# Optimal contracts based on subjective evaluations and reciprocity

**Alexander Sebald, Markus Walzl**

**Working Papers in Economics and Statistics**

2012-16

**University of Innsbruck**
**Working Papers in Economics and Statistics**

The series is jointly edited and published by

- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact Address:
University of Innsbruck
Department of Public Finance
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

# Optimal contracts based on subjective performance evaluations and reciprocity

July 18, 2014

**Abstract**

As demonstrated in a recent laboratory experiment [see Sebald and Walzl (2014)], individuals tend to sanction others who subjectively evaluate their performance whenever this assessment falls short of the individuals' self-evaluation. Interestingly, this is the case even if the individuals' earnings are unaffected by the subjective performance appraisal. Hence, performance feedback which falls short of agents' self-evaluations can be interpreted as an unkind act that triggers a negatively reciprocal response not only if the assessment determines agents' earnings but also when it lacks monetary consequences. We propose a principal-agent model formalizing that agents might engage into conflict in response to ego-threatening performance appraisals and show that these conflicts stabilize principal-agent relationships based on subjective performance evaluations. In particular, we identify conditions for a positive welfare effect of increasing costs of conflict and a negative welfare effect of more capable agents.

*Keywords:* Contracts, Subjective Evaluations, Self-Esteem, Ego-Threats, Reciprocity.

*JEL classification:* D01; D02; D82; D86; J41.

# 1 Introduction

While performance assessments are an integral part of any pay-for-performance incentive scheme, it is very often impossible to objectively measure workers' and especially managers' individual contributions to the success of projects or enterprises. Therefore it is widely prevalent to (also) take into account subjective evaluations in performance pay. Already in 1981 the Bureau of National Affairs reports, for example, that pay for performance systems involving subjective measures are more common than those involving only objective performance signals [see e.g. Milkovich and Wigdor (1991), Levin (2003), or Gibbs *et al.* (2004) for more evidence on the wide usage of subjective performance appraisal systems in performance pay in e.g. investment banks, law firms, human resource management, and consultancy].

As long as subjective performance appraisals are truthful and resemble an unbiased, informative signal, their inclusion certainly enhances the efficiency of the underlying incentive scheme. The corresponding incentive problem for the evaluator, however, may limit the scope of subjective assessments for several reasons. First, performance pay on the basis of subjective performance evaluations may lack credibility. If the evaluator is the residual claimant and labor contracts specify payments on the basis of the principals' subjective appraisals, principals have an incentive to claim that performance was poor according to their perception in order to establish low wages. As a consequence, inefficiently low effort may be spend by agents unless principals can credibly commit to an honest revelation of their subjective information as, for instance, in repeated interaction or with a credible payment to a third party [see e.g. Levin (2003) or MacLeod (2003)]. Second, subjective performance evaluations may generate welfare reducing conflicts initiated by the agent due to possibly diverging assessments by principals and agents. Diverging performance appraisals might generate conflict as agents have the tendency to protect their self-image and self-esteem [see e.g. Greenwald (1980)].

We propose a principal-agent model formalizing an agent's willingness to engage into (costly) conflict in response to ego-threatening feedback and show how these conflicts might stabilize principal-agent relationships based on subjective performance evaluations. We consider a principal who wants to motivate an agent to spend effort on a project – a complex good or service – where neither the effort of the agent nor the success of the project are publicly observable. The principal and the agent only receive private, i.e. subjective, signals regarding the success of the project. The subjective signals that both receive are imperfectly correlated with each other and to the actual effort level. To motivate the agent to spend positive effort, a contract has to specify payments which increase in the subjective signal of the principal.

As in Bénabou and Tirole (2002) and Suvorov and van de Ven (2009) we assume that the success of the project depends not only on the effort of the agent but also on his capability – only effort spend by a capable agent can result in a successful project. The capability of the agent is initially unknown to the agent as well as the principal. Both only know the ex-ante probability with which the agent is capable in successfully completing the specific project and up-date their belief about the agent's capability when they receive their subjective performance signal. Hence, a payment contingent on the principal's subjective performance signal does not only affect the agent's earnings, it also communicates information regarding the agent's capability. We assume that the agent builds his self-esteem on his own subjective performance signal and the corresponding update regarding his capability. Specifically, the agent suffers if the principal communicates a lower subjective signal (i.e. a lower capability) than the agent's own signal – the principal's signal and capability information threatens his self-esteem. If the agent creates conflict to protect his self-esteem in this situation, it is no longer optimal for the principal to always downgrade the subjective signal to save on wage payments. We demonstrate that there are conflict costs of the principal imposed by the agent such that principal and agent sign the same contract as if their subjective performance signals where verifiable.

If, however, conflict costs are too low the principal cannot credibly commit to sufficiently high-powered incentives as the agent expects her to downgrade her subjective assessment. In this situation of a binding truth-telling constraint, increasing costs of conflict may be welfare enhancing. Likewise, if conflict costs are too large, the principal has an incentive to always announce a favorable performance assessment as this avoids any conflict. As a conflict can only arise if the agent considers himself capable, a higher ex-ante capability of the agent may well be welfare detrimental in this case.

Our base-line assumption that agents create conflict in response to performance feedback that does not meet their own performance assessment has been motivated by a recent laboratory study [see Sebald and Walzl (2014)]. We matched participants into pairs and randomly assigned them to the role of a principal or an agent. The agent had to click away boxes on screen and the principal could observe the agent working. The principal's payoff was determined by the percentage of boxes clicked away by the agent. After the agent worked on the task, the principal gave performance feedback that determined the share of the principal's payoff given to the agent in an *incentive treatment* in which the feedback determined the agent's wage and a *flat treatment* in which the agent received a fixed payment independent of his actual performance and independent of the principal's feedback. As both the agent and the principal were only informed about the true performance of the agent after the end of the experiment, their decisions during the experiment were based on subjective assessments. After receiving the principal's feedback, the agent had the opportunity to reduce the principal's payoff at a cost for himself. In this experiment we find that agents' reactions to the principals' feedback strongly depend on their self-perceptions. Agents reduce payoffs of principals, if the principal's feedback is below their self-perception, but accept the feedback and refuse to reduce payoffs if the feedback confirms/is higher than their own evaluation.

This pattern is observed in both treatments, i.e. when the agent's material benefit depends on the principal's feedback (incentive treatment) as well as when it is independent of the feedback (flat treatment). The willingness to reduce payoffs in the incentive treatment can be explained by existing models of reciprocity [e.g. Rabin (1993), Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006) and Hart and Moore (2008)].[1] These contributions argue that agents act reciprocally towards principals whenever their *payoffs* fall short of or exceed certain reference values against which they judge the kindness of the principals' actions and beliefs. However, the results of the flat treatment suggest that there also exists another motivation for payoff reductions that is independent of the payoff consequences of the principal's feedback.

Social psychologists explain the creation of conflict in response to negative feedback

[1] See Netzer and Schmutzler (2013) for a discussion of (mutually) unkind equilibrium behavior in the presence of intention based preferences.

through the individuals' eagerness to actively maintain and protect positive self esteem/self-perceptions [e.g. Greenwald (1980), Bushman and Baumeister (1998), Baumeister (2005)].[2] First, people protect their self-esteem by systematically taking credit for success and denying blame for failure. Second, people have a tendency to uncritically accept positive feedback and eagerly search for flaws/faults in other's criticism [e.g. Baumeister (2005), Greenwald (1980)]. Third and most importantly for our investigation, psychologists have found that conflicts and aggression tend to result from positive self-images that are challenged or threatened [e.g. Baird (1977), Raskin *et al.* (1991), Bushman and Baumeister (1998)]. It is argued that hostile aggression is an expression of the self's rejection of ego-threatening evaluations received from other people [e.g. Baumeister *et al.* (1996)]. People with high self-esteem usually hold confident and highly favorable ideas about themselves, i.e. they exhibit ego-involvement, and react belligerently to ego-threatening feedback from others [e.g. Baird (1977), Shrauger and Lund (1975) and Korman (1970)].

In recent years also economists have started to acknowledge the importance of self-esteem and self-image in decision making and strategic interactions [e.g. Santos-Pinto and Sobel (2005), Köszegi (2006), Bénabou and Tirole (2002), Compte and Postlewaite (2004) and Ellingsen and Johannesson (2008)]. It is argued that people strive for positive self-perceptions because it entails a consumption, signaling and motivational value. Köszegi (2006), for example, endows individuals with 'ego-utility' and demonstrates the effects on choice between more or less ambitious tasks. In particular, this model explains

[2] Self-esteem refers to peoples' self-evaluations or, in other words, the belief they hold about their self-worth. Everywhere people seem to care about it, try to enhance, maintain and protect it. Anything that gives a boost in self-esteem is almost universally welcome. People feel good when their self-perception is high and rising, and people feel bad when it is low or dropping. Hardly anyone enjoys events that constitute a blow or a loss to their self-esteem [Baumeister (2005)].

the phenomenon of overconfidence by individuals who update beliefs according to Bayes' rule. Bénabou and Tirole (2002) and Compte and Postlewaite (2004), on the other hand, center on the motivational value of self-confidence. In Bénabou and Tirole (2002), for example, agents can undertake a project whose payoff depends on the agents' ability. Agents initially hold incomplete information regarding their own ability, but might engage into actions through which their ability is revealed. Given this setting, Bénabou and Tirole (2002) identify, for example, under which circumstances agents protect their self-image by staying willfully ignorant regarding their real ability. Staying ignorant helps the agents to stay motivated to complete the project. Their analysis shows that confidence in one's ability and efficacy can help individuals to undertake more ambitious goals. When people have imperfect knowledge about their own ability and/or when effort and ability are complements, then more self-confidence enhances peoples' motivation to act [see (Bénabou and Tirole, 2002, 873)]. Next to this motivational value of self-esteem within a given contract our contribution emphasizes the ability of self-esteem and ego-protection to facilitate contracts.

The organization of the paper is as follows: In Section 2 we present the principal-agent relation and the psychological payoff structure. As a benchmark, Section 3.1 analyzes the situation of pure moral hazard and determines the optimal effort choice and comparative statics of social welfare in the absence of binding truth-telling constraints. Section 3.2 continues with an analysis of the impact of binding truth-telling constraints on optimal effort choice and social welfare. Sections 4 and 5 discuss and conclude our analysis with some remarks on the practical implications of our model and its robustness.

## 2 The model

In this section we introduce the principal-agent relationship and present a psychological payoff structure which formalizes self-esteem and conflict creation in reaction to ego-threats. We characterize the first best solution and present auxiliary results on the optimality of simple bonus contracts.

