# Bonuses and Managerial Misbehavior

Caspar Siegert

March 31, 2010

**Abstract**

Repeatedly, bonuses and variable compensation of managers have been criticized for encouraging non-compliance and thereby ultimately harming a firm's interest. This argument has in particular been made in the wake of the recent financial crisis, claiming that banker's bonus plans induced excessive risk taking. Yet, this reasoning does not explain why companies found it optimal to offer fairly high-powered incentives in the first place. We show that large bonuses may in fact prevent agents from "gaming" incentive schemes and are insofar an optimal response to the threat of non-compliance. This finding has strong policy implications and sheds a new light on the recent proposals to regulate bonuses in the banking industry. Our model has also implications for optimal anti-trust policy and shows that corporate fines may in certain situations be preferable to personal liability.

JEL-Classification: D-82, D-86, K-14, K-42
Keywords: contracts, incentives, governance

# 1 Introduction

In the wake of the recent financial crisis, excessive bonuses for bankers were frequently blamed as a source of irresponsible behavior: In order to be awarded those tempting rewards, bankers were supposedly prepared to engage in behavior that was not in the interest of banks themselves: In particular, bankers had incentives to take excessive risks, since they could reap the benefits in the case of success and were protected by limited liability in case of failure (e.g. Bebchuk, 2009).

Yet, a question that is rarely asked is why banks found it optimal to offer those contracts in the first place, if they gave incentives to misbehave. We show that comparatively high-powered incentives are in many cases not only robust to potential undesirable behavior, but the possibility of non-compliance may even increase the optimal bonus an agent is offered. This is consistent with the puzzling observation that average pay-performance sensitivities for CEOs in compliance-intensive industries such as Financial Services or Natural Resources do not seem to be noticeably smaller than in industries where compliance is less of an issue, such as Utilities or Industrial Companies (Conyon & Murphy, 2000; Murphy, 1999; Zhou, 2000). In a study of incentives in the banking sector, John and Qian (2003) find that when controlling for leverage,[1] banks offer incentives that are not significantly different from the ones given in other industries.[2] Furthermore, our model is in line with the observation by Fahlenbach and Stulz (2009) that the size of cash bonuses a bank paid was not negatively correlated with performance during the recent financial crisis.

We look at a standard moral-hazard model where the agent is risk-neutral but protected by limited liability and enlarge the agent's action space by assuming that he can not only choose how much effort to exert, but also whether (and how much) to engage in undesirable behavior. This unwanted behavior (or non-compliance) increases the contractable profit signal and hence the agents variable compensation, but it is nevertheless against the principal's interest.[3] If the agent carries out an undesirable action, the principal will find out about this non-compliance with positive probability and will be able to punish the manager. In our analysis we will compare the optimal incentive the principal

[1] John and John (1993) offer a theoretical explanation why pay-performance ratios should depend on the debt ratio of a company.

[2] Our argumentation does not dispute the general assessment that managerial incentives are surprisingly low-powered on average. However, we would expect incentives to be lower in compliance-intensive industries than in other sectors, whereas in reality this does not seem to be the case.

[3] A model where principal and agent unanimously pursue some (socially) undesirable action is analyzed by Garoupa (2000).

chooses in this setting to a standard single-tasking problem where the level of compliance is exogenously given. This allows us to examine how incentives *change* if the principal wants to induce the agent to comply.

We find that in those cases where the agent's effort is not very important, the principal will react to the possibility of undesirable behavior by lowering incentives: Marginally lowering the bonus has little effect on the effort choice of the agent. But at the same time it reduces the gains from non-compliance and therefore the level of undesirable behavior the agent is going to choose. However, in those cases where effort is sufficiently important, the opposite is true: The higher the incentives for effort, the higher the bonus the agent is going to loose out on in case misbehavior is discovered (in which case he receives zero wage payments). Hence, with very high incentives he will be less prepared to jeopardize these wage payments by taking undesirable actions and will be more likely to comply. For this reason the principal will set higher incentives than he would do in the single-tasking case. In both cases, the principal may find it optimal to complement bonuses with fixed wages that are paid regardless of the firm's profit as long as no evidence on misbehavior surfaces in order to enhance compliance. Our model takes the somewhat extreme view that all rents the agent receives are given to ensure incentive compatibility. They are neither due to collusive practices in the determination of pay (Bebchuk and Fried, 2006), nor do they reflect scarcity of prospective employees (Gabaix and Landier, 2008). While this view is unlikely to fully describe the reality of executive compensation, it gives nevertheless some interesting insights. In particular, fixed wages can not only be used in order to satisfy participation constraints, but they may also be an additional disciplining device, comparable to efficiency wages (Shapiro and Stiglitz, 1984).

Our results imply that for top-management positions high incentives may in fact be a method to induce compliance, whereas lower ranks in a firm's hierarchy will receive very performance-inelastic pay in order to achieve the same goal. The model also explains why non-compliance seldom originates from the very top levels of a company's hierarchy where incentives are typically extremely high already due to the importance of effort. The analysis does not only apply to the remuneration of executives, but can equally explain the pay of portfolio managers, traders etc., who typically receive very high-powered incentives and are also able to engage in undesirable behavior, e.g. excessive risk taking that only becomes evident in adverse states of the world. Our model brings us one step closer towards understanding the surprisingly high powered incentives in these jobs.[4]

[4] For example, in 2007 the average Wall Street bonus amounted to $180,420 (http://www.washingtonpost.com/wp-dyn/content/story/2008/01/29/ST2008012900465.html).

The model contribute to the current discussion on imposing legal caps on bonuses by showing that such legal limits on incentive-pay may be counterproductive: If caps are large (yet binding), they may increase misbehavior by depriving principals of a cost-effective mechanism to foster compliance. At the same time, highly restrictive caps on bonuses may be effective at reducing misbehavior, but will clearly destroy incentives for managers to exert effort.

