Moral hazard with risk neutrality but limited liability

A project requires capital investment K and yields returns R if successful.
The probability of success depends on the agent's actions. If the agent behaves well then the probability of success is p_H, but if he misbehaves then the probability of success is p_L.
Let C denote the agent's collateral assets which could be seized in case of failure.
The agent's outside option pays w_0. Misbehavior yields hidden benefits worth B to the agent.
We assume that $p_L < p_H$, $p_H R > K > p_L R + B$, and $K > C$, so that the project is worthwhile only if agent behaves appropriately, but the agent cannot undertake the project alone.
Everyone is risk neutral, but the agent cannot pay more than his collateral C, so the principal's net wage cannot be less than $-C$.

We will consider two versions of this model. In version 1, losing C is the worst possible outcome for the agent. So the the principal chooses the agent's net wages x_S and x_F, for the cases of success and failure respectively. In version 2, we admit the possibility that, in case of failure, the principal couls may also subjected the agent to punishment that cost the agent $z \geq 0$ with no benefit to the principal. So in the extended version 2, the principal's problem is to choose (x_S, x_F, z) to

maximize $V = p_H(R-x_S) + (1-p_H)(-x_F) - K$ subject to

$p_H x_S + (1-p_H)(x_F - z) \geq w_0$ (participation constraint; λ),

$p_H x_S + (1-p_H)(x_F - z) \geq B + p_L x_S + (1-p_L)(x_F - z)$ (strategic incentive constraint; μ),

$x_F \geq -C$, $x_S \geq -C$, and $z \geq 0$ (limited liability).

The principal's problem in version 1 is the same except that drop the variable z (or require z = 0).

The principal's expected profit is $V = p_H R - K - [p_H x_S + (1-p_H)x_F] = p_H R - K - E\mathbf{x}$.
The participation constraint implies $E\mathbf{x} = p_H x_S + (1-p_H)x_F \geq w_0 + (1-p_H)z$.
The incentive constraint implies $(p_H - p_L)(z + x_S - x_F) \geq B$,
and so $E\mathbf{x} = p_H x_S + (1-p_H)x_F \geq Bp_H/(p_H - p_L) + x_F - p_H z$.
The principal wants to minimize his expected wage bill $E\mathbf{x} = p_H x_S + (1-p_H)x_F$, so it will be
$E\mathbf{x} = \text{maximum}\{w_0 + (1-p_H)z,\ Bp_H/(p_H - p_L) + x_F - p_H z\}$.
In any case, this formula can be minimized by making x_F as small as possible, so $x_F = -C$.

Now consider version 1, where punishment is not allowed, so z=0.
When $w_0 \geq Bp_H/(p_H - p_L) - C$, the participation constraint is binding,
and the optimal solution is $x_F = -C$, $x_S = [w_0 + (1-p_H)C]/p_H$, yielding $E\mathbf{x} = w_0$.
But when $w_0 < Bp_H/(p_H - p_L) - C$, the incentive constraint is binding,
and the optimal solution is $x_F = -C$, $x_S = B/(p_H - p_L) - C$, yielding $E\mathbf{x} = Bp_H/(p_H - p_L) - C$.

Now consider version 2, where punishment $z \geq 0$ is allowed.
When $C + w_0 \geq Bp_H/(p_H - p_L)$, the participation constraint is binding,
and the optimal solution remains $x_F = -C$, $x_S = [w_0 + (1-p_H)C]/p_H$, $z = 0$, yielding $E\mathbf{x} = w_0$.
But when $C + w_0 < Bp_H/(p_H - p_L)$, increasing z above 0 reduces the expected wage bill until we get $w_0 + (1-p_H)z = Bp_H/(p_H - p_L) - C - p_H z$, so that both constraints bind, and the optimal solution is
$x_F = -C$, $z = Bp_H/(p_H - p_L) - C - w_0$, $x_S = w_0 + (1-p_H)B/(p_H - p_L)$,
yielding $E\mathbf{x} = p_H w_0 + Bp_H(1-p_H)/(p_H - p_L) - (1-p_H)C$.

[Notation of LM 4.3: $\pi_1 = p_H$, $\pi_0 = p_L$, $\psi = B$, $\underline{S} = -K$, $\bar{S} = R - K$, $l = C$, $\bar{t} = x_S$, $\underline{t} = x_F$, $0 = w_0$.]

