WORKING PAPERS IN ECONOMICS

No. 07/06

ODDVAR M. KAARBØE AND TROND E. OLSEN

DISORTED PERFORMANCE MEASURES AND DYNAMIC INCENTIVES

Department of Economics
UNIVERSITY OF BERGEN

Distorted Performance Measures and Dynamic Incentives*

Oddvar M. Kaarbøe† and Trond E. Olsen‡

June 9, 2006

Abstract

Incentive contracts must typically be based on performance measures that do not exactly match agents' true contribution to principals' objectives. Such misalignment may impose difficulties for effective incentive design. We analyze to what extent implicit dynamic incentives such as career concerns and ratchet effects alleviate or aggravate these problems. Our analysis demonstrates that the interplay between distorted performance measures and implicit incentives implies that career and ratchet effects have real effects, that stronger ratchet effects or more distortion may increase optimal monetary incentives, and that bureaucratic promotion rules may be optimal.

*We thank two anonymous referees, the editor and Ola Kvaløy for helpful comments.

†Programme for Health Economics (HEB), Department of Economics, University of Bergen. Fosswinckelsg. 6, N-5007 Bergen. Norway. Email: oddvar.kaarboe@econ.uib.no
URL http://www.econ.uib.no/stab/oddvar.html

‡Department of Finance and Management Science, Norwegian School of Economics and Business Administration, and Programme for Health Economics (HEB). Email: trond.olsen@nhh.no

1 Introduction

A general problem for designing incentive schemes is that available performance measures seldom capture precisely agents' true contributions to principals' objectives. Performance measures are typically influenced by stochastic factors that agents can't control, and they often do not reflect all aspects that principals care about. For instance, quantitative performance measures often neglect important qualitative (soft) aspects of an agent's performance. Such measures are distorted from, or 'not well aligned with', the principal's true objectives. As is well known, such misalignments may impose severe difficulties for effective incentive design. (Holmstrom and Milgrom 1991; Baker 1992; Feltham and Xie 1994; and Baker 2002.)

Baker (2002) argues that an understanding of how distorted performance measures affect the design of incentive contracts may explain several issues and puzzles in the literature; including (i) why high-signal-to-noise ratio performance measures may receive low weight in an incentive scheme, (ii) how the distinction between paying for "inputs" versus paying for "outputs" can be interpreted, and (iii) why seemingly informative performance measures degrade, (Baker 2002, pp. 738-40). The latter issue is illustrated by a school system that administers standardized tests to its students, but does not use the scores to motivate teachers. A reason for not including these seemingly informative test scores as a performance measure in an incentive system, is that teachers will then have incentives to "teach to the test", and may thus engage in dysfunctional behavior that increases the performance measure without increasing the school's real objective.

We want to point out that, while it certainly is true that incentives to "teach to the test" are affected by direct monetary rewards, it may nevertheless well be the case that teachers face incentives to engage in this kind of behavior even if such direct monetary rewards are absent. Good test scores may give the school administration a signal that the teacher's talent is high, and result in future salary increases. Or, test scores may be used as a criterion to allocate teachers to different classes. A complete understanding of how distorted performance measures affect overall incentive design requires that implicit incentives are also taken into account.

In this paper we analyse the interplay between implicit dynamic incentives and explicit incentives when performance measures are distorted. This latter issue has been studied extensively in the accounting literature (Feltham and Xie 1994; Datar, Kulp, and Lambert 2001 and Huges, Zhang, and Xie 2005), but implicit incentives have not been examined in such a setting. Implicit incentives arise when explicit contracts can be renegotiated as time unfolds. Hence, implicit incentives reflect the fact that future periods' pay depends on today's performance. If today's performance improves the agent's position in the labor market, career concerns are present, (Fama 1980; Holmstrom 1982; and Gibbons and Murphy 1992). Ratchet effects are present if better performance today implies a tougher performance standard tomorrow, (Weitzman 1976). Meyer and Vickers (1997) analyze dynamic incentives with a non-distorted performance measure. We combine the two themes by adding information about the agent's talent given through distorted performance measures to

Meyer and Vickers (1997), and by adding a talent factor to Feltham and Xie (1994).

First we consider the case where the principal can provide incentives on a verifiable, but distorted, performance measure (z). In addition some information about the agent's talent and performance is provided to the principal (and the market) through the non-verifiable value measure (y) that reflects how the agent's talent and performance contribute to the principal's true objective. In this case implicit incentives are related both to the distorted and the undistorted performance measures (and hence the degree of misalignments between them).

By using this model we show that both career and ratchet effects do have real effects; neither can costlessly be neutralized by monetary incentives. Furthermore we find, contrary to what is found in models with non-distorted performance measures (e.g. Gibbons 1987; Meyer, Olsen, and Torsvik 1996; and Meyer and Vickers 1997), that stronger ratchet effects may increase optimal monetary incentives. The intuition behind the latter result is that ratchet effects on the true value measure (y) reduce the agent's incentives to exert effort and is optimally compensated by a stronger monetary incentives on z. These results are partly driven by the fact that performance measures are distorted, and partly by the fact that the principal uses several signal for updating beliefs about the agent's talent (as opposed to one signal in e.g. Meyer and Vickers). We disentangle these influences, and show for instance that career effects on the true value measure y have real effects on equilibrium effort only when the verifiable measure z is distorted from y if the agent is risk-neutral.

Finally we notice that this dynamic model reproduces some of the results of the static version (Baker 2002); that better alignment between the performance measures increases monetary incentives, and that the better aligned the performance measure is with the true value, the higher is the (total) surplus among the principal and the agent. The first of these two results is however not trivial since better alignments strengthens the ratchet elements. This effect tends to lower monetary incentives. We can however show that this latter effect will never dominate the direct effect of better alignments, and hence that monetary incentives do increase with better alignments.

It is often the case that in addition to verifiable (and distorted) performance measures, there are other non-verifiable measures that may yield valuable information about the agent's talent and performance. A typical case is one where quantity aspects are verifiable but quality aspects are not, yet some information about these quality aspects is observable for the relevant parties. Such information may be hard or impossible to verify in a court, but may be used by principals and peers to asses agents' abilities and performance, and hence induce implicit incentives for agents to exert effort.

We also consider such a setting, and show that some new issues arise.[1] In particular, we point out the following features. First we show that incentives may

[1] In addition to measures y and z, there is now a third non-verifiable measure q. We may think of y reflecting the true mix of quality and quantity aspects that the principal cares about, z being a verifiable measure of quantity aspects, and q being a non-verifiable measure of quality aspects.

increase with more distortion in the (verifiable) performance measure, second that it may well be advantageous (in terms of total surplus) that the verifiable performance measure is distorted relative to the measure of true value. Finally we show that distortion may be one explanation behind the observation that organizations often use bureaucratic promotion rules.

To understand these results notice that the availability of the extra information signal will affect how the principal updates her beliefs about the agent's ability. This changes implicit incentives, and thus the monetary incentive provided by the principal. Consider now the case where there are career incentives on the quality signal, but all measures are well aligned. In this case the principal can reduce monetary incentives by much since career incentives ensure that the agent exerts valuable effort. If, on the other hand, the quality signal is distorted so effort on it is not valuable for the principal, she has to keep monetary incentives high to steer attention away from it. These are the cases were monetary incentives increase with distortion.

The intuition behind the second result (that total surplus may increase with distortion) is also related to the way the market updates its estimate of the agent's talent. If some non-verifiable measure of quality aspects is not aligned with the true value, and implicit incentives on this measure induce the agent to focus on these quality aspects, then it may be advantageous that explicit incentives can be used to induce efforts on quantity aspects rather than on a balanced mix of both aspects. This is just to say that it may be advantageous that the verifiable measure is not perfectly aligned with the measure of true value.

The choice of promotion rule is also related to which information is used by the principal to update her estimate of the agent's ability. Consider two different promotion rules: promote the agent with the highest expected talent given all available information, or promote the agent with best verifiable performance in the first period. The point we want to make is that promoting based on expected talent creates implicit incentives which may steer the agent's effort towards activities that have low value for the principal. By committing to promote exclusively on verifiable information, the principal effectively removes such implicit incentives. But this commitment comes at a cost, since the principal runs the risk of promoting the least fitted. If this cost is relatively low, promoting based on the ex post inefficient bureaucratic rule may be preferable. This result thus extends the result in Milgrom and Roberts (1988) and Prendergast (1999) that bureaucratic promotion rules may be preferable also when activities are valuable (and not just to avoid non-productive influence activities).

The paper is organized as follows. In section 2 the basic model is outlined, while the optimal contracts are derived in section 3. In section 4 we consider the case where an additional non-verifiable measure provides some information about the agent's performance. First we show that incentives may increase with more distortion (section 4.1). Then in section 4.2 we show that it may be advantageous (in terms of total surplus) that the verifiable performance measure is distorted relative to the measure of true value. The choice of promotion rule is analysed in section

4.3. Section 5 provides some concluding remarks.

2 The Model

There is one agent, n tasks, and two periods. In each period the agent privately supplies effort $\mathbf{a}_t = (a_{t1}, a_{t2}, ..., a_{tn})$ on the n tasks. The agent's choice of efforts determines the agent's *total* contribution to the principal, denoted by y_t. That is, y_t reflects everything the principal cares about, except for wages, in period t. We assume that no contract on y can be enforced in court because it is prohibitively costly to specify this outcome ex ante in such a way that it can be verified by a third party ex post. We do however assume that all parties–insiders as well as outsiders–observe the $y-$signal ex post, and favorable realizations of this signal improves the agent's standing on the job market.

Let

$$y_t = h'\eta + \mathbf{f}\mathbf{a}_t + \varepsilon_t,$$

where $\mathbf{f} = \{f_1, f_2, ..., f_n\}$ is an n-dimensional vector of marginal products of the agent's efforts, $\mathbf{f}\mathbf{a}_t = f_1 a_{t1} + ... + f_n a_{tn}$ denotes the scalar product, and $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ represents random effects. η is the agent's unknown ability. The ability η is drawn at the beginning of the first period from an independent normal distribution with mean $E\eta = m_0$ and variance σ_η^2. The agent's ability has productivity h' for the principal.

Let z be a verifiable performance measure, so monetary incentives can be provided through this signal. Incentives on this signal serves as a means to increase the agent's total contribution for the principal. Let

$$z_t = \eta + \mathbf{g}\mathbf{a}_t + \zeta_t,$$

where $\mathbf{g} = \{g_1, g_2, ..., g_n\}$ is an n-dimensional vector of the marginal products of actions on the verifiable performance measure and $\zeta_t \sim N(0, \sigma_\zeta^2)$ is the effect of uncontrollable events. Let ζ be independent of ε and of η.

The agent which is risk-averse privately chooses actions a_{ti}, $i = 1, ..., n$. The private cost of effort in monetary units is denoted $c(\mathbf{a}_t)$, and is (for simplicity) assumed to be a quadratic expression, i.e. $c(\mathbf{a}_t) = \sum_{i=1}^n \frac{a_{ti}^2}{2}$. The agent's utility function is exponential, and there is no discouning. We assume that the agent is offered a linear wage contract $w_t = A_t + \alpha_t z_t$.[2] With linear compensation, exponential utility, and normal random variables, the agent's certainty equivalent is $CE = \sum_t [Ew_t - c(\mathbf{a}_t)] - \frac{r}{2} var(w_1 + w_2)$, where E is the expectation operator, and $r \geq 0$ measures the agent's risk aversion.

The principal is risk neutral and has net benefit in period t given by $y_t - w_t$. She can observe neither the actions taken by the agent nor his ability. She only observes

[2] The focus on linear contracts can be justified by appeal to a richer dynamic model in which linear payments are optimal Holmstrom and Milgrom (1987).

the signals $\mathbf{x}_t = (y_t, z_t)$ and may use it in every period to update her beliefs about the agent's ability.

The parties cannot commit not to renegotiate contracts. The second-period contract will therefore be efficient, given the information available at that time.