Undesirable effort plays not only a role in the banking industry, but is always an issue if agents have the possibility to game incentive schemes by engaging is actions that are harmful to the company. Our model is sufficiently general to apply to a variety of situations and is in no way confined to incentive problems in the financial industry. Examples for undesirable behavior in other industries may in particular include illegal actions such as setting up a cartel or paying bribes in order to be awarded a contract: While sufficiently high expected fines guarantee that such actions are typically not in the principal's interest, a manager may nevertheless be tempted to engage in such behavior in order to be awarded a bonus. Another example brought forward by Fischer and Huddart (2008) is permissive behavior by auditing firms in order to secure consulting deals from the same customer: Such conduct will increase the auditor's personal remuneration but will nevertheless most likely be against the company's interest.[5]

If the unwanted behavior creates negative externalities that affect others than the two contracting parties, a welfare-maximizing policymaker may want to step in and discourage such behavior by the threat of punishing one of the parties if evidence on undesirable actions emerges. Using our framework we can make some statements about whom a policymaker should punish in order to efficiently discourage undesirable (respectively illegal) behavior: Even if the policymaker faces less strict limited-liability constraints vis-a-vis the agent than the principal does,[6] he may nevertheless choose to punish the principal rather than the agent for undesirable actions. As long as effort is sufficiently important, this will induce the principal to increase incentives, which is socially desirable for two reasons: First of all, it reduces undesirable behavior. But at the same time it

[5]Our only key assumption is that the agent doesn't fully internalize the negative consequences of his actions. This is due to two reasons: a) The agent's limited liability and b) The imperfect observability of undesirable actions. Imperfect observability may arise because often, the negative consequences of undesirable effort only emerge in the distant future. Alternatively, it may be impossible to condition the agent's remuneration on certain outcomes, e.g. a drop in a firm's reputation, due to verifiability constraints. In this case the principal can only impose punishments in case additional evidence on misbehavior is found.

[6]In particular, the policy maker has the possibility to impose jail sentences, which a private company is in general unable to do. Alternatively, it may be easier for a policy maker to impose fines that exceed current wages (but not current wealth) in certain constituencies.

increases (valuable) effort which reduces the distortions caused by the non-contractibility of effort.

In looking at a two-dimensional moral hazard model, our work is clearly related to the multi-tasking literature initiated by Holmström and Milgrom (1991). Yet, by explicitly incorporating limited liability constraints that are an important issue when punishing managers for misconduct we reach quite different conclusions:[7] While in the traditional multi-tasking model the introduction of additional tasks typically reduces optimal incentives, in our setting the opposite can be true: Incentive problems in the second dimension are mitigated by *increasing* incentives in the first dimension.

The idea that monetary incentives may trigger undesirable behavior is not new and has for example been studied in the empirical literature on earnings management (Healey, 1985; Holthausen et al., 1995 ) or in the context of the optimal scope of a firm by Fischer and Huddart (2008). However, this literature does typically not derive the optimal structure of incentives. Spagnolo (2000, 2005) has looked at the question if incentive contracts can make collusion harder to sustain while abstracting from the effect a given incentive has on the agent's choice of effort. The most closely related work is by Inderst and Ottaviani (2009) who look at optimal contracts if sales agents must be induced to search for potential customers, but not to sell to unsuitable customers. When a sales agent has found an unsuitable customer he can *only* earn a bonus by misbehaving. This implies that unlike in our setting higher incentives will never increase compliance.

Our result that incentives should be either very high or very low if compliance is an issue resembles the predictions of behavioral models where extrinsic incentives crowd out intrinsic motivation (Gneezy and Rustichini 2000a, 2000b).

# 2 The model:

A risk-neutral principal employs a risk-neutral but wealth constrained agent. The agent's wealth is initially 0 and has to be non-negative in all states of the world. The project can make either high profits which are denoted $\overline{\pi}$ or low profits $\underline{\pi}$ with $\overline{\pi} > \underline{\pi}$. The agent's effort determines the probability $q$ that high profits arise, low profits occur hence with a probability of $1 - q$. The agent's effort cost for working on the project is denoted by $C(q)$.

Furthermore the agent has the possibility to increase the probability of high profits

[7] For a discussion of the problems involved in imposing large punishments on managers, even in cases of outright fraud, see Bebchuk et al., 2006.

by $d \in [0, \overline{d}]$ with $\overline{d} < 1$ at a private cost of $K(d)$ if he engages in actions that are not in the interest of the principal. The notion that this action is "undesirable" can be operationalized by assuming that it imposes non-verifiable costs of $\tau(d)$ on the principal, where $\tau'(d) > \overline{\pi} - \underline{\pi} \quad \forall d$. That is, the cost of undesirable effort outweighs the increase in the likelihood of high profits from the principal's point of view. With probability $p(d)$ the principal gains hard information that the agent has been engaging in undesirable behavior and can punish him by paying him any wage that does not violate the limited-liability constraint. We assume that $p(d)$ can take values in the interval $[0, \overline{p}]$ where $\overline{p} < 1$.

In what follows we will assume that

$$
\begin{aligned}
&i) \quad C(0) = 0, C(q)' > 0, C(q)'' > 0, \lim_{q \to (1-\overline{d})} C'(q) = \infty \\
&ii) \quad K(0) = 0, K'(d) > 0, K''(d) > 0 \\
&iii) \quad p'(d) > 0, p''(d) \geq 0
\end{aligned}
$$

Assumption $i)$ is standard and says that effort cost is an increasing and convex function of $q$. Furthermore, the last part of i) makes sure that probability of high profits stays bounded away from 1 even if the agent exerts maximal effort and engages in as much undesirable behavior as possible. We assume that the agent incurs no costs if he engages neither in desirable, nor in undesirable actions. Assumption $ii)$ says that the marginal cost of undesirable behavior is increasing and convex in $d$; iii) ensures the same for the implicit cost of undesirable behavior, i.e. the risk of being caught.

## 2.1 The agents decision problem

The contract the agent can be offered is given by the tuple $(w_f, w, w_{f,d}, w_d)$ where the agent receives a base salary $w_f$ regardless of profits whenever no evidence on non-compliance is found. In case of high profits and no detected misbehavior he gets an additional bonus of $w$. In the event of detected misbehavior he receives a base salary of $w_{f,d}$ and an additional bonus $w_d$ if high profits occur. For most of our analysis we will assume that all payments except for $w$ are zero. In section 2.2.1 we will show that it is indeed always optimal to set $w_{f,d} = w_d = 0$. Our assumption that $w_f = 0$ on the other hand mainly allows us to concentrate on the main effects and will be relaxed in section 2.2.2.

The agent's utility given a contract $(0, w, 0, 0)$ is given by

$$U(w) = (q + d)\Big(\big(1 - p(d)\big)w\Big) - C(q) - K(d)$$

After having been offered a bonus $w$ by the principal, the agent will simultaneously decide how much desirable and undesirable effort to exert with the corresponding first order conditions being

$$\frac{\partial U}{\partial q} = w(1 - p) - C'(q) = 0 \tag{1}$$

$$\frac{\partial U}{\partial d} = w(1 - p) - K'(d) - p'w(q + d) = 0 \tag{2}$$

where for notational simplicity we no longer stress that $p = p(d)$. We will assume that the Hessian matrix of the optimization problem is negative definite, so the first order conditions uniquely characterize any interior optimum.[8]