Becker-Stigler-Shapiro-Stiglitz efficiency wages

Consider a similar moral-hazard problem where the agent is risk neutral, and there are only two possible observations: y=1 denotes normal business, and y=0 denotes an accident occurring.
We consider a short interval of time ε, in which the probability of an accident is $\alpha\varepsilon$ if the agent chooses to be diligent a_H, and $\beta\varepsilon$ if the agent chooses to shirk a_L, where $\beta > \alpha$.
Choosing a_L also yields a hidden benefit of $D\varepsilon$ to the agent in this period.
Suppose that participation constraints apply ex-post: after the outcome is observed, the agent cannot be made worse off than his outside option of $\bar{v}$, which is the net present-discounted value of his lifetime income in the competitive labor market. The principal wants to minimize the expected cost subject to the ex-post participation constraints and the moral-hazard incentive constraint that the agent should not shirk. Let U_1 and U_0 denote the agent's expected total payoff after observing y=1 or y=0 respectively. So the principal's problem is

choose (U_1,U_0) to minimize $(1-\alpha\varepsilon)U_1 + \alpha\varepsilon U_0$ subject to

$(1-\alpha\varepsilon)U_1 + \alpha\varepsilon U_0 \geq D\varepsilon + (1-\beta\varepsilon)U_1 + \beta\varepsilon U_0$, $U_1 \geq \bar{u}$, $U_0 \geq \bar{u}$.

We could add an ex-ante participation constraint $(1-\alpha\varepsilon)U_1 + \alpha\varepsilon U_0 \geq \bar{w}$, but if the payoff $\bar{w}$ that must be promised to recruit the agent is the same as the payoff $\bar{u}$ that he can get by quitting later, then this constraint is redundant with the ex-post participation constraints $U_1 \geq \bar{u}$ and $U_0 \geq \bar{u}$.
The moral-hazard constraint implies $U_1 - U_0 \geq D/(\beta-\alpha)$.
So the optimal solution is $U_0 = \bar{u}$, $U_1 = \bar{u} + D/(\beta-\alpha)$.

So the agent's expected reward $(1-\alpha\varepsilon)U_1 + \alpha\varepsilon U_0 = \bar{u} + (1-\alpha\varepsilon)D/(\beta-\alpha)$ for a short ε-period of service must be greater than the outside option $\bar{u}$ by a positive bonus, even as $\varepsilon \to 0$. In a dynamic model, if this is repeated with different agents each ε-period then the principal's cost will be huge, but the cost can be reduced by using benefits of future employment be part of current incentive-pay.
Consider a stationary solution: if y=1 this period then the agent will be paid εw and rehired for next period, but if the y=0 this period then the agent is dismissed to the outside option $\bar{u}$.
So with discount rate r, we get the recursion equation: $U_1 = w\varepsilon + (1-r\varepsilon)[(1-\alpha\varepsilon)U_1 + \alpha\varepsilon\bar{u}]$.
With $U_1 = \bar{u} + D/(\beta-\alpha)$, this implies $w = r\bar{u} + (r + \alpha - \varepsilon r\alpha)D/(\beta-\alpha)$.
Let $w_0 = r\bar{u}$ be the outside wage rate corresponding to the net present value $\bar{u}$.
As $\varepsilon \to 0$, this optimal stationary plan in continuous time pays the agent an <u>efficiency wage</u> rate w that exceeds the competitive wage rate w_0 by $(r+\alpha)D/(\beta-\alpha)$, that is, $w = w_0 + (r+\alpha)D/(\beta-\alpha)$,
but the agent is dismissed when an accident occurs. The agent's expected discounted value of his future wages U_1 satisfies the continuous-time recursion equation $rU_1 = w + \alpha(\bar{u}-U_1)$,
and so (with $w_0=r\bar{u}$), we get $U_1 = \bar{u}+(w-w_0)/(r+\alpha) = \bar{u} + D/(\beta-\alpha)$.
The agent's expected net present value drops by $D/(\beta-\alpha)$ when he is fired after an accident.

As $\varepsilon \to 0$ this problem becomes <u>controlling a Poisson process</u> where accidents occur as a Poisson process with the low rate α when the agent is diligent, but the high rate β when the agent shirks.
When accidents occur in a <u>Poisson process with rate λ</u>, the number of accidents $\tilde{K}_\delta$ between a time t and time t+δ (δ>0) is a Poisson random variable with mean $\lambda\delta$, and it is independent of the number of accidents before time t. In this interval, $P(\tilde{K}_\delta=k) = e^{-\lambda\delta}(\lambda\delta)^k/k!$ for k=0,1,2,...
$E(\tilde{K}_\delta) = \lambda\delta = Var(\tilde{K}_\delta)$, $Stdev(\tilde{K}_\delta) = (\lambda\delta)^{0.5}$. When δ is small, $P(\tilde{K}_\delta=0) \approx 1-\lambda\delta$, $P(\tilde{K}_\delta=1) \approx \lambda\delta$.

