

Incentive Design under Loss Aversion

David de Meza {LSE} and David C Webb {LSE}

16 May 2006

Abstract

Compensation schemes often reward success but do not penalize failure. Fixed salaries with stock options or bonuses have this feature. Yet the standard principal–agent model implies that pay is normally monotonically increasing in performance. This paper shows that, under loss aversion, there will be intervals over which pay is insensitive to performance, with the use of carrots but not sticks is frequently optimal, especially when risk aversion is low and reference income is endogenous. A further benefit of capping losses, for example through options, is to discourage reckless behavior by executives seeking to resurrect their fortunes. (JEL: F3, F4)

1. Introduction

Executive compensation schemes and pay packages more generally tend to reward success but do not attach financial penalties to failure.¹ Perhaps the most striking example is motivating CEOs via base salary plus stock options.² At first sight, the explanation is that stock options align the interests of the CEO and shareholders and so mitigate agency costs. However, protection from downside loss is not a feature of standard principal–agent theory.³ As shown by Holmstrom (1979) and Grossman and Hart (1983), if the monotone likelihood property and some other weak assumptions are satisfied then pay is everywhere increasing in measured performance.

A rival to the incentive theory is that high-powered compensation, particularly the award of stock options, is an expropriation device. According to this “managerial power” view—which has grown in prominence following the Enron, WorldCom, and other scandals—CEOs loot the companies over which they rule. Stock options facilitate the robbery because, being nominally performance based

Acknowledgements: We would like to thank Heskey Bar Isaac, Kostas Koufopoulos, Jean Charles Rochet, Timothy Van Zandt, seminar participants at the Universities of Birmingham, Durham, Oxford, and Warwick, and especially two anonymous referees for helpful comments and advice.

Email addresses: d.de-meza@lse.ac.uk (Meza); d.c.webb@lse.ac.uk (Webb).

1. There is much evidence that firms rely mostly on bonuses and promotions to motivate rather than fines and demotions. Medoff and Abraham (1980) document that, when employees are evaluated for discretionary bonuses, only 5.5% of the 4,788 managers (in firm A of their two-firm study) are rated worse than “good” while 20% are rated higher. The reference evaluation and resultant bonus is therefore that associated with “good”, so almost nobody is in the loss space while one fifth are in the gain domain.

2. Early discussions of the prevalence of this type of compensation are Jensen and Murphy (1990) and Hall and Lieberman (1998). Murphy (2002) reports that, in the year 2000, stock options constituted more than 50% of the median pay of CEOs of the S&P 500 companies.

3. The paper by Hall and Murphy (2001) appears to be inconsistent with this claim, but the absence of a regular participation constraint means that they do not analyze a standard principal–agent model.

(even if typically repriced or “reloaded” when under water), they appear more defensible to investors than would a sky-high salary. In the terminology of Bebchuck et al. (2001) and Bebchuck and Fried (2003), stock options are less likely to violate the shareholders’ “outrage constraint” than more transparent forms of remuneration, including restricted stock. Yermack (1995) and Bertrand and Mullainathan (2000) provide some evidence for this rent extraction view.⁴

This paper adapts principal–agent theory to explain the tendency for good performance to be rewarded and for poor performance to go unpunished. Our approach is to incorporate loss aversion, as introduced by *prospect theory* (Kahneman and Tversky 1979). The key features are that utility depends upon changes in wealth and not just on its level, with falls in income involving greater disappointment than the elation created by corresponding increases.⁵

Our approach involves first-order risk aversion. That is, preferences are such that there exists a better-than-fair gamble that is rejected even as the income difference between winning and losing tends to zero. As noted by Segal and Spivak (1990), this behavior is consistent with expected utility theory if the utility function is not differentiable near the “certainty equivalent” income. The existence and location of the kink is determined by our further assumption of reference-dependent utility. Realized utility is a function not only of realized income but also of its relation to an accustomed or aspirational reference income. Moreover, the psychological pain of falling below the reference income is greater than any pleasure from surpassing it by an equal amount. This phenomenon of loss aversion does not require first-order risk aversion, but it is natural to assume a kink at the reference income with the utility function steeper immediately below this level. Such a kink is one aspect of prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992), a collection of features intended to capture behavioral aspects of choice. It encompasses first-order risk aversion and loss aversion but assumes a special form for the value function: utility depends only on the deviation in income from the reference level. Kahneman and Tversky (1979, pp. 277–278) remark that, “strictly speaking”, income level should also matter but believe nonetheless that their formulation “generally provides a satisfactory approximation”. This may be true in some circumstances, but it is unlikely that willingness to take a specified gamble will be independent of income. Even more to the point in our setting, suppose an incentive scheme involves basic pay of \$100,000. High effort yields a 10% chance of a bonus of \$10,000. The reference income is then likely to be \$100,000. Consider next a scheme in which basic pay is \$1m (a million dollars) and the bonus \$10,000. The second scheme is unlikely to create the same incentive to work hard as the first owing to the usual

4. Murphy (2002) proposes that stock options are best explained by a “perceived cost bias” whereby managers genuinely underestimate the true cost of their award.

5. Empirical tests suggest that people commonly value a loss more than an equal gain by a factor of 2 to 4 or more (see e.g. Kahneman et al. 1990). Evidence of such reference effects has come from both experimental and nonexperimental settings (Samuelson and Zeckhauser 1988; Kahneman et al. 1991). A typical case is Kachelmeier and Shehata’s (1992) finding that subjects in one experiment were willing to pay, on average, \$5.60 for a 50% chance to win \$20 while demanding an average of \$10.87 to give up the identical chance to win a \$20 prize. Evidence indicates that the phenomenon is not primarily the result of income or wealth effects or even of transaction costs (see e.g. Kahneman et al. 1990; Knetsch et al. 2001).

diminishing marginal utility of income. The prospect theory formulation does not capture this effect, so we adopt a generalization. As discussed herein, our formulation resembles the value function derived axiomatically by Gul (1991) and Sugden (2003).

At first sight it seems obvious that, if people have a strong dislike to feeling “down”, then remuneration schemes should reduce such occurrences by providing carrots, not sticks. Yet this is not straightforward. The purpose of performance pay is to change behavior, so a heavy penalty for failure is potentially effective.⁶ Moreover, since option-type pay implies that performance incentives are concentrated in the gain space, where the marginal utility of pay is low, it follows that the cost to the principal is high. Hence it may seem that considerable loss aversion is required for this type of scheme to be optimal, but this proves not to be the case. Although loss aversion can be viewed as a tax on low payments, avoiding the tax—by boosting the return to poor performance and to good performance while lowering that to middling outcomes—preserves incentives without adding much to the principal’s outlay.

This paper considers both exogenous and endogenous formulations of reference income. With exogenous reference income, earnings in earlier periods determine whether current income is satisfactory. Loss aversion then implies that, over some interval, pay may be insensitive to performance. Depending on the level of the reference income, the flat segment may be at the bottom, the top, or in the middle of the incentive schedule. Thus carrots, sticks, or both may be optimal. When (as much evidence suggests) people are risk loving on the downside, there is always a flat segment at the bottom of the incentive schedule. It is possible that the whole schedule consists of two flat segments with a discontinuity: fixed pay unless performance falls short of a threshold, in which case the agent is dismissed.

Though it provides a useful benchmark, an exogenous reference income is not always plausible. Suppose a manager accepts a job in which the best pay possible is \$1m (but inferior performance receives less). It does not seem likely that disappointment would be associated with obtaining \$1m even if the manager had earned \$2m in the past. The implication is that the reference income is forward looking and depends on the distribution of outcomes, with losses assigned if the individual does not do as well as expected. Preferences of this sort are axiomatized by Gul (1991), in which the reference income is the certainty equivalent of the income distribution. Routledge and Zin (2003) generalize this approach, making reference income some fraction of the certainty equivalent. The resulting design problem looks messy because changing the incentive scheme potentially alters the reference income. In fact, since the participation constraint fixes the certainty equivalent income and hence the reference income, the analysis is equivalent to the exogenous case with one exception: the reference income is never higher than the certainty equivalent and so a flat segment at the top is ruled out.

The certainty equivalent may not be the natural representation when the ref-

6. Mirrlees (1997) gives an example where such a scheme can approximate the first-best outcome arbitrarily closely.

erence income is forward looking. Suppose that 99 times out of 100 a manager is paid \$1m but that he has a 1-in-a-100 chance of earning \$2m. The certainty equivalent formulation implies that loss is experienced whenever the \$1m is earned. In practice, most people would surely regard \$1m as the norm. Elation may be experienced if the \$2m payment is received but no severe disappointment results if the lower amount is realized. This suggests that losses are registered only if it was odds-on (more than a 50% chance) that income would be higher. Preferences of this sort imply that a finite bonus is paid for performance that beats the median and that flat segments occur only in the bottom half of the distribution. Pay fixed at the reference income for all states below the median requires that loss aversion exceed some threshold, which may be absolutely low but is higher when risk aversion is low.

The design of incentive schemes has implications for the choice of risk as well as for effort. The standard view is that option-based contracts create incentives for risk taking. Only recently has Ross (2004) shown that it is possible for the opposite to occur. Loss aversion introduces a new reason why stock options may lessen reckless behavior. Options, by capping the extent of losses, mean that individuals are not delivered into the zone where they are vulnerable to such behavior. In order to see the implications of manager discretion over project risk, we modify our basic model with endogenous effort in a simple way. In particular we assume that, after the effort decision has been made, managers may receive forewarning of how the firm will perform and can take a gamble. Given prospect theory preferences, if events are unfolding in an unsatisfactory way (relative to the reference level of income) then gambling for resurrection is a well-documented phenomenon. However, we show that option-based compensation may make the manager less inclined to take risky actions than under share-based compensation. Whether this is a merit or not depends upon whether the action raises or lowers the expected value of the company. Both possibilities are consistent with the claim that option-based compensation provides relatively less incentive to take risks.