Any interior optimum is hence implicitly defined by

$$F \equiv w(1 - p) - K'(d) - p'w\Big(G\big(w(1 - p)\big) + d\Big) = 0 \tag{3}$$

with $G(\cdot\,) = C'^{-1}(\cdot\,)$. Note that we have assumed the effort cost to be additively separable in the two dimensions and there are hence no technological complementarities between the two dimensions. Nevertheless, we see that the two dimensions are strongly intertwined: On the one hand, a higher level of non-compliance will increase $p$ and thereby reduce incentives for effort as can bee seen in equation (1). On the other hand, the bonus $w$ will not only affect desirable effort, but also the level of compliance as determined by (2): Given a detection probability $p$ an increase in $w$ makes misbehavior more attractive since in increases the gains to undesirable actions by $(1 - p)$. Yet, non-compliance will also increase the probability of detection by $p'$, which looms larger the higher the bonus the agent misses out on. In order to determine the overall effect an increase in the bonus $w$

[8]Note that our initial assumptions i) to iii) are not sufficient for the Hessian to be negative definite. This is due to the fact that joint deviations may be very attractive: When the agent decides to increase non-compliance, we will also reduce effort. A more detailed discussion of this assumption is given in the appendix.

will have on the agent's choice of $d$ we have to look at[9]

$$\frac{dd}{dw} = -\frac{\frac{\partial F}{\partial w}}{\frac{\partial F}{\partial d}} = -\frac{(1-p) - p'\Big(G\big(w(1-p)\big) + d\Big) - wp'(1-p)G'\big(w(1-p)\big)}{-2wp' - w\Big(G\big(w(1-p)\big) + d\Big)p'' - K''(d) + [p'w]^2 G'\big(w(1-p)\big)}. \tag{4}$$

The denominator is negative by the negative definiteness of the Hessian matrix, so sgn($\frac{dd}{dw}$) = sgn($\frac{\partial F}{\partial w}$). Starting from a point where the principal doesn't offer any bonus at all, a marginal increase in the bonus will always increase the level of undesirable behavior:

$$\left.\frac{\partial F}{\partial w}\right|_{w=0} = 1$$

since by (1) and (2) $w = 0$ implies $q = 0$ and $d = 0$. So by continuity we know that for any sufficiently small $w$ we have $\frac{dd}{dw} > 0$: Since the agent hardly receives any wage payment even if the firm makes high profits, he doesn't worry about loosing those if he increases undesirable behavior. However, as the principal increases incentives this has a first-order effect on the gains from undesirable behavior *given a certain* $p$ which induces the agent to choose a higher level of non-compliance.

Yet, it is unclear if this holds true for all $w$, since for large values of $w$ the agent will no longer ignore the impact an increase in $d$ is going to have on the probability of loosing out on a bonus. In order to understand how the marginal impact of an increase in $w$ on $d$ changes, we can look at the partial derivative of $\frac{\partial F}{\partial w}$ with respect to $w$:[10]

$$\frac{\partial^2 F}{\partial w^2} = -2p'(1-p)G'\big(w(1-p)\big) - wp'(1-p)^2 G''\big(w(1-p)\big) \tag{5}$$

Some basic manipulations give us the following condition which, if satisfied, guarantees that $\frac{\partial F}{\partial w}$ is globally decreasing:

**Assumption 2.1** *There exists some $f < 2$ such that*

$$\frac{-G''(v)v}{G'(v)} < f \qquad \forall v \in (0, \infty)$$

Under Assumption 1 the numerator of equation (4) decreases in $w$ and will become

[9]Throughout the paper, we will concentrate on interior solutions. I.e. we will assume that the principal always finds it optimal to offer contracts that implement $q > 0$ and $\overline{d} > d > 0$.

[10]Clearly, as any optimum is characterized by $F(w, d) = 0$, $d$ is itself a function of $w$. However, we show in the appendix that this does not change our results and we ignore this effect here for expositional simplicity.

negative for any sufficiently high $w$, implying that for large bonuses $w$ compliance does in fact *increase* with higher bonuses.

To understand the forces involved more precisely we can decompose the total change in the optimal $d$ into two separate effects: First of all, the bonus $w$ determines how large the return to undetected misbehavior is, since by misbehaving the agent increases the chance of getting an expected payment of $w(1-p)$. Furthermore, this return increases approximately at a constant rate as the principal increases $w$.[11]

The other effect is something that we can call the "intramarginal effect": By engaging in a certain level of misbehavior, the agent affects the risk that his behavior is detected and that he looses out on the expected wage that he would otherwise get. The strength of this effect obviously depends on $p'(d)$, the marginal probability of being caught, as well as on $w$, the bonus that he is going to loose out on. Moreover, it is going to depend on $q = G(w(1-p))$, since this determines the probability that the agent actually gets the bonus, given that no hard evidence on non-compliance emerges. Clearly, this effect increases in $w$ since $G'(w(1-p)) \geq 0$. Moreover, under Assumption 1 it increases at an increasing rate, which implies that this effect will eventually dominate the first effect.[12] Probably the most intuitive way to describe the second effect is to say that by setting a higher bonus $w$ the principal effectively relaxes the limited liability constraint of the agent in case of detected misbehavior. Since the principal will always find it optimal to impose maximal punishments in this state, this relaxation of the limited liability constraint goes hand in hand with an increase in punishment, hence reducing incentives to misbehave.

Note that Assumption 1 guarantees that the cost of effort is not too convex, or equivalently, that the probability of high profits given a certain level of $d$ is not too concave in $w$. More specifically, the assumption imposes an upper bound on what would corresponds to the coefficient of relative risk aversion if $G(\cdot)$ was a von Neumann-Morgenstern Utility function.

**Proposition 2.2** *Under Assumption 1 there will always exist some $w^*$ such that*

- *if $w < w^*$ an increase in the bonus $w$ will lead to more undesirable effort: $dd/dw \geq 0$*
- *if $w \geq w^*$ an increase in the bonus $w$ will lead to less undesirable effort: $dd/dw \leq 0$.*

*Furthermore, an increase in $w$ will always increase desirable effort $q$.*

[11] This holds only true as an approximation, since the expected bonus that can be reached by making large profits is a function of $p(d)$. So as $w$ increases the expected bonus will be affected directly, as well as indirectly through a change in $p$.

[12] Again, our verbal argumentation ignores the impact of $w$ on $p$.