**Production Technology** Assume there is a risk-neutral principal, $P$, who decides upon undertaking a project which generates a value of $\phi > 0$ if successful and zero if not. The project requires effort of a capable agent, $A$. A-priori the principal and the agent expect the agent to be capable with probability $\kappa \in (0, 1)$. Whether a *given* agent is actually capable is unobservable to both parties and can only indirectly be learned from project success. If an incapable agent spends any effort on the project, the project will be unsuccessful. On the other hand, if a capable agent spends effort $p \in [0, 1]$, the project will be successful (create value $\phi$) with probability $p$. The project is a complex good or service and its success is neither verifiable nor observable, i.e. contracts contingent on the generation of $\phi$ are not feasible.

**Information Technology** So neither principal nor agent can observe whether the project is successful or not, but both form an opinion about the agent's performance (and thereby his capability) during the production process. Formally, they receive private signals about the agent's performance. The principal receives $s_P \in S_P$, where $S_P = \{L, H\}$, i.e. the principal's opinion can be such that he regards the agent's performance as either high ($H$) or low ($L$). Analogously, the agent receives $s_A \in S_A$ with $S_A = \{L, H\}$. The signals

$s_P$ and $s_A$ are non-verifiable private pieces of information of the principal and the agent, respectively.

The signals are informative with respect to the success of the project. If the project is not successful (which happens with probability $(1-p)$ if the agent is capable and with probability 1 if the agent is incapable), principal and agent receive the signal $s_P = s_A = L$. If the project is successful (which happens with probability $p$ if the agent is capable and with probability zero if he is incapable), the principal and the agent may receive conflicting signals. Specifically, the principal receives the signal $s_P = H$ with probability $g$, the agent receives the same signal with probability $\rho$ and receives $s_A = H$ as an independent signal with probability $x$ in this case. Hence, $g$ measures the quality of the principal's signal, $\rho$ indicates the correlation between the agent's and the principal's signal - or the counter-probability of an independent judgment - and $x$ quantifies the quality of the agent's signal if he forms an independent judgment (*i.e.,* we adopt the specification of the information technology in (MacLeod, 2003, 228)). We assume that the principal's and the agent's signal are imperfect, i.e., $g \in (0,1)$ and $x \in (0,1)$, and positively but imperfectly correlated, i.e., $\rho \in (0,1)$.

For further reference, we denote by $\gamma_{kl}$ the conditional probability that $s_P = k$ and $s_A = l$ given that the project is a success (and the agent is therefore necessarily capable). Then, the ex-ante probability for the signal pair $s_P = L$ and $s_A = H$, for instance, will be[3] $\kappa \cdot p \cdot \gamma_{LH} = \kappa \cdot p \cdot (1-g) \cdot (1-\rho) \cdot x$.

[3] All $\gamma_{kl}$ as functions of $g$, $\rho$, and $x$ can be found in Appendix 6.1.

**The agent's effort costs** For an effort of $p$ the agent incurs costs $v(p)$ with $v \in C^2$, $v(0) = 0$, $v'(0) = 0$, $v''(p) > 0$ and $\lim_{p \to 1} v(p) = \infty$.

**The Game** The timing of the game is as follows:

1. The principal offers a contract to the agent and the agent decides upon acceptance. Upfront payments are arranged.

2. The agent decides upon effort $p$.

3. The principal receives $s_P$ and the agent receives $s_A$. The principal and the agent report (not necessarily truth-fully) on $s_P$ and $s_A$. Denote the reports by $t_P$ and $t_A$, respectively. $t_P$ and $t_A$ are verifiable.

4. The payments contingent on $t_P$ and $t_A$ are arranged.

5. Contingent on $s_A$ and the received payment, the agent decides upon retaliation (with effort $q$).

6. The project generates value $\phi$ for the principal with probability $\kappa \cdot p$ and 0 otherwise.

**First Best Effort Level** Had the principal access to the agent's production technology (and would she herself also be capable with probability $\kappa$), her effort choice would solve $v'(p) = \kappa \cdot \phi$. For further reference, we will denote the first best effort level by $p_{FB}$ and the respective surplus by $\Pi_{FB}$. Our assumptions on $v(p)$ ensure that $p_{FB} \in (0, 1)$.

**Psychological payoffs, conflict and the agent's utility** We assume the agent is risk-neutral and creates conflict if his own performance assessment exceeds the one of the principal [see lab behavior in Sebald and Walzl (2014)]. I.e. if the agent receives the signal $s_A = H$ and the principal sends the signal $t_P = L$, the agent spends effort $q$ on conflict creation. For example, the agent might start-up a law suit, refuses to cooperate on other tasks, steals from the workplace etc. A simple specification of the agent's preferences that induces this behavior is given by the following utility function.

$$U = w - v(p) - Y(s_A, t_P)(1 - q) - c(q). \tag{1}$$

where $w$ denotes the wage payment, $Y(s_A, t_P)$ represents the agent's psychological payoff for signal $s_A$ and report $t_P$, $q$ is the level of conflict (or retaliation) created by the agent and $c(q)$ is the agent's cost for the level of conflict $q$ with $c \in C^2$, $c(0) = 0$, $c'(0) = 0$, $c''(q) > 0$ and $\lim_{q \to 1} c(q) = \infty$. To capture that the agent retaliates if and only if $s_A = H$ and $t_P = L$, we set $Y(H, L) \equiv Y > 0$ and $Y(s_A, t_P) = 0$ for all other combinations of $s_A$ and $t_P$. The assumptions on $c(\cdot)$ and $Y(\cdot, \cdot)$ ensure that the agent spends a retaliation effort $q^* > 0$ if and only if $s_A = H$ and $t_P = L$. Note that the higher the psychological costs created by the difference in the principal's and agent's evaluation ($Y$), the higher the level of conflict $q^*$.

This ad-hoc specification of the agent's preferences can also be regarded as a reduced form representation of a general preference structure that captures a propensity to protect self-esteem against ego threatening information through the creation of conflict as described in the social psychological literature (see Baird (1977), Shrauger and Lund (1975), and Korman (1970)). To see this recall that the agent knows that he is capable whenever he receives $s_A = H$ and that he up-dates the probability with which he expects himself to be capable to

$$\widetilde{\kappa}_A = \kappa \cdot \frac{1 - p(1 - \gamma_{HL} - \gamma_{LL})}{1 - \kappa \cdot p(1 - \gamma_{HL} - \gamma_{LL})} < \kappa \tag{2}$$

if he receives $s_A = L$. Following social psychology literature, we expect "ego-involvement" after the agent received $s_A = H$. In other words, learning that he is capable boosts the agent's self-esteem and increases his sensitivity to the principal's assessment. Following Baumeister (2005) and Greenwald (1980), if the agent's own evaluation is $s_A = H$ the agent is expected to uncritically accept a positive feedback by the principal ($t_P = H$) but to suffer if the principal puts his capability into question by announcing $t_P = L$ as $s_P = L$ updates the agent's expected capability from the principal's perspective to

$$\widetilde{\kappa}_P = \kappa \cdot \frac{1 - p(1 - \gamma_{LH} - \gamma_{LL})}{1 - \kappa \cdot p(1 - \gamma_{LH} - \gamma_{LL})} = \kappa \cdot \frac{1 - p \cdot g}{1 - \kappa \cdot p \cdot g} < \kappa. \tag{3}$$

The agent simply knows that he is capable but the principal's feedback (and the associated wage) communicates (perhaps also to the outside world) that his capability is below the expected capability of a randomly chosen individual. This suggests the following psychological cost structure in case $s_A = H$: $Y(H, H) = 0$ and $Y(H, L) > 0$. In contrast, if $s_A = L$, the agent does not exhibit ego-involvement and holds low self-esteem as he himself perceives his own capability as below average. Given this, he may – psychologically – neither suffer nor gain from the principal's feedback: $Y(L, H) = 0$ and $Y(L, L) = 0$.[4]

In contrast to the standard economic models of reciprocity [see Rabin (1993), Dufwen-

[4] Alternatively, one could assume that the agent enjoys a positive feedback by the principal or suffers from a bad feedback when $s_A = L$. Both assumptions do not alter our central finding that self-esteem and ego-protection facilitates the implementation of positive efforts, but alter agency costs. A positive psychological payoff from a positive feedback is a substitute for the bonus paid by the principal and therefore reduces agency costs for a given effort level. In turn, $Y(H, L) > Y(L, L) > 0$, i.e. lower but non-zero psychological costs from a low performance feedback even if $s_A = L$ tightens the agent's participation constraint and thereby increases agency costs.

berg and Kirchsteiger (2004), Falk and Fischbacher (2006), and Hart and Moore (2008)], the agent in our setting does not measure the principal's kindness by the monetary consequences of the principal's actions but by the induced psychological costs. To be specific, suppose the agent received $s_A = H$ and let him regard a positive feedback neither as kind nor unkind – e.g. because he expects the principal to send a positive feedback (and pay a bonus) only because she is afraid of retaliation – but let him consider $t_P = L$ as unkind (because the principal implicitly communicates a capability of $\kappa_P < \kappa$ even though he considers it likely that the agent knows his capability and suffers from this ego-threat). Then, the agent would create no conflict after positive feedback but reciprocate a negative feedback with conflict creation as also featured by the preferences in Eq. 1.

**Principal** In contrast to the agent, the principal only cares about his profit

$$\Pi = \kappa \cdot p \cdot \phi - E\{w\} - E\{q\} \cdot \psi, \tag{4}$$

where $\kappa \cdot p \cdot \phi$ is the expected benefit generated by the agent, $E\{w\}$ are the expected wage cost of employing the agent, and $E\{q\}\psi$ are the expected costs of conflict due to retaliation.

**Contracts** In our setting with unobservable effort and subjective measures of performance, a contract $\Gamma$ can only be contingent on the reported subjective opinions of the principal and the agent. Hence, a contract fixes payments for all configurations of reports $t_P$ and $t_A$ and reads $\Gamma = \{w_{kl} \mid k \in S_P, l \in S_A\}$. The agent accepts a contract if he expects a (weakly) positive gain from it relative to his outside option (individual rationality) and chooses $p$ as to maximize his utility (incentive compatibility). Moreover, principal and agent report their opinions, i.e. signals, truthfully if and only if they weakly benefit from doing so. If a contract $\Gamma$ is individually rational and the agent chooses effort $p$, we say that $\Gamma$ *implements* $p$.

**Cost Minimizing Contracts** How do optimal contracts look like given that effort is unobservable, performance measures are subjective and agents try to protect a positive self-image through the creation of conflict? A standard application of the revelation principle implies that we can restrict ourselves to simple bonus contracts without any loss of generality.