3 Optimal Contracts

We first characterize the optimal contract in the second, and last period. There are no career incentives in this period, and the optimal incentives in period 2 correspond to the optimal bonus in the one-period model.

Second period. The true expected value for the principal is $E(y_2 \mid \mathbf{x}_1) = h'E(\eta \mid \mathbf{x}_1) + \mathbf{fa}_2$, and the expected value of the verifiable measure is $E(z_2 \mid \mathbf{x}_1) = E(\eta \mid \mathbf{x}_1) + \mathbf{ga}_2$, where $E(\eta \mid \mathbf{x}_1)$ reflects the updated belief about the agent's ability[3], and is given by

$$E(\eta \mid \mathbf{x}_1) = E\eta + \rho_z(z_1 - Ez_1) + \rho_y(y_1 - Ey_1). \tag{1}$$

The exact expressions for the regression coefficients $\rho_i = \frac{\partial}{\partial i} E(\eta \mid \mathbf{x}_1)$, $i = y, z$ are contained in Appendix A. Here we simply note that $\rho_i \in [0, 1]$ and depends on the noise terms σ_i^2, $i = \eta, y, z$, as well as the productivity parameter of ability h'. Furthermore we note that the if the $z-$signal is more noisy than the $y-$signal (i.e. $\sigma_\zeta^2 > \sigma_\varepsilon^2$), more weight is put on y relative to z in estimating the agent's ability.

The certainty equivalent for the agent in period 2 is $CE_2 = A_2 + \alpha_2 E(z_2 \mid \mathbf{x}_1) - c(\mathbf{a}_2) - \frac{r}{2}\sigma_{2c}^2$ where the first two terms reflects the expected wage Ew_2 and $\sigma_{2c}^2 := var(z_2 \mid \mathbf{x}_1) = var(\eta \mid \mathbf{x}_1) + var(\zeta)$. (See Appendix A for the exact expression.) The agent chooses effort $\mathbf{a}_2$ to maximize this certainty equivalent, and this yields $\mathbf{a}_2 = \alpha_2 \mathbf{g}$. Total expected surplus in period 2 is

$$TCE_2 = h'E(\eta \mid \mathbf{x}_1) + \mathbf{fa}_2 - c(\mathbf{a}_2) - \frac{r}{2}\alpha_2^2\sigma_{2c}^2. \tag{2}$$

By maximizing this w.r.t. α_2, taking into account the agent's response, we obtain the optimal incentive for period 2. It is given by

$$\alpha_2^* = \frac{\mathbf{fg}}{|\mathbf{g}|^2 + r\sigma_{2c}^2} = \frac{|\mathbf{f}||\mathbf{g}| \cos\theta_{fg}}{|\mathbf{g}|^2 + r\sigma_{2c}^2}. \tag{3}$$

In this expression $\mathbf{fg} = f_1 g_1 + ... + f_n g_n$ denotes scalar product, $|\mathbf{g}|$ denotes length of the vector (so $|\mathbf{g}|^2 = \sum_{i=1}^n g_i^2$) and θ_{fg} is the angle between vectors $\mathbf{f}$ and $\mathbf{g}$, defined by $\cos\theta_{fg} = \frac{\mathbf{fg}}{|\mathbf{f}||\mathbf{g}|}$. We follow Baker (2002) and use the angle θ_{fg} as a measure of distortion (or alignment).

The performance measure and the principal's valuation of the marginal products are perfectly aligned when $\mathbf{g} = \gamma\mathbf{f}$, $\gamma > 0$. All else equal the optimal incentive α_2^* on z is lower the more distorted is the performance measure (the larger is θ_{fg} and the

[3] The expectation is also conditional on the agent's assumed (and equilibrium) effort, say $\hat{\mathbf{a}}_1$.

lower is $\cos\theta_{fg}$). This follows since more distortion makes paying on z less effective for increasing y.

Naturally, the optimal incentive is decreasing in the agent's risk aversion (r) and in the variance of outcome (σ_{2c}^2). All else equal (including $|\mathbf{g}|$ and θ_{fg}), it is increasing in the length of the vector of marginal products $\mathbf{f}$. This follows since the length is a measure of the contribution of the agent's action to the principal's value relative to the contribution of noise in the production function. Note also that in the presence of risk costs ($r\sigma_{2c}^2 > 0$) the optimal incentive on z is non-monotone in the length $|\mathbf{g}|$ of the vector of marginal products for this measure; α_2^* is increasing for $|\mathbf{g}| < \sqrt{r}\sigma_{2c}$ and otherwise decreasing. This reflects two opposing considerations; first, the smaller is $|\mathbf{g}|$, the larger one wants to make α_2 to keep efforts the same (a pure scaling effect), and second, a desire to keep α_2 low because of risk costs.

The sharing of the total surplus TCE_2 will be determined by the parties' bargaining strength (and the terms specified in the initial contract). We assume that the agent has some bargaining power and hence can obtain some share of the surplus. If the agent can negotiate for himself some share s of the expected surplus TCE_2 (at the start of period 2), then the fixed wage component A_2 will be adjusted to reflect the information ($\mathbf{x}_1$) revealed in period 1 about the agent's ability as follows:

$$A_2 = (h - \alpha_2^*)E(\eta \mid \mathbf{x}_1) + const$$

where $h = sh'$ and the constant represent terms that do not depend on $\mathbf{x}_1$. The formula follows from the fact that in equilibrium we must have $sTCE_2 = CE_2$.

The second-period wage contract offered to the agent is:

$$w_2(\mathbf{x}_1) = (h - \alpha_2^*)E(\eta \mid \mathbf{x}_1) + \alpha_2^* z_2 + const$$

where the updated expected ability $E(\eta \mid \mathbf{x}_1)$ for the agent is given by (1).

First period. Since the second-period compensation depends on the first-period signals, $\mathbf{x}_1 = (y_1, z_1)$, the agent has incentives to exert effort in the first period to increase his market value. The agent thus chooses effort according to

$$\max_{\mathbf{a}_1} \{\alpha_1 \mathbf{g}\mathbf{a}_1 - c(\mathbf{a}_1) + (h - \alpha_2^*)E(\eta \mid \mathbf{x}_1) + const\}$$
$$\Rightarrow \quad \mathbf{a}_1 = (\alpha_1 + \beta_z)\mathbf{g} + \beta_y \mathbf{f}, \text{ where } \beta_z = (h - \alpha_2^*)\rho_z, \quad \beta_y = (h - \alpha_2^*)\rho_y. \quad (4)$$

In the last expression β_i is the implicit incentive on signal $i = y, z$. We see that this consists of a positive career element ($h\rho_i$) and a negative ratchet element ($\alpha_2^*\rho_i$). The net implicit incentive β_i may be positive or negative, depending on the sign of $h - \alpha_2^*$. We note that $\beta_y < 1$, since $h'\rho_y < 1$, see Appendix A.

To characterize optimal first-period incentives consider the total intertemporal surplus, and note that the variance of total wages may be written as

$$\begin{aligned} var(w_1 + w_2) &= var(\alpha_1 z_1 + \beta_y y_1 + \beta_z z_1 + \alpha_2^* z_2) \\ &= var((\tilde{\alpha}_1 + \alpha_2^*\rho_z)z_1 + h\rho_y y_1 \\ &\quad + \alpha_2^* \left[z_2 - \rho_y y_1 - \rho_z z_1\right]) \end{aligned}$$

where $\tilde{\alpha}_1 = \alpha_1 + \beta_z$ is the effective incentive on the $z-$variable. The stochastic variables in the two last lines are uncorrelated, and the variance of the latter (in square brackets) is σ^2_{2c}. The total intertemporal surplus may then be written as $TCE = TCE_1 + TCE_2^*$, where TCE_2^* is the (equilibrium) second period surplus and TCE_1 is given by

$$TCE_1 = Ey_1 - c(\mathbf{a}_1) - \frac{r}{2}\left[(\tilde{\alpha}_1 + \alpha_2^*\rho_z)^2\sigma^2_{1z} + (h\rho_y)^2\sigma^2_{1y} + 2(\tilde{\alpha}_1 + \alpha_2^*\rho_z)h\rho_y\sigma_{1yz}\right] \quad (5)$$

where $\sigma^2_{1z} = var(z_1)$, $\sigma^2_{1y} = var(y_1)$ and $\sigma_{1yz} = cov(y_1, z_1)$.[4]

Maximizing this expression, taking account of the agent's effort choice, $\mathbf{a}_1 = \tilde{\alpha}_1\mathbf{g} + \beta_y\mathbf{f}$, we see that the optimal effective incentive in period 1 is given by

$$\begin{aligned} \tilde{\alpha}_1^* &= \alpha_1^* + \beta_z = \frac{\mathbf{fg}}{|\mathbf{g}|^2 + r\sigma^2_{1z}} - \rho_y \frac{h\left[\mathbf{fg} + r\sigma_{1yz}\right] - \alpha_2^*\mathbf{fg}}{|\mathbf{g}|^2 + r\sigma^2_{1z}} - \rho_z \frac{r\alpha_2^*\sigma^2_{1z}}{|\mathbf{g}|^2 + r\sigma^2_{1z}} \\ &= \frac{\mathbf{fg}}{|\mathbf{g}|^2 + r\sigma^2_{1z}}\left(1 - \rho_y(h - \alpha_2^*)\right) - \frac{r}{|\mathbf{g}|^2 + r\sigma^2_{1z}}\left(\rho_y h\sigma_{1yz} + \rho_z\alpha_2^*\sigma^2_{1z}\right) \end{aligned} \quad (6)$$

The first line of the formula shows that effective incentives on z can be seen as consisting of three components: the monetary incentive (first term), the implicit incentive associated with the y-variable (second term), where both career and ratchet effects are present, and finally the implicit incentive on the z-variable, where only a ratchet effect is present.

The second line decomposes effective incentives on z in two components; one component (the first) that depends on the degree of alignment between z and the true value y (via the term $\mathbf{fg}$), and a second component that reflects how incentives are adjusted in response to risk costs. The latter is not (for a given second-period bonus α_2^*) directly influenced by the degree of alignment between z and y. This yields the following.

Proposition 1 *Let* $\mathbf{fg} > 0$.
(i) If the measures z and y are perfectly aligned ($\mathbf{g} = \gamma\mathbf{f}$), then implicit incentives have real effects only if the agent is risk averse. For $r > 0$ risk considerations imply that effective incentives are influenced by the career element in y and by the ratchet element in z, and both influence effective incentives negatively.

(ii) If z and y are not perfectly aligned ($\mathbf{g} \neq \gamma\mathbf{f}$), implicit incentives have real effects also if $r = 0$. Apart from adjustments associated with risk costs, effective incentives on z are then affected negatively by the net implicit incentives on y, so that a higher career (ratchet) element in y reduces (increases) effective incentives on z.

Corollary 1 *Career incentives (on y) and monetary incentives (on z) are substitutes independent of the agent's preferences towards risk.*

[4] We have $\sigma^2_{1z} = \sigma^2_z + \sigma^2_\eta$, $\sigma^2_{1y} = \sigma^2_y + \sigma^2_\eta$ and $\sigma_{1yz} = \sigma^2_\eta$ in the given specification.

As emphasized in the second line of (6), effective incentives are adjusted for costs partly associated with risk and partly with misalignment. To consider the latter in isolation, suppose first that the agent is risk neutral. Part (i) of the proposition then follows since the first-best is attainable when z is perfectly aligned with y and the agent is risk-neutral (for $\mathbf{g} = \gamma\mathbf{f}$ and $r = 0$ we get $\tilde{\alpha}_1^* = (1 - \beta_y)/\gamma$ and $\mathbf{a}_1 = \mathbf{f}$). All implicit incentives are then completely neutralized. Note however that *monetary* incentives are affected since implicit incentives substitute monetary incentives on z. Since the ratchet element in y reduces implicit incentives (on y) it is optimally compensated with stronger monetary incentives on z. Similarly, a higher career element in y increases the implicit incentives (on y) and monetary incentives on z are thus lowered. If $\mathbf{g} \neq \gamma\mathbf{f}$, (ii) this compensation is not one-to-one since distortion makes paying on z less effective for increasing y, and implicit incentives on y will then have real effects. Implicit incentives on z will however have no real effects, since both the career element and the ratchet element in z can be costlessly neutralized by monetary incentives when the agent is risk-neutral.

