Performance Targets, Effort and Risk-Taking

Matthew D. Rablen*
Brunel University

May 2010

Abstract

Growing economic and psychological evidence documents effects of target setting on levels of effort and risk-taking, even in the absence of a monetary reward for attaining the target. I explore a principal-agent environment in which the principal sets the agent a performance target, and the agent's intrinsic motivation to work is influenced by their performance relative to the target. When the agent has prospect theory preferences relative to the target I show that a performance target can induce greater effort, but, when set too high, it eventually induces lower effort. Also, the agent's preferences for risk-taking hinge on whether the target is set above or below expected output. I find that the principal's optimal target exceeds expected output.

JEL Classification: J28; D81; D86; D21
PsycINFO Classification: 2360; 3630; 3660
Keywords: Performance targets; Intrinsic motivation; Prospect theory; Principal-agent theory; Reference-dependent preferences.

*Correspondence: MJ265, Department of Economics and Finance, Brunel University, Uxbridge, UB8 3PH. E-mail: matthew.rablen@brunel.ac.uk

1 Introduction

Performance targets are a pervasive feature of the modern corporation. In standard principal-agent theory a performance target can act as an incentive device if a monetary reward is linked to achievement of the target. As the salience of the target level derives entirely from the monetary reward for its achievement, a performance target absent such a monetary reward would have no implications for behavior. However, this approach is inconsistent with an established psychological literature - reviewed in Locke and Latham (2002) - and an emerging economic literature (Camerer et al., 1997; Clark and Oswald, 1998; Mas 2006; Rablen, 2008; Rizzo and Zeckhauser, 2003) - that documents how a performance target in itself can influence economic behavior, even in the absence of a monetary reward contingent upon its achievement.

A reconciliation of the principal-agent approach with the above evidence is possible if pay is recognized as only one of many determinants of an individual's motivation to work. For instance, a substantial literature in psychology argues that an individual's motivation to work can be decomposed into intrinsic and extrinsic components (e.g. Deci, 1971, 1975; Deci and Ryan, 1985). While the extrinsic component of motivation includes monetary rewards, the idea of intrinsic motivation is that work can also provide its own inherent rewards.

In economics the idea that extrinsic rewards are not the only instrument through which individuals can be motivated to exert effort has taken longer to emerge, but is now receiving growing attention. Frey (1997a,b) uses the concept of intrinsic motivation to propose a theory in which individuals may 'perform work for work's sake.' The managers interviewed in Bewley (1999) emphasize that relying solely on wage motivation is unwise. Consistent with this, Sen (1977, p. 101) argues that 'to run an organization entirely on incentives to personal gain is pretty much a hopeless task.' Studies of happiness are making it increasingly clear that work itself can be utility enhancing: unemployment (not employment) is associated with high levels of mental distress (Clark, 2003; Clark and Oswald 1994; Di Tella et al., 2003). Last, Gneezy and Rustichini (2000a,b) present empirical evidence consistent with the idea that individuals can have an intrinsic motivation to work, but that such motivation can be crowded-out by extrinsic rewards.[1]

[1] Further indicative evidence is provided by Jensen and Murphy (1990), who establish empirically

This paper examines the behavioral implications of a performance target in a simple principal-agent setting. In particular I consider how, by influencing the agent's intrinsic motivation to work, the target can affect preferences for effort and risk-taking. I go on to examine how, given the behavior of the agent, the principal should optimally set the performance target.

There are several ways a performance target might influence intrinsic motivation. The cognitive evaluation theory of Deci (1975) argues that perceived competence is an important determinant of intrinsic motivation. Consistent with this, individuals who perceive themselves to be performing well in their job tend to have higher intrinsic motivation and report higher levels of job satisfaction (Barrick and Mount, 1991; Judge et al., 2001). One possibility is that when a performance target is set, it becomes a yardstick that can importantly affect an individual's perceived competence. Achieving the target can act as a signal of competence, while failure to achieve the target can act as a signal of a lack of competence. Additionally, individuals who achieve the target level may have their intrinsic motivation buttressed by praise from superiors, while failure to achieve the target might reduce intrinsic motivation through negative feedback (Deci, 1975; Vallerand and Reid, 1984).

The above arguments suggest that performance relative to the target is an important determinant of intrinsic motivation to work. For a level of output q and a target level of output t, I therefore write an individual's intrinsic motivation to work as $I\left[q-t,\Psi\right]$, where Ψ is a vector containing all other determinants of the intrinsic motivation to work. As I take intrinsic motivation to be reference-dependent with respect to t, I suppose that $I\left[\cdot\right]$ satisfies the properties of prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). In particular, the prospect theory functional implies diminishing sensitivity and loss aversion with respect to outcomes above and below the target level.

Allowing for loss aversion and diminishing sensitivity can be justified on a number of grounds. First, there is now an abundance of empirical and experimental evidence confirming these two properties of the prospect theory functional (see e.g. Abdellaoui, 2000; Laughhann et al., 1980; Mezias, 1988; Payne et al., 1981; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996, 1999). Second, these properties can explain

that pay-for-performance is applied less broadly, and with less intensity, than proposed by principal-agent theory.

otherwise puzzling empirical phenomena such as the reflection effect (Kahneman and Tversky, 1979); the disparity between measures of willingness to pay and willingness to accept (Bateman et al., 1997); and the goal gradient effect, whereby agents expend more effort as they approach a target (Heath et al., 1999; Hull, 1932; Kivetz et al., 2006). Third, I show that if $I[\cdot]$ exhibits loss aversion and diminishing sensitivity the present model is able to fit a range of psychological evidence on effort, risk-taking and target-setting, but not if $I[\cdot]$ is globally concave as is conventionally assumed in economic theory.

With this approach I predict an inverse-$\cup$ shaped relationship between the target level and effort - consistent with findings in the psychological literature on target setting. This effect obtains even though I assume no monetary reward from achieving the target. I also find that increasing the target level of performance increases the agent's preference for risk-taking - also consistent with the empirical findings of the psychological literature. This effect obtains even though the model assumes a complete decoupling of risk-taking from expected output.

Additionally, I find that the relationship between the target level and risk-taking hinges on whether the target level lies above or below expected output: below expected output the agent requires a positive risk premium to bear risk, but above expected output the agent will find it optimal to bear some degree of risk for no compensatory risk premium.

If the agent is allowed to choose simultaneously a preferred level of effort and a preferred level of risk, I find that, when the target level exceeds expected output, effort and risk-taking are substitutes: the agent responds to further increases in the performance target by reducing her effort, and increasing her exposure to risk. The principal's optimal choice of the target level is shown to exceed the equilibrium level of expected output, as the principal can exploit the agent's loss aversion to below-target outcomes.