**Lemma 1.** ***Reduced Form Contracts***

*Suppose there exists a contract $\Gamma$ which implements $p > 0$. Then, there always exists a contract $\hat{\Gamma}$ which implements $p$ at weakly lower costs and*

*(i) Principal and agent tell the truth.*

*(ii) $w_{kl} = w_{km} \equiv w_k$ for all $k \in S_P$ and $l, m \in S_A$.*

*(iii) $w_H > w_L$.*

*Proof.* See Appendix 6.2. □

For convenience, we define $w_H = f + b$, $w_L = f$ and $\Gamma = (b, f)$. By Lemma 1(iii), $b > 0$.[5]

[5] $f$ can be interpreted as an up-front payment or a franchise fee with a payment of zero at Stage 4 if the principal reports $t_P = L$ and a payment of $b$ (a bonus) if she reports $t_P = H$.

# 3 Optimal Contracts

In this section we derive optimal contracts and analyze the corresponding implications for social welfare. The principal's problem is to design a contract $\Gamma = (f, b)$ that maximizes his profit given the problem of moral hazard and truth-telling that is inherent to our setting. First, a contract implements an effort $p$ only if choosing effort $p$ is incentive compatible and individually rational for the agent as the effort is unobservable to the principal (moral hazard problem). Second, for a contract to credibly offer a bonus $b$ in case the principal receives the signal $s_P = H$, it has to be incentive compatible for the principal to truthfully reveal his signal (truth-telling problem). We will proceed with an analysis of the moral hazard problem (under the assumption that truth-telling is credible) and subsequently analyze in how far the corresponding optimal contracts contain bonuses that induce truth-telling by the principal or have to be modified because of the truth-telling problem. Finally, we will investigate how characteristics of the project ($\phi$), the information technology ($g, \rho, x$), conflict ($\psi$), and the agent's expected capability ($\kappa$) influence optimal contracts and social welfare.

## 3.1 Pure Moral Hazard Problem

Throughout this subsection we assume that the principal truthfully reveals his signal $s_P$. Hence, we focus on the moral hazard problem.

**Incentive Compatibility** For a given contract $\Gamma = (f, b)$, the agent chooses effort $p$ so as to maximize his utility (see Eqn. 1) while anticipating the generation of ex-post conflict at level $q^*$. This means, he maximizes

$$U(p) = \kappa \cdot p \cdot (\gamma_{HH} + \gamma_{HL}) \cdot b + f - v(p) - \kappa \cdot p \cdot \gamma_{LH} \cdot (Y(1-q^*) + c(q^*))$$

which induces the first order condition[6]

$$\begin{aligned} b(p) &= \frac{\frac{1}{\kappa} \cdot v'(p) + \gamma_{LH} \cdot (Y(1-q^*) + c(q^*))}{\gamma_{HH} + \gamma_{HL}} \\ &= \frac{1}{g} \cdot (\frac{1}{\kappa} \cdot v'(p) + (1-g) \cdot (1-\rho) \cdot x \cdot (Y(1-q^*) + c(q^*))). \end{aligned} \tag{5}$$

Note that $\frac{d^2U(p)}{dp^2} = \frac{v''(p)}{\kappa \cdot g} > 0$. Eqn. (5) shows that the incentive compatible bonus that the principal pays to the agent in case she receives the signal that the agent did a good job has to overcome marginal effort costs *and* marginal psychological costs. If the principal wants to induce a positive effort level, he has to offer a positive bonus. Note, however, that the required bonus does not vanish in the limit of small efforts, because marginal psychological costs do not vanish for $p = 0$. No bonus below $\underline{b} \equiv b(p = 0) = \frac{1-g}{g} \cdot (1-\rho) \cdot x \cdot (Y(1 - q^*) + c(q^*)) > 0$ can ever induce a positive effort.

For further reference observe that the incentive compatible bonus $b(p)$ increases in the target effort $p$ and in the expected psychological costs of conflict but decreases in the a-priori probability with which the agent is capable, $\kappa$, and the quality of the principal's signal $g$ (i.e. the probability with which a successful project is recognized by the principal) as both parameters increase the agent's return to effort.[7]

[6] We denote a bonus which implements an effort level of $p$ by $b(p)$.

[7] For a formal account of the comparative statics of bonuses see Section 6.3.

**Individual Rationality** The agent accepts a contract $\Gamma = (f, b)$ whenever his expected utility from it is weakly larger than his outside option. If we normalize the outside option to zero, this condition reads

$$\kappa \cdot p \cdot (\gamma_{HH} + \gamma_{HL}) \cdot b + f - v(p) - \kappa \cdot p \cdot \gamma_{LH} \cdot (Y(1-q^*) + c(q^*)) \geq 0.$$

To maximize her profits, the principal sets the upfront payment for a given bonus $b$ to

$$f(b) = -\kappa \cdot p \cdot (\gamma_{HH} + \gamma_{HL}) \cdot b + v(p) + \kappa \cdot p \cdot \gamma_{LH} \cdot (Y(1-q^*) + c(q^*)).$$

Observe that the upfront-payment can well be negative (i.e. a franchise fee) as the agent is not protected by limited liability. Note in particular that $f(b)$ can always be fixed such that the agent does not receive any rents from the relationship and we can consider the principal's profits as social welfare.

What are now the principal's costs to implement an effort level $p > 0$ on the basis of these incentive compatibility and individual rationality constraints?

**The principal's costs of effort** To implement effort $p > 0$ the principal's costs are $C(p) = f + \kappa \cdot p \cdot g \cdot b(p) = v(p) + \kappa \cdot p \cdot \gamma_{LH} \cdot ((1-q^*)Y + c(q^*))$. Note that $C(p)$ is convex and that $C(0) = 0$.

**Optimal Effort** The principal's profit now reads

$$\Pi(p) = \kappa \cdot p \cdot \phi - \kappa \cdot p \cdot \gamma_{LH} \cdot q^* \psi - C(p)$$

which is zero for $p = 0$ and concave for $p > 0$. We denote the unique maximum of $\Pi(p)$ on $[0,1]$ by $\widetilde{p}$, the corresponding bonus $b(\widetilde{p})$ by $\widetilde{b}$, and the corresponding profit for the principal by $\widetilde{\Pi}$, i.e. $\widetilde{p}$, $\widetilde{b}$, and $\widetilde{\Pi}$ are equilibrium efforts, bonuses and profits in the pure moral hazard case.

**Proposition 1.** ***Pure Moral Hazard***

*(i) For* $\phi > \underline{\phi} \equiv \gamma_{LH}(q^*\psi + ((1-q^*)Y + c(q^*)))$, $\widetilde{p} > 0$; *for* $\phi \leq \underline{\phi}$, $\widetilde{p} = 0$.

*(ii) Suppose* $\phi > \underline{\phi}$. *Then,* $\frac{d\widetilde{p}}{d\phi} > 0$, $\frac{d\widetilde{p}}{d\psi} < 0$, $\frac{d\widetilde{p}}{dg} > 0$, $\frac{d\widetilde{p}}{d\rho} > 0$, $\frac{d\widetilde{p}}{dx} < 0$, *and* $\frac{d\widetilde{p}}{d\kappa} > 0$.

*(iii) Suppose* $\phi > \underline{\phi}$. *Then,* $\frac{d\widetilde{\Pi}}{d\phi} > 0$, $\frac{d\widetilde{\Pi}}{d\psi} < 0$, $\frac{d\widetilde{\Pi}}{dg} > 0$, $\frac{d\widetilde{\Pi}}{d\rho} > 0$, $\frac{d\widetilde{\Pi}}{dx} < 0$, *and* $\frac{d\widetilde{\Pi}}{d\kappa} > 0$.

*Proof.* See Appendix 6.4 □

Part (i) indicates that a relationship with positive effort level (i.e. $\widetilde{p} > 0$) can only be established if the value of the project exceeds the expected costs of retaliation for the principal and the expected compensation for the psychological costs of the agent. If a relationship is established because the value of the project is above this threshold, the comparative statics of the optimal effort level $\widetilde{p}$ and profits $\widetilde{\Pi}$ are straightforward. Increasing the value of the project $\phi$, decreasing the probability of conflict $\gamma_{LH} = (1-g) \cdot (1-\rho) \cdot x$, or decreasing agency costs $g \cdot b(p)$ (via an increase of $g$ or $\rho$, or a decrease of $x$) increases the marginal return to effort or decreases the marginal costs of effort and thereby enhances optimal profits.

Likewise, a higher level of conflict reduces optimal efforts (as conflict only arises if the project has been successful) and profits. In turn, a higher ex-ante capability of agents $\kappa$ increases the (expected) value of the project $\kappa \cdot \phi$ but also amplifies the probability of conflict (as conflict only arises when agents are capable) and agency costs (as only capable agents retaliate). However, as long as $\phi > \underline{\phi}$ the positive effect on expected returns to effort is stronger than the negative effect on expected conflict and agency costs such that a higher frequency of capable agents unambiguously enhances optimal efforts and profits in the pure moral hazard case.

As the agent does not receive any rents in the optimal contract, the principal's profit also measures the surplus of the relationship. Hence, in the case of non-binding truth-telling constraints, conflicts (i.e. their likelihood $\gamma_{LH}$ and size $q^* \cdot \psi$) only have a welfare detrimental effect while a higher capability always enhances welfare. Any property of the information technology which reduces conflict (i.e. an increase in $g$ or $\rho$) is welfare-enhancing, while an increase in the quality of the agent's independent judgment $x$ induces the adverse effect.

In this section we have abstracted from the truth-telling problem, i.e. we have concentrated on the case of non-binding truthtelling constraints, to isolate the impact of moral hazard. In the following section, we analyze the robustness of these findings in the presence of truth-telling constraints.

## 3.2 Truth-Telling Problem

The principal's profit contingent on the agent's signal and the principal's report can be represented in the following table (with the principal's report depicted in the rows and the agent's signal depicted in the columns). We denote the expected project value (contingent on the principal's own signal and the (truthfully revealed) signal of the agent by $E\{\phi|s_P, s_A\}$.

| | H | L |
|---|---|---|
| $H$ | $E\{\phi \vert s_P, s_A\} - f - b$ | $E\{\phi \vert s_P, s_A\} - f - b$ |
| $L$ | $E\{\phi \vert s_P, s_A\} - f - q^*\psi$ | $E\{\phi \vert s_P, s_A\} - f$ |

**Credible Bonuses** Suppose $s_P = H$, i.e. the principal knows that the agent is capable. Then, the principal tells the truth, whenever her payoff from doing so (which reads $p\phi - f - b$) is larger than her payoff from reporting $t_P = L$ (which reads p$\phi - f - Pr(s_A = H \mid s_P = H)q^*\psi$). This means the principal reports $t_P = H$ if

$$b \leq \frac{\gamma_{HH}}{(\gamma_{HH} + \gamma_{HL})} \cdot q^* \cdot \psi = (\rho + (1 - \rho) \cdot x) \cdot q^* \cdot \psi \equiv b^{max}. \tag{6}$$

The principal can only credibly promise a bonus $b$ that does not exceed $b^{max}$. The upper bound to credible bonuses thereby increases in the probability that $s_A = H$ whenever $s_P = H$ (i.e. $\frac{db^{max}}{d\rho,x} > 0$) – the more likely this event, the higher the probability of conflict if the principal reports $t_P = L$ instead of $t_P = H$. Moreover, the larger the level of conflict $q^* \cdot \psi$ the higher is the maximal credible bonus as the principal tells the truth only if she is sufficiently afraid of the possible conflict this creates. Note for further reference that $b^{max}$ is independent of the quality of the principal's signal $g$ and the likelihood of the agent being capable $\kappa$ (as the principal received a good signal and the agent is capable whenever $s_P = H$).