Consider now risk aversion ($r > 0$). In addition to the adjustments considered above, incentives on z will then also be adjusted in response to risk costs. From equation (6) we see that risk costs affect effective incentives through two channels. First, risk costs make it costly to neutralize ratchet effects on z. Secondly, the career element in y (the term related to h, and thus to the fixed salary part A_2) contributes to the variance of payments, and since the $y, z-$variables are positively correlated ($\sigma_{1yz} = cov(y_1, z_1) = \sigma_\eta^2 > 0$), this variance can be reduced by reducing the explicit incentives on the $z-$variable. The latter effect implies that career effects on y have real effects both when z is distorted from, and when z is perfectly aligned, with y. Combining parts (i) and (ii) proves the Corollary.

The fact that career elements have real effects is at variance with results from Meyer and Vickers (1997), who show that career effects can be neutralized when there is only one signal, the performance measure is non-distorted (i.e., $\mathbf{g} = \gamma\mathbf{f}$) and the agent is risk averse. The above discussion highlights two mechanisms that can modify this result. The first is that distortion reduces the value of the surrogate measure z, so career effects are not compensated one-to-one. This mechanism applies both for risk-neutral and risk-averse agents. The other mechanism is related to the fact that the principal uses two signals for updating beliefs about the agent's talent (as opposed to one signal in Meyer and Vickers 1997). Since the career element in y contributes to the variance of payments, and the principal cannot write explicit contracts on y, career effects cannot be costlessly neutralized when the agent is risk-averse, even if z is perfectly aligned with y.

From equations (4) and (6) we see that the first-best is attainable when the performance measure is non-distorted and the agent is risk neutral. This resembles the result stated in Proposition 1 in Feltman and Xie (1994): if the true value for the principal is not contractible information, then risk neutrality or a noiseless performance measure is not sufficient to achieve the first-best allocation. The performance measure must also be perfectly aligned with the true value. Note however that when career effects are present ($\beta_y > 0$) and the performance measures are to some extent aligned ($fg > 0$), the presence of implicit incentives reduces the welfare loss due to distorted performance measures. This follows since career incentives on the principal's true value attract the agent's attention and effort towards y. In the extreme case when $\mathbf{f}$ and $\mathbf{g}$ are orthogonal (and $r = 0$), optimal monetary incentives are zero, but the agent exerts effort $\mathbf{a}_1 = \beta_y \mathbf{f} > \mathbf{0}$.

In a static setting –and here for period 2– one sees that incentives on the performance measure z are stronger the better aligned are the two measures. In a dynamic setting the relationship is more complicated. From the formula (6) for the effective first period incentive $\tilde{\alpha}_1^*$, we see that there are both direct and indirect effects associated with better alignment. First there is a direct positive effect in that $\mathbf{fg}$ gets larger. (We keep $|\mathbf{f}||\mathbf{g}|$ fixed and consider only parameters that yield non-negative incentives.) But second, there are indirect effects working via the ratchet elements. Although there are opposing effects generated by dynamic implicit incentives, it turns out that better alignment does in fact increase effective incentives also in the first period, at least for all parameters that yield non-negative effective incentives in this model. We verify this in Appendix B. Note that this implies that monetary incentives must also increase (and by even more) since the implicit incentive β_z is reduced.

In Appendix B we also verify the intuitively reasonable result that the equilibrium surplus (TCE) is also higher the better aligned is the performance measure with the measure of true value. Thus we have:

Proposition 2 *For parameters that yield non-negative effective incentives we have: As the performance measure z gets better aligned with (less distorted from) the measure of value y, (i) optimal effective and monetary incentives in both periods increase, and (ii) the total surplus increases.*

These results show that to the extent that design of performance measures is feasible, it is (all else equal) optimal to construct or choose a measure that is least distorted relative to the measure of true value. As we shall see in the next section, this is however generally true only when such performance measures are verifiable.

4 Additional Non-Verifiable Measures

It is often the case that in addition to verifiable (and distorted) performance measures, there are other non-verifiable measures that may yield valuable information about the agent's performance. A typical case is one where quantity aspects are verifiable but quality aspects are not, yet some information about these quality aspects is observable for the relevant parties. Such information may be hard or impossible to verify in a court, but may be used by principals and peers to asses agents' abilities and performance, and hence induce implicit incentives for agents to exert effort.

We now consider such a setting, and show that some new issues arise. In particular, we point out three new features: (i) that incentives may increase with more distortion in the (verifiable) performance measure, and (ii) that it may well be advantageous (in terms of total surplus) that the verifiable performance measure (z) is distorted relative to the measure of true value (y), and (iii) that distortion has implication for the choice of promotion rules. To simplify the exposition in this section, we consider risk-neutral agents. Furthermore, we focus on the case where career concerns are present. Before we consider the three features mentioned above, we derive the optimal incentives in the extended model.

Suppose there is an additional non-contractible 'quality' variable

$$q_t = \mathbf{k}\mathbf{a}_t + \eta + \kappa_t$$

where $\mathbf{k}$ is an n-dimensional vector, and $\kappa_t \sim N(0, \sigma_\kappa^2)$ is a noise term.

In the second period there are no implicit incentives, and the optimal incentive on z is thus, from (3), given by

$$\alpha_2^* = \frac{|\mathbf{f}|}{|\mathbf{g}|} \cos(\theta_{fg}) > 0.$$

In period 1 the agent faces career incentives on the three measures y, z, q. These incentives are given by $\beta_i = (h - \alpha_2^*)\rho_i \geq 0$, $i = y, z, q$, where the ρ_i's are the regression coefficients for the conditional expectation of ability (η), given first period observations (y_1, z_1, q_1).[5] The agent optimally chooses efforts such the vector of marginal costs equals the vector of marginal benefits, i.e.

$$\mathbf{a}_1 = \tilde{\alpha}_1 \mathbf{g} + \beta_y \mathbf{f} + \beta_q \mathbf{k}, \qquad \tilde{\alpha}_1 = \alpha_1 + \beta_z \tag{7}$$

As before α_1 and $\tilde{\alpha}_1$ denote explicit (monetary) and effective incentives, respectively, on the verifiable performance measure z. Maximization of the first-period surplus $\mathbf{f}\mathbf{a}_1 - c(\mathbf{a}_1)$ with $\mathbf{a}_1$ given by (7), yields the optimal first-period effective incentive

$$\tilde{\alpha}_1^* = \alpha_1^* + \beta_z = \alpha_2^* - \beta_y \frac{|\mathbf{f}|}{|\mathbf{g}|} \cos(\theta_{fg}) - \beta_q \frac{|\mathbf{k}|}{|\mathbf{g}|} \cos(\theta_{kg}), \tag{8}$$

where θ_{kg} is the angle between $\mathbf{k}$ and $\mathbf{g}$; defined by $\cos\theta_{kg} = \frac{\mathbf{g}\mathbf{k}}{|\mathbf{g}||\mathbf{k}|}$. From equation (8) it follows that when there are net career concerns on q, ($\beta_q > 0$), effective incentives (on z) are adjusted down to account for the fact that the agent has incentives to exert effort to promote q. The strength of the incentive correction depends on how well aligned the quality measure is to the verifiable performance measure z (which is equal to the principal's true value in the non-distorted case). If the measures are well aligned (θ_{kg} 'small', so $\cos(\theta_{kg}) \approx 1$), effective incentives are reduced by much since q is then a valuable surrogate measure for z (and thus for y in the case of a perfect measure, $\mathbf{g} = \mathbf{f}$). If the quality measure is highly distorted (θ_{kg} 'large', so $\cos(\theta_{kg}) \approx 0$), q is not a valuable measure for z so effective incentives on z are kept high to steer attention in the direction of z.

[5] Note that the regression coefficients are different now compared to those outlined in the former sections. In Appendix A we give the exact formulas.

The above discussion highlights the role of the additional performance measure q. It affects first-period optimal incentives, and the reason is that the availability of q changes the way the principal updates her beliefs about the agent's ability. This changes implicit incentives, and thus the monetary incentives provided by the principal. We shall see that this implies that there is a real difference between the non-verifiable measures q and y in the way incentives are affected.

4.1 Distortion and Incentives

Consider first the relationship between distortion and incentive strength in the extended model. We want to make the point that in this setting it may well be the case that *first-period incentives become stronger when the performance measure becomes more distorted.* Note that this is at odds with result (i) in Proposition 2, and hence illustrates that there is a real difference between the non-verifiable measures q and y in the way they affect incentives. To illustrate this we consider the following example.

Example 1. *Consider the case illustrated in Figure 1, where $\mathbf{f} = (1, 1)$, $\mathbf{g} = (0, g_2)$ and $\mathbf{k} = (k_1, 0)$. Here z and q measure different aspects of the agent's performance (e.g. quantity and quality, respectively), and the two measures are orthogonal; $\cos(\theta_{kg}) = 0$. To simplify matters, suppose further that y is non-informative ($\sigma_\varepsilon^2 = \infty$), so $\beta_y = \rho_y = 0$. To avoid obvious scaling effects, assume also that $|\mathbf{g}| = |\mathbf{f}|$. Then effective first-period incentives are $\tilde{\alpha}_1^* = \alpha_2^* = \cos(\theta_{fg}) = 1/\sqrt{2}$. Note that these incentives are here not affected by career effects since (i) there is by assumption no such effects on y, and (ii) the orthogonality of z and q implies that career incentives on q does not affect the agent's actions in the $z-$dimension.*

[Figure 1 about here]

Next consider the case of perfect alignment between z and y, i.e. $\mathbf{g} = \mathbf{f}$. Effective first-period incentives are then given by $\tilde{\alpha}_1^{*P} = 1 - \beta_q^P \frac{|\mathbf{k}|}{|\mathbf{f}|} \cos(\theta_{kf}) = 1 - \rho_q (h - 1) \frac{|\mathbf{k}|}{|\mathbf{f}|} / \sqrt{2}$. In this case z and q are to some extent aligned, and career effects on q will then also to some extent provide incentives for actions that promote z. Hence monetary incentives on z are adapted and adjusted for these implicit incentives. The stronger are the implicit incentives on q, the smaller are now the effective incentives on z. Note also that effective incentives remain positive as long as $\rho_q(h-1)\frac{|\mathbf{k}|}{|\mathbf{f}|} < \sqrt{2}$.

We see that effective incentives are smallest in the non-distorted case ($\tilde{\alpha}_1^{*P} < \tilde{\alpha}_1^*$) when the implicit career incentives on q are strong and/or there are scaling effects such that $\frac{|\mathbf{k}|}{|\mathbf{f}|}$ is large; more precisely when $\rho_q(h - 1)\frac{|\mathbf{k}|}{|\mathbf{f}|} > \sqrt{2} - 1$. There is thus a range of parameters where all effective incentives are positive, and where they are smallest in the non-distorted case.

We state the more general condition for effective incentives to be smaller in the non-distorted case in the following proposition. To avoid obvious scaling effects we compare verifiable measures with the same 'length' $|\mathbf{g}|$, and normalize this to be equal to $|f|$.