In psychology, an earlier study by Wu et al. (2004) examines goal driven behavior under prospect theory preferences, but under conditions of certainty, and not within a principal-agent framework. The authors find that "easy" goals can improve performance and "hard" goals can worsen performance - a result mirrored in this paper

- but using somewhat different assumptions (agents are assumed to be myopic in a certain sense).

In economics, contributions that incorporate elements of prospect theory in a principal-agent setting include Herweg et al. (2008), de Meza and Webb (2007), Dittmann et al. (forthcoming) and Iantchev (2005).[2] These studies examine the implications for the optimal incentive scheme if agents are loss averse over compensation amounts that fall below their reference level. Unlike these studies, here I focus not on the extrinsic motivation to work captured by the optimal incentive scheme, but on how the target level itself can influence behavior through intrinsic motivation, and the related question of how the principal should respond to this behavior in setting the performance target.

The analysis is also related to a wider literature that incorporates prospect theory preferences with an endogenous reference level: Barberis et al. (2001) apply the model to asset prices; Berkelaar et al. (2004) to portfolio choice; Kanbur et al. (2008) to optimal taxation; and Dhami and al-Nowaihi (2007) to tax compliance. However, the present model differs from the above in the sense that the target level is only endogenous with respect to the principal, not to the agent.

The plan of the paper is as follows: Sections 2 and 3 outline a simple principal-agent model based on that of Akerlof (1982). Section 4 explores the predictions of the model for the effort and risk-taking behavior of the agent, and the principal's optimal choice of the performance target. Section 5 concludes.

2 Targets, Effort and Risk-Taking

A novel feature of the present analysis is that, as well as allowing the agent to make a choice of effort, I allow the agent also to choose among production strategies that involve differing degrees of risk. Underlying this approach is the idea that in a world of incomplete contracts and informational asymmetries, the agent is able to exercise discretion over certain aspects of her behavior. Risk-taking behavior by managers in the corporate environment is typically constrained from both sides: risk cannot typically

[2] Studies by Rayo and Becker (2007) and Rablen (2009) also consider the choice of a target or reference level in a principal-agent framework. However, these studies lack a ready interpretation in the context of the firm.

be eliminated entirely as managers face an element of systematic risk: firms operate in an inherently risky environment. Equally, systems are normally in place to limit discretionary risk-taking by managers. Nevertheless, within these two constraints, managers can exercise discretion as to the riskiness of the strategies they employ. In this sense it is possible to view the agent as being able to expose the principal to a degree of unsystematic risk in addition to the systematic risk she necessarily faces.

A further feature of the model is that the principal is able to set the agent's performance target. My idea here is that in corporate environments, performance targets are typically imposed hierarchically, with each layer in the hierarchy responsible for setting targets for the level below. For instance, upon entering a university an academic will typically be informed by the Director of Research of the quality of publications expected of them. Although in some instances sub-ordinates may be permitted to exert a degree of influence upon the targets set by their superior - indeed this is considered best practice - the targets set are ultimately at the discretion of the superior. Therefore, allowing the principal complete discretion over the agent's performance target should be seen as a simplification, but one that I argue is a close approximation to a more realistic specification.

As the above discussion should make clear, my interpretation of the target level differs from that of the earlier literature. For instance, recent studies of the labor supply of taxi-drivers (Camerer et al., 1997; Chou, 2002; Farber, 2008; Fehr and Goette, 2007) require a different interpretation of the target level, for as self-employed individuals, taxi-drivers are able to determine their own performance targets. The studies by Herweg et al. (2008), de Meza and Webb (2007) and Iantchev (2005) also view the reference level as being determined by the agent, with the principal able only to influence the agent's choice indirectly through her choice of the incentive scheme. These studies therefore do not capture the hierarchical process of target setting within the firm that I have in mind.

3 A Model

My model is a simple principal-agent setting that loosely follows the framework of Akerlof (1982). In the first period, the principal decides the agent's performance target t. To focus on the effect of the performance target on intrinsic motivation, I remove any extrinsic considerations by supposing that achievement of the target carries no conditional monetary reward. This also allows the model to replicate the setting employed in the psychological literature on target setting. However, it should be noted that the presence of a monetary reward only sharpens the incentive effects with respect to effort that I go on to describe. In the second period, the agent chooses a level of effort and a preferred level of risk, taking as given the performance target set by the principal.

3.1 Preferences

Following arguments made in the Introduction, I suppose the agent has an intrinsic motivation to work function given by $I[q-t,\Psi]$, where q is realized output, and Ψ captures all other determinants of the agent's intrinsic motivation to work. Since I hold all elements of Ψ constant in what follows, herein I write intrinsic motivation as simply $I[q-t]$. An obvious deficiency of this specification is that it ignores the possibility that the agent may also derive intrinsic motivation in some part from their absolute performance. However, I choose to work with the present form so as to focus attention on the role of the target level, and to retain the overall simplicity of the presentation.

Following prospect theory I make the following assumptions on $I[\cdot]$:

A0. $I[x]$ is continuous for all x, twice differentiable for $x \neq 0$, and $I[0] = 0$.
A1. $I[x]$ is strictly increasing.
A2. $I'[x] < I'[-x]$ for $x > 0$.
A3. $I''[x] < 0$ for $x > 0$ and $I''[x] > 0$ for $x < 0$.

Assumptions A0 and A1 are standard technical assumptions. Assumption A2 is loss aversion, and implies that the loss of intrinsic motivation from a below-target outcome exceeds the gain in intrinsic motivation from an equivalent above-target outcome. Assumption A3 is diminishing sensitivity, which requires that marginal intrinsic motivation is a decreasing function in distance from the target level. Diminishing sensitivity implies risk seeking preferences over below-target outcomes and risk averse preferences over above-target outcomes. Together, loss aversion and diminishing sensitivity imply a kink-point at the target level of output.

In the light of widespread evidence of risk aversion amongst firms (see e.g. Hubbard,

1998) I suppose that the principal derives utility from profit according to the function $V[\pi]$, where $V' > 0$ and $V'' < 0$.

3.2 Production Environment

The agent expends an amount of effort e which yields an uncertain level of output $q \equiv e + \varepsilon$, where ε is a random output shock.[3] Effort is assumed to be costly, which is captured by a cost function $c[e]$, where it is assumed that $c' > 0$ and $c'' > 0$.