The paper is organized as follows. Section 2 provides the basic ingredients of the problem. Sections 3 and 4 solve the problem for the three different formulations of the reference income. Section 5 compares the magnitude of loss aversion required for constant rewards for all states below the reference income as variously specified and then undertakes comparative statics. Section 6 looks at the choice of risk, and conclusions are drawn in Section 7.

2. The Model

We assume that stock price or profit is the only verifiable performance indicator for the CEO.⁷ Given the optimal (unobservable) action of the CEO, the perfor-

7. Prendergast (1999, 2000), for example, has examined the relationship between the risk of the environment and incentive provision. He argues that high-powered incentive contracts are often observed in high-risk environments (such as banking and high-tech) where company earnings are relatively unstable and that monitoring is observed in more stable environments, where it will be more efficient at discovering performance. In more stable environments, workers have incentive to exhibit good performance to the monitor. In less stable environments, in which the benefits to monitoring are weak,

mance of the company is realized as the outcome of an effort-dependent probability distribution. Writing reference income as Y^R , realized income as Y , and effort as $e \in \{\underline{e}, \bar{e}\}$, the utility function of the CEO is given by

$$U(Y) - Z(Y^R, Y) - e, \quad (1)$$

where

$$Z(Y^R, Y) = \begin{cases} 0 & \text{if } Y > Y^R, \\ g(U(Y^R) - U(Y)) & \text{if } Y \leq Y^R. \end{cases}$$

We assume nonsatiation so $U' > 0$ and risk aversion so $U'' < 0$. Also, in order to ensure that consumption is positive in all states, we impose the Inada condition: $\lim_{Y \rightarrow 0} U'(Y) = \infty$. The loss aversion function has $g' > 0$, including $g'(0) > 0$; hence there is a kink in the utility function at the reference income Y^R . If loss aversion is linear then $Z(Y^R, Y) = l(U(Y^R) - U(Y))$ for $l > 0$ if $Y \leq Y^R$.⁸ This case provides a benchmark from which we may easily extend results to risk loving on the downside.

The state of the company s is distributed on the compact interval $[\underline{s}, \bar{s}]$ with distribution function $F(s|e)$ and positive conditional density function $f(s|e)$. Specifically, $\{f(\underline{s}|\underline{e}), f(\underline{s}|\bar{e})\} > 0$.⁹ Let $\Delta f(s|e) = f(s|\bar{e}) - f(s|\underline{e})$ denote the effect of higher effort on the probability of state realization s . We make the standard assumption that the monotone likelihood ratio property (MLRP) is satisfied: $\Delta f(s|e)/f(s|e)$ is increasing in the state realization.

The firm's owners are risk neutral and seek to minimize the expected payment to the CEO, subject to the individual rationality (IR) constraint (the CEO has reservation utility V^*) and the incentive compatibility (IC) constraints.¹⁰

Thus the principal's problem, if \bar{e} is to be elicited, is

$$\min \left\{ \int_{\underline{s}}^{\bar{s}} Y(s) f(s|\bar{e}) ds \right\}, \quad (2)$$

the link between effort and output is noisy and so incentive pay must be higher to induce high effort.

8. In their work on asset pricing, Barberis et al. (2001) further specialize this formulation to $U(Y) = Y$ and write $Z = l(Y^R - Y_L)$, $l > 0$.

9. If the lower support increases in effort and if low income imposes a sufficient penalty then, whether or not there is loss aversion, it is optimal to pay the minimum possible in the states that are realized only when effort is low. In all other states there is a fixed payment. High effort is then chosen without exposing the agent to risk, for the penalty states never occur. This result is in the spirit of Mirrlees (1997), who identifies circumstances leading to the use of sticks not carrots. He shows that, with utility and performance unbounded below and with the likelihood ratio tending to $-\infty$ as states become very bad, the principal's first-best payoff can be approximated arbitrarily closely by imposing massive penalties for very poor performance yet paying a fixed amount elsewhere.

10. The assumption that the principal is risk neutral (or able to pool large numbers of independent risks) is not as innocuous here as in Holmstrom (1979), where adding a risk-averse principal leads to a simple modification of the first-order conditions. Depending on the nature of the reference income, shielding the agent from loss may or may not expose the principal to loss.

subject to

$$\int_{\underline{s}}^{\bar{s}} (U(Y(s)) - Z(Y^R, Y))f(s|\bar{e}) ds - \bar{e} \geq V^*, \quad (3)$$

$$\int_{\underline{s}}^{\bar{s}} (U(Y) - Z(Y^R, Y))\Delta f(s|e) ds \geq \Delta e. \quad (4)$$

In the absence of loss aversion, it is a standard result (see Holmstrom 1979) that MLRP implies that the optimal incentive scheme is strictly increasing in performance. The next three sections provide solutions when there is loss aversion for three different specifications of the reference income.

3. Exogenous Reference Income

3.1. Linear Loss Aversion

Determining the reference income is the most problematic aspect of prospect theory. One approach is to suppose that the reference income reflects an accustomed standard of living. Then the individual's income history matters, but the reference income is exogenous in the current period.¹¹ We first consider this formulation.

The reference income is given by \bar{Y}^R and losses are incurred if $Y(s) \leq \bar{Y}^R$. We restrict attention initially to linear loss aversion. The first-order conditions, obtained from (2)–(4) with $Y^R = \bar{Y}^R$, are:

$$\int_{\underline{s}}^{\bar{s}} (U(Y(s)) - \theta l(U(\bar{Y}^R) - U(Y(s))))f(s|\bar{e}) ds - \bar{e} \geq V^*, \quad (5)$$

$$\int_{\underline{s}}^{\bar{s}} (U(Y(s)) - \theta l(U(\bar{Y}^R) - U(Y(s))))\Delta f(s|e) ds \geq \Delta e, \quad (6)$$

$$\begin{aligned} & f(s|\bar{e}) - \gamma U'(Y(s))(f(s|\bar{e}) + x(s)l f(s|\bar{e})) \\ & - \lambda U'(Y(s))(\Delta f(s|e) + x(s)l \Delta f(s|e)) \geq 0. \end{aligned} \quad (7)$$

Here θ is an indicator function taking the value unity if $Y(s) \leq \bar{Y}^R$ and zero otherwise.¹² To handle the kink in the utility function at $Y(s) = \bar{Y}^R$ (where left and right derivatives are not equal), we have introduced a variable $x(s) \in [0, 1]$ in condition (7) that multiplies l . If $Y(s) < \bar{Y}^R$, then loss aversion applies and $x(s) = 1$; if instead $Y(s) > \bar{Y}^R$, then small variations in income leave the agent in the gain space and so $x(s) = 0$. For states in which it is optimal to set $Y(s) = \bar{Y}^R > 0$, (7) holds with equality when $0 \leq x(s) \leq 1$. The $x(s)$ multiplier eliminates the

11. Thaler et al. (1997) and Brunnermeier (2004) have addressed the implications—of learning and the frequency with which risks are experienced—for the stability of reference-level income and for risk-taking behavior.

12. In solving the optimization problem for different specifications of the reference income, we look only at first-order conditions. It can, however, be shown by a transformation of the problem (see Mas-Colell et al. 1995, p. 484) that this problem is convex, so that the first-order conditions are necessary and sufficient for a solution.

slackness in the first-order condition when the solution is held at the kink. It can be shown that the IR and IC constraints are satisfied as equalities, so (5) and (6) hold with equality and the multipliers γ and λ are positive.¹³ The Inada condition means that the $Y(s)$ all are strictly positive and that (7) is an equality.

In order to show the possible shapes of the optimal incentive scheme, we “reverse engineer”. That is, we assume positive values of the multipliers γ and λ that, when inserted into (7), determine optimal incomes $Y(s)$ with the properties we are seeking. Given these values, the left-hand sides of the IR and IC constraints follow, so reservation utility and the incremental cost of effort can be set at the levels that solve conditions (5) and (6). Thus a full solution consistent with the assumed value of the multipliers is obtained. In implementing this procedure, we rewrite (7) as

$$\frac{1}{U'(Y(s))} = (1 + x(s)l) \left(\gamma + \lambda \frac{\Delta f(s|e)}{f(s|\bar{e})} \right). \quad (8)$$

PROPOSITION 1. *Given MLRP and loss aversion, four shapes of the optimal incentive scheme, all continuous, are possible.*

- (i) *Strictly increasing in performance.*
- (ii) *Smoothly increasing up to some threshold, beyond which \bar{Y}^R is paid.*
- (iii) *Smoothly increasing \bar{Y}^R is paid for all performance levels below some threshold, but above this the pay is smoothly increasing in performance; if this case arises then there is a level of loss aversion l^* such that a complete bottom flat segment exists if and only if $l > l^*$.*
- (iv) *\bar{Y}^R is paid for some finite, compact performance interval, but above and below this interval the pay is strictly increasing.*

Parts (ii), (iii), and (iv) of the proposition are illustrated in Figures 1a, 1b, and 1c, respectively.

Figures 1a, b and c here

Proof. (i) This case arises in two circumstances. First, \bar{Y}^R may be so low that it does not affect the solution. This requires that $Y(\underline{s}) \geq \bar{Y}^R$ and so $x(\underline{s}) = 0$; then (8) implies $1/U'(\bar{Y}^R) \leq \gamma + \lambda \Delta f(\underline{s}|e)/f(\underline{s}|\bar{e})$. Alternatively, \bar{Y}^R may be so high that, by (8), all realizations are in the loss space. Then $x(s) = 1$ for all s , and $1/U'(\bar{Y}^R) \geq (1 + l)(\gamma + \lambda \Delta f(\bar{s}|e)/f(\bar{s}|\bar{e}))$.