Proposition 3 *Suppose the agent experiences career incentives and that all signals are informative. Comparing verifiable measures of equal 'length'* $|\mathbf{g}|$*, and normalized to* $|\mathbf{g}| = |\mathbf{f}|$*, we have: While second-period (and static) incentives* ($\alpha_2^* = \cos(\theta_{fg})$) *are maximal when the verifiable performance measure (z) is non-distorted (i.e. when* $\theta_{fg} = 0$*), first-period incentives are lower for a non-distorted compared to a distorted measure when*

$$(1-\alpha_2^*)-\rho_y\left[(h-1)-(h-\alpha_2^*)\alpha_2^*\right]-\rho_q\frac{|\mathbf{k}|}{|\mathbf{f}|}\left[(h-1)\cos(\theta_{kf})-(h-\alpha_2^*)\cos(\theta_{kg})\right]<0. \quad (9)$$

The formula follows from (8), noting that $\alpha_2^* = 1$ in the non-distorted case. The first component of the formula is the difference in monetary incentives, the second part is the difference in the implicit incentives associated with the y-variable, while the last component is the difference in the implicit incentives associated with the q-variable. First we note that if q is non-informative ($\rho_q = 0$) then according to Proposition 2(i) the condition will never be satisfied, given that parameters are such that effective incentives are non-negative. We see that the condition can hold only if q is informative ($\rho_q > 0$) and better aligned with the true value measure y than with the distorted performance measure z; i.e. $\cos(\theta_{kf}) > \cos(\theta_{kg})$.

We see that effective incentives on z are adjusted down to account for the fact that the agent has career incentives on q, and that the strength of the incentive correction depends on how distorted q is relative to z (which is equal to y in the non-distorted case). The statement in the proposition holds true when q is 'close' to y (θ_{kf} small) so effective incentives on z can be reduced by much in the case of a perfect measure, but q is distorted away from z (θ_{kg} large) so incentives on z are kept high to steer attention towards z.

The fact that monetary incentives are adjusted to balance implicit incentives is common in agency models where dynamic effects are present. This follows simply because the first-best is not attainable, so the principal's goal is to use all available instruments to tie the agent's efforts as close to the first-best efforts as possible. What is left, i.e. the loss relative to the first-best value induced by effort not being aligned with f is what Datar et al. (2001) define (in a static setting) as performance measure incongruity. With two verifiable performance measures they define incongruity (in our notation) as $|\mathbf{f}-(\alpha_1\mathbf{g}_1+\alpha_2\mathbf{g}_2)|^2$, i.e. as $|\mathbf{f}-\mathbf{a}|^2$, since effort is $\mathbf{a} = \alpha_1\mathbf{g}_1+\alpha_2\mathbf{g}_2$ in their setting. In our setting with one verifiable performance measure and implicit incentives, this measure of incongruity is given by $|\mathbf{f}-\mathbf{a}|^2 = \left|\mathbf{f}(1-\beta_y)-\left((\alpha_1+\beta_z)\mathbf{g}+\beta_q\mathbf{k}\right)\right|^2$. The monetary incentive (α_1) is chosen to minimize this loss.

The following proposition shows that it is in fact typically the case that first period effective incentives are highest for some measure that is not perfectly aligned with $\mathbf{f}$.

Proposition 4 *Consider given normalized vectors* $\mathbf{f}$ *and* $\mathbf{k}$ *(each of length* 1*) with* $\mathbf{fk} = \cos(\theta_{fk}) := c \in [0,1)$*. Suppose* $h \neq 1$*. Then among all performance measures with normalized vector* $|g| = 1$*, the first-period bonus is largest for some measure that is not aligned with* $\mathbf{f}$*.*

Proof. See Appendix B

4.2 Non-Distorted Performance Measure Is Not Optimal

In this section we consider the relationship between distortion and total surplus. We consider variations in the performance measure (z), and in particular variations in its degree of distortion from the true value, as measured by the angle θ_{fg} between vectors $\mathbf{g}$ and $\mathbf{f}$. In a static case it will be optimal to have a verifiable performance measure (z) that is completely aligned with the true value (y), i.e. such that $\theta_{fg} = 0$, or equivalently $\mathbf{g} = \gamma\mathbf{f}$, $\gamma > 0$. The first-best can then be achieved under risk neutrality (by setting $\alpha = \gamma^{-1}$).

For the dynamic case we want to point out that, unless the non-verifiable 'quality measure' q is completely aligned with the true value y, it will not be optimal to have $\mathbf{g}$ completely aligned with $\mathbf{f}$. Thus, *in the presence of dynamic implicit incentives it will in most cases not be optimal to have a 'perfect' verifiable performance measure.*

The intuition is fairly simple; when there are (say) career incentives on q, the agent's attention is drawn in the direction defined by vector $\mathbf{k}$. Ideally the agent's efforts should be aligned with $\mathbf{f}$. When $\mathbf{k}$ and $\mathbf{f}$ are not aligned, monetary incentives on $\mathbf{g}$ should ideally draw the agent's attention towards $\mathbf{f}$, and this will generally not be least costly to do when $\mathbf{g}$ is perfectly aligned with $\mathbf{f}$.

For given performance measures the optimal surplus can be written as[6]

$$\begin{aligned} TCE^* &= TCE_1^* + TCE_2^* \\ &= \max_{\tilde{\alpha}_1} \left[\mathbf{f}'\mathbf{a}_1(\tilde{\alpha}_1) - c(\mathbf{a}_1(\tilde{\alpha}_1))\right] + \max_{\alpha_2} \left[\mathbf{f}'\mathbf{a}_2(\alpha_2) - c(\mathbf{a}_2(\alpha_2))\right] \end{aligned}$$

where

$$\mathbf{a}_2(\alpha_2) = \alpha_2\mathbf{g}, \qquad \mathbf{a}_1(\tilde{\alpha}_1) = \tilde{\alpha}_1\mathbf{g} + \beta_y\mathbf{f} + \beta_q\mathbf{k}$$

Consider a marginal variation in the component g_i; this yields

$$\begin{aligned} \frac{\partial TCE^*}{\partial g_i} &= (f_i - a_{1i})\tilde{\alpha}_1^* + (f_i - a_{2i})\alpha_2^* \\ &= ((1-\beta_y)f_i - \tilde{\alpha}_1^* g_i - \beta_q k_i)\tilde{\alpha}_1^* + (f_i - \alpha_2^* g_i)\alpha_2^* \end{aligned}$$

We see that for $\mathbf{g} = \gamma\mathbf{f}$ (perfect alignment) the second term in the above expression vanishes, but the first term does not, and hence such perfect alignment will not be optimal.

In fact, the formula shows that some linear combination, say $\mathbf{g} = \mu\mathbf{f} + \lambda\mathbf{k}$, will be optimal. In practice it will hardly be possible to fine-tune performance measures to

[6] For notational simplicity we drop the terms capturing the talent factor η.

find the optimal balance, so to characterize the optimum may not be so interesting. The point we want to make is that some 'distorted' performance measure may in these cases well be better than a non-distorted one.

Proposition 5 *When there is a distorted non-verifiable performance measure (q) that generates implicit incentives, it is not optimal that the verifiable performance measure (z) is non-distorted.*

Proposition 5 is related to results in Datar, Kulp, and Lambert (2001). That paper studies how the principal should weigh different verifiable performance measures when agents are risk averse and the performance measures can be distorted. The main insight from the paper is that the weights are chosen to minimize the incongruity between the firm's outcome and the agent's overall performance, and to minimize the risk premium that must be paid to the agent to compensate him for bearing the risk. In our paper the issue of balancing performance measures does not arise (since only one performance measure is verifiable). However, Datar, Kulp, and Lambert (2001) do also consider a situation where one performance measure is aligned and one is distorted. In this case they show that the aligned measure does not receive all weight in the contract so that the agent's overall performance measure ($\alpha_1 z_1 + \alpha_2 z_2$ in our notation) is not perfectly aligned with y. The intuition behind this result is that the principal can trade-off some of the congruity to save on risk costs. Our result is however not driven by risk considerations, but by the fact that a distorted performance measure can counteract implicit incentives so as to maximize the overall alignment between k, z, and y.

4.3 Distortion and the Choice of Promotion Rules

In this section we analyze how distortion affects the desirability of two different promotion rules. The first rule is the ex post efficient promotion rule: promote the agent with the highest expected talent, given all information at date 2. The second rule is an ex post inefficient (bureaucratic) rule: promote the agent with best verifiable performance in period one. (Thus we assume here that the principal can commit to such a rule.) The point we want to make is that promoting according to rule 2 may improve the incentives for agents to allocate their effort correctly (seen from the principal's view), but the principal runs the risk of promoting the least fitted since not all information is used to update the estimates of agents' abilities. The two promotion rules thus illustrate an important trade-off. Promoting based on expected talent creates implicit incentives since the principal will use all information available to update her estimate of agents' talent. This includes information created from agents' activities on tasks that may be of low value for the principal. By committing to promote exclusively on verifiable information, the principal effectively removes such implicit incentives. But this commitment comes at a cost, since the principal runs the risk of promoting the least fitted.

Prendergast (1999) point out that bureaucracy, where rules rather than discretion are used to allocate resources, is a central feature of organizations. A prominent example is the use of seniority in promotion and layoff decisions, independent of profitability considerations. Prendergast shows that some degree of bureaucracy can be desirable in a setting where the agent can corrupt a subjective performance signal by unproductive influence activities. Here we complement this analysis by showing that a bureaucratic rule can also be preferable when all of the agent's activities are valuable for the principal[7].

The structure of this section is the following. First we calculate the agent's gain of being promoted (the prize in the tournament). Then we describe the contest, derive the agents' effort choices, calculate the optimal incentives, and characterize the equilibrium first-period value. The first-period value is the first-best value minus the loss due to implicit incentives along the vectors **k**, **f** and **g**. Clearly, promotion based on the bureaucratic rule will eliminate the first-period loss due to implicit incentives on **k** and **f**, but it will yield some loss in period two compared to the ex post efficient rule. This loss is related to how valuable the extra information given by the information signals y and q is for updating the beliefs about the agents' ability, and to the productivity of talent.

The prize. The principal and the promoted agent, say agent 1, negotiate the wage in period 2. Given the same negotiation process as before, the agent gets a wage[8]

$$w_2(\mathbf{x}_1^1) = (h - \alpha_2^*)E(\eta^1 \mid \mathbf{x}_1^1) + const,$$

where superscript refers to the agent, and

$$\begin{aligned} E(\eta^1 \quad &\mid \quad \mathbf{x}_1^1) = \xi_1^1 + \mathbf{m}(\mathbf{a}_1^1 - \hat{\mathbf{a}}_1^1) + const., & (10) \\ \xi_1^1 \quad &: \quad = (\rho_z + \rho_y h' + \rho_q)(\eta^1 - E\eta) + (\rho_z \zeta_1^1 + \rho_y \varepsilon_1^1 + \rho_q \kappa_1^1), & (11) \\ \mathbf{m} \quad &: \quad = (\rho_z \mathbf{g} + \rho_y \mathbf{f} + \rho_q \mathbf{k}), \end{aligned}$$

and $\hat{\mathbf{a}}_1^1$ is the anticipated (and equilibrium) effort choice.

The contest. Consider two symmetric agents. According to the ex post efficient rule, agent 1 will win iff $E(\eta^1 \mid \mathbf{x}_1^1) > E(\eta^2 \mid \mathbf{x}_1^2)$. In equilibrium the agents' efforts will be equal ($\hat{\mathbf{a}}_1^1 = \hat{\mathbf{a}}_1^2$), hence we see that agent 1 will win when $\xi^1 + \mathbf{m}\mathbf{a}_1^1 > \xi^2 + \mathbf{m}\hat{\mathbf{a}}_1^2$. The agent's expected reward in the contest is thus (for given $\hat{\mathbf{a}}_1^2$) a function of $\mathbf{m}\mathbf{a}_1^1$, say $W(\mathbf{m}\mathbf{a}_1^1)$. As shown in Appendix B, the marginal return to effort is (in equilibrium) given by

$$\frac{\partial}{\partial a_{1i}^1} W(\mathbf{m}\mathbf{a}_1^1) = \left[(h - \alpha_2^*)\frac{1}{2} + R\phi_2(0) \right] m_i = \omega m_i, \tag{12}$$

[7] Prendergast (1999, p 38) states that "[t]o illustrate the incentive to act bureaucratically, two ingredients are necessary. First, some measures of performance must be corruptible." This statement and the ensuing analysis may leave the impression that some form of unproductive activity or influence on the part of the agent is necessary to explain bureaucracy as a rational response to measurement problems in agency. Our model shows that such an explanation is valid under a much wider set of circumstances.