The total uncertainty in production is measured by the parameter A, which is composed of a systematic element, measured by $\alpha \leq A$, and an unsystematic element, measured by $A - \alpha \geq 0$. To capture this idea as simply as possible I assume that ε is uniformly distributed on the interval $[-A, A]$. The agent is able to choose A, subject to the constraint that $A \geq \alpha$. By choosing ε to have a zero mean, the model implies a complete decoupling of risk-taking from expected output. I do this to make clear that any preference for risk-taking observed in the model cannot be explained via the standard trade-off between risk and expected return.

3.3 Agent's Problem

I assume a simple production environment that introduces uncertainty into the earlier framework of Wu et al. (2004). The problem facing the agent is to choose a level of effort, and an optimal level of unsystematic risk. Although $I[\cdot]$ is not differentiable at the origin, under assumptions A0-A3 it is integrable. The agent's problem is therefore given by

$$\max_{A,e} \frac{1}{2A} \int_{-A}^{A} I[e + \varepsilon - t] \; d\varepsilon - c[e] \qquad \text{s.t. } A \geq \alpha. \tag{1}$$

However, it is often more instructive to work with the unconstrained version of (1) given by

$$\max_{a,e} \frac{1}{2a} \int_{-a}^{a} I[e + \varepsilon - t] \; d\varepsilon - c[e], \tag{2}$$

[3] I have also investigated allowing for productivity shocks that interact multiplicatively with effort. However, this only complicates the results without changing the qualitative conclusions.

where $a \geq 0$ can be interpreted as the agent's preferred level of risk. If $(a, e[a])$ is a solution to the unconstrained problem in (2) then a solution to the agent's constrained problem in (1) is given by:

$$A = \begin{cases} \alpha & a < \alpha \\ a & a \geq \alpha \end{cases} \qquad e[A] = \begin{cases} e[\alpha] & a < \alpha \\ e[a] & a \geq \alpha \end{cases}. \tag{3}$$

It can be seen that when the level of systematic risk exceeds the agent's preferred level of risk there is no incentive for the agent to generate additional unsystematic risk. However, if the agent's preferred level of risk exceeds the systematic risk, she will respond by creating additional unsystematic risk.

To facilitate analysis I can rewrite the unconstrained problem in (2) as:

$$\max_{a,e} \frac{1}{2a} \int\limits_{Y[a,e,t]}^{Z[a,e,t]} I[\varphi] \; d\varphi - c[e], \tag{4}$$

where

$$Y[a,e,t] \equiv e - a - t; \qquad Z[a,e,t] \equiv e + a - t. \tag{5}$$

The first-order conditions are then:

$$e \;:\; \frac{I[Z[a,e,t]] - I[Y[a,e,t]]}{2a} = c'[e]; \tag{6}$$

$$a \;:\; a\left(I[Z[a,e,t]] + I[Y[a,e,t]]\right) = \int\limits_{Y[a,e,t]}^{Z[a,e,t]} I[\varphi] \; d\varphi. \tag{7}$$

The second-order conditions (in this instance sufficient for an interior maximum) are:

$$\frac{I'[Z] - I'[Y]}{2a} - c''[e] < 0; \tag{8}$$

$$-2a\left(I'[Y] - I'[Z]\right) < 0; \tag{9}$$

where, as is the case throughout, the derivatives of $I[\cdot]$ are defined for $Y, Z \neq 0$. Because diminishing sensitivity implies convex intrinsic motivation over below-target outcomes, the second-order condition for effort in (8) is not guaranteed to hold. If (8) is not satisfied, then the optimal effort is either zero, or it is unbounded above. From a positive standpoint, neither of these possible outcomes is attractive. However, so long as the cost function is sufficiently convex I may proceed under the more plausible assumption of an interior maximum for effort. The second-order condition for risk in (9) shows that the agent's preferred risk level must satisfy the property that $I'[Y] > I'[Z]$.

3.4 Principal's Problem

The problem of the principal is to choose the agent's output target t to maximize expected utility, subject to the behavioral response of the agent, as summarized by (3), (6) and (7). The principal's expected utility is given by:

$$\frac{1}{2A[t]} \int_{-A[t]}^{A[t]} V\left[(p-w)\left(e[t]+\varepsilon\right)\right] \, d\varepsilon, \tag{10}$$

where w is the piece-rate paid per unit of output, $p \geq w$ is the market price of a unit of output, and $(A[t], e[t])$ are respectively the agent's optimal choice of risk and effort.

4 Analysis

4.1 Targets and Effort

Before proceeding to analyze the simultaneous choice of effort and unsystematic risk by the agent it is first instructive to examine these two choices separately. In particular, it is helpful to understand the role of the target level in influencing these choices.

A large psychological literature examines the question of the relationship between targets and effort on costly tasks. On both physical (Bandura and Cervone, 1983; Ness and Patton, 1979) and cognitive (Bryan and Locke, 1967; Locke et al., 1970) tasks, subjects asked to achieve a target level of performance outperform subjects simply told to 'do their best'. These results are found both inside and outside the laboratory environment: woods workers given specific targets recorded significantly higher levels of productivity than did those in a 'do your best' condition (Latham and

Baldes, 1975; Latham and Kinne, 1974; Ronan et al., 1973).[4] Levels of persistence on difficult tasks are found to increase as the target level is raised (Hall et al., 1987; LaPorte and Nath, 1976; Stevenson et al., 1984). Consistent with these findings, Terpstra and Rozell (1994) find a positive relationship between reported profitability and use of target setting in questionnaire data from one thousand US employers.[5]

However, the psychological literature also finds that when the target level becomes excessively high, further increases in the target result in reduced levels of effort (Atkinson, 1958; Erez and Zidon, 1984; Locke and Latham, 1990). The psychological evidence therefore points to an asymmetric inverse-$\cup$ shaped relationship between the target level and effort: over most of the domain, effort rises with the target, but eventually begins to fall. This literature attributes the eventual decline of effort to a lapse in commitment to the target level, once it is seen as unattainable. I, however, offer an explanation based on the conventional concept of marginal utility.

These findings can be investigated in my model through the agent's first-order condition for effort in (6), which indicates that the optimal effort is a function of the slope of the chord through $I[Y]$ and $I[Z]$. To focus on effort, suppose here that I can treat $a > 0$ as an exogenous variable. In that case, by differentiating (6) with respect to t, I obtain the response of effort to a change in the target level (subscripts denote partial derivatives) as:

$$e_t = -\frac{(I'[Y] - I'[Z])}{I'[Z] - I'[Y] - 2ac''[e]}. \tag{11}$$

Since the denominator of (11) is negative by the assumption of an interior maximum, I have the following Proposition (all proofs being in the Appendix):

Proposition 1 *For a fixed level of risk the relationship between effort and the target level is positive at 'low' target levels and negative at 'high' target levels:*

$$\begin{array}{ll} \textit{Low } t: & Z > Y > 0 \Rightarrow e_t > 0 \\ \textit{Intermediate } t: & Z > 0 > Y \Rightarrow e_t \gtrless 0 \Leftrightarrow I'[Y] - I'[Z] \gtrless 0 \\ \textit{High } t: & 0 > Z > Y \Rightarrow e_t < 0 \end{array}$$

[4] Mento et al. (1987) and Tubbs (1986) provide supportive meta-analyses from a wide range of further studies.