(ii) Pick γ and λ such that $1/U'(\bar{Y}^R) = \gamma + \lambda \Delta f(\bar{s}|e)/f(\bar{s}|\bar{e})$ and $Y(\bar{s}) = \bar{Y}^R$. For $l \geq 0$ but sufficiently low, there exists some s' , with $\underline{s} < s' < \bar{s}$, satisfying $1/U'(\bar{Y}^R) = (1 + l)(\gamma + \lambda \Delta f(s'|e)/f(s'|\bar{e}))$. By (8) and MLRP, the optimal scheme is then increasing for all $s < s'$ and is flat at \bar{Y}^R for all $s \geq s'$.

(iii)–(iv) Pick γ and λ such that, for some intermediate state \hat{s} , $1/U'(\bar{Y}^R) = \gamma + \lambda \Delta f(\hat{s}|e)/f(\hat{s}|\bar{e})$. For all states better than \hat{s} , MLRP and (8) imply that $x(s) =$

13. An intuitive proof that the IR and IC constraints are satisfied as equalities (so that $\gamma > 0$ and $\lambda > 0$) follows. Suppose that $\gamma = 0$; then, because $F(s|\bar{e})$ first-order stochastically dominates $F(s|e)$, there exists a set of states with $\Delta f(s|e)/f(s|\bar{e}) < 0$. But given $\lambda > 0$, this would mean that $U'(Y(s)) \leq 0$, which is impossible. If $\lambda = 0$ then by MLRP the compensation is not increasing in performance, but this would mean that the manager will choose e rather than \bar{e} , violating the IC constraint.

0 and that pay is increasing in performance. This scheme is flat for all states $s \in [s', \hat{s}]$, where s' satisfies $1/U'(\bar{Y}^R) = (1+l)(\gamma + \lambda \Delta f(s'|e)/f(s'|\bar{e}))$. Using (8) and MLRP, we see that the optimal scheme involves $Y(\underline{s}) < \bar{Y}^R$ and is increasing for all $s < s'$.

Given the multipliers γ and λ , we must find the value of loss aversion l^* that solves for $s' = \underline{s}$. The optimal scheme is then constant at \bar{Y}^R for all $s \leq \hat{s}$. To show that l^* is the threshold level of risk aversion that will generate a flat segment for all states below \hat{s} , keep the values of V^* and Δe fixed—so that the multipliers become endogenous—and then consider the effect of varying l . If $l > l^*$ but the solution remains unchanged, then the first-order conditions yield the same value of the multipliers (remember that in this case the loss space is not entered) and so $1/U'(\bar{Y}^R) < (1+l)(\gamma + \lambda \Delta f(\underline{s}|e)/f(\underline{s}|\bar{e}))$, confirming the unchanged solution. Now suppose that $l < l^*$. If a flat segment remained at the bottom of the distribution (even if it were to end at a different state than under l^*) then this scheme could have been adopted when $l = l^*$, in which case it would have yielded the same payoff since the loss space is not entered. Similarly, the allegedly optimal scheme when $l < l^*$ is feasible and has the same payoff when $l = l^*$, a contradiction. Hence, if $l < l^*$, the optimal scheme does not have a flat segment for all lower states. \square

As long as reference income is neither very high nor very low relative to the optimal scheme in the absence of loss aversion, there will be an interval over which pay does not depend on performance. There may be a threshold below which the compensation scheme is constant—an option-like scheme. The intuition for not punishing poor performance can be seen by first constraining the problem so that no payments are made in the loss space. When reference income is relatively low, this is not too costly. Now move the payment for the worst performance into the loss space. The discontinuous drop in marginal utility as the reference income is crossed means that, in order to maintain expected utility, expected payments in the gain space must increase by a high multiple of the saving in outlay at the bottom of the schedule. The incentive effect depends on the extent to which total utility for poor outcomes compares to that for good outcomes, an extent that changes proportionately less in response to the change in the schedule than do marginal utilities. Thus, entering the loss space is an expensive way to create incentives.

This same line of reasoning also applies to the more surprising case of constant pay at the top of the incentive scheme (sticks not carrots). If reference income is relatively high, then it is not very costly to require that no payment exceed the reference. Raising the payment in the best state so that it is barely in the gain space does not allow for much reduction in payment below the reference income if the participation constraint is not to be violated. So the principal's expected cost goes up—but, as before, there is not much effect on incentives. It may be best to create all incentives in the loss space.

3.2. Downside Risk Loving

Linear loss aversion, as assumed in Proposition 1, introduces a kink in the utility function at the reference income but preserves risk-avoiding behavior on the loss side. Kahneman and Tversky (1979) argue that much behavior under uncertainty can be explained only if people are risk loving below the reference income. We now extend the analysis to convex utility on the downside.

We first assume that limited liability puts a lower bound on the income of \underline{Y} . Then we note that, rather than pay an income between the reference level and \underline{Y} , it would be better to offer a lottery with the same expected income. In effect, optimal lotteries linearize the utility function in the loss domain. There is now no risk premium and so, if the optimal incentive scheme pays anything less than \bar{Y}^R , then the best income to pay is \underline{Y} because doing so creates the most incentives. Hence there are flat segments at \bar{Y}^R and at \underline{Y} , and a lottery is not actually paid. The possible shapes of the optimal incentive scheme are shown in the summary diagram appearing as Figures 2a, b, and c.

Figures 2 a,b and c here.

Convexifying utility below \bar{Y}^R requires that the general formulation of (1) has the following property: $U''(Y(s)) - g''(\cdot)U'(Y(s))^2 + g'(\cdot)U''(\cdot) > 0$.¹⁴ We also assume that $g'(0) > 0$ and so there is a kink at \bar{Y}^R . Convexity implies that, if the agent receives a fixed payment $Y^* \in (\underline{Y}, \bar{Y}^R)$, then the same utility could be delivered to the agent at lower cost to the principal by a lottery paying two prizes, \underline{Y} and \bar{Y}^R , whose respective probabilities p and $1 - p$ are set to yield an expected income of Y^* . The expected utility of the lottery is

$$(1 - p)U(\bar{Y}^R) + p(U(\underline{Y}) + g(U(\bar{Y}^R) - U(\underline{Y}))), \quad (9)$$

where

$$Y^* = (1 - p)\bar{Y}^R + p\underline{Y}. \quad (10)$$

Substituting (10) into (9) yields expected utility as a function of expected income generated by the optimal lottery:

$$u(Y^*) = U(\bar{Y}^R) + \frac{Y^* - \bar{Y}^R}{\bar{Y}^R - \underline{Y}} (U(\bar{Y}^R) - U(\underline{Y}) + g(U(\bar{Y}^R) - U(\underline{Y}))), \quad (11)$$

which is linear in Y^* .

LEMMA 1. *With convex utility in the loss space, it is optimal to have no payments in the interval $(\underline{Y}, \bar{Y}^R)$.*

14. It is possible that, at a sufficiently low income, the marginal utility of consumption becomes so high that concavity of the overall function resumes below this level. Then the optimal scheme has a discontinuity at \bar{Y}^R and a flat segment that may not extend to the worst states.

Proof. Using the utility function (11) for the loss region, the principal–agent problem yields the following first-order condition for a payment $Y^* \in (\underline{Y}, \bar{Y}^R)$ to be optimal:

$$\frac{1}{u'(Y^*)} = \gamma + \lambda \frac{\Delta f(s|e)}{f(s|\bar{e})}. \quad (12)$$

Suppose this condition is satisfied for state s^* . For states worse than s^* , the left-hand side exceeds the right-hand-side, so it follows that expected income should be reduced. This does not change marginal utility (by (11)) and so, in these states, income must be reduced as far as possible—that is, to \underline{Y} . For states better than s^* , expected income should exceed Y^* , which will entail benefits until income is at least \bar{Y}^R . Even for state s^* , income can be equally well set at either extreme. Therefore, it is optimal to have no payments between the two extremes. \square

Using Lemma 1 together with

$$u' = (U(\bar{Y}^R) - U(\underline{Y}) + g(U(\bar{Y}^R) - U(\underline{Y}))) (\bar{Y}^R - \underline{Y})^{-1} > U'(\bar{Y}^R)$$

(by (11)), we can adapt Proposition 1 to cover risk loving on the loss side.

The case of Figure 2(a), constant pay at \bar{Y}^R for all states below $\hat{s} < \bar{s}$, arises if

$$\frac{1}{u'} < \gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})} < \gamma + \lambda \frac{\Delta f(\hat{s}|e)}{f(\hat{s}|\bar{e})} = \frac{1}{U'(\bar{Y}^R)}. \quad (13)$$

The first inequality implies that, in the very worst state, it is better to pay \bar{Y}^R than \underline{Y} ; the equality implies that \hat{s} is the best state in which \bar{Y}^R paid. It thus follows that \bar{Y}^R is paid in all states $[\underline{s}, \hat{s}]$ and that, beyond \hat{s} , pay is monotonically increasing in performance.

Using the same line of reasoning, there are flat segments of pay at \underline{Y} and at \bar{Y}^R , as in Figure 2(b), if the following relations hold for some $s'' < \bar{s}$:

$$\frac{1}{U'(\bar{Y}^R)} = \gamma + \lambda \frac{\Delta f(\hat{s}|e)}{f(\hat{s}|\bar{e})} > \gamma + \lambda \frac{\Delta f(s''|e)}{f(s''|\bar{e})} = \frac{1}{u'} > \gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}. \quad (14)$$

The implications are that: above \hat{s} , pay is monotonically increasing in performance; between s'' and \hat{s} , pay is \bar{Y}^R ; and below s'' , pay is \underline{Y} .