[8] Assuming independent talent variables, there is no extra information in the observation that the agent has won.

where ω is defined by the last equality as the expression in square brackets, R is the prize obtained by the winner of the contest, and $\phi_2(\cdot)$ is the density of the distribution of $\xi^2 - \xi^1$.

The marginal return to effort in the contest thus consists of two parts. The first part is the marginal value of being allotted the right to negotiate (with the principal) over the fixed salary part. This occurs with probability $\frac{1}{2}$ in equilibrium, given symmetric agents. The second part is the increased probability of winning the contest multiplied with the prize awarded to the winner. This prize is the added value for the agent of being promoted, thus $R = sTCE_2 - u_2$, where $sTCE_2$ is the share of the expected surplus that the agent can negotiate for himself when promoted, and u_2 is his reservation value if not promoted. The increased probability of winning the contest is given by the density $\phi_2(0) = 1/(2\pi V_x)^{1/2}$, where $V_x = var(\xi^2 - \xi^1)$.

First-period equilibrium efforts. Given explicit incentives α on z, first-period efforts are given by

$$\alpha g_i + \frac{\partial}{\partial a_{1i}^1} W(\mathbf{m}\mathbf{a}_1^1) - c'(a_{1i}^1) = 0 \qquad i = 1, 2, ..., n$$

where $c'(a_{1i}^1) = a_{1i}^1$. In addition to explicit first-period incentives we here have implicit incentives (the second term) induced by the contest for being promoted in period 2. Substituting for these implicit incentives and for $\mathbf{m}$ as defined above we see that equilibrium effort is given by

$$\hat{\mathbf{a}}_1^1 = (\alpha + \omega\rho_z)\mathbf{g} + \omega\rho_y\mathbf{f} + \omega\rho_q\mathbf{k}. \tag{13}$$

Equilibrium effort is thus a mix of the usual vectors, just with other implicit incentives.

Optimal first-period incentives and surplus. The principal maximizes first-period surplus (per agent)

$$TCE_1 = \mathbf{f}\mathbf{a}_1^1 - c(\mathbf{a}_1^1).$$

Maximizing this expression taking account of the agent's effort choice $\hat{\mathbf{a}}_1^1$, we derive the optimal first-period incentive:

$$\widetilde{\alpha}_1^* = \alpha_1^* + \omega\rho_z = \alpha_2^* - \omega\rho_y \frac{|\mathbf{f}|}{|\mathbf{g}|}\cos(\theta_{fg}) - \omega\rho_q \frac{|\mathbf{k}|}{|\mathbf{g}|}\cos(\theta_{kg}). \tag{14}$$

We see that the optimal incentive has the same structure as before, just with other implicit incentives.

We now compare the total surplus for period 1 for the cases (i) where there are implicit incentives on all information signals and (ii) when there are implicit incentives only on the verifiable signal. The former case corresponds to the ex post efficient promotion rule (based on x) and the latter corresponds to the bureaucratic rule (based on z). Efforts in the two cases can be written as $\widetilde{\alpha}\mathbf{g}+t\mathbf{n}$, where $\mathbf{n} = \omega\rho_y\mathbf{f} + \omega\rho_q\mathbf{k}$, and $t = \{0, 1\}$ depending on which promotion rule that is used. The TCE in period 1 can be written

$$TCE_1 = h'E\eta + \frac{1}{2}|\mathbf{f}|^2 - \frac{1}{2}|\mathbf{f} - \widetilde{\alpha}\mathbf{g} - t\mathbf{n}|^2.$$

The last term in this formula for TCE is the loss relative to the first-best value induced by effort not being aligned with $\mathbf{f}$. (As noted above, this loss is taken by Datar et al. (2001) as a measure of performance measure incongruity.)

Lemma 1 *Let TCE_{1x} and TCE_{1z} be the optimal first-period surpluses when promotion is based on x and on z, respectively. We have:*
(i) $TCE_{1z} > TCE_{1x}$ if $|\mathbf{k}|$ is sufficiently large.
(ii) $TCE_{1z} > TCE_{1x}$ if $\mathbf{g}$ is sufficiently well aligned with $\mathbf{f}$ (i.e θ_{gf} is sufficiently small).
(iii) $TCE_{1z} > TCE_{1x}$ if y is uninformative and the distortion between $\mathbf{k}$ and $\mathbf{f}$ is large.
(iv) $TCE_{1z} < TCE_{1x}$ if $|\mathbf{k}|$ is sufficiently small or q is sufficiently uninformative (i.e. if $|\mathbf{k}|\,\rho_q$ is sufficiently small).

We refer the reader to Appendix B, which gives the exact expression for the difference in the first-period surpluses for the two cases. Intuitively if the loss due to the implicit incentives along $\mathbf{k}$ is large, then promoting based on z will yield a larger first-period surplus. This occurs when the vector of marginal products of effort in the non-verifiable measure q has large components, so that it is 'easy' for the agent to obtain high realizations of this measure. When promotion is based in part on this measure, the agent has then strong incentives to divert his efforts in this direction. When this diversion effect is sufficiently strong, it induces a large first-period loss. From a first-period perspective it would then be better to totally remove the incentives on q by basing the promotion solely on z.

Promotion based on z will on the other hand yield some loss in period 2, compared to promotion based on highest expected talent. We now consider this loss.

Second-period values. The expected second-period value associated with talent, given promotion based on z, is

$$N_z = h' E \max\left\{E(\eta^1 \mid z_1^1), E(\eta^2 \mid z_1^2)\right\}$$

where $z_1^i = \eta^i + \mathbf{g}\tilde{\mathbf{a}}_1 + \zeta_1^i$, $\tilde{\mathbf{a}}_1$ is the common equilibrium effort, and $E(\eta^i \mid z_1^i) = \tilde{\rho}_z(\eta^i + \zeta_1^i - E\eta) + E\eta = \mu^i$. Similarly, the expected second-period value when promotion is based on x_1 is

$$N_x = h' E \max\left\{E(\eta^1 \mid \mathbf{x}_1^1), E(\eta^2 \mid \mathbf{x}_1^2)\right\}$$

where we according to equation (10) and (11) have, in equilibrium $E(\eta^i \mid \mathbf{x}_1^i) = (\xi_1^i - E\xi_1^i) + E\eta$.

Using properties of the normal distribution, we can find exact expressions for the values N_x and N_z.

Lemma 2 *Let $V_z := var(\mu^2 - \mu^1) = 2\tilde{\rho}_z^2(\sigma_\eta^2 + \sigma_\zeta^2)$ and*
$V_x := var(\xi^2 - \xi^1) = 2\left[(\rho_z + \rho_y h' + \rho_q)^2\sigma_\eta^2 + \rho_z^2\sigma_\zeta^2 + \rho_y^2\sigma_\varepsilon^2 + \rho_q^2\sigma_\kappa^2\right].$

(i) The second-period gain induced by basing promotion on x rather than on z is given by

$$N_x - N_z = (\sqrt{V_x} - \sqrt{V_z})h' \int_0^\infty (1 - \Phi_0(x))dx,$$

where $\Phi_0(\cdot)$ is the CDF of the standard normal distribution.
(ii) The variance V_x is decreasing in σ_ε^2, σ_κ^2 and σ_ζ^2, and increasing in σ_η^2 and in h'. Moreover, we have $\lim_{\sigma_\kappa^2,\sigma_\varepsilon^2\to\infty} V_x = 2\frac{(\sigma_\eta^2)^2}{\sigma_\eta^2+\sigma_\zeta^2} = V_z$

Proof. See Appendix B.

The second-period gain must be compared to the first-period loss (with the exact magnitude given in Appendix B). The latter depends on, and is increasing in, the length $|\mathbf{k}|$, while the former does not. Hence for large $|\mathbf{k}|$ it is clearly better to promote based on z. On the other hand, the second-period gain is increasing in the productivity of talent (h'), while the first-period loss is bounded in this parameter. Hence, if this productivity is large, it is better to promote based on x. Thus we have the following

Proposition 6 *For $|\mathbf{k}|$ sufficiently large it is better to promote on the basis of the verifiable (and distorted) measure z only. For $|\mathbf{k}|$ small and/or q sufficiently uninformative (i.e. for $|\mathbf{k}|\,\rho_q$ sufficiently small), promotion based on z is inferior to the ex post efficient promotion rule based on x. The latter is true also for h' (the productivity of talent) sufficiently large.*

We end this section by noting that the second-period gain from promoting based on highest expected talent depends on the variance of ability. Intuitively, if this variance is large, the second-period gain of promoting based on x rather than z is low. In fact, the next Lemma shows that when the variance of ability becomes sufficiently large, there is no extra information about the agent's talent contained in a promotion rule based on x relatively to a promotion rule based on z.

Lemma 3 $\lim_{\sigma_\eta^2\to\infty} (N_x - N_z) = 0$

Proof. See Appendix B.

For large uncertainty about ability, it is thus the relative magnitudes of the first period surpluses that become decisive for the choice of promotion rule. Using Lemma 1 we therefore have the following:

Proposition 7 *For sufficiently large ability variance σ_η^2 we have: promotion based on z only is better than the ex post efficient promotion rule if (a) $\mathbf{g}$ is sufficiently well aligned with $\mathbf{f}$, or (b) y is uninformative and the distortion between $\mathbf{k}$ and $\mathbf{f}$ is large.*

5 Conclusion

A general problem for designing incentive schemes is that available performance measures seldom capture precisely agents' true contributions to principals' objectives. In this paper we have analyzed to what extent implicit dynamic incentives such as career concerns and ratchet effects may alleviate or aggravate these problems.

First we considered the case where the principal provides incentives on a verifiable, but distorted, performance measure, and in addition some information about the agent's performance is provided to the principal (and the market) through a non-distorted but non-verifiable measure of true value. In this case implicit incentives are related both to the distorted and the non-distorted performance measures (and hence the degree of misalignments between them).

The analysis demonstrated that implicit dynamic incentives have important real effects in such settings, and that these effects are in several respects different from the corresponding effects in settings where a single non-distorted performance measure is available. In particular, we found that both career and ratchet elements have real effects; neither can costlessly be neutralized by monetary incentives, and that stronger ratchet effects may increase optimal monetary incentives. The main mechanism behind these results is that the verifiable performance measure serves as a surrogate measure for the true value. Hence if implicit incentives on the true value is weakened, the principal has to boost monetary incentives to ensure that the agent maintain the same level of effort.

The second model we present captures the fact that in addition to a verifiable (and distorted) performance measure, there are often other non-verifiable measures that may yield valuable information about the agent's performance. The availability of these measures will specifically affect the way the principal and the market update their estimates about the agent's talent. A typical case is one where quantity aspects are verifiable but quality aspects are not, yet some information about these quality aspects is observable for the relevant parties. In this setting we show that some new issues arise. Among other things show that explicit incentives may increase with more distortion in the (verifiable) performance measure. This effect occurs when more distortion induces the principal to increase monetary incentives in order to maintain an appropriate balance of the agent's effort among tasks. Hence in this case, and opposed to what is suggested in Kerr (1975), it may be appropriate to "reward for A, while hoping for B".