**Proof** In the Appendix.

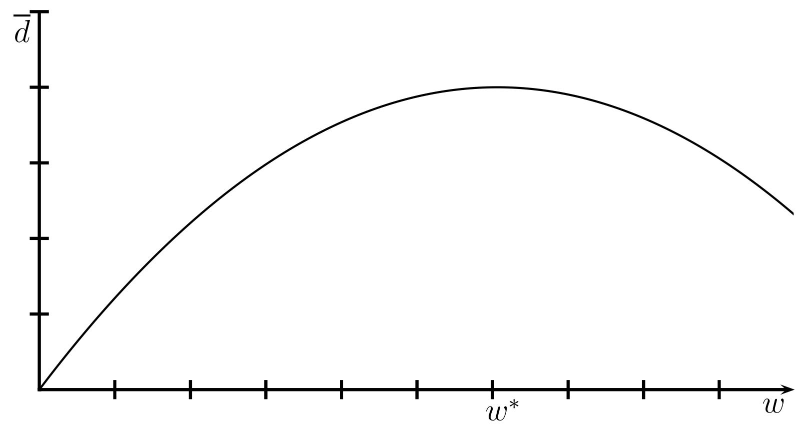


Figure 1

## 2.2 The optimal contract

The principal has two objectives that he wants to pursue by setting a bonus: On the one hand he wants to give the agent incentives to exert desirable effort; on the other hand he doesn't want him to engage in undesirable behavior. Anticipating the agent's choice of $q$ given a bonus payment $w$, we can state the principal's objective function as

$$\Pi = \Big(d + G\big(w(1-p)\big)\Big)\Big(\overline{\pi} - (1-p)w\Big) + \Big(1 - d - G\big(w(1-p)\big)\Big)\underline{\pi} - \tau(d) \tag{6}$$

where $d$ is a function of $w$. Henceforth, we will impose the standard assumption that the inverse of the marginal cost function, $G(.)$, is concave, which ensures concavity of the moral hazard problem when compliance is exogenously given.

**Assumption 2.3**

$$C'''(q) \geq 0 \Leftrightarrow G''(.) \leq 0$$

Let us first construct our benchmark case. This is a case where the level of compliance is exogenously given and the principal's only objective is to induce the agent to exert (desirable) effort. To keep our results comparable, we will assume that in this case, the principal still decides to pay the agent $w_d = 0$ in case of high profits and evidence on undesirable behavior and $w$ if no such evidence emerges. Even though the principal does not gain from doing so as long as the level of $d$ is exogenously given, it can be interpreted as a lottery that is equivalent to paying the agent a bonus of $w(1-p)$ for sure in case of high profits. (Remember that the agent is risk neutral.) In the benchmark case the principal will choose $w$ such that

$$\frac{\partial \Pi}{\partial w} = 0 \tag{7}$$

This "myopic" optimum is equivalent to the solution to a standard single-tasking problem. We will call the $w$ that solves this problem $\tilde{w}(p)$.

In practice however, the principal will take into account the effect his choice of $w$ is going to have on the level of undesirable effort and will hence choose $w$ such that the following first order condition is satisfied:

$$\frac{d\Pi}{dw} = \frac{\partial \Pi}{\partial w} + \frac{\partial \Pi}{\partial d}\frac{dd}{dw} = 0 \tag{8}$$

Note that the complete optimization problem need not be globally concave if compliance is sufficiently important, since compliance makes very large and very small bonuses more attractive. So the solution to the first order condition may not be unique. Henceforth, we will just assume that the problem's global optimum is interior[13] and that the global optimum is unique. The second assumption abstracts from the non-generic case where two (local) optima yield exactly the same level of profits.

It turns out that the principal will increase incentives above the bonus in the benchmark case if and only if $\tilde{w}(p) > w^*$ and reduce incentives whenever $\tilde{w}(p) < w^*$: The principal will always adjust the bonus $w$ in order to exploit the positive effects this has on compliance. If $\tilde{w}(p)$ is sufficiently small, he increases compliance by marginally reducing incentives. If on the other hand the bonus he would set in the benchmark case is very large, a marginal increase in $w$ will increase compliance and the principal will choose to set higher incentives than he would do otherwise.

**Proposition 2.4** *If the optimal incentive in the benchmark case $\tilde{w}(p)$ lies below a threshold $w^*$, the principal will lower incentives relative to the benchmark in order to increase compliance. If the incentive in the benchmark case is above this threshold, the principal will increase incentives in order to enhance compliance.*

[13] We call an optimum interior if it implements some $\overline{d} > d > 0$ and $q > 0$.

**Proof** If $\partial\Pi/\partial w \geq 0$, we must have $dd/dw \geq 0$ since $\partial\Pi/\partial d < 0$: A reduction in $d$ is always beneficial since i) it reduces the cost of undesirable effort (which outweighs the benefits from increases in the likelihood of high profits), ii) it reduces the expected wage payments and iii) it increases incentives for desirable effort by increasing $(1-p)$. If on the other hand $\partial\Pi/\partial w < 0$ than the second term must be positive which may be either because $dd/dw < 0$ and $\partial\Pi/\partial d < 0$ or because $dd/dw > 0$ and $\partial\Pi/\partial d > 0$. We can reformulate the problem and get

$$\frac{d\Pi}{dw} = \frac{\partial\Pi}{\partial w}\left(1 - \frac{dd}{dw}\frac{wp'}{(1-p)}\right) + \left(\overline{\pi} - \underline{\pi} - \tau'(d) - (1-p)w\right)\frac{dd}{dw}$$

Since $dd/dw \leq (1-p)/wp'$ (see Appendix) we know that the first term is negative. So the second term must be positive which requires that $dd/dw < 0$ since $\overline{\pi} - \underline{\pi} - \tau'(d) - (1-p)w < 0$.

### 2.2.1 Optimal punishments

As noted before, it turns out that any optimal contract will always have $w_d = 0$. The intuition for this result is straightforward: Reducing $w_d$ will always reduce the level of undesirable effort the agent chooses, which is unambiguously beneficial. At the same time a decrease in $w_d$ will reduce the incentives for desirable effort $q$. However, this effect can be offset by adjusting $w$ in a way that leaves the expected wage payment *given high profits* constant, which leaves us only with the negative impact on $d$. Also, it can never be optimal to punish the agent in case of high profits: The same $q$ and a weakly lower undesirable effort $d$ could also be obtained by offering the agent no wages at all, strictly reducing the expected wage cost.

**Lemma 2.5** *Any contract offered by the principal that induces $d \in (0, \overline{d})$ will pay no wages if undesirable behavior is detected: $w_d = w_{f,d} = 0$. Furthermore, the expected wage payment in case of high profits will never be smaller than the expected wage payment in case of low profits: $w \geq 0$.*

**Proof** First, observe that it can never be optimal to have a contract $(w_f, w, w_{f,d}, w_d)$ with a resulting probability of detecting non-compliance $p$ and $(1-p)w + pw_d < 0$: In this case $q = 0$, which could also be implemented by a contract $(0, 0, 0, 0)$ that has strictly lower wage costs and induces weakly lower undesirable effort $d = 0$. Therefore $(1-p)w + pw_d < 0$ can never be optimal.

