Basic Facts about Risk Aversion

Consider an individual with a twice-continuously-differentiable utility function u(•).
Suppose this individual has wealth x plus a gamble that will pay a small random amount $\boldsymbol{\varepsilon}$,
such that $E(\boldsymbol{\varepsilon}) = 0$.
Let δ be the maximum that he would pay to insure against this gamble. So $u(x-\delta) = E(u(x+\boldsymbol{\varepsilon}))$.
Assuming that all possible values of $\boldsymbol{\varepsilon}$ are near 0, Taylor series approximations yield
$u(x) - \delta\, u'(x) \approx E[u(x) + u'(x)\, \boldsymbol{\varepsilon} + u''(x)\, \boldsymbol{\varepsilon}^2/2] = u(x) + u''(x)\, \mathrm{Var}(\boldsymbol{\varepsilon})/2$.
and then $\delta \approx -[u''(x)/u'(x)]\, \mathrm{Var}(\boldsymbol{\varepsilon})/2$.
<u>Fact.</u> There must exist some x_1 and x_2 in the (small) interval of possible values of $x+\boldsymbol{\varepsilon}$ such that
$u(x) - \delta u'(x_1) = u(x - \delta) = Eu(x+\boldsymbol{\varepsilon}) = u(x) + 0.5\, u''(x_2)\, \mathrm{Var}(\boldsymbol{\varepsilon})$,
so $\delta = -0.5\, \mathrm{Var}(\boldsymbol{\varepsilon})\, u''(x_2)/u'(x_1)$, where x_2 and x_1 are close to x when the range of $\boldsymbol{\varepsilon}$ is small.

The individual's <u>Arrow-Pratt risk-aversion</u> index at wealth x is $r(x) = -u''(x)/u'(x)$.
So we find that the individual's value for a small zero-expected-value gamble is approximately half of the variance of the gamble times the individual's risk-aversion index.
Notice that this risk-aversion index r(x) would not change if we changed how i's utility is measured to some other equivalent scale $\hat{u}(x) = Au(x)+B$ where $A>0$ and B are constants.
The reciprocal of the risk aversion index $\tau(x) = 1/r(x)$ is called the <u>risk tolerance</u> index,
which has the advantage of being measured in the same units as money (dollars).

An individual whose risk aversion is <u>constant</u>, independent of his given wealth x, must have a utility function that satisfies the differential equation $-u''(x)/u'(x) = R$, for some constant R.
This differential equation is equivalent to $d[LN(u'(x))]/dx = -R$.
For $R>0$, the solutions to this differential equation with $u'>0$ are $u(x) = B-Ae^{-Rx}$,
where $A>0$ and B are arbitary scale constants.

An individual whose risk tolerance is <u>proportional</u> to his wealth would have $r(x) = \alpha/x$ with $\alpha>0$,
so $u(x) = B+Ax^{1-\alpha}/(1-\alpha)$, or $u(x) = B+A\, LN(x)$ if $\alpha=1$.

Suppose an individual with constant risk aversion R ($u(x) = -e^{-Rx}$) will get a random income **Y** drawn from a Normal distribution with mean μ and standard deviation σ.
The <u>certainty equivalent</u> (CE) of this lottery **Y** for this individual is the sure amount of money W that individual would be willing to accept instead of this lottery **Y**.
Thus, the certainty equivalent W satisfies the equation $u(W) = E(u(\mathbf{Y}))$.
<u>Fact</u>: Suppose u(•) is a utility function with constant risk aversion, $\mathbf{Y_1}$ and $\mathbf{Y_2}$ are independent random variables, and $u(W_i) = E(u(\mathbf{Y_i}))$ for each i in {1,2}. Then $u(W_1+W_2) = E(u(\mathbf{Y_1}+\mathbf{Y_2}))$.
<u>Fact</u>: For an individual with constant risk aversion R (risk tolerance $\tau=1/R$), a Normal lottery with mean μ and standard deviation σ has certainty equivalent $W = \mu - 0.5R\sigma^2 = \mu - 0.5(\sigma/\tau)\sigma$.

That is, $-e^{-R(\mu-0.5R\sigma^2)} = \int_{-\infty}^{+\infty}\left(-e^{-Ry}\right)\frac{e^{-0.5((y-\mu)/\sigma)^2}}{\sqrt{2\pi}\,\sigma}dy$. (Proof: use $\int_{-\infty}^{+\infty}\frac{e^{-0.5((y-[\mu-R\sigma^2])/\sigma)^2}}{\sqrt{2\pi}\,\sigma}dy = 1$.)

(This can also be derived from the local approximation formula using the fact that a Normal random variable can be written as the limit of sums of many small independent random variables.)

Efficient Risk Sharing in a Syndicate

Let N denote the set of members of an investment partnership or syndicate.

They hold assets which will yield random returns **Y** that have some given probability distribution.

Each individual i in N has a given utility function for money $u_i(\bullet)$.

Let $x_i(y)$ denote the individual i's planned payoff when the syndicate earns **Y**=y.

To be feasible, we must have $\sum_{i\in N} x_i(y) = y,\ \forall y\in\mathbb{R}$.

Let us consider an efficient allocation rule $(x_i(\bullet))_{i\in N}$ that maximizes $\sum_{i\in N} \lambda_i\, E(u_i(x_i(\mathbf{Y})))$, subject to this feasibility constraint, for some given positive utility weights $(\lambda_i)_{i\in N}$.

To simplify the characterization of the efficient rule, we assume all appropriate differentiability.

For each outcome y, the marginal weighted utility $\lambda_i u_i'(x_i(y))$ must be equal across all i in N.

That is, there must exist some function $\nu(y)$ such that $\lambda_i\, u_i'(x_i(y)) = \nu(y),\ \forall i\in N,\ \forall y\in\mathbb{R}$ (1).

Differentiating with respect to y, we get $\lambda_i\, u_i''(x_i(y))\, \partial x_i/\partial y = \nu'(y)$ (2).

Equation (1) implies $\lambda_i = \nu(y)/u_i'(x_i(y))$.

Substituting this into (2), we get $[u_i''(x_i(y))/u_i'(x_i(y))]\, \partial x_i/\partial y = \nu'(y)/\nu(y)$.

The bracketted term here is just the Arrow-Pratt index of risk aversion times -1.

Let $\tau_i(x_i)$ denote the reciprocal of the Arrow-Pratt risk-aversion index, which is the risk tolerance of individual i when i gets payoff $x_i(y)$. That is $\tau_i(x_i) = -u_i'(x_i)/u_i''(x_i)$.

Then we get $\partial x_i/\partial y = -(\nu'(y)/\nu(y))\, \tau_i(x_i)$.

This derivative $\partial x_i/\partial y$ is i's marginal share of the variable risks held by the syndicate.

Feasibility implies $1 = \sum_{j\in N} \partial x_j/\partial y = -(\nu'(y)/\nu(y)) \sum_{j\in N} \tau_j(x_j(y))$.

Thus, we get $\partial x_i/\partial y = \tau_i(x_i(y))/[\sum_{j\in N} \tau_j(x_j(y))]$. (3)

That is, individual i's share of the syndicate's risks should be proportional to i's risk tolerance.

Constant absolute risk aversion. Suppose that each individual i has utility function $u_i(w) = -e^{-w/T(i)}$ for some given parameter T(i). Then $u_i'(w) = (e^{-w/T(i)})/T(i)$, $u''(w) = -(e^{-w/T(i)})/(T(i))^2$.

So $T(i) = -u_i'(w)/u_i''(w)$ is i's constant risk tolerance.

Let $T^* = \sum_{j\in N} T(j)$. Let $u^*(y) = -e^{-y/T^*}$ be a utility function with constant risk tolerance T*.

So in an efficient sharing rule: $\partial x_i/\partial y = T(i)/\sum_{j\in N} T(j) = T(i)/T^*,\ \forall y$,

and so $x_i(y) = x_i(0) + y\, T(i)/\sum_{j\in N} T(j) = x_i(0) + y\, T(i)/T^*$.

By (1), $\nu(y) = \lambda_i u_i'(x_i(y)) = [\lambda_i e^{-xi(0)/T(i)}/T(i)]e^{-y/T^*}$. (The factor in [] is independent of y and i.)

With this efficient sharing rule, each individual i gets the expected utility

$E[u_i(x_i(\mathbf{Y}))] = E[-e^{-(xi(0) + \mathbf{Y}T(i)/T^*)/T(i)}] = -u_i(x_i(0))\, E[u^*(\mathbf{Y})]$.

Let W_i and W^* be certainty equivalents satisfying $u_i(W_i) = Eu_i(x_i(\mathbf{Y}))$ and $u^*(W^*) = Eu^*(\mathbf{Y})$.

With exponential utilities, $u_i(W_i) = -u_i(x_i(0))\, u^*(W^*)$ implies $W_i = x_i(0) + (T(i)/T^*)W^*$.

So individuals who have constant risk tolerance should share risks linearly in proportion to their risk tolerances, and each wants their syndicate to evaluate risks according to a collective utility function u* that has a constant risk tolerance equal to the sum of their individual risk tolerances.

If all individuals have the same utility function with proportional risk tolerance $u_i(x) = x^{1-\alpha}/(1-\alpha)$ with $\alpha>0$ (or $u_i(x) = LN(x)$ for $\alpha=1$) then (1) implies $\lambda_i\,(x_i(y))^{-\alpha} = \nu(y)$ and so $x_i(y) = \lambda_i^{1/\alpha}\,\nu(y)^{-1/\alpha}$. Then $y = \sum_j x_j(y) = \nu(y)^{-1/\alpha} \sum_j (\lambda_j)^{1/\alpha}$, so $\nu(y)^{-1/\alpha} = y/[\sum_j (\lambda_j)^{1/\alpha}]$ and $\nu(y) = [\sum_j (\lambda_j)^{1/\alpha}]^\alpha\, y^{-\alpha}$. So each i must get a constant fraction of the total y: $x_i(y) = k_i\, y$, where $k_i = (\lambda_i)^{1/\alpha}/[\sum_j (\lambda_j)^{1/\alpha}]$.