We finally analyze how the availability of distorted performance measures affects the desirability of different promotion rules. Bureaucracy, where rules rather than discretion are used to allocate resources, is a central feature of organizations, and a prominent example is the use of seniority in promotion and layoff decisions, independent of profitability considerations. This motivated the analysis of two different promotion rules: promote the agent with the highest expected talent, given all information at date 2, or promote the agent with best verifiable performance in period one. We show that promoting according to the bureaucratic rule 2 may improve the incentives for agents to allocate their effort correctly (seen from the principal's view), but the principal runs the risk of promoting the least fitted since not all information is used to update the estimates of agents' abilities. The two promotion rules thus illustrate an important trade-off. Promoting based on expected talent creates implicit incentives since the principal will use all information available to update her estimate of agents' talent. This includes information created from agents' activities on tasks that may be of low value for the principal. By committing to promote exclusively on verifiable information, the principal effectively removes such implicit incentives. But this commitment comes at a cost, since the principal runs the risk of promoting the least fitted.

References

Baker, G. (1992). Incentives contracts and performance measurement. *Journal of Political Economy 100*(3), 598–614.

Baker, G. (2002). Distortion and risk in optimal incentive contracts. *Journal of Human Resources 37*(3), 728–751.

Datar, S., S. Kulp, and R. Lambert (2001). Balancing performance measures. *Journal of Accounting Research 39*.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Fama, E. F. (1980). Agency problems and the theory of the firm. *Journal of Political Economy 88*, 288–307.

Feltham, G. and J. Xie (1994). Performance measure congruity and diversity in multitask/agent relations. *The Accountion Review 69*.

Gibbons, R. (1987). Piece-rate incentives schemes. *Journal of Labor Economics 5*, 413–29.

Gibbons, R. and K. J. Murphy (1992). Optimal incentives in the presence of career concerns: Theory and evidence. *Journal of Political Economy 100*, 468–505.

Holmstrom, B. (1982). Managerial incentive schemes–a dynamic perspective. In *Essays in economics and management in honour of Lars Wahlbeck*. Svenska Handelshogskolan, Helsinki.

Holmstrom, B. and P. Milgrom (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica 55*, 303–328.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization 7*, 24–52.

Huges, J. J., L. Zhang, and J.-Z. J. Xie (2005). Production externalities, congruity of aggregate signals, and optimal task assignment. *Contemporary Accounting Research 22*, 393–408.

Kerr, S. (1975). On the folly of rewarding for A while hoping for B. *Academy of Management Journal 18*(4), 769–783.

Meyer, M. A., T. E. Olsen, and G. Torsvik (1996). Limited intertemporal commitment and job design. *Journal of Economic Behavior and Organization 31*, 401–417.

Milgrom, P. and J. Roberts (1988). An economic approach to influence in organizations. *The American Journal of Sociology 94*, S154–S179.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economics Literature 37*, 7–63.

Weitzman, M. (1976). The ratchet principle and performance incentives. *Bell Journal Of Economics 11*, 302–308.

Appendices

A Technicalities

In this appendix we provide more details regarding some of the calculations in this paper.

A.1 Regression Coefficients

We first consider the case outlined in section 2 and 3. In this case the information signals are

$$\begin{aligned} y_t &= h^{'}\eta + \mathbf{fa}_t + \varepsilon_t \\ z_t &= \eta + \mathbf{ga}_t + \zeta_t. \end{aligned}$$

We seek $E(z_2 \mid z_1, y_1)$ and $E(\eta \mid z_1, y_1)$.The covariance matrixes (z_2, z_1, y_1) and (η, z_1, y_1) are

$$\begin{bmatrix} \sigma_\eta^2 & \sigma_\eta^2 & h'\sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\zeta^2 & h'\sigma_\eta^2 \\ h'\sigma_\eta^2 & h'\sigma_\eta^2 & (h')^2\sigma_\eta^2 + \sigma_\varepsilon^2 \end{bmatrix},$$

$$\begin{bmatrix} \sigma_\eta^2 + \sigma_\zeta^2 & \sigma_\eta^2 & h'\sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\zeta^2 & h'\sigma_\eta^2 \\ h'\sigma_\eta^2 & h'\sigma_\eta^2 & (h')^2\sigma_\eta^2 + \sigma_\varepsilon^2 \end{bmatrix}.$$

By inverting and applying well-known formulas (e.g. DeGroot 1970) we obtain:

$$\begin{aligned} \sigma_{2c}^2 &= Var(z_2 \mid z_1, y_1) = \frac{\sigma_\eta^2\sigma_\varepsilon^2\sigma_\zeta^2}{\sigma_\eta^2\sigma_\varepsilon^2 + \sigma_\zeta^2 h'^2\sigma_\eta^2 + \sigma_\zeta^2\sigma_\varepsilon^2} + \sigma_\zeta^2 \\ \rho_z &= \frac{\partial}{\partial z}E(\eta \mid \mathbf{x}_1) = \left[\frac{\sigma_\eta^2}{\sigma_\eta^2\sigma_\varepsilon^2 + \sigma_\zeta^2 h'^2\sigma_\eta^2 + \sigma_\zeta^2\sigma_\varepsilon^2}\right]\sigma_\varepsilon^2 \\ \rho_y &= \frac{\partial}{\partial y}E(\eta \mid \mathbf{x}_1) = \left[\frac{h'\sigma_\eta^2}{\sigma_\eta^2\sigma_\varepsilon^2 + \sigma_\zeta^2 h'^2\sigma_\eta^2 + \sigma_\zeta^2\sigma_\varepsilon^2}\right]\sigma_\zeta^2. \end{aligned}$$

We note that $h'\rho_y < 1$.

Consider the case outlined in Section 4. In this case the information signals are

$$\begin{aligned} y_t &= h'\eta + \mathbf{fa}_t + \varepsilon_t \\ z_t &= \eta + \mathbf{ga}_t + \zeta_t \\ q_t &= \eta + \mathbf{ka}_t + \kappa_t. \end{aligned}$$

The covariance matrix for (η, y_1, z_1, q_1) is now

$$\begin{bmatrix} \sigma_\eta^2 & h'\sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 \\ h'\sigma_\eta^2 & h'^2\sigma_\eta^2 + \sigma_\varepsilon^2 & h'\sigma_\eta^2 & h'\sigma_\eta^2 \\ \sigma_\eta^2 & h'\sigma_\eta^2 & \sigma_\eta^2 + \sigma_\zeta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & h'\sigma_\eta^2 & \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\kappa^2 \end{bmatrix}$$

By inverting and applying well-known formulas we get

$$\begin{aligned} \rho_y &= \frac{\partial}{\partial y} E(\eta \mid y_1, z_1, q_1) = \frac{h'\sigma_\eta^2\sigma_\zeta^2\sigma_\kappa^2}{h'^2\sigma_\eta^2\sigma_\zeta^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\eta^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\eta^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\kappa^2} \\ \rho_z &= \frac{\partial}{\partial z} E(\eta \mid y_1, z_1, q_1) = \frac{\sigma_\eta^2\sigma_\varepsilon^2\sigma_\kappa^2}{h'^2\sigma_\eta^2\sigma_\zeta^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\eta^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\eta^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\kappa^2} \\ \rho_q &= \frac{\partial}{\partial q} E(\eta \mid y_1, z_1, q_1) = \frac{\sigma_\eta^2\sigma_\varepsilon^2\sigma_\zeta^2}{h'^2\sigma_\eta^2\sigma_\zeta^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\eta^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\eta^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\kappa^2}. \end{aligned}$$

B Proofs

Proof of Proposition 2. Recall that first- and section period incentives are given by

$$\tilde{\alpha}_1^* = \alpha_1^* + \beta_z = \frac{\mathbf{fg}}{|\mathbf{g}|^2 + r\sigma_{1z}^2} - \rho_y \frac{h\left[\mathbf{fg} + r\sigma_{1yz}\right] - \alpha_2^*\mathbf{fg}}{|\mathbf{g}|^2 + r\sigma_{1z}^2} - \rho_z \frac{r\alpha_2^*\sigma_{1z}^2}{|\mathbf{g}|^2 + r\sigma_{1z}^2}, \tag{6}$$

$$\alpha_2^* = \frac{\mathbf{fg}}{|\mathbf{g}|^2 + r\sigma_{2c}^2}. \tag{3}$$

From (3) we have $\mathbf{fg} = \alpha_2^* K'$, where $K' = |g|^2 + r\sigma_{2c}^2$. Better alignment will thus increase α_2^*. Defining $K = |g|^2 + r\sigma_{1z}^2$, we have moreover from (6)

$$\begin{aligned} \tilde{\alpha}_1^* K &= (1-\beta_y)\alpha_2^* K' - r\alpha_2^*\rho_z\sigma_{1z}^2 - rh\rho_y\sigma_{1yz} \\ &= \left[(1-(h-\alpha_2^*)\rho_y)K' - r\rho_z\sigma_{1z}^2\right]\alpha_2^* - rh\rho_y\sigma_{1yz} \end{aligned}$$

where in particular the square bracket must be positive. Differentiation yields

$$\frac{\partial\tilde{\alpha}_1^*}{\partial\alpha_2^*} K = \left[(1-(h-\alpha_2^*)\rho_y)K' - r\rho_z\sigma_{1z}^2\right] + \alpha_2^*\rho_y K'$$

This is positive for all parameter values that yield non-negative $\tilde{\alpha}_1^*$. This proves the first part of the proposition.

Consider next the equilibrium total surplus $TCE_1^* + TCE_2^*$. From the envelope property, and taking into account that the equilibrium first period effort is $\mathbf{a}_1^* = \tilde{\alpha}_1^* g + \beta_y f$, we obtain from (5):

$$\begin{aligned} \frac{\partial TCE_1^*}{\partial g_i} &= f_i\tilde{\alpha}_1^* - a_{1i}^*\tilde{\alpha}_1^* - r\left((\tilde{\alpha}_1^* + \alpha_2^*\rho_z)\rho_z\sigma_{1z}^2 + \rho_z h\rho_y\sigma_{1yz}\right)\frac{\partial\alpha_2^*}{\partial g_i} \\ &= ((1-\beta_y)f_i - \tilde{\alpha}_1^* g_i)\tilde{\alpha}_1^* - r\left((\tilde{\alpha}_1^* + \alpha_2^*\rho_z)\sigma_{1z}^2 + h\rho_y\sigma_{1yz}\right)\rho_z\frac{\partial\alpha_2^*}{\partial g_i} \end{aligned}$$

where
$$\frac{\partial \alpha_2^*}{\partial g_i} = \frac{\partial}{\partial g_i}\frac{\Sigma f_i g_i}{\Sigma g_i^2 + r\sigma_{2c}^2} = \frac{f_i}{\left(\Sigma g_i^2 + r\sigma_{2c}^2\right)} - \frac{2\alpha_2^* g_i}{\left(\Sigma g_i^2 + r\sigma_{2c}^2\right)}$$

In a similar way we obtain from (2):

$$\frac{\partial TCE_2^*}{\partial g_i} = f_i\alpha_2^* - a_{2i}^*\alpha_2^* = (f_i - g_i\alpha_2^*)\alpha_2^*$$

All in all we thus have, for the equilibrium total surplus

$$\frac{\partial TCE^*}{\partial g_i} = Af_i - Bg_i$$

where A, B are independent of i. The formula shows that the surplus is maximal when the vector g is perfectly aligned with the vector $\mathbf{f}$. ■

Proof of Proposition 4. Consider a vector $\mathbf{g}$ that is a linear combination of $\mathbf{f}$ and $\mathbf{k}$, i.e. that lies in the plane spanned by $\mathbf{f}$ and $\mathbf{k}$:

$$\mathbf{g} = \lambda\mathbf{f} + \mu\mathbf{k}, \qquad \lambda, \mu \in R$$

We have $|\mathbf{g}|^2 = \lambda^2 + 2\lambda\mu(\mathbf{fk}) + \mu^2$, so the normalization $|\mathbf{g}|^2 = 1$ requires that

$$\lambda^2 + 2\lambda\mu c + \mu^2 = 1 \tag{15}$$

Note that we here have

$$\alpha_2^* = \mathbf{fg} = \lambda + \mu\mathbf{fk} = \lambda + \mu c$$

since $c := \mathbf{fk} = \cos(\theta_{fk})$. By substituting this in (15), note that the restriction on admissible values for λ, μ is equivalent to

$$(\alpha_2^*)^2 + \mu^2(1 - c^2) = 1$$

This equation can be considered as defining α_2^* as a function of μ, i.e.

$$\alpha_2^*(\mu) = \sqrt{1 - \mu^2(1 - c^2)}$$

This relation describes how the optimal static bonus $\alpha_2^* = \mathbf{fg}$ depends on the scalar μ, given that we require $\mathbf{g}$ to lie in the plane spanned by $\mathbf{f}$ and $\mathbf{k}$, and to be of length 1.