Let us now show that a contract including $w_d > 0$ never dominates a contract with $w_d = 0$. Consider any contract $(w_f, w, w_{f,d}, w_d)$ with a resulting probability of detecting non-compliance $p$ and $w_d > 0$. Now consider an alternative contract $(\hat{w}, 0, \hat{w}_{f,d}, \hat{w}_f)$ and the resulting probability of detecting non-compliance $\hat{p}$ that satisfies $w(1-p) + w_d p = \hat{w}(1-\hat{p})$. By continuity of $d$ and $p(d)$ such a contract always exists. Clearly such a contract leaves the incentives for $q$ as given by equation (1) unchanged.

Assuming $\hat{p} > p$ we must have $\hat{w} > w$ which means that the new contract globally reduces $\partial U/\partial d$ as given in equation (2), hence weakly reducing $d$. This implies that $\hat{p} \leq p$. which is a contradiction. So the new contract must have $\hat{p} \leq p$ which by monotonicity of $p(d)$ implies that the level of undesirable effort will be weakly lower which is beneficial since $\tau'(d) > \overline{\pi} - \underline{\pi}$. At the same time the expected wage payment is reduced: If $w_f = \hat{w}_f = 0$, this follows directly from the fact that the (positive) expected bonus for high profits has to be paid less often as $d$ decreases and since the expected payments contingent on high or low profits stay constant. (If $w_f > 0$, the principal can additionally adjust $w_f$ such that $q$ and $p$ are left unchanged but the expected wage payments decrease.) If the solution to the agent's optimization problem is interior, the new contract will implement a strictly lower $d$. In this case any contract with $w_d > 0$ is dominated.

The proof that in any interior optimum $w_{f,d} = 0$ is given in the appendix.

### 2.2.2 Bonuses vs. Efficiency wages

So far, we have ignored the option to pay the agent a profit-independent wage component $w_f$. We can think of such a wage payment as an efficiency wage, since it is not needed to satisfy the agent's participation constraint and it is not contingent on profits.[14] Clearly, a fixed wage component is never optimal in a standard moral-hazard problem with limited liability: Since there is a trade-off between giving rents to the agent and inducing effort, it can never be optimal to give additional rents to the agent that do not induce additional effort. In our setting this is less clear: The motivation for increasing incentives above the benchmark level $\tilde{w}(p)$ is to give higher expected wages to the agent that he might loose if he decides to engage in undesirable behavior. Alternatively, the principal could achieve this by paying efficiency wages to the agent that take the form of a fixed wage component $w_f$ that will be paid out whenever no undesirable behavior is observed, irrespective of

[14]Of course, our setting differs considerably from the literature on efficiency wages since in our model the punishment an agent faces if he fails to get the efficiency wage is not given by a labor market equilibrium, but by payments specified in the contract (namely by $w_{f,d} = w_d = 0$)

the firm's profits. On the one hand increased incentives have the advantage that they not only increase compliance, but at the same time induce extra effort. But on the other hand they also increase the returns to undetected non-compliance. It is therefore ex-ante not clear which measure will be favored by the principal. [15]

In order to determine if a given decrease in $d$ will be implemented by adjusting the fixed wage component $w_f$ or the bonus $w$, we can compare the cost of reducing $d$ by one unit via an increase in $w_f$

$$\frac{d\Pi}{dw_f} = \frac{\partial\Pi}{\partial w_f} + \frac{\partial\Pi}{\partial d}\frac{dd}{dw_f}$$
$$\Leftrightarrow \frac{\left(\frac{d\Pi}{dw_f}\right)}{\left(\frac{dd}{dw_f}\right)} = -\frac{(1-p)}{p'}\frac{\partial F}{\partial d} + \frac{\partial\Pi}{\partial d} \tag{9}$$

to the cost of decreasing $d$ by one unit via an increase (decrease) in the bonus $w$:

$$\frac{d\Pi}{dw} = \frac{\partial\Pi}{\partial w} + \frac{\partial\Pi}{\partial d}\frac{dd}{dw}$$
$$\Leftrightarrow \frac{\left(\frac{d\Pi}{dw}\right)}{\left(\frac{dd}{dw}\right)} = -\frac{\partial\Pi}{\partial w}\frac{\left(\frac{\partial F}{\partial d}\right)}{\left(\frac{\partial F}{\partial w}\right)} + \frac{\partial\Pi}{\partial d} \tag{10}$$

Consider first the case where $\tilde{w}(p) > w^*$, i.e. the bonus the principal would set in the benchmark case is sufficiently high. If the principal chose $w = \tilde{w}(p)$ he could costlessly decrease $d$ by marginally increasing $w$ since at $w = \tilde{w}(p)$ we have $\partial\Pi/\partial w = 0$. This means that the principal will always adjust the bonus relative to the benchmark case. More generally, we can show that the principal will always choose to increase the bonus $w$ instead of the fixed wage $w_f$ to achieve a given decrease in $d$ as long as the negative effect of an increase in incentives *given* a certain level of undesirable behavior is not too large:

**Proposition 2.6** *Assume that $\tilde{w}(p) > w^*$. Than the principal will never solely use efficiency wages in order to enhance compliance but will also increase the bonus $w$ relative to $\tilde{w}(p)$. More generally he will always increase incentives rather than efficiency wages*

[15] This intuition only applies to the case where $\tilde{w}(d) > w^*$. In the opposite case reducing the bonus increases compliance and at the same time reduces the expected wage payment that is left to the agent, while it is costly since it distorts $q$ downwards. Again, it is unclear whether adjusting the bonus or increasing $w_f$ is a preferable means to enhance compliance.

*in order to decrease $d$ if and only if*

$$G'\big(w(1-p)\big)(\overline{\pi}-\underline{\pi}) > \frac{(1-p)}{p'}. \tag{11}$$

**Proof** The principal will achieve a given increase in $d$ by increasing $w$ rather than $w_f$ whenever

$$\begin{aligned} -\frac{\partial \Pi}{\partial w}\frac{\left(\frac{\partial F}{\partial d}\right)}{\left(\frac{\partial F}{\partial w}\right)} + \frac{\partial \Pi}{\partial d} &\leq -\frac{(1-p)}{p'}\frac{\partial F}{\partial d} + \frac{\partial \Pi}{\partial d} \\ \Leftrightarrow \frac{\partial \Pi}{\partial w} &\geq \frac{(1-p)}{p'}\frac{\partial F}{\partial w} \end{aligned}$$

since $\partial F/\partial d < 0$ by negative definiteness of the Hessian matrix and $\partial F/\partial w < 0$ by the assumption that $w > w^*$. Substituting (4) and (7) and simplifying gives us the condition stated in the Proposition.

Note that this condition depends on $w_f$ only via the level of undesirable effort $d$ chosen by the agent: Conditional on a given $d$ the cost of efficiency wages to achieve a certain decrease in $d$ are constant. Furthermore, since by assumption 2.3 $G''(w(1-p)) < 0$ we know that the condition is not only necessary but also sufficient for the principal to pay efficiency wages: Since the marginal cost of distorting incentives is increasing, there is a unique threshold above which the principal will no longer increase incentives.

