_2(t) \leq L_1(t)$, and, therefore, $k_H \geq k_L$.

The case of $\Delta a_1 < 0$ is considered in the same way. One should consider $|k_H| - |k_L|$ instead of $k_H - k_L$, and (16) will be rewritten as

$$|k_H| - |k_L| = p(1-p)\frac{|\Delta a_1|}{\sigma_1^2 + \omega^2} \iint\limits_{\Theta, \mathcal{E}} \frac{G(\theta, \varepsilon)}{L^2(\theta, \varepsilon)} K(\theta, \varepsilon)\varphi_\varepsilon(\varepsilon)\, d\theta d\varepsilon$$

The two other differences are that $K(t) < 0$ and that $f\left(\overline{\theta} - \frac{\Delta a_1}{2} + t\right) \leq f\left(\overline{\theta} + \frac{\Delta a_1}{2} + t\right)$, leading to a different comparison between $L_2(t)$ and $L_1(t)$: $L_2(t) \leq L_1(t)$ for $p \geq 1/2$, and $L_2(t) \geq L_1(t)$ for $p \leq 1/2$. $\qquad \square$

Proof of Lemma 6

Proof. Compare $(1 - \beta_L)|k_L|$ and $(1 - \beta_H)|k_H|$. Clearly, $(1 - \beta_L)|k_L| - (1 - \beta_H)|k_H|$ is a continuous function of p. It is also clear that $1 - \beta_L > 1 - \beta_H$. According to Lemma 5, $k_L > k_H$ for $p > \frac{1}{2}$, then

$$(1 - \beta_L)|k_L| > (1 - \beta_H)|k_H| \tag{17}$$

always hold for $p > \frac{1}{2}$. The inequality 17 also holds for $p = \frac{1}{2}$, since in this case $k_L = k_H$. This means that the inequality will continue to hold at least for some values $p < \frac{1}{2}$. This finishes the proof.

$\square$

Proof of Lemma 7

Proof. The first part of the Lemma on the belief update in period $t + 1$ based on the period-t belief and output is simply a restatement of lemma 1 for the two-period case.

The second (retrospective) part of the lemma is obtained from the cyclical substitution of the expressions for σ_{t+1} and λ_{t+1}. Indeed, by using the fact that $\frac{1}{\sigma_{t+1}^2} = \frac{1}{\omega^2} + \frac{1}{\sigma_t^2}$, one can obtain: $\frac{1}{\sigma_2^2} = \frac{1}{\omega^2} + \frac{1}{\sigma_1^2}$, $\frac{1}{\sigma_3^2} = \frac{1}{\omega^2} + \frac{1}{\sigma_3^2} = \frac{2}{\omega^2} + \frac{1}{\sigma_2^2}$ etc., leading to the required formula $\frac{1}{\sigma_{t+1}^2} = \frac{t}{\omega^2} + \frac{1}{\sigma_1^2}$.

To obtain the formula for λ, notice that $\lambda_t = \frac{\sigma_t^2}{\sigma_t^2 + \omega^2}$ can be rewritten as $\frac{1}{\lambda_t} = 1 + \frac{\omega^2}{\sigma_t^2}$, and the "accumulated" expression $\frac{1}{\lambda_t} = t + \frac{\omega^2}{\sigma_1^2}$ is obtained in the same way as the expression for σ_{t+1}^2.

Finally, by substituting cyclically the expressions for $\overline{\theta}_{t+1}$, one can express $\overline{\theta}_{t+s}$ through $\overline{\theta}_{t+s-1}$, then through $\overline{\theta}_{t+s-2}$ etc.

We will prove it by applying the principal of mathematical induction.

Indeed, the formula holds for $s = 1$.

Let us assume that the formula holds for s: $\overline{\theta}_{t+s} = \sum\limits_{q=0}^{s-1} \Lambda_{t+q,t+s}(y_{t+q} - a_{t+q}^\mu) + \prod\limits_{q=0}^{s-1}(1 - \lambda_{t+q}) \cdot \overline{\theta}_t$, then one can obtain the expression for $s + 1$:

$$\overline{\theta}_{t+s+1} = \lambda_{t+s}(y_{t+s} - a_{t+s}^\mu) + (1 - \lambda_{t+s})\overline{\theta}_{t+s} =$$

$$= \lambda_{t+s}(y_{t+s} - a_{t+s}^\mu) + (1 - \lambda_{t+s})\left[\sum\limits_{q=0}^{s-1} \Lambda_{t+q,t+s}(y_{t+q} - a_{t+q}^\mu) + \prod\limits_{q=0}^{s-1}(1 - \lambda_{t+q}) \cdot \overline{\theta}_t\right] =$$