Now consider the optimal first period bonus, given by (8). Taking into account that all vectors are normalized to be of length 1, and substituting for the implicit incentives β_y, β_q we have

$$\tilde{\alpha}_1^* = (1 - [h - \alpha_2^*]\,\rho_y)\alpha_2^* - [h - \alpha_2^*]\,\rho_q \cos(\theta_{kg}),$$

where now $\cos(\theta_{kg}) = \mathbf{kg} = \lambda(\mathbf{kf}) + \mu = \lambda c + \mu$. Recall from above that $\alpha_2^* = \lambda + \mu c$, and that we hence my write $\cos(\theta_{kg}) = \lambda c + \mu = \alpha_2^* c + \mu(1 - c^2)$. Thus we have the following expression for the optimal first period bonus

$$\tilde{\alpha}_1^* = (1 - [h - \alpha_2^*]\,\rho_y)\alpha_2^* - [h - \alpha_2^*]\,\rho_q\left[\alpha_2^* c + \mu(1 - c^2)\right] \equiv F(\alpha_2^*, \mu)$$

Taking into account the dependence of α_2^* on μ, this equation describes how the parameter μ influences $\tilde{\alpha}_1^*$ partly directly via its influence on $\cos(\theta_{kg})$, and partly indirectly through its influence on α_2^*.

From the last equation we have

$$\frac{d\tilde{\alpha}_1^*}{d\mu} = F_\alpha(\alpha_2^*, \mu)\frac{d\alpha_2^*}{d\mu} + F_\mu(\alpha_2^*, \mu)$$

Evaluating this derivative at $\mu = 0$ we see that we have $\frac{d\alpha_2^*}{d\mu} = 0$ and hence (since $\alpha_2^*(0) = 1$)

$$\frac{d\tilde{\alpha}_1^*}{d\mu}\,|_{\mu=0} = F_\mu(\alpha_2^*(0), 0) = -\,[h - 1]\,\rho_q(1 - c^2)$$

Thus we see that the first period bonus $\tilde{\alpha}_1^*$ will increase by increasing μ when $h < 1$ and reducing μ when $h > 1$. Since $\mathbf{g} = \mathbf{f}$ for $\mu = 0$, this shows that the bonus is largest for some $\mathbf{g} \neq \mathbf{f}$. ■

Proof of formula (12). To simplify notation we assume in this proof that $E\eta = 0$. To emphasize the distinction between assumed and actual efforts we also write $E(\eta^1 \mid \mathbf{x}_1^1, \mathbf{a}_1^1) = \xi^1 + \mathbf{m}(\mathbf{a}_1^1 - \hat{\mathbf{a}}_1)$

Given that the agent wins, he will negotiate with the principal. The principal assumes $\mathbf{a}_1^1 = \hat{\mathbf{a}}_1$. The fixed wage A_2 is adjusted so that $sTCE_2 = CE_2$, which implies

$$\begin{aligned} A_2(\mathbf{x}_1^1) &= (h - \alpha_2^*)E(\eta^1 \mid \mathbf{x}_1^1, \hat{\mathbf{a}}_1) + C_2 = (h - \alpha_2^*)\xi^1 + C_2 \\ C_2 &= [s(\mathbf{fa}_2^* - c(\mathbf{a}_2^*)) - \alpha_2^*\mathbf{ga}_2^* + c(\mathbf{a}_2^*)] \end{aligned}$$

If the agent works $\mathbf{a}_1^1$ (instead of $\hat{\mathbf{a}}_1$) in period 1, he will get the following reward in period 2:

$$\begin{aligned} W_2 &= A_2(\mathbf{x}_1^1) + \alpha_2^*(E(\eta^1 \mid \mathbf{x}_1^1, \mathbf{a}_1^1) + \mathbf{ga}_2^*) - c(\mathbf{a}_2^*) \\ &= \left[(h - \alpha_2^*)\xi^1 + C_2\right] + \alpha_2^*(\xi h1 + \mathbf{m}(\mathbf{a}_1^1 - \hat{\mathbf{a}}_1) + \mathbf{ga}_2^*) - c(\mathbf{a}_2^*) \\ &= h\xi^1 + (h - \alpha_2^*)\mathbf{m}(\mathbf{a}_1^1 - \hat{\mathbf{a}}_1) + s(\mathbf{fa}_2^* - c(\mathbf{a}_2^*)) \\ &= h\xi^1 + (h - \alpha_2^*)\mathbf{ma}_1^1 + \hat{C} \end{aligned}$$

where $\hat{C}$ is defined by the last equality. Let u_2 be the reward if the agent loses. The agent wins if $\xi^1 - \xi^2 > \mathbf{m}(\mathbf{a}_1^2 - \mathbf{a}_1^1)$, where in equilibrium $\mathbf{a}_1^2 = \hat{\mathbf{a}}_1^2 = \hat{\mathbf{a}}_1$. Let $\Phi(\xi)$ with density $\phi(\xi)$ be the CDF for ξ^i. The expected reward in the contest is then

$$W(\mathbf{ma}_1^1) = \int_{-\infty}^{\infty} \phi(\xi^1)\left[h\xi^1 + (h - \alpha_2^*)\mathbf{ma}_1^1) + \hat{C} - u_2\right]\Phi\left(\xi^1 + \mathbf{m}(\mathbf{a}_1^1 - \mathbf{a}_1^2)\right)d\xi^1 + u_2$$

Letting $\Phi_2(\cdot)$ be the CDF for the normal variable $\xi^2 - \xi^1$ we thus have

$$\begin{aligned} W(\mathbf{m}\mathbf{a}_1^1) &= h\int_{-\infty}^{\infty} \xi^1 \Phi\left(\xi^1 + \mathbf{m}(\mathbf{a}_1^1 - \mathbf{a}_1^2)\right)\phi(\xi^1)d\xi^1 \\ &\quad + \left[(h-\alpha_2^*)\mathbf{m}\mathbf{a}_1^1 + \hat{C} - u_2\right]\Phi_2(\mathbf{m}(\mathbf{a}_1^1 - \mathbf{a}_1^2)) + u_2 \end{aligned}$$

Differentiating this, and evaluating the derivative at $\mathbf{a}_1^1 = \mathbf{a}_1^2$ (which will hold in equilibrium) we get

$$\begin{aligned} \frac{\partial W}{\partial a_{1i}^1}(\mathbf{m}\mathbf{a}_1^1) &= h\int_{-\infty}^{\infty} \xi^1 \phi^2\left(\xi^1\right) d\xi^1 m_i \\ &\quad + (h-\alpha_2^*)m_i\Phi_2(0) + \left[(h-\alpha_2^*)\mathbf{m}\mathbf{a}_1^1 + \hat{C} - u_2\right]\phi_2(0)m_i \end{aligned}$$

The first integral is zero, since $\phi^2\left(\xi^1\right)$ is symmetric around $E\xi^1 = 0$. Substituting for $\hat{C}$ we see that in equilibrium (for $\mathbf{a}_1^1 = \hat{\mathbf{a}}_1$) the last square bracket equals $s(\mathbf{f}\mathbf{a}_2^* - c(\mathbf{a}_2^*)) - u_2 = R$. This proves formula (12). ■

Proof of Lemma 1. Let $\Delta = 2(TCE_{1z} - TCE_{1x})$. We will show that

$$\begin{aligned} \Delta &= -\gamma(2-\gamma)\left|\mathbf{f}\right|^2\left(1-\cos^2\theta_{fg}\right) + \lambda^2\left|\mathbf{k}\right|^2\left(1-\cos^2\theta_{gk}\right) \\ &\quad + 2\lambda(1-\gamma)\left|\mathbf{f}\right|\left|\mathbf{k}\right|\left(\cos\theta_{fg}\cos\theta_{kg} - \cos\theta_{kf}\right) \end{aligned}$$

where $\gamma = \omega\rho_y$ and $\lambda = \omega\rho_q$. The statements in the lemma follow from this formula.

To prove the formula consider (to simplify notation we set $E\eta = 0$ in this proof)

$$TCE_1 = \mathbf{f}\mathbf{a} - \frac{1}{2}\mathbf{a}^2 = \frac{1}{2}\mathbf{f}^2 - \frac{1}{2}(\mathbf{f}-\mathbf{a})^2 = \frac{1}{2}\mathbf{f}^2 - \frac{1}{2}(\mathbf{f} - \tilde{\alpha}\mathbf{g} - t\mathbf{n})^2$$

The optimal (effective bonus) $\tilde{\alpha}$ is given by

$$(\mathbf{f} - \tilde{\alpha}\mathbf{g} - t\mathbf{n})\mathbf{g} = 0 \quad \text{i.e.} \qquad \tilde{\alpha} = \frac{1}{\mathbf{g}^2}(\mathbf{f}\mathbf{g} - t\mathbf{n}\mathbf{g})$$

Hence

$$\begin{aligned} 2\cdot TCE_1 &= \mathbf{f}^2 - (\mathbf{f} - \tilde{\alpha}\mathbf{g} - t\mathbf{n})(\mathbf{f} - \tilde{\alpha}\mathbf{g} - t\mathbf{n}) \\ &= \mathbf{f}^2 - (\mathbf{f} - \tilde{\alpha}\mathbf{g} - t\mathbf{n})(\mathbf{f} - t\mathbf{n}) - 0 \\ &= \mathbf{f}^2 - (\mathbf{f} - t\mathbf{n})^2 + \tilde{\alpha}\mathbf{g}(\mathbf{f} - t\mathbf{n}) \\ &= \mathbf{f}^2 - (\mathbf{f} - t\mathbf{n})^2 + \frac{1}{\mathbf{g}^2}(\mathbf{f}\mathbf{g} - t\mathbf{n}\mathbf{g})^2 \end{aligned}$$

We have $\Delta/2 = TCE_1(t=0) - TCE_1(t=1)$ and thus

$$\begin{aligned} \Delta &\equiv (\mathbf{f}-\mathbf{n})^2 - (\mathbf{f})^2 + \left(\frac{1}{\mathbf{g}^2}(\mathbf{f}\mathbf{g})^2 - \frac{1}{\mathbf{g}^2}(\mathbf{f}\mathbf{g} - \mathbf{n}\mathbf{g})^2\right) \\ &= -\mathbf{n}(2\mathbf{f} - \mathbf{n}) + \frac{(\mathbf{n}\mathbf{g})}{\mathbf{g}^2}(2\mathbf{f} - \mathbf{n})\mathbf{g} \end{aligned}$$