If $s_P = L$, the principal up-dates the probability that the agent is capable to $\widetilde{\kappa}_P < \kappa$. Given this, she tells the truth, whenever her payoff from doing so (which reads $\widetilde{\kappa}_P \cdot p \cdot \phi - f - Pr(s_A = H \mid s_P = L)q^* \cdot \psi$) is larger than her payoff from reporting $t_P = H$ (which reads $\widetilde{\kappa}_P \cdot p \cdot \phi - f - b$). Hence, the principal reports $t_P = L$ if

$$b \geq \frac{\widetilde{\kappa}_P \cdot p \cdot \gamma_{LH}}{1 - g \cdot \widetilde{\kappa}_P \cdot p} \cdot q^* \cdot \psi = \frac{\widetilde{\kappa}_P \cdot p \cdot (1-g) \cdot (1-\rho) \cdot x}{1 - g \cdot \widetilde{\kappa}_P \cdot p} \cdot q^* \cdot \psi \equiv b^{min}. \tag{7}$$

The principal can also not promise to pay arbitrarily low bonuses as he has an incentive to evade conflict through an 'unconditional bonus'. By paying the bonus independently of his signal, the principal avoids any conflict with an agent who is prepared to protect his positive self-image. The minimal credible bonus is thereby increasing in the probability with which the principal expects the agent to receive $s_A = H$ when she herself received $s_P = L$ (i.e. $\frac{db^{min}}{dg,\rho} < 0$ and $\frac{db^{min}}{dx} > 0$) – the more likely this event, the higher the probability of conflict if the principal reports $t_P = L$ instead of $t_P = H$.

Also $b^{min}$ is increasing in the probability that the agent is capable $\kappa$ and the agent's effort $p$ as only successful projects can lead to the signal configuration $s_A = H$ and $s_P = L$ that tempts the principal to pay an unconditional bonus. Moreover, the larger the level of conflict $q^* \cdot \psi$ the higher is the minimal credible bonus as the principal tells the truth only if she is not too afraid of the possible conflict this creates.

For the discussion of optimal contracts, we summarize the following properties of $b^{min}$ and $b^{max}$.

**Lemma 2.** ***Comparative Statics of*** $b^{max}$ ***and*** $b^{min}$

*Suppose* $p > 0$. *(i)* $b^{min} > 0$. *(ii)* $b^{max} > b^{min}$. *(iii)* $\Delta b \equiv b^{max} - b^{min}$ *is monotone increasing in* $q^*$ *and* $\psi$. *(iv)* $b^{min}$ *is monotone increasing in* $q^*$, $\psi$, $\kappa$ *and* $x$ *and monotone decreasing in* $\rho$ *and* $g$. *(v)* $b^{max}$ *is monotone increasing in* $q^*$, $\psi$, $\rho$ *and* $x$.

*Proof.* See Appendix 6.3 □

**Welfare** For the pure moral hazard case, Proposition 1 indicates that the principal offers a bonus $b(\widetilde{p}) = \widetilde{b} > 0$ whenever the project value exceeds $\underline{\phi}$ (and offers no contract otherwise). For $\phi > \underline{\phi}$, the optimal bonus $\widetilde{b}$ is larger than the minimal bonus $\underline{b} = \frac{1-g}{g} \cdot (1-\rho) \cdot x \cdot (Y(1-q^*) + c(q^*))$ that is necessary to compensate the agent for his psychological costs.

Denote the maximal bonus for which profits are non-negative by $\overline{b}$. Plotting the principal's profit in the pure moral hazard case as a function of the bonus payments $b$ therefore results in a continuous function that is positive and monotone increasing on $(\underline{b}, \widetilde{b})$, maximal at $\widetilde{b}$, positive and and monotone decreasing on $(\widetilde{b}, \overline{b})$, and non-positive elsewhere.

Depending on how the minimal and maximal credible bonuses $b^{min}$ and $b^{max}$ relate to the minimal and maximal bonuses that allow positive profits for the principal $\underline{b}$ and $\overline{b}$, and to the optimal bonus $\widetilde{b}$, the truth-telling problem has a negative or no impact on welfare as the following proposition indicates. We denote the optimal bonus in the presence of the truth-telling problem by $b^*$ and the corresponding profit by $\Pi^*$.

**Proposition 2.** ***Moral Hazard and Truth-Telling***

*Let $\phi > \underline{\phi}$. Then, optimal contracts can be described as follows.*

*Case 1: Breakdown of the relationship:*

*If $b^{max} \leq \underline{b}$ or $b^{min} \geq \underline{b}$, then $b^* = 0$ and $\Pi^* = 0 < \widetilde{\Pi}$.*

*Case 2: No impact of truth-telling:*

*If $\widetilde{b} \in [b^{min}, b^{max}]$, then $b^* = \widetilde{b}$ and $\Pi^* = \widetilde{\Pi}$.*

*Case 3: Truth-telling distortions:*

*If* $b^{max} \in (\underline{b}, \widetilde{b})$, $b^* = b^{max}$ *and* $\widetilde{\Pi} > \Pi^* > 0$;

*If* $b^{min} \in (\widetilde{b}, \overline{b})$, $b^* = b^{min}$ *and* $\widetilde{\Pi} > \Pi^* > 0$.

*Proof.* See Appendix 6.5 □

As $b^{max} > b^{min} > 0$, the list of relative positions between credible bonuses ($b^{min}$ and $b^{max}$), the optimal bonus in the pure moral hazard case $\widetilde{b}$, and bonuses that ensure positive profits ($\underline{b}$ and $\overline{b}$) in Proposition 2 is exhaustive. As $\widetilde{b}$ (and $\overline{b}$) take any value larger than $\underline{b}$ depending on $\phi$, any of the Cases 1-3 can become relevant for given $b^{min}$ and $b^{max}$.

[Figures 1-3 here]

Case 1 is relevant if either the conflict level $\psi$ is too small (i.e. sufficiently close to zero) such that there is no credible bonus that is large enough to cover the agent's psychological costs ($b^{max} \leq \underline{b}$); or if the conflict level is too large (i.e. $\psi$ is large but $\phi$ is sufficiently close to $\underline{\phi}$) such that there is no credible bonus that is small enough to ensure a positive profit for the principal ($b^{min} \geq \overline{b}$). In this case, the principal cannot implement a positive effort by the agent and cannot be better off than by offering the null contract $\Gamma_0 = (f = 0, b = 0)$. As $\phi > \underline{\phi}$, the truth-telling problem therefore strictly reduces welfare compared to the pure moral hazard situation.

If $b^{max} \leq \underline{b}$ the principal would benefit from a sufficient increase of $b^{max}$, i.e. more conflict $q^* \cdot \psi$, more correlated signals, or a better independent judgement of the agent. If $b^{min} \geq \overline{b}$ the principal would benefit from a sufficient decrease of $b^{min}$, i.e. more correlated signals or a better signal for herself, but less conflict $q^*\psi$, a worse independent judgement of the agent $x$, and a lower expected capability $\kappa$.

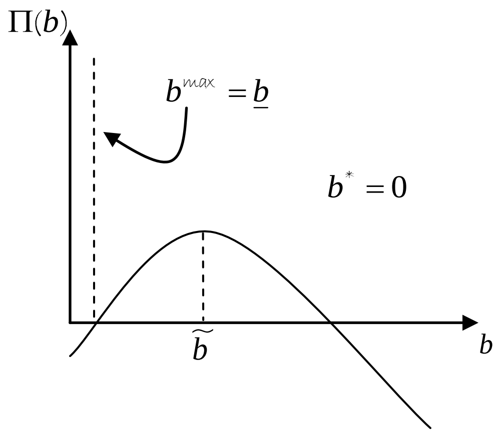


Figure 1: Breakdown of the relationship (Case 1)

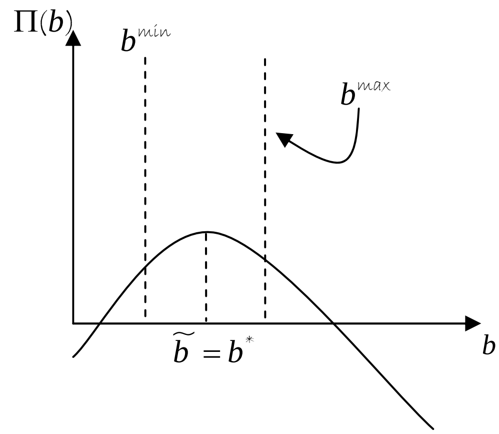


Figure 2: No impact of truth-telling (Case 2)

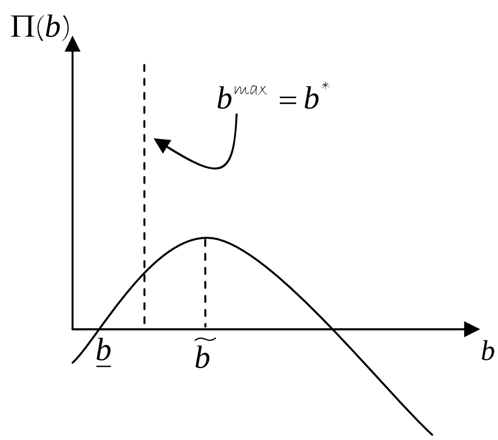


Figure 3: Truth-telling distortions (Case 3)

For intermediate levels of conflict either Case 2 or 3 become relevant. In Case 2, the truth-telling problem has no impact on optimal contracts and welfare as the optimal bonus for the pure moral hazard case $\widetilde{b}$ is a credible bonus. Then, the impact of conflict, capability, and the information technology on optimal contracts and welfare is as displayed in Proposition 1.