[5] Much practical literature on personnel management and motivation (e.g. Hiam, 1999; Spitzer, 1995) also advocates the setting of performance targets to employees as a non-monetary form of motivation, as do organizations such as Business Link, the UK government's business advisory service.

Proposition 1 shows that the model predicts the inverse-∪ shaped relationship between the target level and effort found in the psychological literature: effort is increasing for low target levels (in which optimal effort is sufficiently high that the target is exceeded even if the worst shock realizes) and decreasing for high target levels (in which the target level is not achieved even if the highest shock realizes). The effort maximum occurs at some intermediate target level which is attained with a probability strictly between zero and one at the optimum effort. The key to the result is diminishing sensitivity: when output is above the target level (low t) raising t moves output closer to the target level, so increasing marginal intrinsic motivation. Conversely, when output is below the target level (high t) raising t moves output further from the target level, so decreasing marginal intrinsic motivation. Note that if $I[\cdot]$ is a standard concave function, I do not generate this prediction. Rather, it always holds that $I'[Y] > I'[Z]$, with the consequence that effort is everywhere increasing in the target level.

4.2 Targets and Risk-taking

In this section I now treat effort as exogenous and examine the relationship between risk-taking and the target level. A closely related psychological literature to that on targets and effort has studied the relationship between targets and risk-taking. Experimental evidence from psychology finds that higher targets induce higher levels of risk-taking (Knight et al., 2001; Larrick et al., 2009). More generally, there is also evidence that the further agents are from achieving the target, the more they are willing to take risks. Studies of race track betting document the common phenomenon that bettors, when losing, tend to bet more and more on longer odds horses (Hausch et al., 1981; McGlothlin, 1956). Managers claim to take more risks when their firm is missing performance targets than when it is meeting them (MacCrimmon and Wehrung, 1986; Shapira, 1995). Consistent with this, Bowman (1980, 1982) finds that less profitable firms within industries exhibit higher variances in their operations and profits.[6]

There are also findings that traders and fund managers who perform poorly in the first half of their regular performance cycle tend to increase the riskiness of their

[6] There are parallels too in the behavior of birds, where studies show that they become increasingly risk prone as food levels are manipulated downward below energy expenditure levels (Caraco and Lima, 1985).

portfolio in the second half of the cycle (Brown et al., 1996; Shapira, 2002), although in these studies attainment of the performance target is linked to payouts.

These findings can be investigated in the model through the agent's first-order condition for risk in (7). As the parameter a is a mean-preserving spread, expected output is independent of the level of risk. Therefore, as is well-known, the preferred risk level of a risk averse principal is $a = 0$: the principal would chose to eliminate all risk were it possible to do so. If the agent instead has a preferred risk level $a > 0$, I say that the agent has a preference for 'excess risk', in the sense that her preferred level of risk is excessive from the perspective of the principal.

If I fix the effort level of the agent at $e > 0$ then the first-order condition (7) leads to the following Proposition:

Proposition 2 *For a fixed level of effort it holds that:*

$$\begin{array}{rl} i) & \textit{If } t \leq e \textit{ then } a = 0; \\ ii) & \textit{If } t > e \textit{ then } a > 0; \\ iii) & \textit{If } a > 0 \textit{ then } Z > 0 > Y. \end{array}$$

Noting that e is expected output, part (i) of Proposition 2 shows that for target levels below expected output intrinsic motivation is locally concave so the agent does not have a preference for excess risk. This conforms to the prediction of standard economic theory.

However, part (ii) of the Proposition shows that if the target exceeds expected output, intrinsic motivation is locally convex, so $a = 0$ is never optimal: the agent has a preference for excess risk. Note, however, that the agent's optimal risk level is finite, so behavior is different from that implied by risk-seeking preferences in standard theory, whereby the agent would choose an unbounded level of excess risk. Here the existence of a non-zero, but finite, optimal level of excess risk arises from the assumption of diminishing sensitivity, which implies alternative risk attitudes over outcomes above and below the target level.

A possible example of part (ii) is the British trader Nick Leeson who brought down Barings Bank in 1995 after losing around \$1.3 billion while attempting to eradicate hidden debts. Leeson's target appears to have been the break-even level. As his position moved further from this target he entered into a series of increasingly speculative gambles worth more than the entire reserves of the Bank. Serious incidents of a similar nature have occurred in other major financial institutions, although their exact frequency is unknown (see e.g. Shapira, 2002).[7]

Part (*iii*) is essentially a corollary of part (*ii*), stating that when the optimal value of a is positive, it must be that $Z > 0 > Y$. To see this, note that if $Y, Z > 0$ then all outcomes fall on the concave interval of intrinsic motivation, so an interior maximum is never optimal. Conversely, a finite level of risk is never optimal if $Y, Z < 0$ as all outcomes fall on the convex interval of intrinsic motivation. Therefore, for the optimal level of risk to be positive but finite, it must be that some outcomes lie on the convex interval of intrinsic motivation and others lie on the concave interval of intrinsic motivation.

It is important to note that the preference for risk-taking when the target exceeds expected output arises even in a model where risk-taking is completely decoupled from expected output. Accommodating a positive relationship between expected output and risk-taking would only strengthen this effect.

Loss aversion acts as a restraint on risk-taking. An increase in a increases the probability of achieving the target level, but simultaneously exposes the agent to greater downside risk. This downside risk looms larger under loss aversion than would be the case if gains and losses were treated symmetrically, thereby checking the preferred level of risk.

I now investigate the comparative static properties of the optimal risk level for the case when $a > 0$. These I summarize in the following Proposition:

Proposition 3 *For interior solutions to the problem given by (2) it holds that:*

$$a_t > 1; \qquad a_e < 0.$$

Proposition 3 shows that the agent's preferred risk level is increasing in the target level. Moreover, I am able not only to sign this relationship, but also to show that the

[7] Proposition 2 is also consistent with evidence in Thaler and Johnson (1990) that the same individual can be observed to sometimes accept a fair gamble, and other times reject it. Here this phenomena can be explained by the additional contextual dimension provided by the target level. By contrast, standard theory would suggest the gamble is always rejected.

response of the optimal risk level to a change in the target level is greater than one-for-one. This result is a strong and testable prediction, for the environment specified by the model is amenable to experimental replication. However, owing to the simplifying assumptions I employ to maintain tractability, direct testing against empirical data would require great caution.