Hidden efforts to control the mean of a Normal distribution

Suppose that an agent's hidden effort d controls the mean $\mu(d)$ of the principal's gross profit $\tilde{y}$, which are drawn from a Normal distribution with mean $\mu(d)$ and standard deviation σ.
We are assuming that the standard deviation does not depend on the effort, only the mean does.
Let C(d) denote the agent's net cost of choosing effort d.
The agent has constant risk tolerance T for wage income minus effort-cost, and his outside option would offer him net income worth w_0. The principal is risk neutral for gross profit less wage-costs.

Let us consider incentive plans where the agents wage w depends on the gross profit $\tilde{y}$ according to some linear formula of the form $w(\tilde{y}) = A + B\tilde{y}$.
If the agent chooses effort d, then his wage is a Normal random variable with mean $A+B\mu(d)$ and standard deviation $B\sigma$, and so his net certainty equivalent is $A+B\mu(d) - (0.5/T)(B\sigma)^2 - C(d)$.
The principal's net return is $\tilde{y}-(A+B\tilde{y})$, which has expected value $\mu(d)-(A+B\mu(d))$.

The principal wants to choose A and B and the recommended effort d so as to
maximize $E(\tilde{y} - w(\tilde{y})) = (1-B)\mu(d) - A$ subject to the constraints
$A + B\mu(d) - (0.5/T)(B\sigma)^2 - C(d) \geq w_0$ [participation]
$A + B\mu(d) - (0.5/T)(B\sigma)^2 - C(d) \geq A + B\mu(f) - (0.5/T)(B\sigma)^2 - C(f)$ for all f [moral-hazard].

The base wage A does not affect the moral-hazard constraint. Participation implies that it should be $A = w_0 + C(d) + (0.5/T)(B\sigma)^2 - B\mu(d)$. So in a plan that designate effort d and gives the agent B share of gross profits, the optimal base wage A makes the the principal's expected net profit equal to $(1-B)\mu(d) - [w_0 + C(d) + (0.5/T)(B\sigma)^2 - B\mu(d)] = \mu(d) - w_0 - C(d) - (0.5/T)(B\sigma)^2$.
For any given d, the principal wants to use the smallest B^2 that satisfies the moral hazard constraint.

Suppose that the agent has just two feasible effort levels d_L and d_H, which have hiddens costs $C(a_L) = c_L < c_H = C(a_H)$ and which yield expected gross profits $\mu(d_L) = m_L < m_H = \mu(d_H)$.
To motivate the high effort level d_H, the moral-hazard constraint is
$A + Bm_H - (0.5/T)(B\sigma)^2 - c_H \geq A + Bm_L - (0.5/T)(B\sigma)^2 - c_L$,
which implies $B \geq (c_H - c_L)/(m_H - m_L)$.
So the principal's optimal linear incentive plan to motivate effort d_H is $w(\tilde{y}) = A_H+B_H\tilde{y}$, where $B_H = (c_H - c_L)/(m_H - m_L)$ and $A_H = w_0 + c_H + (0.5/T)(B_H\sigma)^2 - B_Hm_H$.
The principal's expected net profit with the agent choosing d_H is $m_H-w_0-c_H-(.5/T)(B_H\sigma)^2$.

On the other hand, the principal can motivate the low effort d_L with $B_L=0$, because $c_L < c_H$.
So the principal's optimal linear incentive plan to motivate the low effort d_L is $w(\tilde{y}) = A_L = w_0+c_L$, and then the principal's expected net profit with the agent choosing d_L is $m_L-w_0-c_L$.

The principal should use the plan to motivate d_H if $m_H-w_0-c_H-(.5/T)(B_H\sigma)^2 \geq m_L-w_0-c_L$.
Otherwise, the principal should offer the optimal fixed wage w_0+c_L that motivates d_L.

What if we consider nonlinear incentive plans? In a simple one-period model where the agent chooses his effort once and then the gross profit $\tilde{y}$ is drawn from the resulting Normal distribution, linear wages are not optimal. The principal could do better by a nonlinear plan where the agent pays a very large penalty if gross profit is below very some cutoff. But Holmstrom and Milgrom have shown that linear wages are optimal in a dynamic model where the Normal gross profits are the sum of short-term daily profits over a long period where the agent's effort must be motivated every day.