In Case 3, there are credible bonuses that allow the principal to induce a positive effort by the agent and to generate a positive profit. However, credible bonuses are either lower than the optimal bonus (i.e., $b^{max} < \widetilde{b}$) such that profits are increasing over the range of credible bonuses and the principal pays $b^{max}$; or credible bonuses are larger than the optimal bonus (i.e., $b^{min} > \widetilde{b}$) such that profits are decreasing over the range of credible bonuses and the principal pays $b^{min}$.

In both situations, parameters of the model potentially have a twofold impact on welfare. On the one hand, conflict levels, capability, and the information technology directly affect profits as in the moral hazard case, on the other hand these parameters also alter the set of credible bonuses.

As displayed in the following results, the quality of the principal's signal and the correlation of signals unambiguously raises welfare because both parameters directly enhance profits *and* relax the truth-telling constraint. In contrast, the conflict level $\psi$ and the quality of an independent judgement by the agent $x$ reduce the principal's profit for a given effort by the agent but may relax the truth-telling constraint. Likewise, the capability rate $\kappa$ raises profits for a given effort but a lower expected capability may also enlarge the set of credible bonuses in a profitable way for the principal.

**Impact of the Principal's Signal Quality and Signal Correlation**

**Result 1.** *(i)* $\frac{d\Pi^*}{d\rho,g} \geq 0$*; In Case 2 and 3,* $\frac{d\Pi^*}{d\rho,g} > 0$*. (ii) For* $g = 1$ *or* $\rho = 1$ *or* $x = 0$*,* $p^* = p_{FB}$ *and* $\Pi^* = \Pi_{FB}$ *whenever* $\phi \leq g(\rho + (1-\rho)x)q^*\psi$*.*

*Proof.* The proofs for Results 1-3 are in Appendix 6.6. □

For a given effort level, $g$ and $\rho$ reduce agency costs and thereby enhance the principal's profit. Moreover, $g$ and $\rho$ reduce the minimal credible bonus and $\rho$ raises the maximal credible bonus while $g$ has no impact in this case. Hence, $g$ and $\rho$ weaken the truth-telling problem as they enlarge the set of credible bonuses. This leads to an unambiguously positive effect on welfare. This welfare effect is strictly positive if another contract than the null contract is profitable and credible (Part (i)). Whenever the principal always correctly identifies a successful project ($g = 1$) or the signals are perfectly correlated ($\rho = 1$) or the agent never receives $s_A = H$ independent of the principal's signal ($x = 0$), the signal configuration $s_P = L$ and $s_A = H$ never arises (i.e. $\gamma_{LH} = 0$) and the principal's profit function reads $\Pi = \kappa \cdot p\phi - v(p)$ such that the principal chooses to implement $p_{FB}$ whenever the corresponding bonus $b(p_{FB}) \in [b^{min}, b^{max}]$. As $b^{min} = 0$ whenever either $g = 1$, $\rho = 1$, or $x = 0$, this condition boils down to $b(p_{FB}) \leq b^{max}$.

**Impact of the Agent's Independent Judgement and the Conflict Level**

**Result 2.** *(i) Let* $b^{max} \leq \underline{b}$*, then* $\frac{d\Pi^*}{dx,\psi} \geq 0$*. (ii) Let* $b^{max} > \underline{b}$*, then* $\frac{d\Pi^*}{dx,\psi} \leq 0$ *if* $\phi > \underline{\phi}$ *is sufficiently close to* $\underline{\phi}$*; and* $\frac{d\Pi^*}{dx,\psi} > 0$ *if* $\phi$ *is sufficiently large.*

If credible bonuses are too low to compensate for the agent's psychological costs (Part (i)), the principal never suffers from a better independent judgement of the agent or a larger conflict level as a larger $x$ or $\psi$ may push $b^{max}$ beyond $\underline{b}$ and therefore allow for positive profits. Whenever there are credible bonuses that compensate the agent for his psychological costs (Part (ii)), the principal suffers from larger $x$ or $\psi$ if the project value is so low that $\widetilde{b}$ is credible or $\widetilde{b}$ is too low to be credible. In the former case, the negative impact of $x$ or $\psi$ on profits is driven by enhanced agency costs as in the pure moral hazard situation. In the latter case, the principal would benefit from lower credible bonuses but larger $x$ or $\psi$ only pushes $b^{min}$. In contrast, if project values are sufficiently large, $\widetilde{b}$ is too large to be credible and the principal benefits from $b^{max}$ being increasing in $x$ or $\psi$. If $\phi$ is sufficiently large, profits increase so steeply in the agents effort $p$ such that this effect overcompensates the negative direct impact of $x$ or $\psi$ on profits for a given effort by the agent.

**Impact of Capability** Compared to the likelihood and the level of conflict, the probability of a capable agents $\kappa$ has a rather different impact as it (i) raises profits for a given effort level, (ii) does not alter the maximal credible bonus but (iii) increases the minimal credible bonus.

**Result 3.** *(i) Let $b^{min} \geq \overline{b}$, then $\frac{d\Pi^*}{d\kappa} \leq 0$. (ii) Let $b^{min} < \overline{b}$, then $\frac{d\Pi^*}{d\kappa} > 0$ if $\phi$ is sufficiently large.*

If credible bonuses are too high to allow for positive profits of the principal (Part(i)), the principal never gains from a higher frequency of capable agents. Only a smaller $\kappa$ may reduce $b^{min}$ below $\overline{b}$ and thereby allow for positive profits. In fact, as $b^{min} = 0$ for $\kappa = 0$ there is always a capability level that is small enough (together with a project value $\phi$ that is large enough) to allow for credible bonuses that generate positive profits.

Whenever there are credible bonuses that allow for positive profits (Part (ii)), the principal gains from a higher capability rate $\kappa$ unless increasing $b^{min}$ by a larger $\kappa$ pushes credible bonuses too far beyond the optimal bonus $\widetilde{b}$. This can be excluded for sufficiently large project values such that $b^{max} < \widetilde{b}$.

# 4 Concluding Discussion

The analysis of our model revealed that the individuals' eagerness to protect their self-image in response to ego-threatening feedback regarding their capability may facilitate principal-agent relationships even if performance signals are subjective and no third-party can enforce truth-telling. In particular, we analyzed the impact of the conflict level, the quality of the information technology, and ex-ante capability on optimal effort levels and social welfare.

**Conflict Level** Conflict as modeled in this paper unambiguously reduces optimal effort levels and social welfare in the absence of truth-telling constraints. In the presence of truth-telling constraints, however, we demonstrate that some conflict potential is needed to establish a positive effort by the agent. Furthermore we show that enhanced conflict levels have a positive effect on social welfare in the case of valuable projects which require substantial bonus payments to the agent. Hence, a well-functioning (internal or external) processing of appeals against managerial decision making is not only providing a more peaceful workforce, it may also implement the conflict level needed to make announced bonus payments credible and thereby raise firm profits.

**Information Technology** The different parameters that determine the quality of the principal's and the agent's signal have a diverse impact on welfare in the absence and in the presence of truth-telling constraints. First of all, the principal is advised to use a signal technology which displays a high correlation between her own and the agent's signal. With perfectly correlated signals the probability of conflicting signals is zero such that the agent does not expect any psychological costs. Moreover, the lower (upper) truth-telling constraint is decreasing (increasing) in the signal correlation such that the interval of credible bonuses is maximized for a given conflict level. Whenever the first best bonus is credible due to a sufficiently large conflict level, perfectly correlated signals will allow the principal to implement first best. This lends additional support to the practice of using information for performance evaluation which is not necessarily highly correlated with actual performance but ensures a high correlation with the agent's self-assessment [see Demougin and Garvie (1991)]. Similarly, the probability of conflict will be zero if the principal always identifies a good project ($g = 1$) or the agent does not observe good performance independent of the principal ($x = 0$). Hence, a first best can also be achieved in these situations.

But while the quality of the principal's signal and the correlation of signals improves welfare in the absence and presence of truth-telling constraints, the quality of the agent's independent judgement is not unambiguously welfare reducing. While an independent judgement raises agency costs and thereby only reduces welfare under pure moral hazard, it also increases the probability of conflict if the principal receives a high signal and cheats on the agent. This relaxes the truth-telling constraint in a potentially welfare enhancing manner.

**Capability** The ambiguous welfare effect of a higher ex-ante capability of agents follows a completely different logic. While a higher frequency of capable agents always reduces agency costs and enhances profit expectations in the pure moral hazard case, it has a potentially welfare detrimental effect not because it lowers the conflict probability if the principal receives a good signal and is tempted to downgrade it but because it enhances the conflict probability if the principal receives a bad signal and has an incentive to pay an unconditional bonus to avoid the conflict.

**Sensitivity to Ego-Threats** In our analysis we focused on the impact of conflict levels, the information technology, and the ex-ante capability of an agent as these parameters display all the different ways in which welfare with and without truth-telling constraints can be affected. In comparison, the agents psychological costs display a more compound impact on welfare. First of all, some sensitivity is needed to establish the prospect of conflict for the principal and thereby ensure truth-telling. The more aggressive the agent reacts to ego-threats, the higher the anticipated level of conflict and the less restrictive the upper truth-telling constraint. Hence, a more aggressive agent will induce a welfare improvement in case of valuable projects with associated high bonus payments as discussed above. However, the higher the sensitivity of the agent, the larger the required compensation for anticipated psychological costs. This ceteris paribus enhances necessary bonus payments for a given effort level and thereby reduces the principal's profit and social welfare. The ideal agent from the point of view of a principal who wishes to conduct a very valuable project is therefore someone who reacts very strongly to ego-threats (i.e. who has low costs of retaliation) but does not suffer too much from an ego-threat and the corresponding retaliation (e.g. because $q^*$ is large). This reinforces our above-made appraisal of appeal systems and suggests to ensure low costs of conflict creation for the employee (e.g. low costs of law suits etc). Note, however, that these recommendations only hold for very valuable projects which make the upper truthtelling constraint binding. For non-binding truthtelling constraints, psychological sensitivity and the corresponding conflict remain detrimental to the principal's profits and welfare.

# 5 Robustness and Extensions

We have decided to address the impact of ego-threats on principal-agent relationships in a rather simple model. The information technology is binary and never misidentifies a bad outcome, agents are risk-neutral and not protected by limited liability, psychological costs only depend on the configuration of signals and reports, and the principal has neither other agents at her disposal nor can she delegate the task to monitor and pay the agent to a sub-principal. We have opted for these simplifications in order to provide a framework which allows for an easy identification of the relevant effects as discussed in the previous paragraphs. However, several extensions of our basic model may deserve some attention.