Figure 2(c) requires that

$$\frac{1}{U'(\bar{Y}^R)} > \gamma + \lambda \frac{\Delta f(\bar{s}|e)}{f(\bar{s}|\bar{e})} > \gamma + \lambda \frac{\Delta f(s''|e)}{f(s''|\bar{e})} = \frac{1}{u'} > \gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})}. \quad (15)$$

We may summarize as follows.

PROPOSITION 2. *If reservation income affects the optimal incentive scheme, then either \underline{Y} is paid in the lowest states with a discontinuous jump to \bar{Y}^R (which is paid in some or all higher states) or else \bar{Y}^R is paid in the lowest states beyond which pay is smoothly increasing in performance.*

The scheme of Figure 2(c) can be interpreted as fixed pay except that, if performance falls below some threshold, the agent is fired.

4. Endogenous Reference Income

Koszegi and Rabin (2004) argue that expectations matter in locating the reference income. People code gains and losses according to whether the outcome is better or worse than anticipated. Thus, Mellers et al. (1999) and Breiter et al. (2001) find that both the actual lottery outcome and unattained outcomes affect subjects' satisfaction with their payoffs. The utility of winning a given prize is lower, the greater is the probability of doing better. In the present context, where the task is to design a reward scheme, this means that reference income is endogenous. One possibility is that the reference income is, as in Gul (1991), the certainty equivalent. In the principal–agent setting, this turns out to be equivalent to the exogenous formulation. The reservation utility ties down the reference income.

Making reference income equal or proportional to the certainty equivalent does yield some questionable implications. Suppose two managers differ only in that one feels disappointment more intensely (i.e., has a higher l). Given the incentive scheme, the more loss-averse manager has lower certainty equivalent, so there are realizations at which the less loss-averse manager experiences loss but the more loss-averse manager does not. It seems more plausible that differences in l influence the cost of loss but not the set of incomes for which it is experienced.

An alternative formulation is suggested by the following example. Suppose that pay in a new job is fixed at \$100,000 but, for very rarely achieved exceptional performance, there is a bonus of \$10,000. The natural reference income is then \$100,000. Were the performance target more lax and so achieved (say) 70% of the time, the reference income might well be \$110,000. What this suggests is that there is no disappointment in achieving an income that is rarely bettered, whereas losses are experienced when one fails to beat the odds. The focus is thus on probabilities, not on whether utility is below the expected level. Specifically, we propose that the reference income is the median income. Aside from the psychological appeal, an important consequence is that—because different income distributions can have the same expected utility but different medians—the design of the incentive scheme now extends to the choice of reference income.

A test of these various theories of the reference income can be derived from the experiment reported in Coelho (2004) and Coelho et al. (2004). Subjects engage in a binary outcome task for which there is a prize for success. The design elicits subjective success probabilities and associated certainty equivalents. On average, the ratio of certainty equivalent to expected income is unity, but it is decreasing in the success probability. The relationship is highly significant.

According to expected utility theory, the ratio of certainty equivalent to expected income is monotonically increasing in the probability of success. Loss aversion with an exogenous reference income compounds this tendency. The expected cost of loss, which decreases the certainty equivalent, is smaller for those with a lower probability of failing, whereas expected income is higher. The Gul formulation also has the property that the certainty equivalent steadily approaches expected income as success probability rises.¹⁵ This is not the case if loss is ex-

15. For a binary lottery, let the prize be w obtained with probability p and let the certainty equivalent be c . Normalizing so that $U(0) = 0$, the Gul formulation implies $U(c) = pU(w) - (1 - p)lU(c)$. Thus $U(c)/U(c)pU(w) = 1/(1 + (1 - p)l)$. As p rises, $U(c)$ must increase by a higher percentage.

perienced only when the income is so low that it occurs less than half of the time. Rather, losses are now felt only by those expecting to win more than half of the time. Hence the subjects with higher success probabilities tend to have lower certainty equivalents relative to expected winnings. If there is elation when an unexpectedly good event occurs, this tendency is further augmented.

The next two sections analyze the implications of the certainty equivalent and median specifications for the choice of incentive scheme.

4.1. Gul's Model of Disappointment Aversion

An axiomatic foundation for the utility function (1) is provided by Gul (1991), in which loss aversion is linear and reference income is the certainty equivalent of the prospect, including the cost of loss.¹⁶ The general idea is that disappointment is registered whenever realized utility falls short of its expected level. That is, reference income solves

$$U(Y^R(\bar{e})) = \int_{\underline{s}}^{\bar{s}} (U(Y^R(\bar{e})) - \theta l(U(Y^R(\bar{e})) - U(Y(s)))) f(s|\bar{e}) ds - \bar{e}. \quad (16)$$

At an optimum the IR constraint binds, so $U(Y^R(\bar{e})) = V^*$. At an optimum the IC constraint also binds and so

$$\begin{aligned} & \int_{\underline{s}}^{\bar{s}} (U(Y^R(\bar{e})) - \theta l(U(Y^R(\bar{e})) - U(Y(s)))) f(s|\bar{e}) ds - \bar{e} \\ &= \int_{\underline{s}}^{\bar{s}} (U(Y^R(\underline{e})) - \theta l(U(Y^R(\underline{e})) - U(Y(s)))) f(s|\underline{e}) ds - \underline{e}. \end{aligned} \quad (17)$$

Hence the utility of the prospect net of the cost of low effort, $U(Y^R(\underline{e}))$, must also equal V^* . Therefore, $Y^R(\bar{e}) = Y^R(\underline{e})$. Denote the common value by \bar{Y}^R .

The principal–agent problem with exogenous reference income of \bar{Y}^R is equivalent to the solution with Gul preferences.¹⁷ Equation (16) does, however, locate the reference income and so rules out all but one of the possibilities identified in Proposition 1.

PROPOSITION 3. *When the reference income is the certainty equivalent, there exists a set of intermediate states $\underline{s} < s < \bar{s}$ that all pay the same income.*

Given the normalization, risk aversion implies that the elasticity of $U(c)$ is below unity. Thus c/pw is increasing in p , which is inconsistent with the experimental results.

16. Gul (1991) modifies the independence axiom in a simple way to account for the Allais paradox. The utility function that emerges captures both elation and disappointment; it has one more parameter than the expected utility function and is unique up to an affine transformation. The Gul formulation precludes risk aversion in the gain space and risk loving in the loss space. Loomes and Sugden (1986) use a utility function that is similar to Gul's, but the reference income is the certainty equivalent ignoring loss aversion.

17. Suppose to the contrary that the two solutions differ. Since both must satisfy the same IC and IR constraints, each solution must be feasible for the other problem. Hence, if they deliver different expected costs, then there is a contradiction.

Proof. First we show that any flat segment of pay cannot include \underline{s} . Similar reasoning implies that a flat segment cannot include \bar{s} . Finally, since there are necessarily incomes above and below the certainty equivalent, there must be an intermediate interval of states satisfying

$$\frac{1}{U'(\bar{Y}^R)} = (1 + x(s)l) \left(\gamma + \lambda \frac{\Delta f(s|e)}{f(s|\bar{e})} \right)$$

and $x(s) \leq 1$. □

Routledge and Zin (2003) extend Gul's preference formulation in a simple way that again has an axiomatic foundation. Their modification involves non-central disappointment, which is achieved by shifting the disappointment cutoff to some fraction of the certainty equivalent. This means that only the worst outcomes are viewed as disappointing; otherwise the formulation is identical to that of Gul. Since the reference income is below the certainty equivalent, it is now possible that the flat segment extends to the worst state—an option-like compensation scheme.

4.2. Median Income as Reference Income

We now assume that loss or disappointment occurs at all incomes for which it is odds-on that a higher income would be drawn. At first sight, this implies that the reference income is the median income. However, even though the states are continuously distributed, the chosen incentive scheme may have a mass point (a flat segment in the income schedule). Indeed, our main result is that, with loss aversion, this will always be the case. Suppose for example that payments smoothly increase over the worst 30% of the states to reach \$100, that the next 30% of states are all paid \$100, and that pay smoothly increases from \$120 for the remaining 40% of states. In this example the reference income is \$100. Although only 40% of realizations are higher than \$100, all incomes that are exceeded with probability more than half are below \$100. Thus $Y^R = \arg \max \{G(Y|\bar{e}) \leq 0.5\}$, where $G(Y|\bar{e})$ is the cumulative conditional probability distribution of income.¹⁸ We condense this formulation as “reference income equals median income”.

Given linear loss aversion, we conjecture that the optimal compensation scheme is (weakly) monotonic in performance. This implies that reference income is necessarily paid in the state that is median as long as high effort is chosen. Denoting this state as s^m , we further propose that $Y(s) > Y^R$ for all $s > s^m$ and $Y(s) = Y^R$ for all $s \in [s', s^m]$. Imposing these properties, the choice variables are: Y^R , the income paid in the median state; s' , the worst state for which $Y(s) = Y^R$; and $Y(s)$ for all states. We show that the first-order conditions arising from this problem imply that $s' < s^m$, so there is an interval of constant pay. There is an upward discontinuity in income at state s^m , justifying our assumption that the reference income is not paid above the median. Having found the scheme that is optimal

18. $G(Y(s)|\bar{e})$ is derived from the distribution function $F(s|\bar{e})$ and allows for the same realizations of $Y(s)$ for distinct states. If income realizations are nondecreasing in the state realization, then states that have the same income form a mass point.

subject to these constraints, we finally note that, for all $s \in [s', s^m]$, neither upward nor downward deviations from Y^R are profitable; this establishes that the optimal scheme has been found for the unrestricted problem.