Implementation with a price system

Suppose that **Y** has finitely many possible values y, each with some probability p(y).
The marginal utility equation (1) $\lambda_i\, u_i'(x_i(y)) = \nu(y),\ \forall i\in N,\ \forall y\in\mathbb{R}$,
implies that the efficient allocation can be implemented by a price system,
in which each partner i has a budget I_i to spend on acquiring claims on income in each state of the total returns **Y**, and $\pi(y)$ is the price of one unit of income when **Y**=y.
In such a price system, i would choose the quantities $x_i(y)$ for all possible y to maximize
$\sum_y p(y)u_i(x_i(y))$ subject to $\sum_y \pi(y)x_i(y) \le I_i$.
Let μ_i be the Lagrange multiplier of i's budget constraint (i's marginal utility of budgeted funds).
At an optimal solution, we have $p(y)u_i'(x_i(y)) = \mu_i\pi(y),\ \forall i,\ \forall y$.
This coincides with the efficient marginal utility equation (1) iff
$u_i'(x_i(y)) = \mu_i\pi(y)/p(y)= \nu(y)/\lambda_i,\ \forall i\ \forall y$.
Thus $\mu_i\lambda_i = p(y)\nu(y)/\pi(y)$ must be some constant A that does not depend on i or y.
That is $\pi(y) = p(y)\nu(y)/A$.
This price system allows a risk-free rate of return on budgeted funds $1+\rho = 1/[\sum_y \pi(y)]$
So in terms of the risk-free rate ρ, we must have $(1+\rho)[\sum_y p(y)\nu(y)] = A$,
and so $\pi(y) = p(y)\nu(y)/[(1+\rho) \sum_z p(z)\nu(z)]\ \forall y$.
The price of future income from returns y depends both on the probability p(y) and on $\nu(y)$.
$\nu(y)$ measures the marginal social utility of future returns y, and it is generally decreasing in y, because it is the Lagrange multiplier of the total-returns constraint in the problem:
choose $(x_i)_{i\in N}$ to maximize $\sum_i \lambda_i u_i(x_i)$ subject to $\sum_i x_i \le y$.

References: Robert Wilson, *Econometrica* (1968); Karl Borch, *Econometrica* (1962).

Optimal risk sharing among partners with constant risk tolerance
Consider a group of individuals who have formed a partnership to share the risky profits from some joint venture or gamble Suppose each individual j in this group has a constant risk tolerance, denoted by τ_j. Let T_o denote the sum of all the partners' risk tolerances ($T_o = \sum_j \tau_j$).

These partners can maximize the sum of their certainty equivalents by sharing the risky profits among themselves in proportion to their risk tolerances, with each individual j taking the fractional share τ_j/T_o of the risky profits. The maximal sum of the partners' certainty equivalents that can be achieved by such efficient risk sharing is equal to the certainty equivalent of the whole gamble to an individual who has a constant risk tolerance equal to T_o, the sum of these partners' risk tolerances. Thus, in making decisions about risky investments, the partnership should act as a corporate person with a risk tolerance equal to the sum of its members' risk tolerances.

(Coase theorem) If the partners were planning to share risks according to a sharing rule that does not maximize the sum of the partners' certainty equivalents, then any partner j could propose another sharing rule rule that would increase j's own certainty equivalent and would not decrease the certainty equivalents of any other partners.

Risk-sharing example. Consider a financial asset yielding a return $\tilde{Y}$ that is drawn from a Normal distribution with mean μ = \$35000, σ = \$25000. Investor 1 has risk tolerance $T_1 = 20000$, and investor 2 has risk tolerance $T_2 = \$30000$.
When 1 owns the whole asset, his certainty equivalent is
$CE_1(\tilde{Y}) = 35000-(0.5/20000)*25000^2 = 35000-15625 = \$19375$.
If 2 owned the whole asset, her certainty equivalent would be
$CE_2(\tilde{Y}) = 35000-(0.5/30000)*25000^2 = 35000-10417 = \$24583$
So if 1 sold the asset to 2 for price x then they would get $CE_1(x) = x$, $CE_2(\tilde{Y}-x) = \$24583 - x$.
The transaction makes both better off if $19375 \le x \le 24583$.
In this range, x=24583 would be best for investor 1.
Now consider the possibility that 1 could sell a 50% share to 2 for some price x. Then they get
$CE_1(0.50\tilde{Y}+x) = 0.50*35000 + x - (0.5/20000)*(0.50*25000)^2 = 17500+x - 3906 = \$13594 + x$,
$CE_2(0.50\tilde{Y}-x) = 0.50*35000 - (0.5/30000)*(0.50*25000)^2 = 17500 - x - 2604 = \$14896 - x$.
This transaction makes both better off if $19375-13594 = 5781 \le x \le 14896$.
The sum of their certainty equivalents is 13594+14896 = \$28490
which is higher than either could get from the asset alone.
When the price x is 14896, investor 1 gets all of this certainty equivalent value.
In general, if 1 sells a share θ to investor 2 for a price x, their certainty equivalents are
$CE_1((1-\theta)\tilde{Y}+x) = (1-\theta)*35000 + x - (0.5/20000)*[(1-\theta)*25000]^2$
$CE_2(\theta\tilde{Y}-x) = \theta*35000 - x - (0.5/30000)*(\theta*25000)^2$
The sum of these is maximal when $(1-\theta)^2/T_1 + \theta^2/T_2$ is minimized, at $\theta = T_2/(T_1+T_2) = 0.6$.
When 1 sells the optimal 60% share to investor 2 for a price x, they get
$CE_1(0.40*\tilde{Y}+x) = 0.40*35000 + x - (0.5/20000)*(0.40*25000)^2 =$
$= 14000 + x - 2500 = \$11500 + x$,
$CE_2(0.60*\tilde{Y}-x) = 0.60*35000 - (0.5/30000)*(0.60*25000)^2 =$
$= 21000 - x - 3750 = \$17250 - x$.
This transaction makes both better off if $19375-11500 = 7875 \le x \le 17250$.
The maximized sum of their certainty equivalents is 11500+17250 = \$28750.
When the price x is 17250, investor 1 gets this maximal certainty-equivalent value.

For a fictitious corporate person with risk tolerance $T_o = T_1+T_2 = 50000$, the asset $\tilde{Y}$ would be worth $CE_o(\tilde{Y}) = 35000-(0.5/50000)*25000^2 = 35000-6250 = \$28750$.

Moral hazard with a linear wage rule, Normal risks, and constant risk tolerance.

Suppose that a firm's net profits will be a random variable $\mathbf{y}$ drawn from a Normal distribution with a variance σ^2 but with a mean that depends on the effort of an agent.

If the agent's effort is high, then the agent's effort cost is c_H and profit has mean $E(\mathbf{y}|c_H) = m_H$.

If the agent's effort is low, then the agent's effort cost is c_L and profit has mean $E(\mathbf{y}|c_L) = m_L$.

Suppose $m_H > m_L$ and $c_H > c_L$.

Nobody else can observe the agent's effort, but his wage can depend on the observable profit $\mathbf{y}$. For now, let us assume that firm must specify the agent's wage as a linear function of observed profits, according to any linear formula $w(\mathbf{y}) = \alpha+\beta\mathbf{y}$.

(This assumption is questionable. We will see soon that the firm might prefer a nonlinear formula. But let's start with this linearity assumption because linearity makes things simpler.)

Suppose that the agent has a constant risk tolerance T, but the owners of the firm are risk neutral (meaning that they have linear utility for money or infinite risk tolerance).

The agent's best outside option is to earn $\bar{w}$ elsewhere with zero effort costs.

What linear wage rule (α,β) is best for the firm if the firm wants the agent to choose high effort?

Under an (α,β) linear wage rule, the agent's net certainty equivalent with high effort is
$E(\alpha+\beta\mathbf{y}|c_H) - (0.5/T)\text{Var}(\alpha+\beta\mathbf{y}|c_H) - c_H= \alpha + \beta m_H - (0.5/T)(\beta\sigma)^2 - c_H$.

By choosing his outside option, the agent could instead get certainty equivalent $\bar{w}$.

With low effort, the agent could get the certainty equivalent $\alpha+\beta m_L-(0.5/T)(\beta\sigma)^2-c_L$.

So for the agent to want to work here and exert high effort, (α,β) must satisfy the participation constraint $\alpha + \beta m_H - (0.5/T)(\beta\sigma)^2 - c_H \geq \bar{w}$, and the moral-hazard incentive constraint $\alpha+\beta m_H-(0.5/T)(\beta\sigma)^2-c_H \geq \alpha+\beta m_L-(0.5/T)(\beta\sigma)^2-c_L$.

These constraints imply $\alpha + \beta m_H \geq \bar{w} + c_H + (0.5/T)(\beta\sigma)^2$ and $\beta \geq (c_H-c_L)/(m_H-m_L)$.

The risk-neutral owners of the firm want to minimize their expected wage bill $\alpha+\beta m_L$.

So the optimal linear rule should choose α (given any β) to make the participation constraint bind, and should choose β as small as possible to make the moral-hazard constraint bind.

That is, the optimal rule is $\beta = (c_H-c_L)/(m_H-m_L)$ and $\alpha = \bar{w} + c_H + (0.5/T)(\beta\sigma)^2 - \beta m_H$.

The firm's minimal expected wage bill is $\bar{w} + c_H + (0.5/T)(\beta\sigma)^2$, and the owner's expected net profit is thus $m_H - \bar{w} - c_H - (0.5/T)(\beta\sigma)^2$.

Given that the agent is risk averse and the owners are risk neutral, the optimal sharing rule without moral hazard would have $\beta=0$ (and so $\alpha=\bar{w}+c_H$), if the agent could be forced to choose high effort. So moral hazard provides a fundamental reason why risk-averse agents cannot insure away their professional risks: because their share of such risks is essential to motivate their efforts.