Let us now turn to the case where $\tilde{w}(p) < w^*$. In this case the principal will increase compliance either by decreasing $w$ or by increasing $w_f$. In this case the cost of increasing compliance via reducing incentives is again given by equation (9). For the same reason as above, the principal will never pay any efficiency wages without also lowering the bonus $w$ to enhance compliance. However, the sufficient condition for the principal to find non-infinitesimal adjustments in the bonus optimal is now turned around: The principal will now be prepared to cut back on incentives as a means to enhance compliance if the positive effect of an increase in $w$ on profits given a certain level of $d$ is not too large.

**Proposition 2.7** *Assume that $\tilde{w}(p) < w^*$. Than the principal will never solely use efficiency wages in order to enhance compliance but will also decrease the bonus $w$. More generally he will always decrease incentives rather than increasing efficiency wages in order to decrease $d$ if and only if*

$$G'\big(w(1-p)\big)(\overline{\pi}-\underline{\pi}) < \frac{(1-p)}{p'} \tag{12}$$

**Proof** Analogue to 2.3

These results show that the effect we have shown in sections 2.1 and 2.2 where we abstracted from positive payments $w_f$ will indeed always be present: Distorting incentives in order to increase compliance is not only a last resort if for some exogenous reason no fixed wages can be paid, but it is indeed optimal for a large range of parameters.

# 3 Policy Intervention

In most examples we can think of, undesirable behavior has negative externalities on society as a whole. To cite our motivating example, excessive risk taking may create systemic risk and require public bail-outs. The same holds true for many other forms of misbehavior: Cartel agreements reduce consumer surplus, permissive behavior by auditing firms reduces the reliability of financial statements etc. A natural question to ask is hence how a social planner may want to discourage such behavior. In order to answer this question, we look at two policy instruments: The first one is a legal cap on bonuses. While such a measure has so far not been widely used, it is vividly discussed and therefore merits some closer theoretic examination. An instrument that is widely applied already is to punish firms or managers if evidence on misbehavior surfaces. In fact, in many situations corporate fines are the only reason why undesirable behavior is against the interest of the principal in the first place. The question we will be concerned with in this contest is wether it is generally optimal to impose personal rather than corporate fines.

## 3.1 Caps on bonuses

The pervious discussion also makes clear that the negative consequences of legal restrictions on the size of bonus payments are potentially twofold: Besides the obvious effect of reducing the incentives for effort, depending on the size of incentive payments such a policy may even have adverse effects on compliance and encourage misbehavior.

**Proposition 3.1** *Assume that absent any regulation, the principal sets a bonus $w > w^*$. Than a legal cap $\hat{w} < w^*$ on bonuses will weakly increase non-compliance as long as $w - \hat{w} \leq \epsilon$ for some small $\epsilon$.*

**Proof** Assume that $w_f = 0$. Since the optimum in the case without intervention was unique, a cap that is just binding will make it optimal to set a new bonus at $w = \hat{w}$, leading to a marginally lower $w$. This will increase the level of misbehavior because for all $w > w^*$ we have $\frac{dd}{dw} \leq 0$. If we allow for $w_f > 0$ the level of misbehavior can still not decrease after the introduction of the cap. While the principal can now offset part of the increase in $d$ by increasing $w_f$ the resulting level of non-compliance will still be weakly lower. (By revealed preferences we know that the alternative instrument to achieve compliance is more expensive.) Our results will be reinforced by the fact that if $w$ can not be chosen arbitrarily large, we can no longer guarantee that the principal finds it optimal to set $w_d = 0$.

Proposition 3.1. tells us that what would seem to be cautious regulation may in fact be very harmful: By imposing caps on bonuses that are close to the level of bonuses paid in an unregulated labor market, we destroy incentives for effort and increase managerial misbehavior. Yet, this does not hold for large interventions: If the legal maximum on bonus payments is very small, this may have positive effects on compliance. A cap that is sufficiently close to zero will also result in negligible levels of misbehavior. In fact, even a less stringent cap on bonuses can potentially increase compliance, since another local optimum with a lower level of non-compliance may now yield higher profits than setting $w = \hat{w}$. However, such bold interventions clearly erode incentives for managers to work hard and may not increase social welfare.

## 3.2 Personal vs. Corporate Liability

First, let us look at the case where a social planer decides to punish the principal in case of undesirable behavior. So far we have taken the cost of undesirable behavior to the principal to be exogenously given. However, in many cases these costs comprise legal fines. This implies that a policy maker may choose to increase this cost in case he wants to discourage undesirable behavior. The policy maker has access to some monitoring technology $\mathcal{P}(d)$ and can impose some punishment $\mathcal{T}$ on the principal if he observes illegal behavior. For simplicity we assume that the states of the world in which the policy maker receives evidence on misbehavior are a subset of the states in which the principal does so, which implies that $\mathcal{P}(d) \leq p(d)$. Furthermore, we will assume that the principal is free to impose any fine on either the principal or the agent. While this assumption is very strong, it nevertheless generates some interesting insights: Even if the policy maker does not face any limited liability constraints, he may still decide to punish the principal rather than the agent. This is surprising since we have seen that the agent's limited liability constraint was the only reason why the principal doesn't implement $d = 0$. However, we will shortly see that the policy maker may choose to do so in oder to reduce the distortions associated with the unobservability of effort.

The the new objective function of the principal is given by

$$\begin{aligned} \hat{\Pi} &= \Pi - \mathcal{P}(d)\mathcal{T} \\ G \equiv \frac{d\hat{\Pi}}{dw} &= \tfrac{\partial \Pi}{\partial w} + \left(\tfrac{\partial \Pi}{\partial d} - \mathcal{P}'(d)\mathcal{T}\right)\tfrac{dd}{dw} = 0 \end{aligned}$$

A marginal increase in $\mathcal{T}$, the fine the principal has to pay in case of a demonstrably misbehaving agent, will always induce the principal to invest more into compliance-enhancing incentives. Whenever $w > w^*$ and when the principal does not find it optimal to pay efficiency wages[16] this will take the form of offering a higher bonus $w$ to the agent:

$$\frac{dw}{dT} = \frac{\mathcal{P}'(d)\frac{dd}{dw}}{\frac{\partial G}{\partial w}}$$

which is positive since $\frac{dd}{dw} < 0$ and $\frac{\partial G}{\partial w} < 0$ by (local) concavity.

Since we can show that $w$ will always implement effort that is too low from the point of view of a social planner, this is clearly good news: By punishing the principal the policy maker not only reduces unwanted behavior, but he also reduces the distortions created by the non-observability of effort.