For $\mathbf{n} = \gamma\mathbf{f} + \lambda k$ we get

$$\begin{aligned}
\Delta &= \frac{(\mathbf{ng})}{\mathbf{g}^2}(2\mathbf{f}-\mathbf{n})\mathbf{g} - \mathbf{n}(2\mathbf{f}-\mathbf{n}) \\
&= \frac{1}{\mathbf{g}^2}(\gamma\mathbf{fg}+\lambda\mathbf{kg})((2-\gamma)\mathbf{fg}-\lambda\mathbf{kg}) - (\gamma\mathbf{f}+\lambda\mathbf{k})((2-\gamma)\mathbf{f}-\lambda\mathbf{k}) \\
&= \frac{1}{\mathbf{g}^2}\left(\gamma(2-\gamma)(\mathbf{fg})^2 - \lambda^2(\mathbf{kg})^2 + [\lambda(2-\gamma)-\lambda\gamma]\,(\mathbf{fg})(\mathbf{kg})\right) \\
&\quad -(\gamma(2-\gamma)\mathbf{f}^2 - \lambda^2\mathbf{k}^2 + [\lambda(2-\gamma)-\lambda\gamma]\,(\mathbf{fk})) \\
&= -\gamma(2-\gamma)\left(\mathbf{f}^2 - \frac{(\mathbf{fg})^2}{\mathbf{g}^2}\right) + \lambda^2\left(\mathbf{k}^2 - \frac{(\mathbf{kg})^2}{\mathbf{g}^2}\right) + 2\lambda(1-\gamma)\left(\frac{(\mathbf{fg})(\mathbf{kg})}{\mathbf{g}^2} - \mathbf{fk}\right) \\
&= -\gamma(2-\gamma)\mathbf{f}^2\left(1-\cos^2\theta_{fg}\right) + \lambda^2\mathbf{k}^2\left(1-\cos^2\theta_{gk}\right) \\
&\quad +2\lambda(1-\gamma)\,|\mathbf{f}|\,|\mathbf{k}|\,(\cos\theta_{fg}\cos\theta_{kg} - \cos\theta_{kf})
\end{aligned}$$

This proves the lemma. ■

Proof of Lemma 2. In this proof we set $E\eta = 0$ to simplify notation. Consider

$$N_z = h' E \max\left\{E(\eta^1 \mid z_1^1), E(\eta^2 \mid z_1^2)\right\}$$

where (in equilibrium) $E(\eta^i \mid z_1^i) = \tilde{\rho}_z(z_1^i - Ez_1^i) = \tilde{\rho}_z(\eta^i + \zeta_1^i)$. Here $\tilde{\rho}_z$ is the relevant regression coefficient; $\tilde{\rho}_z = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\zeta^2}$ (obtained e.g. from ρ_z by letting $\sigma_\varepsilon^2, \sigma_\zeta^2 \to \infty$). So we have

$$\begin{aligned}
E\max\left\{E(\eta^1 \mid z_1^1), E(\eta^2 \mid z_1^2)\right\} &= \tilde{\rho}_z E\max\left\{(\eta^1+\zeta_1^1), (\eta^2+\zeta_1^2)\right\} \\
&= E\max\left\{\tilde{\rho}_z(\eta^1-\eta^2+\zeta_1^1-\zeta_1^2), 0\right\}
\end{aligned}$$

Let $\phi_3(\cdot)$ be the density and $\Phi_3(\cdot)$ the CDF for the normal variable $\mu = \tilde{\rho}_z\left(\eta^1-\eta^2+\zeta_1^1-\zeta_1^2\right)$. Then we have

$$N_z/h' = \int_0^\infty \mu\phi_3(\mu)d\mu = \int_0^\infty (1-\Phi_3(\mu))d\mu = \int_0^\infty (1-\Phi_0(x))dx\sqrt{V_z}$$

where $\Phi_0(\cdot)$ is the CDF for a standard normal $N(0,1)$ variable, and the last equality follows from $\Phi_3(y) = \Phi_0(y/\sqrt{V_z})$, $V_z = var(\mu)$. Note that

$$V_z = var(\mu) = 2\tilde{\rho}_z^2(\sigma_\eta^2+\sigma_\zeta^2) = 2\frac{(\sigma_\eta^2)^2}{\sigma_\eta^2+\sigma_\zeta^2}.$$

Analogously we we have

$$N_x/h' = \int_0^\infty (1-\Phi_0(x))dx\sqrt{V_x}$$

where

$$V_x = var(\xi^2 - \xi^1) = 2\left[(\rho_z + \rho_y h' + \rho_q)^2\sigma_\eta^2 + \rho_z^2\sigma_\zeta^2 + \rho_y^2\sigma_\varepsilon^2 + \rho_q^2\sigma_\kappa^2\right].$$

This proves the first part of the lemma. Substituting for ρ_z, ρ_y, ρ_q in the last formula, we see by direct differentiation and limiting operations that the last part of the lemma also holds true. ■

Proof of Lemma 3.

We first note that

$$V_x = var(\xi^2 - \xi^1) = 2\left[(\rho_z + \rho_y h' + \rho_q)^2\sigma_\eta^2 + \rho_z^2\sigma_\zeta^2 + \rho_y^2\sigma_\varepsilon^2 + \rho_q^2\sigma_\kappa^2\right],$$

and

$$V_z = var(\mu) = 2\tilde{\rho}_z^2(\sigma_\eta^2 + \sigma_\zeta^2)$$

where ρ_z, ρ_y, ρ_q are given in Appendix A**,** and we have

$$\tilde{\rho}_z = \lim_{\sigma_\kappa^2, \sigma_\varepsilon^2 \to \infty} \rho_z = \lim_{\sigma_\eta^2, \sigma_\varepsilon^2 \to \infty} \frac{\sigma_\eta^2}{h'^2 \frac{\sigma_\eta^2\sigma_\zeta^2\sigma_\kappa^2}{\sigma_\kappa^2\sigma_\varepsilon^2} + \sigma_\eta^2 + \sigma_\zeta^2 + \frac{\sigma_\varepsilon^2\sigma_\zeta^2\sigma_\eta^2}{\sigma_\kappa^2\sigma_\varepsilon^2}} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\zeta^2}$$

Thus

$$V_z = 2\tilde{\rho}_z^2(\sigma_\eta^2 + \sigma_\zeta^2) = 2\frac{(\sigma_\eta^2)^2}{\sigma_\eta^2 + \sigma_\zeta^2}$$

Note that we have, for $\sigma_\eta^2 \to \infty$:

$$\begin{aligned} \rho_y &\to \frac{h'\sigma_\zeta^2\sigma_\kappa^2}{h'^2\sigma_\zeta^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\kappa^2 + \sigma_\varepsilon^2\sigma_\zeta^2} := \frac{h'\sigma_\zeta^2\sigma_\kappa^2}{A} \\ \rho_z &\to \frac{\sigma_\varepsilon^2\sigma_\kappa^2}{A}, \qquad \rho_q \to \frac{\sigma_\varepsilon^2\sigma_\zeta^2}{A}, \qquad \tilde{\rho}_z \to 1, \\ V_z, V_x &\to \infty \end{aligned}$$

To prove the Lemma it suffices to show that $\lim_{\sigma_\eta^2 \to \infty}\left(\sqrt{V_x} - \sqrt{V_z}\right) = 0$.

By concavity of $\sqrt{V}$ we have $\sqrt{V_x} - \sqrt{V_z} \leq \frac{1}{2\sqrt{V_z}}(V_x - V_z)$. The lemma is thus proved if we show that $(V_x - V_z)$ stays bounded as $\sigma_\eta^2 \to \infty$. We have

$$(V_x - V_z)/2 = \left[(\rho_z + \rho_y h' + \rho_q)^2 - \tilde{\rho}_z^2\right]\sigma_\eta^2 + (\rho_z^2 - \tilde{\rho}_z^2)\sigma_\zeta^2 + \rho_y^2\sigma_\varepsilon^2 + \rho_q^2\sigma_\kappa^2$$

The result follows if we show that $\left[(\rho_z + \rho_y h' + \rho_q) - \tilde{\rho}_z\right]\sigma_\eta^2$ stays bounded. From the expressions above we have

$$\left[(\rho_z + \rho_y h' + \rho_q) - \tilde{\rho}_z\right] = \frac{A\sigma_\eta^2}{A\sigma_\eta^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\kappa^2} - \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\zeta^2} = \frac{\left(A - \sigma_\varepsilon^2\sigma_\kappa^2\right)\sigma_\zeta^2\sigma_\eta^2}{\left(A\sigma_\eta^2 + \sigma_\varepsilon^2\sigma_\zeta^2\sigma_\kappa^2\right)\left(\sigma_\eta^2 + \sigma_\zeta^2\right)}$$

From this expression we see that $\left[(\rho_z + \rho_y h' + \rho_q) - \tilde{\rho}_z\right]\sigma_\eta^2$ does indeed stay bounded, and hence the proof is complete. ■

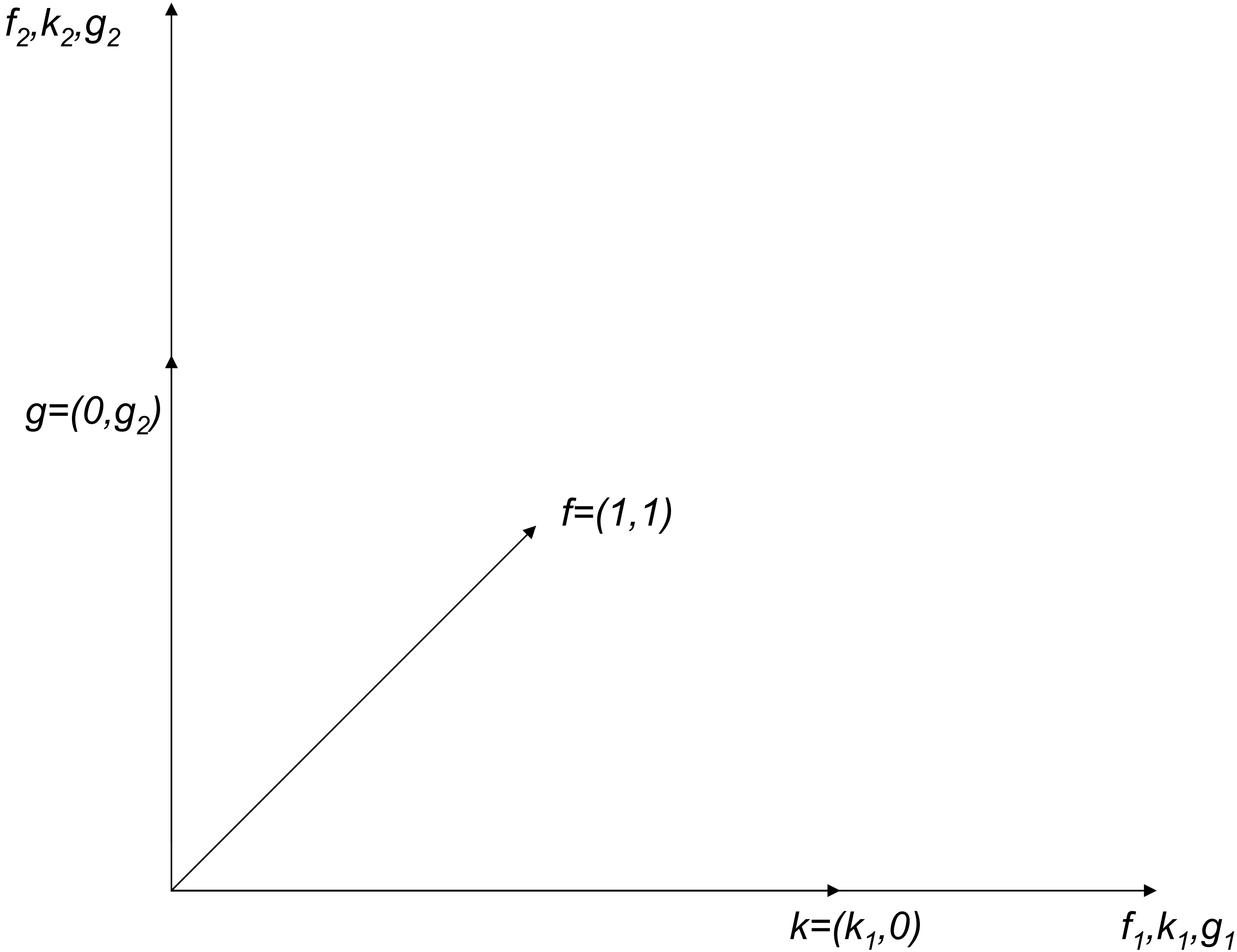

Figure 1

Department of Economics
University of Bergen
Fosswinckels gate 6
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