$$= \underbrace{\lambda_{t+s}}_{\Lambda_{t+s,t+s+1}}(y_{t+s} - a_{t+s}^\mu) + \sum\limits_{q=0}^{s-1}\underbrace{(1 - \lambda_{t+s})\Lambda_{t+q,t+s}}_{\Lambda_{t+q,t+s+1}}(y_{t+q} - a_{t+q}^\mu) + \prod\limits_{q=0}^{s}(1 - \lambda_{t+q}) \cdot \overline{\theta}_t$$

which brings the required expression after changing the indexing: .

$\square$

Proof of Lemma 8

Proof. The agent's optimization in period t is

$$\max_{a_t} \beta a_t - C(a_t) + (1 - \beta) \sum\limits_{s=t+1}^{T} \delta^{s-t} E_t w_s + \{const\ on\ a_t\}$$

where the term $\{const\ on\ a_t\}$ includes all the benefits and costs associated with effort choice in the periods $t+1, \ldots, T$.

Clearly, $E_t w_s = E_t[\overline{\theta}_s + a_s] = \Lambda_{ts} E_t y_t + \{const\ on\ a_t\}$, and a_t enters additively into y_t. Therefore, $\frac{\partial}{\partial a_t}[E_t w_s] = \Lambda_{ts}$. The first order condition for the optimization problem brings then the required expression. □

Proof of Lemma 9

Proof. According to lemma 8,

$$C'(a_t) = \beta + (1-\beta)\Gamma_t \tag{18}$$

where

$$\Gamma_t = \sum_{s=1}^{T-t} \delta^s \Lambda_{t,t+s} \tag{19}$$

We first prove that $\{\Gamma_t\}$ is a decreasing sequence.

First, notice that $\{\lambda_t\}$ is a decreasing sequence. This follows immediately from the expression for λ_t obtained in lemma 7:

$$\frac{1}{\lambda_t} = t + \frac{\omega^2}{\sigma_1^2} \tag{20}$$

Second, it holds that $\lambda_t(1 - \lambda_{t+1}) > \lambda_{t+1}(1 - \lambda_{t+s})$ for all $s > 0$. Indeed, substituting (20) gives $\frac{1}{t+\mu}(1 - \frac{1}{t+1+\mu}) > \frac{1}{t+1+\mu}(1 - \frac{1}{t+s+\mu})$ (here $\mu = \frac{\omega^2}{\sigma_1^2}$), which simplifies to $1 > 1 - \frac{1}{t+s+1+\mu}$, which clearly holds.

Now compare Γ_t and Γ_{t+1} by substituting expressions for Λ_{tq} obtained in lemma 7 into (19):

$$\Gamma_t = \delta\lambda_t \quad +\delta^2\lambda_t(1-\lambda_{t+1}) \quad +\delta^3\lambda_t(1-\lambda_{t+1})(1-\lambda_{t+2}) \quad +\ldots$$
$$\Gamma_{t+1} = \delta\lambda_{t+1} \quad +\delta^2\lambda_{t+1}(1-\lambda_{t+2}) \quad +\delta^3\lambda_{t+1}(1-\lambda_{t+2})(1-\lambda_{t+3}) \quad +\ldots$$

Both sums have terms with $\delta, \delta^2, \cdots, \delta^{T-t-1}$, and the sum for Γ_t has the term with δ^{T-t}. The terms with δ^s in each sum are: for Γ_t: $\delta^s\lambda_t(1-\lambda_{t+1})\cdots(1-\lambda_{t+s-1})$ and for Γ_{t+1}: $\delta^s\lambda_{t+1}(1-\lambda_{t+2})\cdots(1-\lambda_{t+s})$. The number of multipliers is the same in both expressions, and there are only two different multipliers: in the expression for Γ_t these are $\lambda_t(1-\lambda_{t+1})$, and those for Γ_{t+1} are $\lambda_{t+1}(1-\lambda_{t+s})$. As shown above, the first expression is greater. Therefore, every term in the sum for Γ_t is greater than the corresponding term in the sum for Γ_{t+1}, and the former has one extra term with δ^{T-t} (which is positive). Consequently, $\Gamma_t > \Gamma_{t+1}$.

The argument can be applied cyclically, therefore $\Gamma_t > \Gamma_{t+1} > \Gamma_{t+2} > \ldots$.