The policy maker can also discourage undesirable behavior by punishing the agent directly in case of observed misbehavior. However, assuming that the policy is implemented before the principal has offered a contract $(w_f, w, w_{f,d}, w_d)$ to the agent, it is unclear how the bonus $w$ offered by the principal is going to change given a certain punishment.
- RESULTS TO FOLLOW -

Since the legislator can use any of those two punishments to implement a certain level of compliance, their effect on social welfare only differs with regard to the effort level the principal is going to induce *given those punishments*. As long as $w < w^{fb}$ social welfare monotonically increases in $w$. A legislator that is interested in maximizing social welfare should therefore punish the principal rather than the agent, since by doing so he *unambiguously* reduces the distortions created by the unobservability of effort and moves the implemented effort level closer to its first-best level.

**Proposition 3.2** *Assume that $w > w^*$. Then a marginal increase in the punishment the principal faces in case of observed non-compliance will always increase the optimal bonus $w$ that he offers the agent. This increases social welfare in two ways: i) it encourages*

[16] Recall that this is the case whenever $G'\big(w(1-p)\big)(\overline{\pi} - \underline{\pi}) > \frac{(1-p)}{p'}$.

*compliance and ii) it increases (socially valuable) effort. The effect of punishing the agent on the principal's choice of the bonus is ambiguous.*

# 4 Conclusion

Our model shows that large bonus payments may be optimal from a firm's point of view in order to discourage the gambling of incentive schemes by its agent. Applied to the banking industry this implies that large bonuses may have in fact reduced excessive risk-taking. This is consistent with the observation by Fahlenbach and Stulz (2009) that cash bonuses did not correlate with bad performance during the recent crisis. It also explains why pay-performance sensitivities are not lower in the banking industry than in other sectors where compliance is less crucial. While public opinion seems the attribute the size of bonus payments to some sort of market failure, our model suggests that large pay-performance ratios may in fact be an optimal mechanism to increase compliance. This can explain why those seemingly suboptimal contracts have not been eliminated by market forces.

At the same time, our results shed some new light on the discussion to limit the size of bonuses via legislation. We have shown that under some rather general conditions this will have counterproductive effects and reduce compliance, while at the same time diluting incentives for managers to work hard. In particular, this holds true for rather generous legal caps. Highly restrictive caps on the other hand may have some merit in reducing undesirable behavior, which of course comes at the cost of a large decrease in incentives for productive effort.

Our model predicts that undesirable behavior is most prevalent for intermediate incentives, while for sufficiently low or high powered incentives, such behavior is less attractive. This has important normative implications in its own right, as it gives an indication where it is most important to monitor agents. Regulators and other stakeholders on whom non-compliance has negative externalities can hence use observable contracts in order to asses the danger of harmful behavior by agents. Furthermore, the prediction of the model is in line with behavioral models that highlight the hidden cost of intermediate incentives.

Insofar as bonuses may be used to enhance compliance, our paper has further relevant policy implications: Policy makers may want to exploit agency-problems in the compliance dimension in order to reduce welfare-reducing distortions in the provision of effort by an agent: Punishing firms for misbehaving agents (instead of punishing agents directly) will induce the principal to set higher bonuses. This reduces non-compliance and increases effort provision, both of which increases efficiency. These policy implications contradict the standard paradigm in the law and economics literature that the "least cost avoider" should be punished for undesirable behavior. The optimal structure of legal punishments in cases where there are multi-dimensional agency problems within an organization therefore seems to be a fruitful area for future research.

# References

Bebchuk, L.A., & Fried, J. 2006. *Pay without performance: The unfulfilled promise of executive compensation.* Harvard Univ Pr.

Bebchuk, Lucian A., Bachelder, Joseph E., Campos, Roel C., Georgiou, Byron S., Hevesi, Alan G., Lerach, William, Mendelsohn, Robert, Monks, Robert A.G., Myerson, Toby, Olson, John F., Strine, Leo E., & Wilcox, John C. 2006. Director Liability. *Delaware Journal of Corporate Law*, **31**(3), 1011–1045.

Conyon, M.J., & Murphy, K.J. 2000. The prince and the pauper? CEO pay in the United States and United Kingdom. *The Economic Journal*, **110**(467), 640–671.

Fahlenbrach, R., & Stulz, R.M. 2009. Bank CEO Incentives and the Credit Crisis. *NBER Working Paper.*

Fischer, Paul, & Huddart, Steven. 2008. Optimal Contracting with Endogenous Social Norms. *American Economic Review*, **98**(4), p1459 – 1475.

Gabaix, X., & Landier, A. 2008. Why Has CEO Pay Increased So Much?*. *Quarterly Journal of Economics*, **123**(1), 49–100.

Garoupa, Nuno. 2000. Corporate Criminal Law and Organization Incentives: A Managerial Perspective. *Managerial and Decision Economics*, **21**(6), p243 – 252.

Gneezy, Uri, & Rustichini, Aldo. 2000a. A Fine is a Price. *Journal of Legal Studies*, **29**(1), p1 – 17.

Gneezy, Uri, & Rustichini, Aldo. 2000b. Pay Enough or Don't Pay at All. *Quarterly Journal of Economics*, **115**(3), p791 – 810.

Holmstrom, Bengt, & Milgrom, Paul. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, **7**, p24 – 52.

John, K., & Qian, Y. 2003. Incentive features in CEO compensation in the banking industry. *Economic Policy Review*, **9**(1).

John, T.A., & John, K. 1993. Top-management compensation and capital structure. *The Journal of Finance*, **48**(3), 949–974.

Murphy, K.J. 1999. Executive Compensation. *Handbook of Labor Economics*, **Volume 3b**, 2485–2563.

Prendergast, Canice. 1999. The Provision of Incentives in Firms. *Journal of Economic Literature*, **37**(1), p7 – 63.

Shapiro, C., & Stiglitz, J.E. 1984. Equilibrium unemployment as a worker discipline device. *The American Economic Review*, **74**(3), 433–444.

Spagnolo, Giancarlo. 2000. Stock-Related Compensation and Product-Market Competition. *RAND Journal of Economics*, **31**(1), p22 – 42.

Spagnolo, Giancarlo. 2005. Managerial Incentives and Collusive Behavior. *European Economic Review*, **49**(6), p1501 – 1523.

Zhou, X. 2000. CEO pay, firm size, and corporate performance: evidence from Canada. *The Canadian Journal of Economics/Revue canadienne d'Economique*, **33**(1), 213–251.