The principal minimizes

$$\int_{s^m}^{\bar{s}} Y(s)f(s|\bar{e}) ds + \int_{s'}^{s^m} Y^R f(s|\bar{e}) ds + \int_{\bar{s}}^{s'} Y(s)f(s|\bar{e}) ds \quad (18)$$

subject to the IR and IC constraints. These constraints take the same form as (5) and (6) with now Y^R equal to the lowest income for which no more than 50% of incomes are higher.¹⁹ The first-order conditions for the problem are the IR and IC constraints satisfied as equalities (again it can be shown that the multipliers γ and λ are positive). In addition, as in the exogenous reference income case, for all states outside $[s', s^m]$ we have that $Y(s)$ satisfies

$$f(s|\bar{e}) - (1 + x(s)l)U'(Y(s))(\gamma f(s|\bar{e}) + \lambda \Delta f(s|e)) = 0. \quad (19)$$

The variable $x(s) \in [0, 1]$ has the same interpretation as in the exogenous case. Since reference income is endogenous, we have a further condition for the choice of Y^R :

$$\begin{aligned} \int_{s'}^{s^m} f(s|\bar{e}) - U'(Y^R) \int_{s'}^{s^m} (\gamma f(s|\bar{e}) + \lambda \Delta f(s|e)) ds \\ + lU'(Y^R) \int_{\bar{s}}^{s'} x(s)(\gamma f(s|\bar{e}) + \lambda \Delta f(s|e)) ds = 0. \end{aligned} \quad (20)$$

PROPOSITION 4. *The optimal incentive scheme has a flat segment that ends at the median state and may extend back to the worst state. At the median state there is an upward discontinuity beyond which pay is smoothly increasing in performance. There is a threshold level of loss aversion l^* such that the flat segment includes the worst state if and only if $l \geq l^*$.*

Proof. Y^R is paid in the interval $[s', s^m]$, so s' is the worst state at which $Y(s) = Y^R$. Hence, s' satisfies (19) written as

$$\frac{1}{U'(Y(s))} = (1 + x(s)l) \left(\gamma + \lambda \frac{\Delta f(s|e)}{f(s|\bar{e})} \right), \quad (21)$$

where $Y(s) = Y^R$ and $x(s) = 1$. Since (21) and MLRP imply that $x(s) = 1$ and $Y(s) < Y^R$ for states $s \in [\bar{s}, s']$, it follows that (20) can be written as

$$\frac{1}{U'(Y^R)} = \gamma + \lambda \frac{\int_{s'}^{s^m} \Delta f(s|e) ds}{\int_{s'}^{s^m} f(s|\bar{e}) ds} - \frac{l \int_{\bar{s}}^{s'} (\gamma f(s|\bar{e}) + \lambda \Delta f(s|e)) ds}{\int_{s'}^{s^m} f(s|\bar{e}) ds}, \quad (22)$$

which (given l , γ , and λ) solves for Y^R .

19. If $e = \underline{e}$ then the median state will be below s^m , but we assume that it exceeds s' , as will be true if the effect of high effort on probabilities is not too great. It then follows that reservation income is the same under both income levels.

For states better than s^m , from (21) with $x(s) = 0$ we have $Y(s) > Y^R$; by MLRP, $Y(s)$ is increasing in s . From (21) with $x(s) = 0$ and (22), continuity of the reward scheme at s^m requires that

$$\gamma + \lambda \frac{\Delta f(s^m|e)}{f(s^m|\bar{e})} = \gamma + \lambda \frac{\int_{s'}^{s^m} \Delta f(s|e) ds}{\int_{s'}^{s^m} f(s|\bar{e}) ds} - l \frac{\int_{s'}^{s'} (\gamma f(s|\bar{e}) + \lambda \Delta f(s|e)) ds}{\int_{s'}^{s^m} f(s|\bar{e}) ds}.$$

However, the left-hand side exceeds the right-hand side because, by MLRP,²⁰

$$\frac{\Delta f(s^m|\bar{e})}{f(s^m|\bar{e})} \geq \frac{\int_{s'}^{s^m(\bar{e})} \Delta f(s|e) ds}{\int_{s'}^{s^m(\bar{e})} f(s|\bar{e}) ds}.$$

Hence there must be an upward discontinuity at s^m in the compensation scheme. This justifies our initial assertion that the flat segment ends at the median state, s^m .

Next we show that $s' < s^m$. Let δ be the right-hand side of (21) minus the right-hand side of (22) with $x(s') = 1$. As s' tends to s^m from below, δ tends to

$$l \left(\gamma + \lambda \frac{\Delta f(s^m|\bar{e})}{f(s^m|\bar{e})} \right) + l \frac{\int_{s'}^{s^m} (\gamma f(s|\bar{e}) + \lambda \Delta f(s|e)) ds}{f(s^m|\bar{e})} > 0.$$

Since the left-hand sides of (21) and (22) are equal, it follows that s' cannot be close to s^m . Given MLRP, lowering s' reduces δ . If $\delta > 0$ at $s' = \underline{s}$ then the reservation income is paid in all states below the median. Otherwise, the flat segment of the incentive scheme does not extend to the worst states.

To check that deviations from the flat segment are unprofitable, note that in the modified problem it is feasible to increase $Y(s^m)$ and set $s' = s^m$. This increases the reference income, leaving the remainder of the flat segment unaffected. It has been shown, however, that this scheme is worse for the principal than the flat segment extending down from the median state for a finite interval. Now consider raising income for some state $s' < s < s^m$. The consequences are identical to increasing $Y(s^m)$ except that, by MLRP, the incentive effect is lower. Hence, it is not optimal to increase any individual income in the flat segment. Decreases do not change reference income and, since decreases are not optimal for s' , this holds *a fortiori* for better states.

In order to show that there exists a constant pay interval for all states below s^m if and only if loss aversion exceeds some threshold l^* , first impose $s' = \underline{s}$ and set l to solve

$$\frac{1}{U'(Y^R)} = (1 + l) \left(\gamma + \lambda \frac{\Delta f(\underline{s}|e)}{f(\underline{s}|\bar{e})} \right).$$

20. To see this, set $g(s|\bar{e}) \equiv \Delta f(s|e)/f(s|\bar{e})$. Then

$$\frac{\int_{s'}^{s^m} \Delta f(s|\bar{e}) ds}{\int_{s'}^{s^m} f(s|\bar{e}) ds} = \frac{\int_{s'}^{s^m} g(s|\bar{e}) f(s|\bar{e}) ds}{\int_{s'}^{s^m} f(s|\bar{e}) ds}$$

and, by MLRP, for $s \leq s'$ we have

$$g(s|\bar{e}) f(s|\bar{e}) \leq g(s^m|\bar{e}) f(s|\bar{e}) \leq g(s^m|\bar{e}) \frac{\int_{s'}^{s^m} f(s|\bar{e}) ds}{\int_{s'}^{s^m} f(s|\bar{e}) ds} = \frac{\Delta f(s^m|e)}{f(s^m|\bar{e})}.$$

Then, even if one could freely vary income in the worst state, it would be optimally set at the reference level.

If $l > l^*$ but the solution remains unchanged, then the first-order conditions yield the same value of the multipliers (in this case the loss space is not entered), confirming the unchanged solution. Now suppose that $l < l^*$. If a flat segment remained at the bottom of the distribution—even if it ends at a different state than under l^* —then this scheme could have been adopted when $l = l^*$, in which case (as the loss space is not entered) it would have yielded the same payoff. Similarly, the scheme that is allegedly optimal when $l < l^*$ is also feasible and has the same payoff when $l = l^*$, a contradiction. Hence, if $l < l^*$, then the optimal scheme does not have a complete flat segment at the bottom. \square

The two forms of pay schedule implied by Proposition 4 are illustrated in Figures 3a and 3b.

Figures 3a and b here.

When reference income equals the median realization, the optimal incentive scheme has a discontinuity even under concavity on the downside. This is because the principal not only takes into account the “tax” on payments in the loss space, as in the case of exogenous reference income, but also recognizes that whether a particular income is in the loss space depends on the nature of the scheme. Compare the effect of a marginal reduction in pay in a state just above the median to that of a pay cut in the median state. In the latter case, the fall means that pay can also be lowered in somewhat worse states on the flat segment without the loss space being entered. Thus there would be an inconsistency if income were virtually the same at the median and just above it, since the effects of a cut are very different at the two levels. Hence there is a finite bonus for doing better than expected.

5. Loss Aversion, Reference Income, and Risk Aversion

We use a simple discrete formulation of the model to illustrate the magnitude of loss aversion required to generate a complete flat segment at the bottom of the compensation scheme (an option) and to explore how this magnitude depends on the nature of the reference income, the level of loss aversion, and the degree of risk aversion.