To gain some intuition for Proposition 3 note that in the unconstrained problem, for a given level of effort, the probability of at least achieving the target level is given by:

$$P\left(e+\varepsilon\geq t\right)=\begin{cases}0 & Z>Y>0\\ \frac{1}{2}\left(1+\frac{e-t}{a}\right) & Z>0>Y\\ 1 & 0>Z>Y\end{cases}. \tag{12}$$

Note from (12) that, for a fixed level of effort and $Z>0>Y$, if a does not increase in response to an increase in t, then the probability of achieving the target converges to zero, which cannot be optimal. Moreover, a must respond more than one-for-one to increases in t in order to prevent (12) converging to zero.[8] The Proposition also states that risk-taking and effort are substitutes. The intuition for this can again be seen from (12), where both effort and risk enter positively, such that maintaining a constant probability of achieving the target can alternatively be achieved by bearing more risk or increasing effort.

4.3 Agent's Optimum

Having examined effort and risk-taking in isolation, I now consider the agent's simultaneous choice of effort and unsystematic risk. The comparative static properties of the agent's unconstrained optimum can now be summarized in the following Proposition:

Proposition 4 *For the unconstrained problem given by (2), it holds that:*

$$a_t\begin{cases}=0 & t\leq e\\ >0 & t>e\end{cases};\qquad e_t\begin{cases}>0 & t\leq e\\ <0 & t>e\end{cases}.$$

Combining the intuitions behind earlier propositions, Proposition 4 states that when the target level is in excess of expected output, further increases in the target level

[8] This can be seen by setting (12) equal to a constant and differentiating. I then have that $\partial a/\partial t=a/\left(t-e\right)$, which implies that $\partial a/\partial t>1\Leftrightarrow Z>0$, where the r.h.s. is satisfied since $Z>0>Y$.

induce the agent to reduce effort and acquire a stronger preference for excess risk. However, when the target is at or below expected output, the agent does not have a preference for excess risk and responds to increases in the target by increasing effort. This result has potentially important ramifications for the principal, who can therefore benefit from an appropriately set target, but can be exposed to both declining productivity and excessive risk if the performance target is set excessively high.

It remains to deduce the comparative static properties of A, the agent's constrained choice of risk. Combining (3) and Proposition 4 I have that

$$A_t = \begin{cases} 0 & a < \alpha \\ a_t > 0 & a \geq \alpha \end{cases} . \tag{13}$$

From (3) I have that $A = \alpha$ for all $a < \alpha$, so clearly $A_t = 0$ on this interval. For $a \geq \alpha$ I have that $A = a$, so also $A_t = a_t$.

4.4 Optimal Target Level

In this section I now analyze the problem of the principal. I have the following Proposition:

Proposition 5 *At the equilibrium between the principal and agent it holds that:*

i) The principal's choice of t *satisfies*

$$t > e;$$

ii) The agent's choice of (A, e) *satisfies*

$$A = \alpha; \qquad e = e_{\max}.$$

Proposition 5 shows that the principal chooses the performance target to maximize the agent's effort. More interestingly, it shows that the effort-maximizing performance target will lie above expected output. The intuition for Proposition 5 can be seen from the principal's utility in (10). Suppose that the solution to the agent's unconstrained problem satisfies $a < \alpha$. In this case, from (3), the best the agent can do is to set $A = \alpha$. It follows from (13) that $A_t = 0$, so - by Proposition 4 - the agent will respond to an increase in the target level by raising effort, and holding the level of risk constant. The principal's utility is therefore increasing in the target level.

Now suppose $a \geq \alpha$, then the agent sets $A = a$, so from (13) it follows that $A_t > 0$. Then - by Proposition 4 - the agent will respond to an increase in the target level by choosing a greater level of risk and by reducing her effort. A reduction in effort lowers the principal's utility, as must an increase in risk since the principal is risk averse. It follows that the principal's utility is decreasing in the target level.

A straightforward corollary of these two sets of arguments is that the optimal target level must be where $A = a = \alpha$, at which point effort is maximized and - since the agent's preferred risk level matches the level of systematic risk - the principal bears no unsystematic risk. From Proposition 2 it follows that, since the principal's optimal target level satisfies $a > 0$, it must be that $t > e$.

The finding that the principal's optimal target level exceeds expected output arises from the interaction of loss aversion and systematic risk. The presence of systematic risk constrains the ability of the agent to substitute risk for effort, thereby creating an interval of target levels above expected output at which the agent's optimal effort is still increasing in the target level. The size of this interval is dependent upon the degree of loss aversion: strong loss aversion discourages the agent from risk-taking and therefore expands the interval on which effort is increasing. Conversely, weak loss aversion (when outcomes above and below the target are treated close to symmetrically), implies a greater readiness to take risk, which correspondingly reduces the interval on which effort is increasing.

5 Conclusion

Growing economic and psychological evidence suggests that performance targets can act as an incentive device even in the absence of monetary rewards conditional on achieving the target. I model this phenomenon by appealing to the notion of intrinsic motivation to work - which can be importantly affected by targets through their influence on an individual's perceived competence, and their esteem with colleagues. The psychological literature finds that targets lead to increased effort over most of the range, but can lead to decreased effort at high levels. Also, tougher targets induce agents to take greater risks. Traditional decision theory is not consistent with these findings, but I show here that an approach based on the insights of prospect theory is.

In a setting in which the principal can choose the agent's performance target I find that the optimal target level chosen by the principal lies above expected output. This arises because the principal has an incentive to exploit the agent's loss aversion over below-target outcomes in an environment where the presence of systematic risk constrains the agent's ability to substitute risk for effort.

There are a number of avenues for possible further research: the prediction made by the model that risk-taking responds more than proportionately to increases in the target level is amenable to an experimental test. Additionally, the analysis could be extended to a dynamic setting with repeated interactions between principals and agents. There is evidence that target levels set within organizations adapt over time to reflect actual performance (Lant, 1992; Lant and Hurley, 1999). One way such adaptation might be generated in the current model is if agents differ in unobserved productivity, such that the principal must use information from past outcomes to infer an agent's productivity, and hence the agent's optimal target level.

A dynamic setting might also raise issues relating to fairness and trust in the principal-agent relationship. For instance, setting targets that will not on average be achieved in equilibrium could be perceived as manipulative or unethical by the agent. The principal might therefore have an incentive to lower the target so as not to violate fairness norms in the principal-agent relationship. The managers interviewed in Bewley (1999) argue that fairness is important to productivity through its impact on morale. More generally, having agents fail against their targets too regularly might be expected to have a long-run impact on job satisfaction and work morale. Both factors could be expected to reduce productivity and increase labor turnover.