# A Appendix

**Negative Definiteness:** Negative Definiteness of the Hessian Matrix that characterizes the agent's decision problem requires that

$$[-2p'w - w(q+d)p'' - w_f p'' - K''(d)][-C''(q)] - (wp')^2 > 0$$

Note that for large values of $w$ our conditions i) to iii) are not sufficient to ensure concavity of the agent's optimization problem. The reason for this lies in the very nature of the agent's problem: The agent has an incentive to exert less effort if he engages in high levels of misbehavior and vice-versa, making joint deviations relatively attractive. Yet, we can show that regardless of the agent's behavior the principal will always choose some $w < \frac{\overline{\pi}-\underline{\pi}}{1-\overline{p}}$. So we just need $p(d)$, $K(d)$ and $C(q)$ to be sufficiently convex to make sure that the inequality is satisfied for all $w \in [0, \frac{\overline{\pi}-\underline{\pi}}{1-\overline{p}})$.

**Lemma A1:** It can never be optimal to pay a positive wage in case of low profits and observed misbehavior:

Proof: For this proof it will be useful to redefine the wage that is paid in state $[i,j]$ as $w_{i,j}$ where $i \in \{h,l\}$ and $j \in \{d,n\}$ denote wether high profits (h) have been made or not (l) and wether misbehavior has been detected (d) or not (n) . The utility of the agent is than given by

$$\begin{aligned}
U =&(q+d)\big(w_{h,n}(1-p) + w_{h,d}p\big) + (1-q-d)\big(w_{l,n}(1-p) + w_{l,d}p\big) - C(q) - K(q) \\
\frac{\partial U}{\partial q} =&(1-p)w_{h,n} + pw_{h,d} - (1-p)w_{l,n} - pw_{l,d} - C'(q) \\
\frac{\partial U}{\partial d} =&(1-p)w_{h,n} + pw_{h,d} - (1-p)w_{l,n} - pw_{l,d} - K'(d) - p'\Big((q+d)(w_{h,n} - w_{h,d}) \\
&+ (1-q-d)(w_{l,n} - w_{l,d})\Big)
\end{aligned}$$

Now assume that the initial contract $(w_{h,n}, w_{h,d}, w_{l,n}, w_{l,d})$ has $w_{l,d} > 0$ and a corresponding probability of detecting misbehavior $p$. By continuity there exists a contract $(\hat{w}_{h,n}, \hat{w}_{h,d}, \hat{w}_{l,n}, 0)$ that has $\hat{p}$ and satisfies $(1-p)w_{l,n} + pw_{l,d} = (1-\hat{p})\hat{w}_{l,n}$. Assume furthermore that $\hat{p} > p$. If $w_{h,n} \geq w_{h,d}$ the principal can increase $w_{h,n}$ in order to keep the expected wage payments conditional on the firm making high profits constant. If $w_{h,n} < w_{h,d}$ the principal can reduce $w_{h,d}$ to achieve this. In both cases the new contract implements the same $q$ and globally reduces the incentive for $d$, hence weakly reducing $p$. This is a contradiction to the assumption that $\hat{p} > p$. The new contract has therefore

i) the same expected wage payments conditional on the firm making high profits ii) the same expected wage payments conditional on the firm making low profits iii) the same $q$ iv) a weakly lower $d$. Decreasing $d$ is beneficial since $\tau'(d) > \overline{\pi} - \underline{\pi}$ and since the expected wage expenditure in case of high profits must be weakly higher than the expected wage expenditure in case of low profits (as shown in the proof of Lemma 2.3). In any interior optimum, the $d$ implemented by the new contract is strictly lower, so any contract involving a $w_{l,d}$ is dominated.

**Lemma A2:** $\frac{dd}{dw} \leq \frac{(1-p)}{wp'}$. This inequality is strict for all $w > 0$

Proof:

$$\begin{aligned}
& \frac{dd}{dw} = \frac{(1-p) - p'(G(w(1-p)) + d) - wp'(1-p)G'(w(1-p))}{2wp' + w(G(w(1-p)) + d)p'' + K''(d) - [p'w]^2 G'(w(1-p))} \leq \frac{(1-p)}{wp'} \\
\Leftarrow \quad & \frac{(1-p) - wp'(1-p)(G'(w(1-p)) + d)}{2wp' + w(G(w(1-p)) + d)p'' + K''(d) - [p'w]^2 G'(w(1-p))} \leq \frac{(1-p)}{wp'} \\
\Leftrightarrow \quad & wp' - [wp']^2 G'(w(1-p)) \leq 2wp' + K''(d) + p''w(G(w(1-p)) + d) - [wp']^2 G'(w(1-p)) \\
\Leftrightarrow \quad & 0 \leq wp' + K''(d) + p''w(G(w(1-p)) + d)
\end{aligned}$$

**Proof of Proposition 1:** Since $d$ is itself a function of $w$, the derivative of the numerator of equation (4) is given by

$$\frac{d\left(\frac{\partial F}{\partial w}\right)}{dw} = \frac{\partial^2 F}{\partial w^2} + \frac{\partial\left(\frac{\partial F}{\partial w}\right)}{\partial d}\frac{dd}{dw} \tag{13}$$

Case i): $\frac{dd}{dw} \geq 0$:

By Assumption 1 we know that $\frac{\partial^2 F}{\partial w^2} < 0$, furthermore $\frac{dd}{dw} \geq 0$ by assumption. So in order to show that $\frac{d\left(\frac{\partial F}{\partial w}\right)}{dw} < 0$ we have to look for an upper bound on $\frac{\partial\left(\frac{\partial F}{\partial w}\right)}{\partial d}$. Defining $v = w(1-p)$ we have

$$\begin{aligned}
& \frac{\partial\left(\frac{\partial F}{\partial w}\right)}{\partial d} = -p' - p''(G(v) + d) + 2w(p')^2 G'(v) - wp''(1-p)G'(v) + (wp')^2(1-p)G''(v) \\
\Leftrightarrow \quad & \frac{\partial\left(\frac{\partial F}{\partial w}\right)}{\partial d} \leq 2w(p')^2 G'(v) + (wp')^2(1-p)G''(v)
\end{aligned}$$

Using this upper bound on $\frac{\partial\left(\frac{\partial F}{\partial w}\right)}{\partial d}$ we can now get the sufficient condition for (13) to be negative:

$$2p'(1-p)G'(v) + wp'(1-p)^2G''(v) + \left(2w(p')^2G'(v) + (wp')^2G''(v)\right)\frac{dd}{dw} < 0$$

which is fulfilled for all $w > 0$ since $\frac{dd}{dw} < \frac{(1-p)}{wp'}$ by Lemma A2.

Case ii): $\frac{dd}{dw} < 0$:

In this case $\frac{d\left(\frac{\partial F}{\partial w}\right)}{dw}$ will always be negative as long as $\frac{dd}{dw}$ is sufficiently close to zero. The indirect effect an increase in $w$ has on the choice of $d$ and thereby on the agent's reaction to further increases in $w$ may hence limit how negative $\frac{dd}{dw}$ can become, but will never change it's sign.