Suppose three states H, M, L occur (respectively) with probabilities P_H, P_M, P_L when high effort is applied and with probabilities P_H^*, P_M^*, P_L^* when low effort is applied. Then MLRP implies that

$$\frac{\Delta P_L}{P_L} < \frac{\Delta P_M}{P_M} < \frac{\Delta P_H}{P_H}.$$

The contracted payments to the entrepreneur in these states are (respectively) $Y_H, Y_M,$ and Y_L . The principal’s problem is to choose Y_H, Y_M, Y_L to solve

$$\min\{P_H Y_H + P_M Y_M + P_L Y_L\} \tag{23}$$

subject to

$$P_H U(Y_H) + P_M U(Y_M) + P_L (U(Y_L) - l(U(Y^R) - U(Y_L))) - \bar{e} = V^*, \quad (24)$$

$$\Delta P_H U(Y_H) + \Delta P_M U(Y_M) + \Delta P_L (U(Y_L) - l(U(Y^R) - U(Y_L))) = \Delta \bar{e}. \quad (25)$$

5.1. Threshold Loss Aversion

We define the threshold level of loss aversion as the lowest level of loss aversion for which the principal sets $Y_M = Y_L = Y^R$. First, let the reference income be exogenous so that $Y^R = \bar{Y}^R$. The payments Y_H, Y_M, Y_L and the multipliers γ, λ must satisfy the IR and IC constraints as well as

$$P_i - \gamma(P_i U'(Y_i) + P_i x_i l U'(Y_i)) - \lambda(\Delta P_i U'(Y_i) + \Delta P_i x_i l U'(Y_i)) = 0, \quad (26)$$

with $i = H, M, L$. Here $x_i = 0$ for $Y_i > \bar{Y}^R$; $x_i = 1$ for $Y_i < \bar{Y}^R$; and $0 \leq x_i \leq 1$ for $Y_i = \bar{Y}^R$. Solve IR and IC for Y_H and $Y_M = Y_L = Y^*$. For an option it is necessary that $\bar{Y}^R = Y^*$, so we set the reference income at this level. In addition, at l^{E^*} , the threshold level of loss aversion for the option to be optimal, we set $x_M = 0$, $x_L = 1$, $Y_M = \bar{Y}^R$, and $Y_L = \bar{Y}^R$; then the three equations in (26) must solve for γ and λ and l^{E^*} . This yields

$$l^{E^*} = \left(1 + \left(-\frac{\Delta P_M}{P_M} + \frac{\Delta P_L}{P_L} \right) \left(\frac{U'(\bar{Y}^R)}{U'(Y_H)} - 1 \right) \left(\frac{\Delta P_H}{P_H} - \frac{\Delta P_M}{P_M} \right)^{-1} \right)^{-1} - 1. \quad (27)$$

Now let the reference income be endogenous and equal to the median income. Assume that M is the median state whether or not high effort is applied, so $Y^R = Y_M$; hence $x_H = x_M = 0$. At an optimum, Y_H, Y^R, Y_L and the multipliers γ, λ must satisfy the IR constraints and the following first-order conditions:

$$P_H - \gamma P_H U'(Y_H) - \lambda \Delta P_H U'(Y_H) = 0, \quad (28)$$

$$P_M - \gamma(P_M U'(Y^R) - P_L x_L l U'(Y^R)) - \lambda(\Delta P_M U'(Y^R) - \Delta P_L x_L l U'(Y^R)) = 0, \quad (29)$$

$$P_L - \gamma(P_L U'(Y_L) + P_L x_L l U'(Y_L)) - \lambda(\Delta P_L U'(Y_L) + \Delta P_L x_L l U'(Y_L)) = 0. \quad (30)$$

To find the threshold level of loss aversion, first solve the IR and IC conditions for Y_H and $Y_L = Y^R$. Then set $x_L = 1$ and solve (28)–(30) for γ, λ and the threshold level of loss aversion:

$$l^{M^*} = \left(1 + \left(\frac{\Delta P_H}{1 - P_H} + \frac{\Delta P_L}{P_L} \right) \left(\frac{U'(Y^R)}{U'(Y_H)} - 1 \right) \left(\frac{\Delta P_H}{1 - P_H} + \frac{\Delta P_H}{P_H} \right)^{-1} \right)^{-1} - 1. \quad (31)$$

PROPOSITION 5. *The monotone likelihood ratio property is a necessary and sufficient condition for $l^{E^*} > l^{M^*}$.*

Proof. Set the exogenous reference income equal to the level chosen when the reference income is endogenous. Comparing (28) to (24), it follows that $l^{E^*} > l^{M^*}$ if and only if $\Delta P_H / (1 - P_H) = -(\Delta P_M + \Delta P_L) / (P_M + P_L) > -\Delta P_M / P_M$ or $\Delta P_M / P_M > \Delta P_L / P_L$, as is implied by MLRP. \square

To understand this result, suppose that loss aversion were at the threshold for the median formulation. If median income is increased then Y_L is now in the loss space, so this adds to the cost of the change. When the reference income is exogenous, raising Y_M does not precipitate losses for the Y_L realization. Hence, the option scheme cannot be optimal for the exogenous formulation if loss aversion is barely sufficient for an option to be optimal in the endogenous case.

The following stylized numerical example illustrates this point. The manager's utility function is given by $U = Y^r$ with relative risk aversion of $\sigma = 1 - r = 0.5$. The principal's costs are given by C . The three states H, M, L occur with probabilities $(0.25, 0.5, 0.25)$ if effort is high. Without effort the probabilities are $(0.15, 0.5, 0.35)$, so MLRP is satisfied by this formulation. The cost of the effort needed to induce the superior distribution is 0.1. The alternative for the manager is to receive a riskless income of 9 at an alternative effort level normalized to zero. The principal's objective is to minimize the cost of compensation subject to (i) a fixed reservation level of utility from income plus the utility cost of effort of $U^* + \bar{e} = 3.1$ and (ii) the incentive constraint that effort be applied.

The first row of the following table shows results when there is no loss aversion. The second row shows the optimal scheme when, in the median formulation, loss aversion is at the threshold level necessary to induce an option.

H	M	L	C
13	9.6	6.7	9.73
14.8	8.1	8.1	9.8

In this example, $l^{M*} = 0.214$. Thus, if loss aversion increases the marginal utility of downside income by only 21%, then this is sufficient to induce an option-like scheme. Constraining payment to be equal in the bottom two states increases the principal's costs very little, so this is worth doing to avoid even low costs of risk aversion.

If there is an exogenous reference income of 8.1 then the option-type scheme becomes optimal with sufficient loss aversion. The required threshold level of loss is $l^{E*} = 0.542$, but even this level of loss aversion is below that suggested by experimental evidence.

Finally, note that loss aversion raises payments in both the best and the worst states and depresses pay for intermediate performance. Incentives are concentrated in the upper segments of the compensation scheme, which becomes more sensitive to performance.

5.2. Risk Aversion and Loss Aversion

Next we show how risk aversion affects the threshold level of loss aversion required for the optimal contract to set $Y_M = Y_L = Y^R$. The manager's utility function is given by $U = Y^r$ ($0 < r < 1$) with relative risk aversion of $\sigma = 1 - r$. In this formulation we assume that the manager has an outside income opportunity of W that is evaluated with the same utility function.

PROPOSITION 6. *Raising r (lowering risk aversion) raises l^{M*} .*

Proof. The scheme $\{Y_H, Y_M = Y_L = Y^R\}$ solves IC and IR written as

$$P_H Y_H^r + (P_M + P_L)(Y^R)^r - \bar{e} = W^r, \quad (32)$$

$$\Delta P_H Y_H^r + (\Delta P_M + \Delta P_L)(Y^R)^r = \bar{e}. \quad (33)$$

Varying r , we have

$$\frac{dY_H}{dr} = ((\ln W)W^r - (\ln Y_H)Y_H^r)(rY_H^{r-1})^{-1}, \quad (34)$$

$$\frac{dY^R}{dr} = ((\ln W)W^r - (\ln Y^R)Y^R)(r(Y^R)^{r-1})^{-1}. \quad (35)$$

Under our scheme, $Y_H > W$ and $Y_L < W$; thus we have $dY_H/dr < 0$ and $dY^R/dr > 0$. It then follows that

$$\frac{d}{dr} \left(\frac{Y_H}{Y^R} \right)^{1-r} = \left(\left(\frac{dY_H}{dr} Y^R - \frac{dY^R}{dr} Y_H \right) (Y^R)^{-2} \right) \left(\left(\frac{Y_H}{Y^R} \right)^{1-r} \ln \left(\frac{Y_H}{Y^R} \right) \right) < 0. \quad (36)$$

Hence (28) can be written as

$$l^{M*} = \left(1 + \left(\frac{\Delta P_H}{1 - P_H} + \frac{\Delta P_L}{P_L} \right) \left(\left(\frac{Y_H}{Y^R} \right)^{1-r} - 1 \right) \left(\frac{\Delta P_H}{1 - P_H} + \frac{\Delta P_H}{P_H} \right)^{-1} \right)^{-1} - 1, \quad (37)$$

where we have used the fact that $U'(Y^R)/U'(Y_H) = (Y_H/Y^R)^{1-r}$. Using (36) then yields the result that l^{M*} is increasing in r . \square

At first sight this result seems paradoxical, since the role of an option-type compensation scheme appears to protect against risk. In fact, raising the intermediate payment at the expense of the highest is beneficial only to the extent that the marginal utility of income is diminishing—that is, if risk aversion is present. Unlike increasing income in the lowest state, there are no adverse incentive effects (more generally, MLRP means that the incentive effects of raising income are more favorable in higher states). The net gain to abandoning the compensation scheme with $Y_M = Y_L$ is thus greater, the more risk averse is the agent. Thus, if $Y_M = Y_L$ is only just optimal at (r, l^{M*}) , then marginally lowering r means the agent values Y_L more relative to Y_M and so the principal can lower Y_M . Because the marginal utility of high income is now lower, maintaining incentives requires that we raise Y_H even though Y_M has declined. But then $U'(Y_M)/U'(Y_H) = (Y_H/Y_M)^{1-r}$ increases and therefore l^* decreases.

6. Choice of Risk

Option-based compensation contracts are often criticized for inducing “too much” risk taking by managers. In our basic model, managers choose effort for a given project and intrinsic risk is beyond their control. When this is not the case, incentive pay that exposes managers to more company risk than they would normally choose (for portfolio reasons) may induce actions that reduce this risk. Hence, it

is possible that managers will choose projects that are excessively safe from the shareholders' point of view.²¹

Such reasoning is sometimes presented as an argument in favor of using option-like compensation. Eliminating downside risk restores the balance in favor of risk taking. As Ross (2004) points out, the standard argument—that the convexity in the call option's payoff function induces gambling—is not the whole story. The implicit assumption behind the standard argument is that the manager's von Neuman–Morgenstern utility function has the same (absolute) risk aversion over the entire relevant domain of outcomes. But this will generally not be the case: transforming the manager's reward function (with an option contract) will have the well-understood direct effect of convexifying the utility function over outcomes; but it can also shift the domain of outcomes into a region of the utility function with more risk aversion (what Ross calls a translation effect) such that the utility cost of risk bearing actually increases. Ross also identifies a magnification effect: when it is out of the money, the power of the option contract is locally lower than the share-based scheme and reduces the incentive to take risks relative to shares. Overall, then, it is ambiguous whether option-based compensation encourages risk taking.