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

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 earn $\bar{w}$ 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.
[This model differs from standard textbook models in that effort cost here is monetary, subtracted from income before applying the utility function: $u(w-c(a))$ instead of $u(w)-c(a)$.]
The principal can only observe an outcome **y** that is a random variable which depends on the agent's action according to the conditional probability distribution $p(y|a_H)$ or $p(y|a_L)$.
Let Y denote the set of possible values of **y**, which we assume here to be a finite set.
The principal can promise the agent a wage $w(y)$ 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 $w(\bullet)$ 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 $(w(y))_{y\in Y}$ to minimize $\sum_{y\in Y} p(y|a_H)\, w(y)$ subject to
$\sum_{y\in Y} p(y|a_H)\, u(w(y)-c(a_H)) \geq u(w_o)$ (participation constraint: λ),
$\sum_{y\in Y} p(y|a_H)\, u(w(y)-c(a_H)) \geq \sum_{y\in Y} p(y|a_L)\, u(w(y)-c(a_L))$ (moral-hazard constraint: μ).

The participation constraint must be binding, or else the principal could reduce all $w(y)$.
If the moral-hazard constraint were not binding, then the optimal solution would be $w(y)=\bar{w}+c_H$ for all outcomes y, but then the agent would prefer the lower-cost action a_L (with $c(a_L)<c(a_H)$).
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

$$\mathcal{L}(w;\lambda,\mu) = \sum_y p(y|a_H)\, w(y) - \lambda\Big[\sum_y p(y|a_H)\, u(w(y)-c_H) - u(\bar{w})\Big]$$
$$- \mu\Big[\sum_y p(y|a_H)\, u(w(y)-c_H) - \sum_y p(y|a_L)\, u(w(y)-c_L)\Big].$$

The optimality conditions $0 = \partial L/\partial w(y)$ yield
$\forall y\in Y$: $\quad 0 = p(y|a_H) - (\lambda+\mu)\, p(y|a_H)\, u'(w(y)-c_H) + \mu\, p(y|a_L)\, u'(w(y)-c_L)$.
With constant risk tolerance, $u'(w-c) = \text{EXP}(-(w-c)/T)/T = -u(w-c)/T$.
Thus, summing the optimality conditions over all y in Y, we get
$0 = 1 + (\lambda+\mu) \sum_y p(y|a_H)\, u(w(y)-c_H)/T - \mu \sum_y p(y|a_L)\, u(w(y)-c_L)/T$,
which with the binding constraints yields $0 = 1 + (\lambda+\mu)\, u(\bar{w})/T - \mu\, u(\bar{w})/T = 1 + \lambda\, u(\bar{w})/T$.
So the participation constraint's Lagrange multiplier is $\lambda = 1/u'(\bar{w}) = -T/u(\bar{w}) = T\times\text{EXP}(\bar{w}/T)$.

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 $0 = 1 - u'(w(y)-c_H)\big[\lambda + \mu - \mu\, \eta\, p(y|a_L)/p(y|a_H)\big]$, $\forall y\in Y$.
Then the optimal wage $w(y)$ can be determined from the equations:
$T\times\text{EXP}((w(y)-c_H)/T) = 1/u'(w(y)-c_H) = \lambda + \mu - \mu\, \eta\, p(y|a_L)/p(y|a_H)$, $\forall y\in Y$.
Thus, the optimal wage $w(y)$ is monotone decreasing in the likelihood ratio $p(y|a_L)/p(y|a_H)$.
With $\mu \leq \lambda/(\eta \max_{y\in Y} p(y|a_L)/p(y|a_H) - 1)$, μ is determined by the requirement that the moral-hazard constraint must be satisfied as a binding equality.

Credit rationing in a binary moral-hazard model with limited liablity (Stiglitz-Weiss).

An entrepreneur can undertake a project which requires an amount I to be invested this period.
Next period, the project will return the amount RI if it succeeds, or 0 if it fails.
The entrepreneur must borrow the amount I but can offer collateral worth C for the loan.
If the entrepreneur manages the investment well, then its probability of success will be p_H.
But the entrepreneur could act badly, divert an amount γI, and reduce the probability of success to p_L.
This reduction in the probability of success is the only observable consequence of such malfeasance.
The entrepreneur is risk neutral but has limited liability in the sense that he cannot lose more than C.
Let w_0 denote the entrepreneur's alternative wage next period if he does not manage this project,
Let ρ denote the market rate of interest per period for risk free investments elsewhere.
Suppose that $p_H R > 1+\rho > p_L R + \gamma$, so that the project could be worthwhile for risk-neutral investors if the entrepreneur manages it well, but not if he acts badly.

If the entrepreneur is charged an interest rate r to borrow I, then the entrepreneur will get the surplus $(R-(1+r))I$ if the project is a success, but he will default and lose the collateral C if the project fails.
Including the value of defaulted collateral, the investors' expected return is $p_H(1+r)I + (1-p_H)C$.
Per unit invested, the investors' rate of return is $p_H(1+r) + (1-p_H)C/I$.
So the investors' participation constraint is $p_H(1+r) + (1-p_H)C/I \geq 1+\rho$,
which holds iff $1+r \geq (1+\rho)/p_H - (C/I)(1-p_H)/p_H$.
The entrepreneur's participation constraint is $p_H(R-(1+r))I - (1-p_H)C \geq w_0$.
The entrepreneur's participation constraint implies $p_H R - w_0/I \geq p_H(1+r) + (1-p_H)C/I$.
The moral-hazard constraint to deter entrepreneurial malfeasance is
$p_H(R-(1+r))I - (1-p_H)C \geq \gamma I + p_L(R-(1+r))I - (1-p_L)C$.
The moral-hazard constraint yields a lower bound on an agent's stake $C + (R-(1+r))I \geq \gamma I/(p_H-p_L)$,
and an upper bound on investors' rate of return $C/I + p_H R - p_H\gamma/(p_H-p_L) \geq p_H(1+r) + (1-p_H)C/I$.

Suppose that C and w_0 are small so that $C + w_0 < p_H\gamma I/(p_H-p_L)$. [poor agent case]
This poor-agent condition implies that $C/I + p_H R - p_H\gamma/(p_H-p_L) < p_H R - w_0/I$.
Then the upper bound on investors' rate of return is determined by the moral-hazard constraint:
$p_H(1+r) + (1-p_H)C/I \leq C/I + p_H R - p_H\gamma/(p_H-p_L)$.
The maximal interest rate r* that can avoid moral hazard is $1+ r^* = R + C/I - \gamma/(p_H-p_L)$.
Suppose also that R satisfies $R \geq (1+\rho)/p_H + \gamma/(p_H-p_L) - C/(p_H I)$.
Then the investors are willing to lend at this rate r*, because
$p_H(1+r^*)I + (1-p_H)C = p_H[R + C/I - \gamma/(p_H-p_L)]I + (1-p_H)C \geq (1+\rho)I$.
Even at this maximal interest rate, however, the entrepreneur's expected gain is strictly positive:
$p_H(R-(1+r^*))I - (1-p_H)C - w_0 = p_H\gamma/(p_H-p_L) - C - w_0 > 0$.
The quantity $p_H\gamma/(p_H-p_L)$ here may be called the expected moral-hazard rent that the entrepreneur must be allowed in this financial transaction, to deter his malfeasance.

Now suppose that there are many such entrepreneurs, but only a limited supply of funds from investors who understand these projects well enough to enter into such financial deals.
Then the interest rate for such loans can rise only to the maximal rate r*. If the demand from entrepreneurs at this rate exceeds the supply of funds, there must be credit rationing.

Long-term agency relationships

Let us consider a version of the above model, but now the projects are available to firms which must hire agents to manage them. As before, R is the return per unit invested if the project succeeds, p_H is the probability of success if the manager acts appropriately, but the manager could instead divert a fraction γ of the invest funds and reduce the probability of success to p_L. Here $p_L < p_H$.
Agents' alternative wages and collateral are negligible relative to these investments (so $w_0 = 0$, C=0).
An agent can manage a project of various investment sizes I, with an upper bound of $I \leq 1$ (billion\$) (and with some much smaller positive lower bound that is still large relative to wages and collateral).
If firms can profitably hire agents to manage such projects, then many firms will do so, driving down the prices of the goods which are the outputs of these investment projects.
So (assuming such a general equilibrium) let us consider the optimal agency contracts for these firms when R is near the lower bound where such contracts are just barely profitable.
Let us ignore interest over the periods of the project (ρ=0).

Assuming limited liability for the agents, the terms of an agent's contract specify (I,w_S,w_F), where I is the investment size handled by the agent, w_S is his wage if success, w_F is his wage if failure.
Assuming risk-neutrality, the agent's participation constraint is $p_Hw_S + (1-p_H)w_F \geq 0$,
the moral-hazard constraint is $p_Hw_S + (1-p_H)w_F \geq \gamma I + p_Lw_S + (1-p_L)w_F$,
and the limited-liability and investment bound constraints are $w_S \geq 0$, $w_F \geq 0$, $0 \leq I \leq 1$.
(The actual lower bound on I may be positive, but we allow I=0 to represent not hiring the agent.)
The firm wants to maximize its expected profit $(p_HR-1)I - [p_Hw_S + (1-p_H)w_F]$.
The optimal solution has $w_F = 0$, $w_S = \gamma I/(p_H-p_L)$, giving the firm $[p_HR - 1 - p_H\gamma/(p_H-p_L)]I$, and so the investment I = 1 is worthwhile as long as $R \geq 1/p_H + \gamma/(p_H-p_L)$.
Let $G = \gamma/(p_H-p_L)$ denote the moral-hazard rent per unit invested which a successful agent must get, and let $g = p_HG = p_H\gamma/(p_H-p_L)$ denote the expected moral-hazard rent per unit invested.
So with competitive firms, the output-market equilibrium would yield R at or near $R^* = 1/p_H + G$.