References

Abdellaoui, M. (2000). Parameter free elicitation of utility and probability weighting functions. *Management Science*, 46, 1497-1512.

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, 97, 543-569.

Atkinson, J. (1958). Towards experimental analysis of human motivation in terms of motives, expectancies and incentives. In J. Atkinson (Ed.), *Motives in fantasy, action and society* (pp. 288-305). Princeton NJ: Van Nostrand.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology*, 45, 1017-1028.

Barberis, N., Huang, M., & Santos, T. (2001). Prospect theory and asset prices. *Quarterly Journal of Economics*, 116, 1-53.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance. *Journal of Applied Psychology*, 78, 111-118.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., & Sugden, R. (1997). A test of the theory of reference-dependent preferences. *Quarterly Journal of Economics*, 112, 479-505.

Bewley, T. F. (1999). *Why wages don't fall during a recession*. Cambridge MA: Harvard University Press.

Berkelaar, A. B., Kouwenberg, R., & Post, T. (2004). Optimal portfolio choice under loss aversion. *Review of Economics and Statistics*, 86, 973-987.

Bowman, E. H. (1980). A risk-return paradox for strategic management. *Sloan Management Review*, 21, 17-31.

Bowman, E. H. (1982). Risk seeking by troubled firms. *Sloan Management Review*, 23, 33-42.

Brown, K. C., Harlow, W. V., & Starks, L. T. (1996). Of tournaments and temptations: an analysis of managerial incentives in the mutual fund industry. *Journal of Finance*, 51, 85-110.

Bryan, J. F., & Locke, E. A. (1967). Parkinson's law as a goal-setting phenomenon. *Organizational Behavior and Human Decision Processes*, 2, 258-275.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York City cab drivers: one day at a time. *Quarterly Journal of Economics*, 112, 407-441.

Caraco, T., & Lima, S. L. (1985). Survivorship, energy budgets, and foraging risk. In M. Commons, S. J. Shettleworth, & J. Nevin (Eds.), *Quantitative analyses of behavior* (pp. 132-159). Cambridge MA: Ballinger.

Chiang, A. C. (1984). *Fundamental methods of mathematical economics.* (3rd ed.). Singapore: McGraw Hill.

Chou, Y. K. (2002). Testing alternative models of labour supply: evidence from taxi drivers in Singapore. *Singapore Economic Review*, 47, pp. 17-47.

Clark, A. E. (2003). Unemployment as a social norm: psychological evidence from panel data. *Journal of Labor Economics*, 21, 323-351.

Clark, A. E., & Oswald, A. J. (1994). Unhappiness and unemployment. *Economic Journal*, 104, 648-659.

Clark, A. E., & Oswald, A. J. (1998). Comparison-concave utility and following behavior in social and economic settings. *Journal of Public Economics*, 70, 133-155.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social and Social Psychology*, 18, 105-115.

Deci, E. L. (1975). *Intrinsic motivation.* New York: Plenum.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

de Meza, D., & Webb, D. C. (2007). Incentive design under loss aversion. *Journal of the European Economic Association*, 5, 66-92.

Dhami, S., & al-Nowaihi, A. (2007). Why do people pay taxes? Prospect theory versus expected utility theory. *Journal of Economic Behavior and Organization*, 64, 171-192.

Di Tella, R., MacCulloch, R., & Oswald, A. J. (2003). The macroeconomics of happiness. *Review of Economics and Statistics*, 85, 809-827.

Dittmann, I., Maug, E., & Spalt, O. (forthcoming). Sticks or carrots? Optimal CEO compensation when managers are loss averse. *Journal of Finance*.

Erez, M., & Zidon, I. (1984). Effects of goal acceptance on the relationship of goal setting and task performance. *Journal of Applied Psychology*, 69, 69-78.

Farber, H. S. (2008). Reference-dependent preferences and labor supply: the case of New York City taxi drivers. *American Economic Review*, 98, 1069-1082.

Fehr, E., & Goette, L. (2007). Do agents work more if wages are high? *American Economic Review*, 97, 298-317.

Frey, B. S. (1997a). *Not just for the money: an economic theory of personal motivation*. Brooklyn VT: Edward Elgar.

Frey, B. S. (1997b). On the relationship between intrinsic and extrinsic work motivation. *International Journal of Industrial Organization*, 15, 427-439.

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. *Journal of Legal Studies*, 29, 1-17.

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115, 791-810.

Hall, H. K., Weinberg, R. S., & Jackson, A. (1987). Effects of goal specificity, goal difficulty, and information feedback on endurance performance. *Journal of Sport Psychology*, 9, 43-54.

Hausch, D. B., Ziemba, W. T., & Rubenstein, M. (1981). Efficiency of the market for racetrack betting. *Management Science*, 27, 1435-1452.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38, 79-109.

Herweg, F., Müller, D., & Weinschenk, P. (2008). The optimality of simple contracts: moral hazard and loss aversion. Bonn Econ Discussion Paper 17/2008, University of Bonn.

Hiam, A. (1999). *Streetwise motivating and rewarding employees: new and better ways to inspire your people*. Holbrook MA: Adams.

Hubbard, G. R. (1998). Capital-market imperfections and investment. *Journal of Economic Literature*, 36, 193-225.

Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review*, 39, 25-43.

Iantchev, E. P. (2005). Optimal contracts with prospect theory preferences in the presence of moral hazard. Unpublished manuscript, University of Chicago.

Jensen, M. C., & Murphy, K. J. (1990). Performance pay and top-management incentives. *Journal of Political Economy*, 98, 225-264.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: a qualitative and quantitative review. *Psychological Bulletin*, 127, 376-407.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47, 263-292.

Kanbur, R., Pirtillä, J., & Tuomala, M. (2008). Moral hazard, income taxation, and prospect theory. *Scandinavian Journal of Economics*, 110, 321-337.

Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, 43, 39-58.

Knight, D., Durham, C., & Locke, E. (2001). The relationship of team goals, incentives, and efficacy to strategic risk, tactical implementation, and performance. *Academy of Management Journal*, 44, 326-338.

Lant, T. K. (1992). Aspiration level adaptation: an empirical exploration. *Management Science*, 38, 623-644.

Lant, T. K., & Hurley, A. E. (1999). A contingency model of response to performance feedback. *Group and Organization Management*, 24, 421-437.

LaPorte, R., & Nath, R. (1976). Role of performance goals in prose learning. *Journal of Educational Psychology*, 68, 260-264.