Here we argue that the presence of loss aversion provides a further reason why risk taking may be lower with an option contract than with a share or other incentive scheme that does not protect against the downside. Risk loving in the loss space means that a manager once down may behave recklessly. Options prevent the manager from entering the gambling zone, and the firm benefits from the changed behavior.

Assume that, once *ex ante* effort is applied, the manager obtains early warning of the realization. If indications are that returns will be in the loss space, then the manager can exert effort on a risky action with some chance of converting a performance that is below median to one in excess of it, although failure results in a collapse in performance. The action lowers expected firm value, so the risk-neutral shareholders wish to discourage it.

Given the model of Section 3.2, compare a scheme of the form in Figure 2a, which protects against the loss space, with that of Figure 2b or 2c, which do not. In Figure 2a, essentially an option, an unfair gamble that if successful raises performance—from a low level to one just above that at which the reference income is paid—is a one-way bet. Nonetheless, expected income increases little, and over this interval the agent is risk averse. Therefore, the action may be rejected if the effort required is costly. This is a version of the Ross (2004) magnification effect. Under the schemes in Figures 2b and 2c, such a gamble has a large effect on income in a zone where the agent is risk loving. Since the gamble lowers the expected return to shareholders, they will prefer the option-like schedule that offers carrots for success rather than a scheme that involves sticks in the event of failure.

The conclusions are similar in comparing share-based compensation with stock options. Shares imply that pay is sensitive to performance throughout the

21. Carpenter (2000) considers a conventional investment problem for a risk-averse fund manager. She shows in a dynamic setting that, under some conditions, option-based compensation can induce less risk taking. However, options that are deep out of the money induce excess risk taking.

range of performance. There is a cost to the manager if the action fails, but success yields significant gains. Because the gamble is predominantly in the loss space (within which the manager is a risk lover), the expected private gain to the action may be positive. Suppose, for example, that if no action is taken then the manager anticipates being 10 below the reference income. With probability 0.6, the action either further reduces income to 20 below or boosts it to \bar{Y}^R . Convexity may mean that the increase in expected utility is enough for the risk-increasing action to be taken. Under a full option, it never will be.

More explicitly, let the value function be given by (1) with exogenous reference income. As in the general prospect theory formulation, the value function is convex in the loss space. Assume that, after early warning of the realization, a manager can take a costly action that has a chance of converting a below-median realization to one in excess of it; but if the action fails, the firm's income is reduced. The foreseen down state is denoted by d . Suppose the costly action is taken. Then the chance of improvement to a state in the gain space denoted by u is p^+ ; the probability of precipitating an even worse state, denoted by w , is the complementary probability p^- .²² Under an option with strike price \bar{Y}^R , the expected gross utility gain from the action is

$$G^O = p^+(U(Y_u^O) - U(\bar{Y}^R)), \quad (38)$$

where Y_u^O is the manager's income under the option if the action pays off and \bar{Y}^R (the reference level of income) is income when the action fails and the option is out of the money. With shares, the expected gross utility gain is

$$G^S = p^+(U(Y_u^S) - U(Y_d^S) + g(U(\bar{Y}^R) - U(Y_d^S))) \\ + p^-(U(Y_w^S) + g(U(Y^R) - U(Y_w^S)) - U(Y_d^S) + g(U(\bar{Y}^R) - U(Y_d^S))), \quad (39)$$

where Y_d^S , Y_u^S , and Y_w^S denote (respectively) the income of the manager under the share scheme in states d , u , and w . This scheme may have a fixed wage component but will have the property that $Y_u^S > \bar{Y}^R > Y_d^S > Y_w^S$. As in Ross (2004), it is ambiguous whether G^O is greater or less than G^S . Our analysis introduces a new consideration. The convexity of the objective function in the loss space means that the marginal cost of loss is most severe as the loss space is entered. Hence, in the expression for G^S , if the gamble pays off then the value of the saving in utility forgone from loss, $g(U(\bar{Y}^R) - U(Y_d^S))$, is greater per dollar of income gain than the per-dollar utility cost of the extra loss incurred, $g(U(\bar{Y}^R) - U(Y_w^S)) - g(U(\bar{Y}^R) - U(Y_d^S))$, should the gamble fail. With enough loss aversion, the risky action yields greater benefit with the share scheme, since the upside includes the elimination of the magnified cost of loss. This result may be stated as follows.

PROPOSITION 7. *Convexity of loss aversion increases the gain from taking a risky action by more under share-based compensation than under an option contract.*

22. For simplicity and plausibility the reference income is taken as the median prior to any gamble.

Thus, option-like compensation may make the manager less inclined to take risky actions than would be the case with share-based compensation. Whether this is a merit or not depends upon whether the action increases or decreases the expected value of the company. Both possibilities are consistent with the options providing relatively less incentive to take risks.

7. Conclusion

Adding loss aversion to the standard principal–agent model with unobservable effort can explain why bonuses are paid for good performance rather than penalties being exacted for poor performance. The use of option-like compensation in executive remuneration schemes may therefore be efficient. Subjecting a manager to the discomfort of loss relative to their reference income is too costly a stick to efficiently induce good performance, especially when the perception of losses can be reduced by lowering median income.

Loss aversion may account for zones in which pay is insensitive to performance as well as for discontinuous incentives. Depending on the nature of loss aversion and the reference income, the only incentive may be dismissal if performance falls below some threshold or a finite bonus paid when performance is better than expected. Our analysis also shows how options can help to eliminate reckless behavior by managers. Of course, the efficacy of corporate governance, financial reporting standards, and cultural factors must be considered when attempting to understand the general patterns of executive compensation both within and between countries. Our contention is that loss aversion is also important.

References

- [1] Barberis, Nicholas, Ming Huang, and Tano Santos (2001). “Prospect Theory and Asset Prices.” *Quarterly Journal of Economics*, 116, 1–53.
- [2] Bebchuck, Lucien, Jesse Fried, and David Walker (2001). “Executive Compensation in America: Optimal Contracting or Extraction of Rents?” Discussion Paper no. 338, John Ohlin Center for Law, Economics and Business, Harvard University.
- [3] Bebchuck, Lucien and Jesse Fried (2003). “Executive Compensation as an Agency Problem.” *Journal of Economic Perspectives*, 17, 71–92.
- [4] Bertrand, Marianne and Sendhil Mullainathan (2000). “Agents without Principals.” *American Economic Review*, 90, 203–208.
- [5] Breiter, Hans, Itzhak Aharon, Daniel Kahneman, Anders Dale, and Peter Shizgal (2001). “Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains, and Losses.” *Neuron*, 30, 619–639.
- [6] Brunnermeier, Marcus (2004). “Learning to Reoptimize Consumption at New Income Levels: A Rationale for Prospect Theory.” *Journal of the European Economic Association*, 2, 98–114.

- [7] Carpenter, Jennifer (2000). "Does Option Compensation Increase Managerial Risk Appetite?" *Journal of Finance*, 55, 2311–2332.
- [8] Coelho, Marta (2004). "Unrealistic Optimism, Entrepreneurship and Adverse Selection." Ph.D. thesis, London School of Economics.
- [9] Coelho, Marta, David de Meza, and Dianne Reyniers (2004). "Irrational Exuberance, Entrepreneurial Finance and Public Policy." *International Tax and Public Finance*, 11, 391–417.
- [10] Grossman, Sanford and Oliver Hart (1983). "An Analysis of the Principal–Agent Problem." *Econometrica*, 51, 7–45.
- [11] Gul, Faruk, (1991). "A Theory of Disappointment Aversion." *Econometrica*, 59, 667–686.
- [12] Hall, Brian and Jeffery Lieberman (1998). "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, 113, 653–691.
- [13] Hall, Brian and Kevin Murphy (2001). "Optimal Exercise Prices for Executive Stock Options." *American Economic Review*, 90, 209–214.
- [14] Holmstrom, Bengt (1979). "Moral Hazard and Observability." *Bell Journal of Economics*, 10, 74–91.
- [15] Jensen, Michael and Kevin Murphy (1990). "Performance Pay and Top Management Incentives." *Journal of Political Economy*, 98, 225–264.
- [16] Kachelmeier, Steven and Mohamed Shehata (1992). "Examining Risk Preferences under High Monetary Incentives: Experimental Evidence from the People's Republic of China." *American Economic Review*, 82, 1120–1140.
- [17] Kahneman, Daniel, Jack Knetsch, and Richard Thaler (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98, 1325–1348.
- [18] Kahneman, Daniel, Jack Knetsch, and Richard Thaler (1991). "The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 5, 193–206.
- [19] Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47, 263–291.
- [20] Knetsch, Jack, Fang Fang Tang, and Richard Thaler (2001). "The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing?" *Experimental Economics*, 4, 257–269.
- [21] Koszegi, Botond and Mathew Rabin (2004). "A Model of Reference-Dependent Preferences." Working paper, University?
- [22] Loomes, Graham and Robert Sugden (1986). "Disappointment and Dynamic Consistency in Choice under Uncertainty." *Review of Economic Studies*, 53, 272–281.
- [23] Mas-Collel, Andreu, Michael Whinston, and Jerry Green (1995). *Microeconomic Theory*. Oxford University Press.
- [24] Medoff, James and Katharine Abraham (1980). "Experience, Performance, and Earnings." *Quarterly Journal of Economics*, 95, 703–736.
- [25] Mellers, Barbara, Alan Schwartz, and Ilana Ritov (1999). "Emotion-Based Choice." *Journal of Experimental Psychology: General*, 128, 332–345.
- [26] Mirrlees, James (1997). "The Theory of Moral Hazard and Unobservable Behaviour, Part 1." *Review of Economic Studies*, 66, 3–21.
- [27] Murphy, Kevin (2002). "Explaining Executive Compensation: Managerial