To verify the optimality of the solution, notice that the optimization is a linear programming problem.
The solution is supported by the following Lagrange multipliers (dual variables):
λ=0 for the agent's participation constraint, and $\mu = p_H/(p_H-p_L)$ for the moral-hazard constraint.
These multipliers yield $0 = \partial\mathcal{L}/\partial w_S = -p_H + \lambda p_H + (p_H-p_L)\mu$,
$0 > \partial\mathcal{L}/\partial w_F = -(1-p_H) + \lambda(1-p_H) - (p_H-p_L)\mu$,
and $\partial\mathcal{L}/\partial I = p_HR - 1 - \mu\gamma \geq 0$ when $R \geq 1/p_H + G$.
A surplus G is required here to cover the cost of the agent's moral-hazard rent.

Now suppose that an agent can manage one such project in each of T periods in his career, and the agent's responsibilities can depend on his past history of successes or failures.
We will now show that the lowest R at which firms can expect nonnegative profits from such multiperiod agency relationships is $R^* = 1/p_H + G/T$.
Intuitively, the cost of an agent's moral-hazard rent G can now be divided over T periods.

The optimal policy at each period t can be characterized in terms of a recursive optimization problem with two new parameters: v_t represents the expected monetary rewards that were promised to the agent at the end of the previous period, and λ_{t+1} denotes the firm's future net marginal cost of fulfilling promises of future monetary rewards. Then the firm's problem at period t is

choose $(I_t,w_{S,t},w_{F,t},x_t)$ to maximize $(p_HR-1)I_t - \lambda_{t+1}[p_Hw_{S,t} + (1-p_H)w_{F,t}] - x_t + \lambda_tv_t$ (net profit)
subject to $x_t + p_Hw_{S,t} + (1-p_H)w_{F,t} \geq v_t$ (λ_t, promise-keeping)
$p_Hw_{S,t} + (1-p_H)w_{F,t} - [\gamma I + p_Lw_{S,t} + (1-p_L)w_{F,t}] \geq 0$ (μ_t, moral hazard)
$x_t \geq 0$, $w_{S,t} \geq 0$, $w_{F,t} \geq 0$, $0 \leq I_t \leq 1$. (bounds)

Here x_t represents the firm's option of paying its past promises in cash at the start of period t.
For the full multi-period problem, we must add the following conditions: the agent's initial credit is $v_1 = 0$ because the firm has no past obligations to a new agent, each later v_t is either $w_{S,t-1}$ or $w_{F,t-1}$ depending on whether the previous period was a success or failure, the final marginal wage-cost is $\lambda_{T+1} = 1$ because rewards to a retiring agent must be paid in cash, and each earlier period derives its marginal cost of end-of-period wage promises from the Lagrange multiplier of next period's promise-keeping constraint. (The last term in the objective λ_tv_t is an accounting constant, not affected by current decisions, to let us sum profits across periods without double-counting back-loaded wages.)

We will now show that, when $R = R^* = 1/p_H + G/T$, the following dynamic policy is optimal but yields zero expected profits for the firm, and so firms cannot make expected profits at any lower R.
In the last period, if the agent has managed successful projects in all previous periods, then the agent manages the maximum feasible investment $I_T = 1$ and is paid $w_{S,T} = I_TG = G$ if it succeeds.
The possibility of this large final payoff motivates the agent at all previous periods.
At each period t, if the agent has never failed in the past then he manages a project with $I_t = (p_H)^{T-t}$.
Then his promised reward for success at period t will be $w_{S,t} = I_tG = (p_H)^{T-t}G$, which is just the expected value of his potential final reward (G times the probability of T−t future successes).
If the agent fails at any period t, he gets $w_{F,t} = 0$ and does not manage any more projects.
The backloading of all rewards is indicated in the solution by $x_t = 0$ $\forall t$.

Under this plan, the agent invests $I_t = (p_H)^{T-t}$ at time t only if he has never failed before, which has probability $(p_H)^{t-1}$. His expected investment is $(p_H)^{t-1}(p_H)^{T-t} = (p_H)^{T-1}$ at each period t, and his expected wage bill is $(p_H)^TG$, so the firm's expected net profit is $T(p_H)^{T-1}(p_HR-1)-(p_H)^TG$.

So at any t>1, as long as the agent has been successful, we get $v_t = w_{S,t-1} = (p_H)^{T+1-t}G$,
and so the promise-keeping constraint is satisfied with equality by $v_t = p_Hw_{S,t} = p_H\,(p_H)^{T-t}G$.
The moral-hazard constraint is also binding at every t, because $w_{S,t} = I_tG= I_t\gamma/(p_H-p_L)$.
To support this as an optimal solution in which investment is positive but the firm's expected net profits are zero, we need Lagrange multipliers $\mu_t \geq 0$ and $\lambda_t \geq 0$ to satisfy the equations
$0 = \partial\mathcal{L}/\partial I_t = p_HR - 1 - \mu_t\gamma$ and $0 = \partial\mathcal{L}/\partial w_{S,t} = -\lambda_{t+1}p_H + \lambda_tp_H + (p_H-p_L)\mu_t$, for all $t\in\{1,...,T\}$,
and we need $\lambda_1 = 0$ because promise-keeping has slack at t=1 when the agent starts with v_1=0.

For $0 = \partial\mathcal{L}/\partial I_t = p_HR-1-\mu_t\gamma$, at all t we must have $\mu_t = (p_HR-1)/\gamma$.
Then $0=\partial\mathcal{L}/\partial w_{S,t} = \lambda_tp_H +(p_H-p_L)\mu_t - \lambda_{t+1}p_H$ implies $\lambda_t = \lambda_{t+1}-\mu_t(p_H-p_L)/p_H = \lambda_{t+1} - (p_HR-1)/g$.
(Recall $g = p_HG = p_H\gamma/(p_H-p_L)$.) Given λ_{T+1}=1, we get $\lambda_t = 1-(T+1-t)(p_HR-1)/g$, $\forall t$.
Then $\lambda_1 = 0$ is satisfied iff $1 = T(p_HR-1)/g$, which holds when $R = (1+g/T)/p_H = 1/p_H + G/T$.
$\partial\mathcal{L}/\partial x_t = \lambda_t - 1 < 0$ and $\partial\mathcal{L}/\partial w_{F,t} = -(\lambda_{t+1}-\lambda_t)(1-p_H)-(p_H-p_L)\mu_t < 0$ imply x_t=0=$w_{F,t}$ is optimal.

If R is slightly greater than this R*, this plan can still satisfy all recursive optimality conditions with one change: at T, with I_T at the upper bound I_T=1, we can have $\partial\mathcal{L}/\partial I_T$ >0 and $\mu_T < (p_HR-1)/\gamma$.
Keeping other μ_t fixed, we get $\lambda_t = 1 - \mu_T(p_H-p_L)/p_H - (T-t)(p_HR-1)/g$ $\forall t$, and we can get λ_1=0 with $\mu_T = [1-(T-1)(p_HR-1)/g]p_H/(p_H-p_L)$. This fails only when a larger R makes μ_T negative.

This model describes a world where productive investments are only possible when agents have long-term relationships with firms which the agents can trust to reliably pay deferred rewards for service. The temptation to deny long-promised wages by finding fault with the agent's late-career performance may become particularly acute when $R^* < G$ (which can happen when p_H, γ, and T are large).
So this economy relies heavily on firms' reputations for judging and rewarding their agents.

The large moral-hazard rents are also a striking feature of this model. But one might ask whether these are just a result of the limited-liability assumption, which puts a lower bound on agents' payoffs. So let us see how the solution would change if the limited liability constraint were relaxed by allowing **punishment in case of failure**. A punishment $z \geq 0$ would reduce the agent's utility to $-z$ but without generating any benefit anyone else. (This is different from taking valuable collateral.)

The recursive problem at any period t is now
choose $(I_t, w_{S,t}, w_{F,t}, z_t)$ to maximize $(p_H R - 1)I_t - \lambda_{t+1}[p_H w_{S,t} + (1-p_H)w_{F,t}] + \lambda_t v_t$ (net profit)
subject to $p_H w_{S,t} + (1-p_H)(w_{F,t} - z_t) \geq v_t$ (λ_t, participation)
$p_H w_{S,t} + (1-p_H)(w_{F,t} - z_t) \geq \gamma I_t + p_L w_{S,t} + (1-p_L)(w_{F,t} - z_t)$ (μ_t, moral hazard)
$w_{S,t} \geq 0$, $w_{F,t} \geq 0$, $z_t \geq 0$, $0 \leq I_t \leq 1$. (bounds)
with $\lambda_{T+1} = 1$, $v_1 = 0$, and $v_{t+1} = w_{S,t}$ for agents who succeed through time t.

As before, we consider parametric cases where R is at or near the lower bound at which firms can make nonnegative expected profit from agents who could manage a project in up to T periods. Expanding the contract options to include punishment decreases the lowest R at which firms can profitably invest down to $R^{**} = 1/p_H + G/[T + p_H/(1-p_H)]$.
We show now that, at or near this R**, the optimal contract involves no punishment after period 1.

As above, consider a plan in which the agent is paid a moral-hazard rent $w_{S,T} = G$ at period T only if he manages successful projects at all previous periods. The prospect of this moral-hazard rent yields expected rewards worth $w_{S,t} = G(p_H)^{T-t}$ to the agent when he has been successful through period t.
We always have $w_{F,t}=0$. At any time $t>1$, the agent can manage $I_t = w_{S,t}/G = (p_H)^{T-t}$ with $z_t = 0$ (no punishment), and the promise-keeping constraint is also binding with $v_t = w_{S,t-1} = (p_H)^{T+1-t}$.
With $v_1=0$, the promise-keeping/participation constraint would have slack at t=1 if z_1 were zero, but we can make it bind with $z_1 = w_{S,1}p_H/(1-p_H) = G(p_H)^T/(1-p_H)$ Then the moral-hazard constraint will be binding when $\gamma I_1 = (p_H - p_L)(w_{S,1}+z_1) = (p_H - p_L)G(p_H)^{T-1}/(1-p_H)$, so $I_1 = (p_H)^{T-1}/(1-p_H)$.