Moral Hazard with constant risk tolerance, monetary effort cost, 2 actions

Now let's consider a variation where the agent's effort cost is a monetary cost to be subtracted from his income (instead of being a utility value to be subtracted from his utility as in LM Sect 4.5.2).

A risk-neutral principal is designing an incentive plan for a risk-averse agent.
The agent must choose among two unobservable actions: a_H and a_L.
Each action has a cost to the agent, $c(a_L) = c_L < c(a_H) = c_H$.
The agent could have earned w_0 elsewhere.
Suppose the agent's utility from choosing action a and getting wage w would be
$u(w-c(a)) = -\text{EXP}(-(w-c(a))/T)$, where $T>0$ is the agent's constant risk tolerance.
The principal can only observe an outcome $\tilde{x}$ that depends on the agent's action according to the conditional probability distribution $p(x|a_H)$ or $p(x|a_L)$.
Let X denote the set of possible values of $\tilde{x}$, which we assume here to be a finite set.
The principal can promise the agent a wage w(x) that depends on the observable outcome.
Suppose that the principal's expected payoff is much higher when the agent chooses a_H.
So the principal's problem is to design the wage-function w(•) to minimize the expected wage expense, subject to the constraints that the agent should not prefer the outside option $\bar{w}$ or the lower action a_L:

Choose $(w(x))_{x\in X}$ to minimize $\sum_{x\in X} p(x|a_H)\, w(x)$ subject to

$\sum_{x\in X} p(x|a_H)\, u(w(x)-c_H) \geq u(w_0)$ (participation constraint: λ),

$\sum_{x\in X} p(x|a_H)\, u(w(x)-c_H) \geq \sum_{x\in X} p(x|a_L)\, u(w(x)-c_L)$ (moral-hazard constraint: μ).

The participation constraint must be binding, or else the principal could reduce all w(x).
If the moral-hazard constraint were not binding, then the optimal solution would be $w(x)=w_0+c_H$ for all outcomes x, but then the agent would prefer the lower-cost action a_L (with $c_L<c_H$).
So both constraints must be binding at the optimal solution. Let λ and μ denote the Lagrange multipliers of the participation and moral-hazard constraints respectively. The Lagrangean is

$$\mathcal{L}(w;\lambda,\mu) = \sum_x p(x|a_H)\, w(x) - \lambda\big[\sum_x p(x|a_H)\, u(w(x)-c_H) - u(w_0)\big]$$
$$- \mu\big[\sum_x p(x|a_H)\, u(w(x)-c_H) - \sum_x p(x|a_L)\, u(w(x)-c_L)\big].$$

The optimality conditions $0=\partial\mathcal{L}/\partial w(y)$ yield

$\forall x\in X$: $\quad 0 = p(x|a_H) - (\lambda+\mu)\, p(x|a_H)\, u'(w(x)-c_H) + \mu\, p(x|a_L)\, u'(w(x)-c_L)$.

With constant risk tolerance, $u'(w-c) = \text{EXP}(-(w-c)/T)/T = -u(w-c)/T$.
Thus, summing the optimality conditions over all x in X, we get

$0 = 1 + (\lambda+\mu) \sum_x p(x|a_H)\, u(w(x)-c_H)/T - \mu \sum_x p(x|a_L)\, u(w(x)-c_L)/T$,

which with the binding constraints yields $0 = 1 + (\lambda+\mu)\, u(w_0)/T - \mu\, u(w_0)/T = 1 + \lambda\, u(w_0)/T$.

So the participation constraint's Lagrange multiplier is $\lambda = 1/u'(w_0) = -T/u(w_0) = T\times\text{EXP}(w_0/T)$.

With constant risk tolerance T, $u'(w(y)-c_L) = \eta\, u'(w(y)-c_H)$, where $\eta = \text{EXP}(-(c_H-c_L)/T)$.
So the optimality conditions become $0 = 1 - u'(w(x)-c_H)\big[\lambda + \mu - \mu\eta p(x|a_L)/p(x|a_H)\big]$.
Then the optimal wage w(y) can be determined from the equations:

$T\times\text{EXP}((w(x)-c_H)/T) = 1/u'(w(x)-c_H) = \lambda + \mu - \mu\, \eta\, p(x|a_L)/p(y|a_H)$.

Thus, the optimal wage w(y) is monotone decreasing in the likelihood ratio $p(x|a_L)/p(x|a_H)$.
With $\mu \leq \lambda/(\eta\, \max_{x\in X} p(x|a_L)/p(x|a_H) - 1)$, μ is determined by the requirement that the moral-hazard constraint must be satisfied as a binding equality.