First of all, a certain robustness of our results can be expected for more general information technologies. The general impact of conflict and psychological sensitivity in the absence and presence of binding truth-telling constraints does not depend on the exact parametrization of the information technology but rather on the assumption that a tension between the principal's and the agent's signal creates conflict which induces truth-telling by the principal. But depending on the quality of the information technology more or less “fine-tuning” of the conflict level would be needed to establish the implementability of a certain effort. Suppose, for instance, that both, principal and agent, may receive a positive signal even if the project failed. If the agent takes the quality of his information into account, this may reduce his eagerness to retaliate if the principal provides feedback below his own assessment. But it also may increase the incentives of the principal to pay an unconditional bonus if he is afraid of the agent's retaliation. Both effects therefore shrink the interval of credible bonuses and make a breakdown of the principal-agent relationship more likely.

Second, we have chosen to model the agent as risk-neutral and with unlimited liability. While this yields the rather special case of a principal-agent relationship which never leaves a rent to the agent, it clearly helps to isolate the welfare effect of conflict or the information technology etc. when agent's are concerned about self-esteem and engage in ego-protection. With a risk-averse agent, for instance, a change of the information technology does not only affect the probability of conflict but also modifies the agent's need for insurance. Likewise, with an agent who is protected by limited liability, a modified conflict level does not only alter the set of credible bonuses, it also changes the distribution of rents. While ego-protection still facilitates positive effort in these situations, the impact of model parameters certainly becomes less straightforward to analyze.

Third, with our simple signal-dependent specification of psychological costs we do not want to deny that other psychological motives may also play a role in the set-up under consideration. The literature on gift exchange and reciprocity in labor markets [see e.g. Dufwenberg and Kirchsteiger (2000) or Netzer and Schmutzler (2013)] as well as the experimental findings from the incentive treatment in Sebald and Walzl (2014)] suggest that an unconditional bonus – *if* regarded as a gift by the agent – may be superior to an incentive contract in our setting if the agent reciprocates with sufficient effort spending to the gift and retaliates in case of a negative performance report. The difference of our approach to the existing models therefore is not so much the way we in which we model the agent's incentive to retaliate, but rather the fact that the agent's psychological costs in our setting do not depend on the payoff distribution and associated intentions, but on the feedback as such. As suggested by the retaliation pattern we observed in [Sebald and Walzl (2014)] this motive to retaliate adds to the known motives based on gift exchange and (payoff-dependent) reciprocity. It should also be acknowledged that incentive contracts as analyzed in our contribution have the potential to crowd out intrinsic motivations [see e.g. Falk and Kosfeld (2006), Sliwka (2007) and von Siemens (2013)]. While the conflict created by the agent in response to unfavorable performance evaluations can facilitate incentive contracts as demonstrated in our paper, it should also be noted that the corresponding expected costs of conflict for the principal may turn contracts that rely on trust and intrinsic motivations into a more attractive alternative institution.

Finally, we ignored in our model that the principal may delegate the performance evaluation of the agent to a third party, e.g. manager. If he could and the agent continues to create conflict with the principal even though the performance appraisal was given by the third party, the third party may truthfully reveal the signal out of indifference. But as soon the agent also starts to create conflict with the third party, there is a similar incentive problem as the third party will always give a positive feedback to avoid conflict. Instead of employing a third party as a manager, the literature on subjective performance evaluation typically suggests to run a tournament between several agents where the principal commits up-front to pay a certain price for each rank [see e.g. Malcomson (1984), Dur *et al.* (2010), or Berger *et al.* (2013)]. As she committed to the prices, the principal only has to rank agents according to (her subjective impression of) the agent's performance and has no incentive to reveal a false ranking. If agents do not suffer from psychological costs in these kind of tournaments, a first best can be achieved and performance pay as characterized in this paper is never superior. However, it is an empirical question whether tournaments actually lead to lower psychological costs. If self-esteem is threatened fiercely by the explicit announcement that *someone-else* is better, the principal may well face more conflict ex-post. This can lead to an inferiority of such a scheme and promote performance pay.

# References


Baird, L. (1977). Self and superior ratings of performance: As related to self-esteem and satisfaction with supervision. *Academy of Management Journal*, **20** (2), 291–300.

Baumeister, R. (2005). Self-concept, self-esteem, and identity. In J. Derlega, W. Jones and B. Winstead (eds.), *Personality: Contemporary theory and research*, Wadsworth Publishing Company.

—, Smart, L. and Boden, J. (1996). Relation of threatened egotism to violence and aggression: the dark side of high self-esteem. *Psychological review*, **103** (1), 5–33.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, **117** (3), 871–915.

Berger, J., Harbring, C. and Sliwka, D. (2013). Performance appraisals and the impact of forced distribution: An experimental investigation. *Management Science*, **59**, 54–68.

Bureau of National Affairs (1981). *Wage and Salary Administration*. PPF Survey, No. 131, Washington, D.C.

Bushman, B. and Baumeister, R. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of personality and social psychology*, **75** (1), 219–229.

Compte, O. and Postlewaite, A. (2004). Confidence-enhanced performance. *American Economic Review*, pp. 1536–1557.

Demougin, D. and Garvie, D. (1991). Contractual design with correlated information under limited liability. *Rand Journal of Economics*, **22**, 477–489.

Dufwenberg, M. and Kirchsteiger, G. (2000). Reciprocity and wage undercutting. *European Economic Review*, **44** (4-6), 1069–1078.

— and — (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, **47** (2), 268–298.

Dur, R., Non, A. and Roelfsema, H. (2010). Reciprocity and incentive pay in the workplace. *Journal of Economic Psychology*, **31** (4), 676–686.

Ellingsen, T. and Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory. *The American Economic Review*, **98** (3), 990–1008.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, **54** (2), 293–315.

— and Kosfeld, M. (2006). The hidden costs of control. *The American economic review*, pp. 1611–1630.

Gibbs, M., Merchant, K., Van der Stede, W. and Vargus, M. (2004). Determinants and effects of subjectivity in incentives. *Accounting Review*, pp. 409–436.

Greenwald, A. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, **35** (7), 603–618.

Hart, O. and Moore, J. (2008). Contracts as reference points. *Quarterly Journal of Economics*, **123** (1), 1–48.

Korman, A. (1970). Toward an hypothesis of work behavior. *Journal of Applied Psychology*, **54** (1), 31–41.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, **4** (4), 673–707.

Levin, J. (2003). Relational incentive contracts. *The American Economic Review*, **93** (3), 835–857.

MacLeod, W. (2003). Optimal contracting with subjective evaluation. *The American Economic Review*, **93** (1), 216–240.

Malcomson, J. (1984). Work incentives, hierarchy, and internal labor markets. *The Journal of Political Economy*, pp. 486–507.

Milkovich, G. and Wigdor, A. (1991). *Pay for performance: Evaluating performance appraisal and merit pay*. National Academies Press.

Netzer, N. and Schmutzler, A. (2013). Explaining gift-exchange–the limits of good intentions. *Journal of the European Economic Association*.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, **83** (5), 1281–1302.

Raskin, R., Novacek, J. and Hogan, R. (1991). Narcissism, self-esteem, and defensive self-enhancement. *Journal of personality*, **59** (1), 19–38.

Santos-Pinto, L. and Sobel, J. (2005). A model of positive self-image in subjective assessments. *American Economic Review*, pp. 1386–1402.

Sebald, A. and Walzl, M. (2014). Subjective performance evaluations and reciprocity in principal–agent relations. *The Scandinavian Journal of Economics*, **116** (2), 570–590.

Shrauger, J. and Lund, A. (1975). Self-evaluation and reactions to evaluations from others. *Journal of personality*, **43** (1), 94–108.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, pp. 999–1012.

Suvorov, A. and van de Ven, J. (2009). Discretionary rewards as a feedback mechanism. *Games and Economic Behavior*, **67** (2), 665–681.

von Siemens, F. A. (2013). Intention-based reciprocity and the hidden costs of control. *Journal of Economic Behavior & Organization*, **92**, 55–65.

# 6 Appendix

## 6.1 Information Technology

The conditional probabilities $\gamma_{k,l}$ for signal configuration $(s_P = k, s_A = l)$ read

$$\gamma_{HH} = g\left(\rho + (1-\rho)\, x\right) \text{ and } \gamma_{HL} = g(1-\rho)(1-x),$$

$$\gamma_{LL} = (1-g)\left(\rho + (1-\rho)\,(1-x)\right) \text{ and } \gamma_{LH} = (1-g)\,(1-\rho)\, x.$$

Note that $\gamma_{HH}\gamma_{LL} > \gamma_{HL}\gamma_{LH}$.

## 6.2 Proof of Lemma 2

To save on notation, we denote $Y(s_A = k, t_P = l)(1-q^*) - c(q^*) \equiv Y_{kl}$ throughout this proof.

*Part(i).* For a given contract $\Gamma$ and signals $s_P$ and $s_A$, the principal and the agent decide upon their report. Let $\sigma_P : S_P \to \Delta(S_P)$ and $\sigma_A : S_A \to \Delta(S_A)$ be the principal's and agent's reporting strategies (*i.e.,* mappings from the set of signals $S_P$ and $S_A$ to the set of probability distributions over $S_P$ and $S_A$, respectively). Suppose that $(\sigma_P^*, \sigma_A^*)$ is the pair of optimal reporting strategies for contract $\Gamma$. Then, the revelation principle implies that there exists a contract $\hat{\Gamma}$ which implements the same effort at the same costs and induces truthful reports by principal and agent. We will, henceforth, restrict our analysis to this type of (revelation) contracts.