The optimality conditions are almost the same as above. The only changes are that λ_1 can be nonzero but now, for $z_1>0$, we need $0 = \partial\mathcal{L}/\partial z_1 = \mu_1(p_H - p_L) - \lambda_1(1-p_H)$.
Also, to justify $z_t=0$ for all $t>1$, we need $0 \geq \partial\mathcal{L}/\partial z_t = \mu_t(p_H - p_L) - \lambda_t(1-p_H)$.
From $0=\partial\mathcal{L}/\partial I_t$ in the zero-expected-profit solution at R=R**, we again get $\mu_t = (p_H R - 1)/\gamma$ $\forall t$.
From $0=\partial\mathcal{L}/\partial z_1 = \mu_1(p_H - p_L) - \lambda_1(1-p_H)$, we get $\lambda_1 = \mu_1(p_H - p_L)/(1-p_H)$.
As before, $0 = \partial\mathcal{L}/\partial w_{S,t} = -\lambda_{t+1}p_H + \lambda_t p_H + (p_H - p_L)\mu_t$ implies $\lambda_{t+1} = \lambda_t + \mu_t(p_H - p_L)/p_H$ at all t, and so $\lambda t = [p_H/(1-p_H) + t-1]\mu_t(p_H - p_L)/p_H = [p_H/(1-p_H) + t-1](p_H R - 1)/g$ for all t.
Then we get $1 = \lambda_{T+1} = [p_H/(1-p_H) + T](p_H R - 1)/g$ when $R = R^{**} = 1/p_H + G/[T + p_H/(1-p_H)]$.
Finally, to justify $z_t=0$ for $t>1$, we have $\partial\mathcal{L}/\partial z_t = \mu_t(p_H - p_L) - \lambda_t(1-p_H) \leq 0$ for all t because $\partial\mathcal{L}/\partial z_1=0$, μ_t is a constant over t, and λ_t is increasing in t.
This solution could fail only if $I_1>1$, but then we can proportionally decrease all I_t and $w_{S,t}$ until $I_1=1$.
When R is slightly more than R**, we can still show optimality by reducing μ_T if $I_T=1$, or μ_1 if $I_1=1$.

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 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 V_1 and V_0 denote the agent's expected total payoff after observing y=1 or y=0 respectively. So the principal's problem is

choose (V_1,V_0) to minimize $(1-\alpha\varepsilon)V_1 + \alpha\varepsilon V_0$ subject to
$(1-\alpha\varepsilon)V_1 + \alpha\varepsilon V_0 \geq D\varepsilon + (1-\beta\varepsilon)V_1 + \beta\varepsilon V_0$, $V_1 \geq \bar{v}$, $V_0 \geq \bar{v}$.

We could add an ex-ante participation constraint $(1-\alpha\varepsilon)V_1 + \alpha\varepsilon V_0 \geq \bar{w}$, but if the payoff $\bar{w}$ that must be promised to recruit the agent is the same as the payoff $\bar{v}$ that he can get by quitting later, then this constraint is redundant with the ex-post participation constraints $V_1 \geq \bar{v}$ and $V_0 \geq \bar{v}$.
The moral-hazard constraint implies $V_1 - V_0 \geq D/(\beta-\alpha)$.
So the optimal solution is $V_0 = \bar{v}$, $V_1 = \bar{v} + D/(\beta-\alpha)$.

So the agent's expected reward $(1-\alpha\varepsilon)V_1 + \alpha\varepsilon V_0 = \bar{v} + (1-\alpha\varepsilon)D/(\beta-\alpha)$ for a short ε-period of service must be greater than the outside option $\bar{v}$ by a positive bonus even as $\varepsilon-0$.
In a dynamic model, if the problem is repeated with a different agent every ε-period then the principal's cost becomes 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 $\bar{v}$.
So with discount rate r, we get the recursion equation: $V_1 = w\varepsilon + (1-r\varepsilon)((1-\alpha\varepsilon)V_1 + \alpha\varepsilon\bar{v})$.
With $V_1 = \bar{v} + D/(\beta-\alpha)$, this implies $w = r\bar{v} + (r + \alpha - \varepsilon r\alpha)D/(\beta-\alpha)$.
Here $r\bar{v}$ is the outside wage rate corresponding to the present-discounted value $\bar{v}$.
As $\varepsilon\to 0$, this optimal stationary incentive plan pays the agent an efficiency wage rate that exceeds the outside wage rate by $(r+\alpha)D/(\beta-\alpha)$, but dismisses the agent when an accident occurs.

This $\varepsilon\to 0$ is a problem of controlling a Poisson process 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 $\tilde{X}$ is a Poisson random variable with mean λ, $\tilde{X}$ can be any nonnegative integer,
$P(\tilde{X}=k) = e^{-\mu}\,(\mu)^k/k!$ for any $k \in \{0,1,2,...\}$, $E(\tilde{X}) = \mu$, $Var(\tilde{X}) = \mu$, $Stdev(\tilde{X}) = \mu^{0.5}$.
When accidents occur in a Poisson process with rate λ, the number of accidents between any two times t and $t+\delta$ ($\delta>0$) is a Poisson random variable with mean $\lambda\delta$, and it is independent of the number of accidents before time t.
In any short time interval of length ε, the probability of an accident is approximately $\lambda\varepsilon$,
the probability of no accidents is approximately $1-\lambda\varepsilon$, and the probability of two or more accidents is vanishingly smaller of order ε^2.

Holmstrom-Milgrom Control of Brownian Motion (Econometrica 55(2):303-328, 1987):
Suppose the agent is risk-averse with constant risk tolerance T, subject to an ex-ante participation constraint with the alternative reservation wage w_o. But now suppose that the observable is a Normal random variable with mean $\mu(a)$ that depends on his action, and variance σ^2 that does not depend on his action. For two Normals with different means and the same variance, the likelihood ratios go to infinity in the tails, and so the multiplier of our moral-hazard constraint must be $\mu=0$: it is costless!
The optimal solution achieves first-best nonlinearly: pay w_o except for an infinite punishment in an event that has infinitesimal probability, infinitesimally smaller when he chooses high μ. Something seems wrong in this solution.
Holmstrom-Milgrom changed the problem to allow the agent to get feedback as the Normal outcome evolves, by considering the problem of controlling the drift of a Brownian motion.
For such a problem, they showed that linear incentive pay becomes optimal.

For important applications, they explicitly considered multidimensional Brownian motion.
We may (w.l.o.g.) restrict attention to Brownian motion $(\mathbf{Y}_1(t),...,\mathbf{Y}_n(t))$ where the various components $\mathbf{Y}_i(t)$ are independent and have the same volatility:
In any time interval from t to t+δ ($\delta>0$), the changes $\mathbf{Y}_i(t+\delta)-\mathbf{Y}_i(t)$ are Normal random variables, with mean $\mu_i(a)\delta$ and variance $\sigma^2\delta$, independent across i, and independent of the path to time t.
Here the drifts $\mu_i(a)$ are functions of the agent's hidden action (a) during this interval.
The agent also pays a hidden cost of effort at rate c(a) over time.
The agent's choices of action after time t can depend on the past $(\mathbf{Y}_1(t),...,\mathbf{Y}_n(t))$.
We may assume that the process begins at $(\mathbf{Y}_1(0),...,\mathbf{Y}_n(0)) = (0,...,0)$.
A linear incentive plan for the agent over the period $[0,\Omega]$ specifies constants $(A,B_1,...,B_n)$ and promises to pay the agent $w(\mathbf{Y}) = A + \sum_i B_i\mathbf{Y}_i(\Omega)$.
The agent's final utility payoff is $u(w(\mathbf{Y})-C(a))$, where $C(a) = \int_0^\Omega c(a(t))dt$,
and $u(\bullet)$ is a utility function with constant-risk-tolerance T. We may let the ending time be $\Omega=1$, so that, for any constant action a, the $\mathbf{Y}_i(\Omega)$ are independent Normals with mean $\mu_i(a)$ and variance σ^2.
Holmstrom and Milgrom show that such simple linear incentive plans are optimal for the principal to maximize any linear function of the final $(\mathbf{Y}_1(\Omega),...,\mathbf{Y}_n(\Omega))$
subject to an ex-ante participation constraint at time 0 and a moral-hazard constraint that the constant-risk-tolerant agent always chooses an action that maximizes his expected utility:

Choose $w(\bullet)$ and a^* to maximize $E(\sum_i \pi_i \mathbf{Y}_i(\Omega) - w(\mathbf{Y})\,|a^*)$ subject to
$a^* \in \text{argmax}_a E(u(w(\mathbf{Y})-C(a))|a)$, and $E(u(w(\mathbf{Y})-C(a^*))|a^*) \geq u(w_o)$.

Linearity makes $w(\mathbf{Y}) = A + \sum_i B_i\mathbf{Y}_i(\Omega)$ a Normal random variable, so the agent's constraints can be rewritten with the formula for certainty equivalents of Normals with constant risk tolerance:
$a^* \in \text{argmax}_a A+\sum_i B_i\mu_i(a)-C(a)-(0.5/T)\sigma^2\sum_i B_i^2$, $A+\sum_i B_i\mu_i(a^*)-C(a^*)-(0.5/T)\sigma^2\sum_i B_i^2 \geq w_o$.
The binding participation constraint yields $A = w_o + C(a^*) + (0.5/T)\sigma^2\sum_i B_i^2 - \sum_i B_i\mu_i(a^*)$.
Then the principal's expected profit is $E(\sum_i \pi_i\mathbf{Y}_i(\Omega)- w(\mathbf{Y})|a^*) = \sum_i \pi_i\mu_i(a^*) - \sum_i B_i\mu_i(a^*) - A$
$= \sum_i \pi_i\, \mu_i(a^*) - C(a^*) - (0.5/T)\sigma^2\sum_i B_i^2 - w_o$, which we want to maximize
subject to the incentive constraint, which reduces to: $a^* \in \text{argmax}_a \sum_i B_i\mu_i(a) - C(a)$.
For each component a_k of vector a, first-order conditions are $\sum_i B_i\, \partial\mu_i(a^*)/\partial a_k = \partial C(a^*)/\partial a_k$
or, with nonnegativity constraints $a_k \geq 0$, we may have $\sum_i B_i\, \partial\mu_i(a^*)/\partial a_k \leq \partial C(a^*)/\partial a_k$ if a_k=0.