- Power vs. the Perceived Cost of Stock Options.” *University of Chicago Law Review*, 69, 847–869.
- [28] Prendergast, Canice (1999). “The Provision of Incentives in Firms.” *Journal of Economic Literature*, 37, 7–63.
- [29] Prendergast, Canice (2000), “What Trade-off of Risk and Incentives?” *American Economic Review, Papers and Proceedings*, 90, 421–425.
- [30] Ross, Stephen (2004). “Compensation, Incentives and the Duality of Risk Aversion and Riskiness.” *Journal of Finance*, 59, 207–225.
- [31] Routledge, Brian and Stanley Zin (2003), “Generalized Disappointment Aversion and Asset Prices.” Working paper, Carnegie Mellon University.
- [32] Samuelson, William and Richard Zeckhauser (1988). “Status Quo Bias in Decision Making.” *Journal of Risk and Uncertainty*, 1, 7–59.
- [33] Segal, Uzi and Aavia Spivak (1990). “First Order versus Second Order of Risk Aversion.” *Journal of Economic Theory*, 51, 111–125.
- [34] Sugden, Robert (2003). “Reference-Dependent Subjective Expected Utility.” *Journal of Economic Theory*, 111, 172–191.
- [35] Thaler, Richard, Amos Tversky, Daniel Kahneman, and Alan Schwartz (1997). “The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test.” *Quarterly Journal of Economics*, 112, 647–661.
- [36] Tversky, Amos, and Daniel Kahneman (1992). “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty*, 5, 297–323.
- [37] Yermack, David (1995). “Do Corporations Award CEO Stock Options Efficiently?” *Journal of Financial Economics*, 39, 237–269.

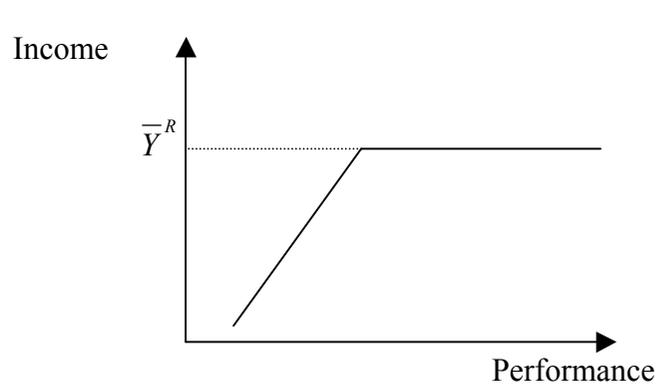


Figure 1a

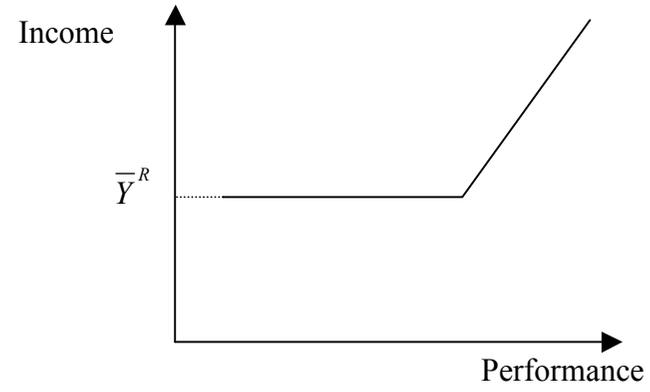


Figure 1b

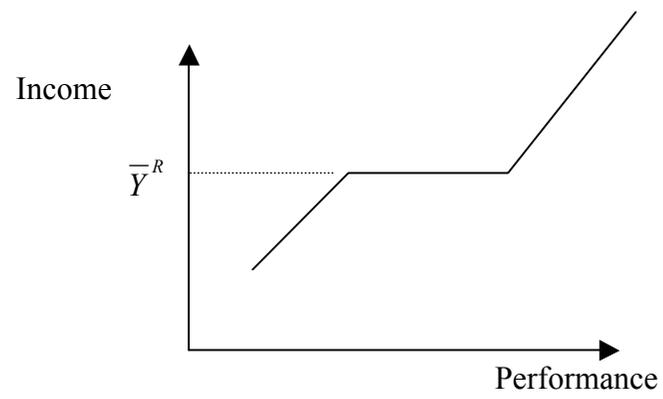


Figure 1c

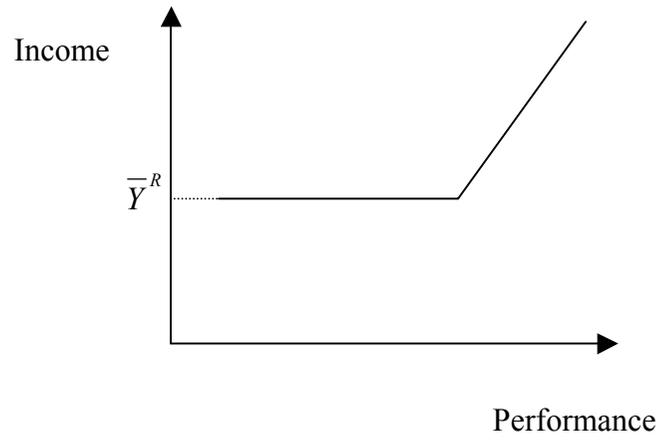


Figure 2a

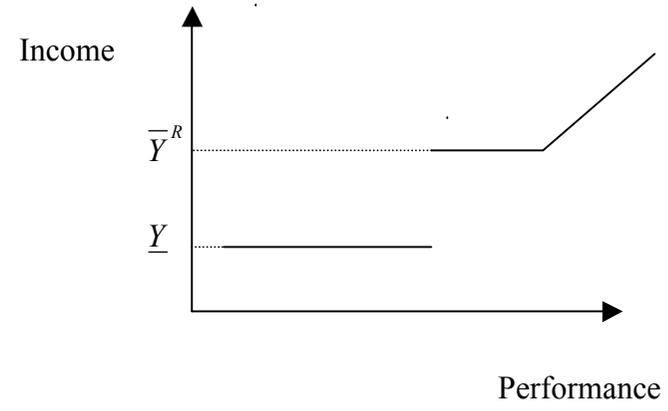


Figure 2b

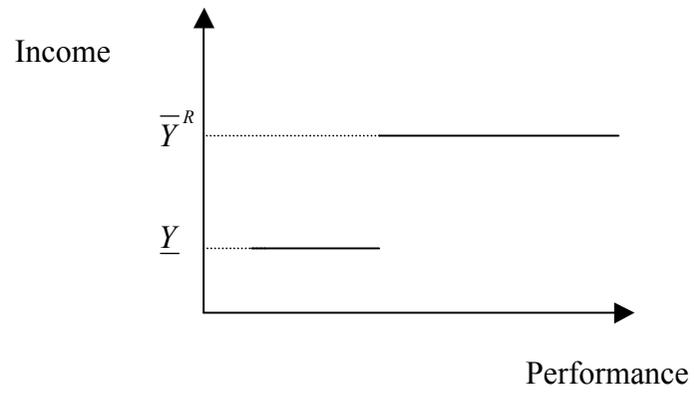


Figure 2c

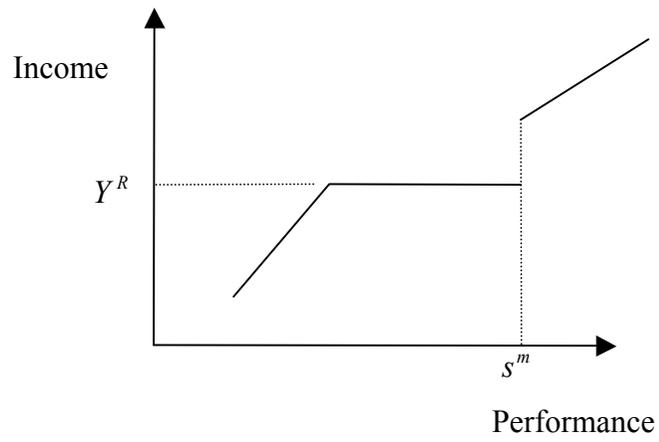


Figure 3a

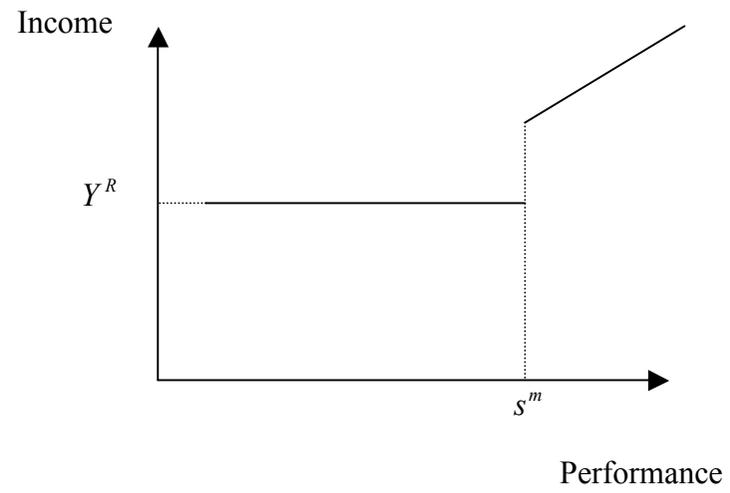


Figure 3b