Larrick, R. P., Heath, C., & Wu, G. (2009). Goal-induced risk-taking in strategy choice. *Social Cognition*, 27, 342-364.

Latham, G. P., & Baldes, J. J. (1975). The “practical significance” of Locke’s theory of goal setting. *Journal of Applied Psychology*, 60, 122-124.

Latham, G. P., & Kinne, S. B. (1974). Improving job performance through training in goal setting. *Journal of Applied Psychology*, 59, 1-9.

Laughhann, D. J., Payne, J. W., & Crumm, R. L. (1980). Managerial risk preferences for below-target returns. *Management Science*, 26, 1238-1249.

Locke, E. A., Cartledge, N., & Knerr, C. (1970). Studies of the relationship between satisfaction, goal-setting and performance. *Organizational Behavior and Human Decision Processes*, 5, 135-158.

Locke, E. A., & Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs NJ: Prentice-Hall.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *American Psychologist*, 57, 705-717.

MacCrimmon, K. R., & Wehrung, D. A. (1986). *Taking Risks: The Management of Uncertainty*. New York: Free Press.

Mas, A. (2006). Pay, reference points, and police performance. *Quarterly Journal of Economics*, 121, 783-821.

McGlothlin, W. H. (1956). Stability of choices among uncertain alternatives. *American Journal of Psychology*, 69, 604-615.

Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal-setting and task performance: 1966-1984. *Organizational Behavior and Human Decision Processes*, 39, 52-83.

Mezias, S. (1988). Aspiration level effects: an empirical investigation. *Journal of Economic Behavior and Organization*, 10, 389-400.

Ness, R. G., & Patton, R. W. (1979). The effect of beliefs on maximum weight-lifting performance. *Cognitive Therapy and Research*, 3, 205-211.

Payne, J. W., Laughhann, D. J., & Crumm, R. L. (1981). Further tests of aspiration level effects in risky choice behavior. *Management Science*, 27, 953-958.

Rablen, M. D. (2008). Relativity, rank and the utility of income. *Economic Journal*, 118, 801-821.

Rablen, M. D. (2009). An evolutionary model of reference-dependent utility. Unpublished manuscript, University of Westminster.

Rayo, L., & Becker, G. (2007). Evolutionary efficiency and happiness. *Journal of Political Economy*, 115, 302-337.

Rizzo, J. A., & Zeckhauser, R. J. (2003). Reference incomes, loss aversion, and physician behavior. *Review of Economics and Statistics*, 85, 909-922.

Ronan, W. W., Latham, G. P., & Kinne, S. B. (1973). Effects of goal setting and supervision on agent behavior in an industrial situation. *Journal of Applied Psychology*, 58, 302-307.

Sen, A. K. (1977). Rational fools: a critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*, 6, 317-344.

Shapira, Z. (1995). *Risk taking: a managerial perspective*. New York: Russel Sage Foundation.

Shapira, Z. (2002). Aspiration levels and risk-taking by government bond traders. Unpublished manuscript, New York University.

Spitzer, D. R. (1995). *Super-Motivation: A Blueprint for Energizing Your Organization from Top-to-Bottom*. New York: AMACOM.

Stevenson, M. K., Kanfer, F. H., & Higgins, J. M. (1984). Effects of goal specificity and time cues on pain tolerance. *Cognitive Therapy and Research*, 8, 415-426.

Terpstra, D. E., & Rozell, E. J. (1994). The relationship of goal setting to organizational profitability. *Group and Organization Management*, 19, 285-294.

Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: the effects of prior outcomes on risky choice. *Management Science*, 36, 643-660.

Tubbs, M. E. (1986). Goal setting: a meta-analysis examination of the empirical evidence. *Journal of Applied Psychology*, 71, 474-483.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323.

Vallerand, R. J., & Reid, G. (1984). On the causal effects of a perceived competence on intrinsic motivation: a test of cognitive evaluation theory. *Journal of Sport Psychology*, 6, 94-102.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, 42, 1676-1690.

Wu, G., & Gonzalez, R. (1999). Nonlinear decision weights in choice under uncertainty. *Management Science*, 45, 74-85.

Wu, G., Heath, C., & Larrick, R. (2004). A prospect theory model of goal behavior. Unpublished manuscript, University of Chicago.

Appendix

Proof of Proposition 1

From (5) I have that $Z > Y$, so by diminishing sensitivity, $Z > Y > 0 \Rightarrow I'[Y] > I'[Z]$ and $0 > Z > Y \Rightarrow I'[Y] < I'[Z]$. From (11) - and since the denominator is negative by assumption - I have that $e_t \gtreqless 0 \Leftrightarrow I'[Y] - I'[Z] \gtreqless 0$. Therefore:

$$\begin{aligned} Z &> Y > 0 \Rightarrow I'[Y] > I'[Z] \Leftrightarrow e_t > 0 \\ Z &> 0 > Y \Rightarrow I'[Y] \gtreqless I'[Z] \Leftrightarrow e_t \gtreqless 0 \\ 0 &> Z > Y \Rightarrow I'[Y] < I'[Z] \Leftrightarrow e_t < 0 \end{aligned}$$

It remains to show that e_t is initially increasing for low values of t and decreasing for high values. This requires that the optimum satisfy $\frac{\partial}{\partial t}(I'[Y] - I'[Z]) < 0$. I have that $\frac{\partial}{\partial t}(I'[Y[t,e[t]]] - I'[Z[t,e[t]]]) = Y_t I''[Y] + Z_t I''[Z]$. Then $I''[Y] > 0$ and $I''[Z] < 0$ and:

$$Y_t[t,e[t]] = Z_t[t,e[t]] = e_t - 1 = -\frac{2ac''[e]}{I'[Y] - I'[Z] + 2ac''[e]} < 0.$$

It follows that $\frac{\partial}{\partial t}(I'[Y] - I'[Z]) < 0$.

Proof of Proposition 2

$i)$ Suppose $e - t \geq 0$. Differentiating the maximand in (2) with respect to a gives

$$\frac{\partial E[I]}{\partial a} = \left(\frac{1}{2a^2}\right)\left(a(I[Z] + I[Y]) - \int_Y^Z I[\varphi]\ d\varphi\right).$$

Using L'Hopital's rule I have that $\lim_{a\downarrow 0} \frac{\partial E[I]}{\partial a} = \lim_{a\downarrow 0}\left(\frac{1}{4}\right)(I'[Z] - I'[Y])$. Note that $e - t \geq 0 \Leftrightarrow Z + Y \geq 0$ and $e - t \geq 0 \Rightarrow Z > 0$, so $Z > 0$ and $Y \in [-Z, Z)$. Therefore, by loss aversion and diminishing sensitivity, it must be that $I'[Z] - I'[Y] < 0$. It follows that $\lim_{a\downarrow 0} \frac{\partial E[I]}{\partial a} < 0$ so the limit at zero is approached from below, and $a = 0$ is a local maximum. The second-order condition at stationary points is given by $\frac{1}{2a}(I'[Z] - I'[Y])$, implying that any stationary point of $\frac{\partial E[I]}{\partial a}$ is a local maximum. As there cannot be two local maxima without an intervening local minimum, $a = 0$ is the only local maximum.