Technical key to their proof: Such a Brownian motion can be viewed as the limit of a linear function of multidimensional Poisson process. Large-mean Poissons are approximately Normal.
Consider some small $\varepsilon>0$. For any action a, let the Poisson processes start at $\mathbf{x}_i(0|a,\varepsilon)=0$ for all i.
For any time interval t to $t+\delta$, let $\mathbf{x}_i(t+\delta|a,\varepsilon)-\mathbf{x}_i(t|a,\varepsilon)$ (= <u>arrivals of type i</u>) be a Poisson random variable with mean $\delta\Lambda_i(a,\varepsilon) = \delta(\sigma^2+\varepsilon\mu_i(a))/\varepsilon^2$, independently of the past before t, and independently across all i=1,...,n.
Recall that the variance of a Poisson random variable is equal to its mean or expected value.
Now let $\mathbf{y}_i(t|a,\varepsilon) = \varepsilon\, \mathbf{x}_i(t|a,\varepsilon) - t\,\sigma^2/\varepsilon$.
So $\mathbf{y}_i(t|\alpha,\varepsilon)$ has frequent discontinuous upward jumps of size ε, but between these jumps it is decreasing at the rate $-\sigma^2/\varepsilon$.
Then $E(\mathbf{y}_i(t+\delta|a,\varepsilon)-\mathbf{y}_i(t|a,\varepsilon)) = \varepsilon\,\delta\,(\sigma^2+\varepsilon\mu_i(a))/\varepsilon^2 - \delta\,\sigma^2/\varepsilon = \mu_i(a)\delta$,
and $\mathrm{Var}(\mathbf{y}_i(t+\delta|a,\varepsilon)-\mathbf{y}_i(t|a,\varepsilon)) = \varepsilon^2\,\delta\,(\sigma^2+\varepsilon\mu_i(a))/\varepsilon^2 = \sigma^2\delta + \varepsilon\mu_i(a)\delta \to \sigma^2\delta$ as $\varepsilon\to 0$.
So as $\varepsilon\to 0$, this **y** process converges to the Brownian motion **Y** that we wanted to study.
So a wage that is linear in the $\mathbf{y}_i$, say $A+\sum_{i=1}^n B_i\mathbf{y}_i$, would also be linear in the $\mathbf{x}_i$, with coefficients:
$A+\sum_{i=1}^n B_i\,\mathbf{y}_i(t|a,\varepsilon) = A+\sum_{i=1}^n B_i\,[\varepsilon\,\mathbf{x}_i(t|a,\varepsilon)-t\sigma^2/\varepsilon] = (A-nt\sigma^2/\varepsilon) + \sum_{i=1}^n (B_i\varepsilon)\,\mathbf{x}_i(t|a,\varepsilon)$

Consider a discrete-time approximation to the multidimensonal Poisson process problem ($\varepsilon>0$).
In any short time interval of length δ (much less than ε^2, say $\delta=\varepsilon^3$), there are n+1 events that could occur: no arrival, or one arrival of some type $i\in\{1,..,n\}$, each with probability $\delta\sigma^2/\varepsilon^2+\delta\mu_i(a)/\varepsilon$.
Two or more arrivals in a short time interval has vanishingly small probability.
The optimal incentive plan in this short time interval, given any participation constraint, would pay some amount for each of these possible events. Changing the participation constraint would only add a constant to all payments, because of constant risk tolerance.
At any point in time, past payments would not affect the agent's preferences over gambles for additional income and effort-cost, again because of constant risk tolerance.
Thus, the optimal incentive plan for the dynamic model can be decomposed to identical incentive problems in each short time interval, plus a constant to meet the ex-ante participation constraint.
But $\mathbf{x}_i(\Omega)$ measures the number of times that a type-i arrival occurred, for each of which the agent is paid the same amount.
So the optimal final payment to the agent is a linear function of $(\mathbf{x}_1(\Omega|a,\varepsilon),...,\mathbf{x}_n(\Omega|a,\varepsilon))$,
and so it is also a linear function of $(\mathbf{y}_1(\Omega|a,\varepsilon),...,\mathbf{y}_n(\Omega|a,\varepsilon))$.
Now take the limit as $\varepsilon-0$, and the limit of the optimal plans pays the agent as a linear function of the Brownian-motion endpoint $(\mathbf{Y}_1(\Omega),...,\mathbf{Y}_n(\Omega))$.

Controlling a Poisson processes with constant risk tolerance, no liability limits

Suppose that an agent with constant risk tolerance T is to be paid $\beta\tilde{X}$,
where $\tilde{X}$ is a Poisson random variable with some mean λ.
Then the agent's expected utility is

$$EU = \sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}(-e^{-k\beta/T}) = -e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{-\beta/T})^k}{k!} = -e^{-\lambda T(1-e^{-B/T})/T}$$

and so the agent's certainty equivalent of this income $\beta\tilde{X}$ is $\lambda T(1-e^{-\beta/T})$.

Now suppose that the agent chooses an action $a \in A$ (where A is his set of feasible actions), and this action affects the means $\Lambda_1(a),...,\Lambda_n(a)$ of n independent Poisson random variables $\tilde{X}_1,...,\tilde{X}_n$.
The agent also pays a personal cost C(a) for his action, and has constant risk tolerance T.
The a action and cost C(a) cannot be observed by anybody except the agent, but the principal can observe the Poisson random variables $\tilde{X}_1,...,\tilde{X}_n$.
A risk neutral principal gets revenue $\tilde{X}_1\pi_1+...\tilde{X}_n\pi_n$ from these Poisson random variables.
Consider incentive plans, where the principal pays the agent a wage w according to a linear formula $w(\tilde{X}_1,...,\tilde{X}_n) = \alpha + \tilde{X}_1\beta_1 + ... + \tilde{X}_n\beta_n$.
With constant risk tolerance, the agent's certainty equivalent for the sum of n independent random payments is equal to the sum of the certainty equivalents of these n random payments
(this additivity condition only holds with constant risk tolerance).
Let $\bar{w}$ denote the best alternative wage that the agent could earn elsewhere.
So the principal's problem can be written: choose $(\alpha,\beta_1,...,\beta_n, a^*)$ to
maximize $\Lambda_1(a^*)\pi_1+...+ \Lambda_n(a^*)\pi_n - [\alpha + \Lambda_1(a^*)\beta_1 + ... +\Lambda_n(a^*)\beta_n]$
subject to $\max_{a\in A}\ \alpha + \Lambda_1(a)T(1-e^{-\beta_1/T}) +...+\Lambda_n(a)T(1-e^{-\beta_n/T}) - C(a)$
$\quad = \alpha + \Lambda_1(a^*)T(1-e^{-\beta_1/T}) +...+\Lambda_n(a^*)T(1-e^{-\beta_n/T}) - C(a^*) \geq \bar{w}$.

Substituting $\alpha = \bar{w}+C(a^*)-\sum_i \Lambda_i(a^*)T(1-e^{-\beta_i/T})$ from the participation constraint,
the principal's expected profit becomes $\sum_i \Lambda_i(a^*)[\pi_i - \beta_i+T(1-e^{-\beta_i/T})] - \bar{w} - C(a^*)$.
If we can differentiate with respect to the action a, first-order conditions of the incentive constraint are $\sum_i \Lambda_i'(a^*)T(1-e^{-\beta_i/T}) = C'(a^*)$.

Extending the Poisson process over time, the number of i-arrivals in a time period of length δ may be denoted by $\tilde{X}_i(\delta)$, and it is a Poisson random variable with mean $\delta\Lambda_i(a)$ (=variance).
Results of separate time intervals are independent. With constant risk tolerance, they can be analyzed independently, so the optimal wage is linear in $\tilde{X}_1(\delta),...,\tilde{X}_n(\delta)$ over the whole period.

Now fix $\mu_i(a)$ and σ^2, and consider small positive numbers ε.
Given any $\varepsilon>0$, let $\Lambda_i(a) = [\sigma^2 + \varepsilon\mu_i(a)]/\varepsilon^2$, and let $\tilde{Y}_i(\delta) = \varepsilon\tilde{X}_i(\delta) - \delta\sigma^2/\varepsilon$.
Then $E(\tilde{Y}_i(\delta)) = \delta\mu_i(a)$, and $Var(\tilde{Y}_i(\delta)) = \varepsilon^2 Var(\tilde{X}_i(\delta)) = \delta[\sigma^2 + \varepsilon\mu_i(a)] \rightarrow \delta\sigma^2$ as $\varepsilon\rightarrow 0$.
$\Lambda_i(a)$ here becomes large as $\varepsilon\rightarrow 0$, and Poisson random variables with large means are approximately Normal. So as $\varepsilon\rightarrow 0$, $\tilde{Y}_i(\delta)$ approaches a Normal random variable,
and $\tilde{Y}_i$ becomes a Brownian-motion process with drift $\mu_i(a)$ and volatility σ.
Wages that are linear in $\tilde{Y}_i(\delta)$, say $A\delta+\sum_{i=1}^{n} B_i\tilde{Y}_i(\delta)$, are also linear in $\tilde{X}_i(\delta)$, with coefficients:
$A\delta+\sum_{i=1}^{n} B_i\tilde{Y}_i(\delta) = A\delta+\sum_{i=1}^{n} B_i\,[\varepsilon\,\tilde{X}_i(\delta)-\delta\sigma^2/\varepsilon] = (A\delta-n\delta\sigma^2/\varepsilon) + \sum_{i=1}^{n} (B_i\varepsilon)\,\tilde{X}_i(\delta)$

First adverse-selection problem: one privately-informed agent with two possible types

An agent (seller) is one of two types, θ_H or θ_L where $\theta_H>\theta_L$. His type is his private information.
A principal (buyer) thinks the probability of θ_H is p_H, and the probability of θ_L is $p_L = 1-p_H$.
The agent's effort q and wage w are both observable numbers which may depend on the agent's reported type, but the agent can misrepresent his type. Effort must be nonnegative $q \geq 0$.
When the agent's type is θ, any wage w and effort $q\geq 0$ yield payoff $w-\theta q$ for the agent and yield profit $\pi(q|\theta)-w$ for the principal. The agent's payoff is his gain from trade over alternatives worth 0.
With prime $'$ for $\partial/\partial q$, suppose $\pi(0|\theta)=0$, $\pi'(q|\theta)>0$, $\pi''(q|\theta)\leq 0$, $\forall\theta, \forall q$, and $\pi'(0|\theta_L)>\theta_L$.