*Part (ii).* Suppose that $\Gamma = \{w_{kl}\}$ is a revelation contract, *i.e.,* the principal and the agent tell the truth under contract $\Gamma$. As $\Gamma$ implements $p > 0$, the incentive compatibility constraint

$$\Sigma_{k\in S_P,l\in S_A}(w_{kl}-Y_{kl})\frac{dPr\{s_P=k,s_A=l\}}{dp}=v'(p)$$

is satisfied. Consider a contract $\hat{\Gamma}$ which fixes payments of $\hat{w}_k=\sum_{l\in S_A}w_{kl}Pr\{s_P=k,s_A=l\}$ if the principal receives signal $s_P=k$, i.e., payments are independent of $s_A$. These payments also satisfy the incentive compatibility constraint (see above).[8] Moreover, the agent weakly benefits from telling the truth. Finally, the principal's truth-telling constraint is also satisfied under $\hat{\Gamma}$. To see this observe that the principal reports $k$ given that he has received $k$ under contract $\Gamma$ if

$$\begin{aligned}&Pr\{s_A=H|s_P=k\}(w_{oH}-w_{kH})+Pr\{s_A=L|s_P=k\}(w_{oL}-w_{kL})\\&\geq Pr\{s_A=H|s_P=k\}((q^*\psi)_{kH}-(q^*\psi)_{oH})+Pr\{s_A=L|s_P=k\}((q^*\psi)_{kL}-(q^*\psi)_{oL})\end{aligned}\tag{8}$$

for all $o\in S_P$ (where $(q^*\psi)_{t_A,t_P}$ denotes the anticipated conflict costs for a reported configuration $(t_A,t_P)$). This set of inequalities holds because $\Gamma$ implements truth-telling by assumption. $\hat{\Gamma}$ implements truth-telling if

$$\hat{w}_o-\hat{w}_k\geq Pr\{s_A=H|s_P=k\}((q^*\psi)_{kH}-(q^*\psi)_{oH})+Pr\{s_A=L|s_P=k\}((q^*\psi)_{kL}-(q^*\psi)_{oL}).$$

holds for all $o,k\in S_P$. Inserting $\hat{w}_k$ and $\hat{w}_o$ yields

$$\begin{aligned}&Pr\{s_A=H|s_P=k\}(w_{oH}-c_{kH})+Pr\{s_A=L|s_P=k\}(w_{oL}-w_{kL})\\&\geq Pr\{s_A=H|s_P=k\}((q^*\psi)_{kH}-(q^*\psi)_{oH})+Pr\{s_A=L|s_P=k\}((q^*\psi)_{kL}-(q^*\psi)_{oL}).\end{aligned}$$

[8] Individual rationality is trivially fulfilled as expected payments for the agent are the same under $\Gamma$ and $\hat{\Gamma}$ and $\Gamma$ is individually rational by assumption.

which coincides with Eqs. 8 and therefore shows that for $\hat{\Gamma}$ the principal's truthtelling constraint is satisfied as well. Hence, any revelation contract $\Gamma$ can be substituted by a revelation contract $\hat{\Gamma}$ with $w_{kl}$ independent of $l$ which also implements $p > 0$ and leaves the principal weakly better off.

*Part (iii).* Suppose by contradiction that $\Gamma$ implements $p > 0$ with $w_H = b$ and $w_L = b + \epsilon$ with $\epsilon \geq 0$. Then, the agent's utility reads

$$U(p) = \kappa p g b + (1 - \kappa p g)(b + \epsilon) - \kappa p \gamma_{LH} Y_{LH} - v(p)$$

and the incentive compatibility constraint $\frac{dU(p)}{dp} = 0$ boils down to

$$\epsilon \kappa g = -\kappa \gamma_{LH} Y_{LH} - v'(p).$$

Observe that the $rhs$ is strictly negative for any $p > 0$ and the incentive compatibility constraint is not satisfied for any $\epsilon \geq 0$. A contradiction.

## 6.3 Comparative Statics of Bonuses

$$b(p) = \frac{\frac{v'(p)}{\kappa} + \gamma_{LH}(Y(1-q^*) + c(q^*))}{\gamma_{HH} + \gamma_{HL}} = \frac{1}{g}(\frac{v'(p)}{\kappa} + (1-g)(1-\rho)x(Y(1-q^*) + c(q^*)))$$

implies the following results.

**Lemma 3.** ***Comparative Statics of*** $b(p)$

*(i) Suppose* $p > 0$*. Then,* $b(p) > 0$*. (ii)* $\lim_{p \to 0} b(p) > 0$*. (iii)* $\frac{db(p)}{dp} > 0$*. (iv)* $\frac{db(p)}{dg} < 0$*. (v)* $\frac{db(p)}{d\rho} < 0$*. (vi)* $\frac{db(p)}{dx} > 0$*. (vii)* $\frac{db(p)}{d\kappa} < 0$*.*

$$b^{max} = \frac{\gamma_{HH}}{(\gamma_{HH} + \gamma_{HL})} q^* \psi = (\rho + (1-\rho)x) q^* \psi$$

$$b^{min} = \frac{\widetilde{\kappa}_P \cdot p \cdot \gamma_{LH}}{1 - \widetilde{\kappa}_P \cdot p \cdot g} \cdot q^* \cdot \psi = \frac{\widetilde{\kappa}_P \cdot p \cdot (1-g) \cdot (1-\rho) \cdot x}{1 - \widetilde{\kappa}_P \cdot p \cdot g} \cdot q^* \cdot \psi$$

implies the following results.

**Lemma 4.** ***Comparative Statics of*** $b^{max}$ ***and*** $b^{min}$

*Suppose* $p > 0$. *(i)* $b^{min} > 0$. *(ii)* $b^{max} > b^{min}$. *(iii)* $\Delta b \equiv b^{max} - b^{min}$ *is monotone increasing in* $q^*$ *and* $\psi$. *(iv)* $b^{min}$ *is monotone increasing in* $q^*$, $\psi$, $\kappa$ *and* $x$ *and monotone decreasing in* $\rho$ *and* $g$. *(v)* $b^{max}$ *is monotone increasing in* $q^*$, $\psi$, $\rho$ *and* $x$.

*Proof.* (i) follows directly from Eqn. 7. (ii) follows from $Pr(s_A = H|s_P = H) > Pr(s_A = H|s_P = L)$. (iii) follows from Eqn. 6 and 7 together with (ii). (iv) and (v) follow directly from Eqs. 6 and 7 and the definition of $\widetilde{\kappa}_P$. □

## 6.4 Proof of Proposition 1

*Part(i).* Consider

$$\Pi(p) = \kappa p \phi - \kappa p \gamma_{LH} q^* \psi - C(p)$$

with $C(p) = v(p) + \kappa p \gamma_{LH}((1-q^*)Y + c(q^*))$. Observe that we can write $\Pi(p) = a\kappa p - v(p)$ with $a = \phi - \gamma_{LH}(q^*\Psi + ((1-q^*)Y + c(q^*)))$. Recall that $v(0) = 0$, $v'(0) = 0$, and $v''(p) > 0$. Then, $\widetilde{p} > 0$ if and only if $a > 0$.

*Part (ii).* We use the first order condition

$$\frac{d\Pi}{dp} = \kappa\phi - \kappa\gamma_{LH} q^* \psi - v'(p) - \kappa\gamma_{LH}(Y(1-q^*) + c(q^*)) = 0. \tag{9}$$

as an implicit function of $\widetilde{p}$ and get

$$\frac{d\widetilde{p}}{d\phi} = -\frac{\kappa}{-v''(\widetilde{p})} > 0,$$

$$\frac{d\widetilde{p}}{d\psi} = -\frac{-\kappa\gamma_{LH}q^*}{-v''(\widetilde{p})} < 0,$$

$$\frac{d\widetilde{p}}{d\gamma_{LH}} = -\frac{-\kappa q^*\psi - \kappa(Y(1-q^*) + c(q^*))}{-v''(\widetilde{p})} < 0,$$

$$\frac{d\widetilde{p}}{d\kappa} = -\frac{\phi - \underline{\phi}}{-v''(\widetilde{p})} > 0.$$

which implies Part (ii) (recall that $\frac{d\gamma_{LH}}{dg} < 0$, $\frac{d\gamma_{LH}}{d\rho} < 0$, and $\frac{d\gamma_{LH}}{dx} > 0$).

*Part (iii).* Follows directly from

$$\frac{\partial\Pi(p)}{\partial\phi} = \kappa p > 0,$$

$$\frac{\partial\Pi(p)}{\partial\Psi} = -\kappa p\gamma_{LH}q^* < 0,$$

$$\frac{\partial\Pi(p)}{\partial g} = -\kappa p\frac{d\gamma_{LH}}{dg}(q^*\Psi + (Y(1-q^*) - c(q^*))) > 0,$$

$$\frac{\partial\Pi(p)}{\partial\rho} = -\kappa p\frac{d\gamma_{LH}}{d\rho}(q^*\Psi + (Y(1-q^*) - c(q^*))) > 0,$$

$$\frac{\partial\Pi(p)}{\partial x} = -\kappa p\frac{d\gamma_{LH}}{dx}(q^*\Psi + (Y(1-q^*) - c(q^*))) < 0$$

$$\frac{\partial\Pi(p)}{\partial\kappa} = p(\phi - \underline{\phi}) > 0$$

for any $p > 0$ and the envelope theorem $\frac{d\widetilde{\Pi}}{dy} = \frac{\partial\Pi}{\partial y}|_{p=\widetilde{p}}$ for a parameter $y$.

## 6.5 Proof of Proposition 2

Observe that for $\phi > \underline{\phi}$, $\widetilde{\Pi}(b) : [\underline{b}, \infty) \to \mathbb{R}$ is continuous, monotone increasing on $(\underline{b}, \widetilde{b})$, monotone decreasing on $(\widetilde{b}, \overline{b})$, and exhibits a unique maximum at $\widetilde{b} > \underline{b}$. Moreover, $\frac{d\widetilde{\Pi}}{db}|_{b=\underline{b}} > 0$ and $\widetilde{\Pi}(b)$ is non-positive for bonuses $b \leq \underline{b}$ and $b \geq \overline{b}$. Then, Cases 1-3 cover all possible relative positions of $b^{min}$, $b^{max}$, $\underline{b}$, $\widetilde{b}$, and $\overline{b}$. Optimal contracts as described in the

Proposition than follow from the shape of $\widetilde{\Pi}(b)$.

### 6.6 Proof of Results 1-3

**Result 1** Recall that $\frac{\partial\Pi}{\partial g,\rho} > 0$ for $p > 0$. Then, Part (i) follows as $\frac{db^{min}}{dg,\rho} < 0$, $\frac{db^{max}}{d\rho} > 0$, and $\frac{db^{max}}{dg} = 0$. Part (ii) follows from $\Pi = \kappa p\phi - v(p)$ for $\gamma_{LH} = 0$. In this case, the optimal effort is determined by $v'(p) = \kappa\phi$ which coincides with the equation that determines $p_{FB}$. The corresponding bonus reads $b(p_{FB}) = \frac{1}{\kappa g}v'(p_{FB}) = \frac{\phi}{g}$. This bonus is credible whenever $\frac{\phi}{g} \leq b^{max}$ (recall that $b^{min} = 0$ for $g = 1$ or $\rho = 1$ or $x = 0$).