Let us now prove the first claim of the lemma. Let $\beta_L < \beta_H$, and $a_\tau(\beta_H) < a_\tau(\beta_L)$, therefore $\beta_H + (1-\beta_H)\Gamma_\tau < \beta_L + (1-\beta_L)\Gamma_\tau$ (these are the right-hand sides in (18)), which is equivalent to $\Gamma_\tau + (1-\Gamma_\tau)\beta_H < \Gamma_\tau + (1-\Gamma_\tau)\beta_L$. One can see that $a_\tau(\beta_H) < a_\tau(\beta_H)$ if and only if $\Gamma_\tau > 1$.

Since $\Gamma_\tau > 1$, for all $t < \tau$ also holds $\Gamma_t > 1$, since the sequence $\{\Gamma_t\}$ is a decreasing sequence. Consequently, the inequality $a_\tau(\beta_H) < a_\tau(\beta_L)$ will hold for all $t < \tau$.

The second claim of the lemma is proved by applying the same argument. In this case $\Gamma_\tau < 1$, therefore, $\Gamma_t > 1$ for $t < \tau$.

$\square$

Proof of Lemma 10

Proof. The belief update on the distribution of β is the result of applying the Bayesian rule. Conditional on β, the distribution of θ is updated in the same way as for the two-period case (see lemma 1). $\square$

Proof of Lemma 11

Proof. The formulas for $k_{t,t+1}^j$ are obtained by differentiation of the formula for p_{t+1} given in lemma 10. The computations repeat those from the proof of lemma 4.

To obtain the recursive formula for k_{t+s+1}^j, we substitute the formula for update for p from lemma 10 into the definition of $k_{t,t+s+1}^j$ in (13):

$$k_{t,t+s+1}^j = \int\limits_{\mathcal{E}_t,\dots,\mathcal{E}_{t+s},\Theta} \cdots \int \frac{\partial}{\partial a_t} \frac{p_{t+s}\gamma_{t+s}^H}{p_{t+s}\gamma_{t+s}^H + (1-p_{t+s})\gamma_{t+s}^L} \, d\Omega(\varepsilon_t,\dots,\varepsilon_{t+s},\theta)$$

In the integrand, p_{t+s} and γ_{t+s}^j are functions of a_t. After differentiating and simplifying the integrand, one obtains

$$\frac{\partial}{\partial a_t} \frac{p_{t+s}\gamma_{t+s}^H}{p_{t+s}\gamma_{t+s}^H + (1-p_{t+s})\gamma_{t+s}^L} =$$

$$= \frac{1}{L^2}\left[p_{t+s}(1-p_{t+s})\left(\gamma_{t+s}^L \frac{\partial\gamma_{t+s}^H}{\partial a_t} - \gamma_{t+s}^H \frac{\partial\gamma_{t+s}^L}{\partial a_t} \right) + \frac{\partial p_{t+s}}{\partial a_t}\gamma_{t+s}^H\gamma_{t+s}^L \right]$$

where $L = p_{t+s}\gamma_{t+s}^H + (1-p_{t+s})\gamma_{t+s}^L$.

By taking into account that $\gamma_{t+s}^j = f_{t+s}^j\left(y_{t+s} - a_{t+s}^j \right)$, where f_{t+s}^j is the density function of the normal distribution $\mathcal{N}\left(\overline{\theta}_{t+s}^j, \sigma_t^2 + \omega^2 \right)$, one obtains

$$\frac{\partial\gamma_{t+s}^j}{\partial a_t} = \gamma_{t+s}^j \frac{y_{t+s} - a_{t+s}^j - \overline{\theta}_{t+s}^j}{\sigma_t^2 + \omega^2} \frac{\partial\overline{\theta}_{t+s}^j}{\partial a_t}$$

It follows from lemma 7 that $\frac{\partial\overline{\theta}_{t+s}^j}{\partial a_t} = \Lambda_{t,t+s}$.

Substituting these expressions into the first term of the integrand, $p_{t+s}(1 - p_{t+s})\left(\gamma_{t+s}^L \frac{\partial\gamma_{t+s}^H}{\partial a_t} - \gamma_{t+s}^H \frac{\partial\gamma_{t+s}^L}{\partial a_t} \right)$, one obtains the term (k^θ) in the lemma.

The term (k^p) is simply the result of integration of the second term of the integrand.

$\square$

Proof of Proposition 3

Proof. The agent's optimization problem at period t is

$$\max_{a_t} \beta a_T - C(a_t) + (1-\beta) \sum_{s=t+1}^{T} \delta^{s-t} E_t w_s + \{const\ on\ a_t\}$$

The term $\{const\ on\ a_t\}$ includes all benefits and costs associated with effort choice in periods $s = t+1, \ldots, T$.