The principal's problem is to choose a contract-menu (q_H,w_H,q_L,w_L) to
maximize $p_H(\pi(q_H|\theta_H)-w_H) + p_L(\pi(q_L|\theta_L)-w_L)$ subject to $w_L \in \mathbb{R}$, $w_H \in \mathbb{R}$, $q_L \geq 0$, $q_H \geq 0$,

$w_L - \theta_L q_L \geq 0,$ [L-participation, λ_L]
$w_H - \theta_H q_H \geq 0,$ [H-participation, λ_H]
$w_L - \theta_L q_L \geq w_H - \theta_L q_H,$ [|L-informational incentive, $\alpha_{|L}$]
$w_H - \theta_H q_H \geq w_L - \theta_H q_L.$ [|H-informational incentive, $\alpha_{|H}$]

(If there were no incentive constraints, the solution would have $\pi'(q_i|\theta_i)=\theta_i$, $w_i = \theta_i q_i$, $\forall i$.)

The incentive constraints (|H first, then |L) imply: $\theta_H(q_L-q_H) \geq w_L-w_H \geq \theta_L(q_L-q_H)$.
With $\theta_H>\theta_L$, this implies $q_L-q_H \geq 0$ and $w_L-w_H \geq 0$, so $q_L \geq q_H$ and $w_L \geq w_H$.
If both incentive constraints were binding, these would all be equalities.
So both incentive constraints can bind only for a <u>pooling plan</u> that satisfies $q_L=q_H$ and $w_L=w_H$.
Note: If $q_L > q_H$ then: $w_H - t\,q_H \geq w_L - t\,q_L \iff t \geq (w_L-w_H)/(q_L-q_H)$. So any cost type $t >\theta_H$ would prefer (q_H,w_H) over (q_L,w_L), while any type $t < \theta_L$ would prefer (q_L,w_L).

The Lagrangean can be written: $\mathcal{L}(w,q;\lambda,\alpha) =$

$$= p_H(\pi(q_H|\theta_H)-w_H) + p_L(\pi(q_L|\theta_L)-w_L) + \lambda_L[w_L-\theta_L q_L] + \lambda_H[w_H-\theta_H q_H]$$
$$+\alpha_{|L}[w_L-\theta_L q_L-w_H+\theta_L q_H] + \alpha_{|H}[w_H-\theta_H q_H-w_L+\theta_H q_L]$$
$$= p_H\pi(q_H|\theta_H) - q_H\,[(\lambda_H+\alpha_{|H})\theta_H - \alpha_{|L}\theta_L] + p_L\pi(q_L|\theta_L) - q_L\,[(\lambda_L+\alpha_{|L})\theta_L - \alpha_{|H}\theta_H]$$
$$+ w_H\,[-p_H + \lambda_H + \alpha_{|H} - \alpha_{|L}] + w_L\,[-(1-p_H) + \lambda_L + \alpha_{|L} - \alpha_{|H}].$$

Lagrange multipliers must be nonnegative $\lambda_H\geq 0$, $\lambda_L\geq 0$, $\alpha_{|H}\geq 0$, $\alpha_{|L}\geq 0$, and satisfy complementary slackness. The first-order Lagrange optimality conditions for $w_L \in \mathbb{R}$ and $w_H \in \mathbb{R}$ are

$0 = \partial\mathcal{L}/\partial w_L = -p_L + \lambda_L + \alpha_{|L} - \alpha_{|H}$, and so $\lambda_L + \alpha_{|L} = p_L + \alpha_{|H}$;
$0 = \partial\mathcal{L}/\partial w_H = -p_H + \lambda_H + \alpha_{|H} - \alpha_{|L}$, and so $\lambda_H + \alpha_{|H} = p_H + \alpha_{|L}$.

So we must have $\lambda_L+\lambda_H = p_L+p_H = 1$. With these "balance" equations, the Lagrangean simplifies to:

$$\mathcal{L} = p_H\{\pi(q_H|\theta_H) - q_H\,[\theta_H + (\theta_H-\theta_L)\alpha_{|L}/p_H]\} + p_L\{\pi(q_L|\theta_L) - q_L[\theta_L + (\theta_L-\theta_H)\alpha_{|H}/p_L]\}.$$

The λ's have dropped out, but we still need $\lambda_L = p_L + \alpha_{|H} - \alpha_{|L} \geq 0$ and $\lambda_H = p_H + \alpha_{|L} -\alpha_{|H} \geq 0$, as well as $\alpha_{|H} \geq 0$ and $\alpha_{|L} \geq 0$ and complementary slackness with the corresponding constraints.

Note: So far, the analysis would be exactly the same if we had no participation constraints but took λ_L and λ_H as parametric weights for the payoffs of the two types of agent in a social-welfare function

$$[p_H(\pi(q_H|\theta_H)-w_H) + p_L(\pi(q_L|\theta_L)-w_L)] + \lambda_L[w_L-\theta_L qL] + \lambda_H[w_H-\theta_H q_H].$$

With $\theta_L < \theta_H$ and $q_H \geq 0$, the L-participation constraint is implied by the H-participation and $|$L-incentive constraints. So L-participation is a redundant constraint, and its multiplier is $\lambda_L = 0$. So $\alpha_{|L} = p_L + \alpha_{|H} > 0$. If the optimal solution is not a pooling plan, then at most one incentive constraint can bind, and so from $\alpha_{|L}>0$ we get $\alpha_{|H}=0$. Nonpooling also implies $q_L > q_H \geq 0$.
Thus, either (case 2) the optimal solution is a pooling plan with $\alpha_{|H} > 0$, or (case 1) the Lagrange multipliers for the optimal solution must be $\lambda_L = 0$, $\alpha_{|H}=0$, $\alpha_{|L} = p_L$, $\lambda_H = p_H + \alpha_{|L} = 1$.

In the nonpooling case 1, we have $\lambda_L = 0$, $\alpha_{|H} = 0$, $\alpha_{|L} = p_L$, $\lambda_H = 1$.
The optimal efforts must maximize the Lagrangean subject only to $q_L \geq 0$ and $q_H \geq 0$:
$\pi'(q_L|\theta_L) = \theta_L$, $\pi'(q_H|\theta_H) \leq \theta_H+(\theta_H-\theta_L)p_L/p_H$ and $q_H \geq 0$ with at least one equality.
Wages are determined by the constraints that must bind: $w_H = \theta_H\, q_H$ (for $\lambda_H>0$), and
$w_L = \theta_L q_L + w_H - \theta_L q_H$ (for $\alpha_{|L}>0$).
So when $q_H>0$, the binding $|$L-incentive constraint induces ex-post inefficiency for type H:
$\pi'(q_H|\theta_H) = \theta_H+(\theta_H-\theta_L)p_L/p_H > \theta_H$ (undersupply of q_H),
and it induces a positive payoff or information rent for type L: $w_L-\theta_L q_L = (\theta_H-\theta_L)q_H > 0$.
This solution is feasible if it satisfies the $|$H-incentive constraint, which will be satisfied if and only if our computed quantities satisfy $q_H \leq q_L$.

Otherwise, we have the pooling case 2 where both incentive constraints bind, and $q_H = q_L = q^* > 0$.
The pooling q^* satisfies $\pi'(q^*|\theta_H) = \theta_H + (\theta_H-\theta_L)\alpha_{|L}/p_H$ and $\pi'(q^*|\theta_L) = \theta_L - (\theta_H-\theta_L)\alpha_{|H}/p_L$.
So q^* can be computed from $p_H\pi'(q^*|\theta_H) + p_L\pi'(q^*|\theta_L) = \theta_H$, because $\alpha_{|L}-\alpha_{|H} = p_L = 1-p_H$.
For the H-participation constraint to bind, total wages must be $w_H = w_L = \theta_H q^*$.
The Lagrange multipliers for this pooling solution are $\alpha_{|H} = p_L[\theta_L - \pi'(q^*|\theta_L)]/(\theta_H-\theta_L)$,
$\alpha_{|L} = \alpha_{|H}+p_L = p_L[\theta_H - \pi'(q^*|\theta_L)]/(\theta_H-\theta_L)$, $\lambda_L=0$, $\lambda_H=1$.
$\alpha_{|H} > 0$ requires $\pi'(q^*|\theta_L) < \theta_L$ for this pooling solution (oversupply by type L at q^*).