**Result 2** For Part (i) observe that $\Pi^* = 0$ if $b^{max} \leq \underline{b}$. For $b^{max} = \underline{b}$ and $\frac{db^{max}}{d\psi,x} > \frac{d\underline{b}}{d\psi,x}$, $\frac{d\Pi^*}{dx,\psi} > 0$. $\frac{db^{max}}{dx} > \frac{d\underline{b}}{dx}$ holds whenever $Y$ and $c(q)$ is sufficiently small and $\frac{db^{max}}{d\psi} > \frac{d\underline{b}}{d\psi}$ holds because $\frac{db^{max}}{d\psi} > 0$ and $\frac{d\underline{b}}{d\psi} = 0$. For Part (ii) note that $\widetilde{b}$ as a function of $\phi$ is a continuous and monotone increasing mapping from $(\underline{\phi}, \infty)$ to $(\underline{b}, \infty)$. Therefore, as long as $b^{max} > \underline{b}$ there will always be $\phi$ such that $\widetilde{b}$ is too low to be credible or credible. As $\frac{\partial\Pi}{\partial x,\psi} < 0$ for $p > 0$, $\frac{d\Pi^*}{dx,\psi} < 0$ whenever $\widetilde{b}$ is credible. If $\widetilde{b}$ is too low to be credible, $\frac{d\Pi^*}{dx,\psi} \leq 0$ because $\frac{\partial\Pi}{\partial x,\psi} < 0$ and $\frac{db^{min}}{dx,\psi} > 0$. But if $\widetilde{b}$ is too large to be credible (note that this will be the case for any $b^{max}$ as soon as $\phi$ is sufficiently large), then $b^* = b^{max}$ and $\frac{\partial\Pi}{\partial x,\psi} < 0$ while $\frac{\partial\Pi}{\partial p}\frac{dp}{dx,\psi} > 0$. But as $\frac{\partial\Pi}{\partial x,\psi}$ is bounded and independent of $\phi$ and $\frac{dp}{dx,\psi}|_{b=b^{max}} \neq 0$, $\frac{\partial\Pi}{\partial p} = \kappa(\phi - \underline{\phi})$ guarantees that $\frac{d\Pi}{dx,\psi} = \frac{\partial\Pi}{\partial\psi,x} + \frac{\partial\Pi}{\partial p}\frac{dp}{dx,\psi} > 0$ for $\phi$ sufficiently large.

**Result 3** For Part (i) observe that $\Pi^* = 0$ if $b^{min} \geq \bar{b}$. For $b^{min} = \bar{b}$ and $\frac{db^{min}}{d\kappa} < \frac{d\bar{b}}{d\kappa}$, $\frac{d\Pi^*}{d\kappa} < 0$ – the principal would benefit from a lower capability $\kappa$. $\frac{db^{min}}{d\kappa} > \frac{d\bar{b}}{d\kappa}$ holds whenever $\psi$ is large. For Part (ii) observe that if $\widetilde{b}$ is too large to be credible (e.g. for sufficiently large $\phi$), $b^* = b^{max}$ and $\frac{d\Pi}{d\kappa} > 0$ because $\frac{db^{max}}{d\kappa} = 0$ and $\frac{\partial\Pi}{\partial\kappa} > 0$.

**University of Innsbruck - Working Papers in Economics and Statistics**
**Recent Papers** can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2012-16 **Alexander Sebald, Markus Walzl:** Optimal contracts based on subjective evaluations and reciprocity

2012-15 **Alexander Sebald, Markus Walzl:** Subjective performance evaluations and reciprocity in principal-agent relations

2012-14 **Elisabeth Christen:** Time zones matter: The impact of distance and time zones on services trade

2012-13 **Elisabeth Christen, Joseph Francois, Bernard Hoekman:** CGE modeling of market access in services

2012-12 **Loukas Balafoutas, Nikos Nikiforakis:** Norm enforcement in the city: A natural field experiment *forthcoming in European Economic Review*

2012-11 **Dominik Erharter:** Credence goods markets, distributional preferences and the role of institutions

2012-10 **Nikolaus Umlauf, Daniel Adler, Thomas Kneib, Stefan Lang, Achim Zeileis:** Structured additive regression models: An R interface to BayesX

2012-09 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** History repeating: Spain beats Germany in the EURO 2012 Final

2012-08 **Loukas Balafoutas, Glenn Dutcher, Florian Lindner, Dmitry Ryvkin:** To reward the best or to punish the worst? A comparison of two tournament mechanisms with heterogeneous agents

2012-07 **Stefan Lang, Nikolaus Umlauf, Peter Wechselberger, Kenneth Harttgen, Thomas Kneib:** Multilevel structured additive regression

2012-06 **Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang:** Bayesian semiparametric additive quantile regression

2012-05 **Eric Mayer, Sebastian Rueth, Johann Scharler:** Government debt, inflation dynamics and the transmission of fiscal policy shocks

2012-04 **Markus Leibrecht, Johann Scharler:** Government size and business cycle volatility; How important are credit constraints?

- 2012-03 **Uwe Dulleck, David Johnston, Rudolf Kerschbamer, Matthias Sutter:** The good, the bad and the naive: Do fair prices signal good types or do they induce good behaviour?
- 2012-02 **Martin G. Kocher, Wolfgang J. Luhan, Matthias Sutter:** Testing a forgotten aspect of Akerlof's gift exchange hypothesis: Relational contracts with individual and uniform wages
- 2012-01 **Loukas Balafoutas, Florian Lindner, Matthias Sutter:** Sabotage in tournaments: Evidence from a natural experiment *forthcoming in Kyklos*
- 2011-29 **Glenn Dutcher:** How Does the Social Distance Between an Employee and a Manager affect Employee Competition for a Reward?
- 2011-28 **Ronald Peeters, Marc Vorsatz, Markus Walzl:** Truth, trust, and sanctions: On institutional selection in sender-receiver games *published in Scandinavian Journal of Economics*
- 2011-27 **Haoran He, Peter Martinsson, Matthias Sutter:** Group Decision Making Under Risk: An Experiment with Student Couples *forthcoming in Economics Letters*
- 2011-26 **Andreas Exenberger, Andreas Pondorfer:** Rain, temperature and agricultural production: The impact of climate change in Sub-Sahara Africa, 1961-2009
- 2011-25 **Nikolaus Umlauf, Georg Mayr, Jakob Messner, Achim Zeileis:** Why Does It Always Rain on Me? A Spatio-Temporal Analysis of Precipitation in Austria *forthcoming in Austrian Journal of Statistics*
- 2011-24 **Matthias Bank, Alexander Kupfer, Rupert Sendlhofer:** Performance-sensitive government bonds - A new proposal for sustainable sovereign debt management
- 2011-23 **Gerhard Reitschuler, Rupert Sendlhofer:** Fiscal policy, trigger points and interest rates: Additional evidence from the U.S.
- 2011-22 **Bettina Grün, Ioannis Kosmidis, Achim Zeileis:** Extended beta regression in R: Shaken, stirred, mixed, and partitioned *published in Journal of Statistical Software*
- 2011-21 **Hannah Frick, Carolin Strobl, Friedrich Leisch, Achim Zeileis:** Flexible Rasch mixture models with package psychomix *published in Journal of Statistical Software*
- 2011-20 **Thomas Grubinger, Achim Zeileis, Karl-Peter Pfeiffer:** evtree: Evolutionary learning of globally optimal classification and regression trees in R

2011-19 **Wolfgang Rinnergschwentner, Gottfried Tappeiner, Janette Walde:** Multivariate stochastic volatility via wishart processes - A continuation

2011-18 **Jan Verbesselt, Achim Zeileis, Martin Herold:** Near Real-Time Disturbance Detection in Terrestrial Ecosystems Using Satellite Image Time Series: Drought Detection in Somalia *forthcoming in Remote Sensing and Environment*

2011-17 **Stefan Borsky, Andrea Leiter, Michael Pfaffermayr:** Does going green pay off? The effect of an international environmental agreement on tropical timber trade

2011-16 **Pavlo Blavatskyy:** Stronger Utility

2011-15 **Anita Gantner, Wolfgang Höchtl, Rupert Sausgruber:** The pivotal mechanism revisited: Some evidence on group manipulation *revised version with authors Francesco Feri, Anita Gantner, Wolfgang Höchtl and Rupert Sausgruber forthcoming in Experimental Economics*

2011-14 **David J. Cooper, Matthias Sutter:** Role selection and team performance

2011-13 **Wolfgang Höchtl, Rupert Sausgruber, Jean-Robert Tyran:** Inequality aversion and voting on redistribution

2011-12 **Thomas Windberger, Achim Zeileis:** Structural breaks in inflation dynamics within the European Monetary Union

2011-11 **Loukas Balafoutas, Adrian Beck, Rudolf Kerschbamer, Matthias Sutter:** What drives taxi drivers? A field experiment on fraud in a market for credence goods

2011-10 **Stefan Borsky, Paul A. Raschky:** A spatial econometric analysis of compliance with an international environmental agreement on open access resources

2011-09 **Edgar C. Merkle, Achim Zeileis:** Generalized measurement invariance tests with application to factor analysis *forthcoming in Psychometrika*

2011-08 **Michael Kirchler, Jürgen Huber, Thomas Stöckl:** Thar she bursts - reducing confusion reduces bubbles *published in American Economic Review*

2011-07 **Ernst Fehr, Daniela Rützler, Matthias Sutter:** The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence

2011-06 **Octavio Fernández-Amador, Martin Gächter, Martin Larch, Georg Peter:** Monetary policy and its impact on stock market liquidity: Evidence from the euro zone

2011-05 **Martin Gächter, Peter Schwazer, Engelbert Theurl:** Entry and exit of physicians in a two-tiered public/private health care system

2011-04 **Loukas Balafoutas, Rudolf Kerschbamer, Matthias Sutter:** Distributional preferences and competitive behavior *published in* *Journal of Economic Behavior and Organization*

2011-03 **Francesco Feri, Alessandro Innocenti, Paolo Pin:** Psychological pressure in competitive environments: Evidence from a randomized natural experiment: Comment

2011-02 **Christian Kleiber, Achim Zeileis:** Reproducible Econometric Simulations *forthcoming in* *Journal of Econometric Methods*

2011-01 **Carolin Strobl, Julia Kopf, Achim Zeileis:** A new method for detecting differential item functioning in the Rasch model

University of Innsbruck

Working Papers in Economics and Statistics

2012-16

Alexander Sebald, Markus Walzl

Optimal contracts based on subjective evaluations and reciprocity

**Abstract**
As demonstrated in a recent laboratory experiment [see Sebald and Walzl (2014)], individuals tend to sanction others who subjectively evaluate their performance when-ever this assessment falls short of the individuals' self-evaluation. Interestingly, this is the case even if the individuals' earnings are unaffected by the subjective performance appraisal. Hence, performance feedback which falls short of agents' self-evaluations can be interpreted as an unkind act that triggers a negatively reciprocal response not only if the assessment determines agents' earnings but also when it lacks monetary consequences. We propose a principal-agent model formalizing that agents might engage into conflict in response to ego-threatening performance appraisals and show that these conflicts stabilize principal-agent relationships based on subjective performance evaluations. In particular, we identify conditions for a positive welfare effect of increasing costs of conflict and a negative welfare effect of more capable agents.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)