One can see that

$$
\begin{aligned}
E_t w_s &= E_t \left[p_s \left(\overline{\theta}_s^H + a_s^H \right) + (1-p_s) \left(\overline{\theta}_s^L + a_s^L \right) \right] = \\
&= E_t \left[p_s \Lambda_{ts} \left(y_t - a_t^H \right) + (1-p_s)\Lambda_{ts} \left(y_t - a_t^L \right) + p_s a_s^H + (1-p_s)a_s^L \right] = \\
&= E_t \left[\Lambda_{ts} y_t - \Lambda_{ts} \left(p_{ts} a_t^H + (1-p_s)a_t^L \right) + p_s a_s^H + (1-p_s)a_s^L \right] = \\
&= \Lambda_{ts} E_t y_t - \Lambda_{ts} E_t[a_t^L + p_s \Delta a_t] + a_s^L + E_t p_s \Delta a_s
\end{aligned}
$$

Substituting this expression in the optimization problem and taking the derivative with respect to a_t, we obtain the first order condition for the agent's optimization:

$$C'(a_t) = \beta + (1-\beta) \sum_{s=t+1}^{T} \delta^{s-t} \left(\Lambda_{ts} - \Lambda_{ts} \frac{\partial(E_t p_s)}{\partial a_t} \Delta a_t + \frac{\partial(E_t p_s)}{\partial a_t} \Delta a_s \right)$$

Substituting now $k_{ts} = \frac{\partial(E_t p_s)}{\partial a_t}$ and $\Gamma_t = \sum_{s=t+1}^{T} \delta^{s-t} \Lambda_{ts}$ and changing the summation index to $s' = s - t$ gives the necessary expression. $\qquad\square$

References

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

Besley, Timothy and Maitreesh Ghatak, "Competition and incentives with motivated agents," *American Economic Review*, 2005, *95* (3), 616–636.

Casas-Arce, Pablo, "Career Tournaments," *Journal of Economics & Management Strategy*, 2010, *19* (3), 667–698.

Delfgaauw, Josse, "Dedicated Doctors: Public and Private Provision of Health Care with Altruistic Physicians," *Tinbergen Institute Discussion Papers*, 2007.

_ **and Robert Dur**, "Incentives and Workers Motivation in the Public Sector," *The Economic Journal*, 2008, *118* (525), 171–191.

Dewatripont, Mathias, Ian Jewitt, and Jean Tirole, "The economics of career concerns, part I: Comparing information structures," *Review of Economic Studies*, 1999, *66* (1), 183–198.

Ellingsen, Tore and Magnus Johannesson, "Pride and Prejudice: The Human Side of Incentive Theory," *American Economic Review*, 2008, *98* (3), 990–1008.

Falk, Armin and Michael Kosfeld, "The hidden costs of control," *American Economic Review*, 2006, *96* (5), 1611–1630.

Fehr, Ernst and Bettina Rockenbach, "Detrimental effects of sanctions on human altruism," *Nature*, 2003, *422*, 137–140.

Francois, Patrick, "Public service motivation'as an argument for government provision," *Journal of Public Economics*, 2000, *78* (3), 275–299.

— **and Michael Vlassopoulos**, "Pro-social motivation and the delivery of social services," *CESifo Economic Studies*, 2008, *54* (1), 22.

Heckman, James J., Jeffrey A. Smith, and Christopher Taber, "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program," *NBER Working Papers No.5535*, 1996.

Holmström, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 1999, *66* (1), 169–182.

— **and Paul Milgrom**, "Multitask principal–agent analyses: incentive contracts, asset ownership, and job design," *Journal of Law, Economics, and organization*, 1991, *7* (special issue), 24–52.

Kőszegi, Botond and Wei Li, "Drive and talent," *Journal of the European Economic Association*, 2008, *6* (1), 210–236.

Penczynski, Stefan P., "Genius or Enthusiast? Additional Heterogeneity in a Model of Career Concerns," 2007.

Rose-Ackerman, Susan, "Altruism, nonprofits, and economic theory," *Journal of Economic Literature*, 1996, *34* (2), 701–728.

Shchetinin, Oleg, "Contracting Under Reciprocal Altruism," *University of Gothenburg Working Papers in Economics*, 2009, *421*.

— , "Altruism and Career Concerns," *University of Gothenburg Working Papers in Economics*, 2010, *427*.

Wilson, James Q., *Bureaucracy: What government agencies do and why they do it*, Basic Books (New York), 1989.