$ii)$ If instead $e-t<0$ then $Z+Y<0$. As $a\downarrow 0$ I have that $Y<0$ and $Z<0$, so by diminishing sensitivity $I'[Z]-I'[Y]>0$. Then $\lim_{a\downarrow 0}\frac{\partial E[I]}{\partial a}>0$ implying that $a=0$ is a local minimum. Therefore, all local maxima must satisfy $a>0$.

$iii)$ Note from (9) that $I'[Y]>I'[Z]$ at the optimum and, from (i,ii), that $a>0\Rightarrow e-t<0\Rightarrow Y+Z<0$. The latter condition implies that Y, Z cannot both be positive. Neither can Y,Z both be negative, for $Z>Y$ from (5), so diminishing sensitivity would imply $I'[Z]>I'[Y]$. It follows that $Z>0>Y$.

Proof of Proposition 3

From assumptions A0 and A3 and the definitions of concavity and convexity (Chiang, 1984; p. 345), I have that $Z>0\Rightarrow I[Z]-ZI'[Z]>0$ and $Y<0\Rightarrow YI'[Y]-I[Y]>0$. Differentiating (7) I have

$$\begin{aligned}a_t &= \frac{2\left(I[Z]-ZI'[Z]+YI'[Y]-I[Y]\right)-(Y+Z)\left(I'[Y]-I'[Z]\right)}{(Z-Y)\left(I'[Y]-I'[Z]\right)}>0,\\ a_e &= -a_t<0.\end{aligned}\tag{A.1}$$

From (A.1) I have that $a_t>1\Leftrightarrow I[Z]-I[Y]-(Z-Y)I'[Y]>0$. Second note that

$$Z_t[t,a[t]]=a_t-1=\frac{2\left(I[Z]-I[Y]-(Z-Y)I'[Y]\right)}{(Z-Y)\left(I'[Y]-I'[Z]\right)},$$

so $Z_t[t,a[t]]>0\Leftrightarrow I[Z]-I[Y]-(Z-Y)I'[Y]>0$. Establishing $a_t>1$ is therefore equivalent to establishing $Z_t[t,a[t]]>0$. From the proof of Proposition 2 I know that all interior optima satisfy $t>e$ (which implies $Z+Y<0$) and from (9) I know that $I'[Y]>I'[Z]$. Initially consider the limiting case where $t\downarrow e$, where it holds that $\lim\limits_{t\downarrow e}Z=\lim\limits_{t\downarrow e}Y=0$. Suppose I now increase t: since $Y_t[t,a[t]]=-a_t-1<0$ I know that Y becomes negative. However, Z cannot also fall, for if both Y and Z were negative diminishing sensitivity would imply $I'[Z]>I'[Y]$. Nor can Z remain constant at zero and still satisfy the first-order condition with respect to a, since $\frac{\partial E[I]}{\partial a}|_{Z=0,Y<0}=-\frac{1}{4a^2}\left(YI[Y]+2\int_Y^0 I[\varphi]\ d\varphi\right)<0$. It follows that Z is increasing in the neighborhood of $t=e$. I now show that $Z_t[t,a[t]]>0$ everywhere. Let $\bar{Z}$ be implicitly defined by $I\left[\bar{Z}\right]-I\left[Y\right]-\left(\bar{Z}-Y\right)I'\left[Y\right]=0$. Note that

$$\bar{Z}_t\left[t,a\left[t\right]\right]=\frac{Y_t\left[t,a\left[t\right]\right]\left\{I''\left[Y\right]\left(\bar{Z}-Y\right)\right\}}{I''\left[Y\right]-I''\left[Z\right]}>0.$$

Since $I\left[Z\right]-I\left[Y\right]-\left(Z-Y\right)I'\left[Y\right]$ is decreasing in Z, using the definition of $\bar{Z}$ gives that $Z<\bar{Z}\Leftrightarrow I\left[Z\right]-I\left[Y\right]-\left(Z-Y\right)I'\left[Y\right]>0$. Previous arguments have established that $Z<\bar{Z}$ holds in the neighborhood of $t=e$. However, as t is allowed to increase it must remain the case that $Z<\bar{Z}$ since $\lim_{Z\uparrow\bar{Z}}Z_t=0<\bar{Z}_t$. Therefore $Z<\bar{Z}$ everywhere, which completes the proof.

Proof of Proposition 4

When $a=0$ the problem (2) collapses to $\max\limits_{e} I\left[e-t\right]-c\left[e\right]$ with first-order condition

$$I'\left[e-t\right]-c'\left[e\right]=0, \tag{A.2}$$

and second-order condition $I''\left[e-t\right]-c''\left[e\right]<0$. For internal optima I use standard comparative static methods on the first-order conditions (6), (7) and (A.2). I then obtain:

$$a_t=\begin{cases}0 & t\leq e\\ \frac{c''[e](\psi+\phi)}{|H|}>0 & t>e\end{cases};\qquad e_t=\begin{cases}-\frac{I''[e-t]}{c''[e]-I''[e-t]}>0 & t\leq e\\ -\frac{4\psi\phi}{(Z-Y)^2|H|}<0 & t>e\end{cases};$$

where $|H|>0$ is the determinant of the Hessian matrix and:

$$\phi\equiv I\left[Z\right]-I\left[Y\right]-\left(Z-Y\right)I'\left[Z\right]>0;\qquad \psi\equiv\phi-\left(Z-Y\right)\left(I'\left[Y\right]-I'\left[Z\right]\right)>0.$$

Proof of Proposition 5

Suppose the unconstrained solution to (2) is such that $a<\alpha$, then $A_t=0$ and $e_t>0$. It follows that the principal's utility (10) is increasing in t. Now suppose $a\geq\alpha$, then $A_t>0$ and $e_t<0$. It follows that the principal's utility is decreasing in t. The principal's utility is therefore maximized where $A=a=\alpha$. Since $a>0$ the optimal target level satisfies $t>e$ by Proposition 2.

Acknowledgements

I thank Andrew Oswald, Gordon Brown and Andrew Clark for their comments and encouragement, and the ESRC for financial support.