Example Suppose the agent has 2 possible cost-types: $\theta_L=1$ or $\theta_H=2$, each with probability 1/2.
The principal's value of effort q from an agent of type θ is $\pi(q|\theta) = (1+\theta)q^{0.5}$, so the principal's gain from trade when paying w for effort q is $2\sqrt{q}-w$ if $\theta=\theta_L$, $3\sqrt{q}-w$ if $\theta=\theta_H$.
(Without incentive constraints, the solution would be $q_L=1$, $w_L=1$, $q_H=9/16=0.5625$, $w_H=1.125$.)
We can apply case 1 from our previous analysis of adverse-selection problem with two types.
The Lagrange multiplier of the $|$L-incentive constraint is $\alpha_{|L} = p_L = 0.5$, and the Lagrangean is:
$\mathcal{L} = 0.5\{3q_H^{0.5} - q_H\,[2+(2-1)0.5/0.5]\} + 0.5\{2q_L^{0.5} - q_L[1]\}$.
To maximize this over (q_H,q_L), we need $0 = 3(0.5)q_H^{-0.5} - 3$ and $0 = 2(0.5)q_L^{-0.5} - 1$,
and so $q_H = 0.25$ and $q_L = 1$.
Because $q_H < q_L$, we know that the case 1 assumptions about binding constraints are OK here.
The values of w_H and w_L are determined by the binding constraint equations:
To make H-participation binding, $w_H = \theta_H q_H = 2\times0.25 = 0.5$.
To make $|$L-incentive binding, $w_L = \theta_L q_L + w_H-\theta_L q_H = 1\times1 + 0.5 - 1\times0.25 = 1.25$.
Then type L gets information rent $U_L = w_L-\theta_L q_L = 0.25$, but $U_H = w_H-\theta_H q_H = 0$.
If we had changed $\pi(q_H|\theta_H)$ to be $A_H\sqrt{q_H}$ with some $A_H > 6$, then the above analysis would have yielded $q_H > q_L$, and so the pooling case 2 would have applied.

A trading example where sellers with a single asset have two possible types

To be specific, let us consider an example where the seller's cost type is either θ_L=20 or θ_H=40, and the buyer's value of the seller's asset is π_L=30 if the seller's type is θ_L, but the buyer's value is π_H=50 if the seller is θ_H. So the asset is always worth 10 more to the buyer than to the seller.
(This is MWG's example 23.F.2.) Our analysis uses the inequalities $\pi_L > \theta_L$ and $\pi_H > \theta_H > \theta_L$.
For now, let's keep the probability of the high type p_H as a parameter, with $p_L = 1-p_H$.

In trading plan (w,q), for each seller-type t, q_t is t's probability of selling, w_t is t's expected revenue.
In such a plan (w,q), a high-type seller's expected gain is $U_1(w,q|H) = w_H - \theta_H q_H$,
a low-type seller's expected gain is $U_1(w,q|L) = w_L - \theta_L q_L$,
and the buyer's expected gain is $U_2(w,q) = p_H(\pi_H q_H - w_H) + p_L(\pi_L q_L - w_L)$.
A plan (w,q) is <u>incentive-compatible</u> iff $w_H - \theta_H q_H \geq w_L - \theta_H q_L$ and $w_L - \theta_L q_L \geq w_H - \theta_L q_H$.
To be <u>feasible</u> here, (w,q) must be incentive compatible and have $0 \leq q_H \leq 1$, $0 \leq q_L \leq 1$.

A trading plan is (strongly) <u>interim (Pareto-)dominated</u> if there is some other feasible plan that would yield higher expected gains to each possible type of each individual (given only his type information).
Here, the seller has two possible types and the buyer has only one possible type,
and so a plan (w,q) would be interim dominated by some other plan $(\hat{w},\hat{q})$ iff
$U_1(w,q|H) < U_1(\hat{w},\hat{q}|H)$, $U_1(w,q|L) < U_1(\hat{w},\hat{q}|L)$, and $U_2(w,q) < U_2(\hat{w},\hat{q})$.
A plan is (weakly) <u>incentive efficient</u> if it is feasible and is not interim dominated by any other feasible plan. (See B. Holmstrom and R.Myerson, *Econometrica*, 1983.)

<u>Fact</u> An incentive-compatible trading plan $(\bar{w},\bar{q})$ is incentive efficient iff there exist nonnegative weights $(\lambda_H, \lambda_L, \lambda_2)$ such that $(\bar{w},\bar{q})$ solves the problem
maximize $\lambda_H U_1(w,q|H) + \lambda_L U_1(w,q|L) + \lambda_2 U_2(w,q)$ subject to (w,q) being feasible
(incentive compatible with $q_t \in [0,1]$ $\forall t$). Without loss of generality, we may let $\lambda_2 = 1 = \lambda_L + \lambda_H$.

The Lagrangean for such an optimization problem in this case can be written:

$$\mathcal{L}(w,q;\lambda,\alpha) = \lambda_H(w_H - \theta_H q_H) + \lambda_L(w_L - \theta_L q_L) + (p_H(\pi_H q_H - w_H) + p_L(\pi_L q_L - w_L))$$
$$+ \alpha_{|H}(w_H - \theta_H q_H - w_L + \theta_H q_L) + \alpha_{|L}(w_L - \theta_L q_L - w_H + \theta_L q_H).$$

For an optimum with finite payments w_H and w_L, we must have:
$0 = \partial\mathcal{L}/\partial w_H = \lambda_H + \alpha_{|H} - p_H - \alpha_{|L}$ and $0 = \partial\mathcal{L}/\partial w_L = \lambda_L + \alpha_{|L} - p_L - \alpha_{|H}$.
These "balance" equations imply $\lambda_L + \lambda_H = p_L + p_H = 1$, and they simplify the Lagrangean to
$\mathcal{L} = p_H q_H\{\pi_H - [\theta_H + (\theta_H - \theta_L)\alpha_{|L}/p_H]\} + p_L q_L\{\pi_L - [\theta_L + (\theta_L - \theta_H)\alpha_{|H}/p_L]\}$.
This Lagrangean must be maximized over $0 \leq q_H \leq 1$, $0 \leq q_L \leq 1$,
and we can only have positive multipliers α for incentive constraints that bind at the solution
These $\alpha_{|H}$ and $\alpha_{|L}$ must also yield $p_H + \alpha_{|L} - \alpha_{|H} = \lambda_H \geq 0$ and $p_L + \alpha_{|H} - \alpha_{|L} = \lambda_L \geq 0$.

For this example, all the incentive-efficient trading plans can be supported by one (λ,α) vector.
(The incentive-efficient frontier is flat.) Let us now see how to construct this "universal" (λ,α).

Any feasible trading plan must have $q_H \le q_L$, and $q_H < 1$ can be a useful signal of the high type.
But $0<q_H<1$ can be an optimal for the Lagrangean if and only if $\pi_H = \theta_H + (\theta_H - \theta_L)\alpha_{|L}/p_H$.
With $\pi_H > \theta_H > \theta_L$, this equality can be achieved by letting
[*] $\alpha_{|L} = p_H(\pi_H - \theta_H)/(\theta_H - \theta_L) = p_H(50-40)/(40-20) = 0.5p_H > 0.$
For any separating plan, only one incentive constraint can bind, so we should try $\alpha_{|H} = 0$;
then the balance equations yield $\lambda_H = p_H + \alpha_{|L} = 1.5p_H$ and $\lambda_L = p_L - \alpha_{|L} = 1 - p_H - \alpha_{|L} = 1 - 1.5p_H$.
These formulas yield valid nonnegative Lagrangean parameters with $\lambda_L \ge 0$
if and only if $p_H \le (\theta_H - \theta_L)/(\pi_H - \theta_L) = (40-20)/(50-20) = 2/3$ here,
So in the low-p_H case when $p_H \le (\theta_H - \theta_L)/(\pi_H - \theta_L) = 2/3$, the Lagrangean with this (λ,α) reduces to
$\mathcal{L} = p_H q_H\{\pi_H - \pi_H\} + p_L q_L\{\pi_L - \theta_L\} = p_L q_L(30-20)$, which is maximized by $q_L = 1$.
So when $p_H \le 2/3$, these (λ,α) values verify the incentive-efficiency of any feasible plan (w,q) such that $q_L = 1$ and the $|L$-incentive constraint is binding $w_L - \theta_L q_L = w_H - \theta_L q_H$.
These are all the incentive-efficient trading plans when $p_H \le 2/3$, and their expected payoffs satisfy
$(p_H + \alpha_{|L})U_1(w,q|H) + (p_L - \alpha_{|L})U_1(w,q|L) + U_2(w,q) = p_L(\pi_L - \theta_L) = (1-p_H)10.$

But when $p_H > (\theta_H - \theta_L)/(\pi_H - \theta_L) = 2/3$ here, this attempt to find incentive-efficient separating plans fails, because it yields would yield $\lambda_L = p_L - \alpha_{|L} \le 0$.
In this high-p_H case when $p_H > 2/3$, the $\alpha_{|L}$ formula [*] can satisfy the Lagrangean conditions for incentive-efficiency only if we also have $\alpha_{|H} > 0$, to get $\lambda_L \ge 0$, and so we can only find incentive-efficient plan that are pooling, with both incentive constraints binding and with $q_H = q_L$.
So in this case the Lagrangean conditions for incentive-efficiency can be satisfied by
$\alpha_{|L} = p_H(\pi_H - \theta_H)/(\theta_H - \theta_L) = 0.5p_H$, $\lambda_L = 0$, $\alpha_{|H} = \alpha_{|L} - p_L = 1.5p_H - 1$, and $\lambda_H = 1$.
Then the Lagrangean simplifies to
$\mathcal{L} = p_H q_H\{\pi_H - \pi_H\} + p_L q_L\{\pi_L - [\theta_L + (\theta_L - \theta_H)\alpha_{|H}/p_L]\} = q_L\{p_H\pi_H + p_L\pi_L - \theta_H\} = q_L\{20p_H - 10\}$,
which is maximized by $q_L = 1$.
So when $p_H > 2/3$, the incentive efficient plans are the pooling plans with $q_L = 1$,
and their expected payoffs satisfy $U_1(w,q|H) + U_2(w,q) = p_H\pi_H + p_L\pi_L - \theta_H = 20p_H - 10$.

In such bilateral trading problems with a single asset and one-sided private information, the uninformed buyer's optimal trading plan is always to offer a fixed take-it-or-leave it bid that is equal to (or slightly more than) one of the possible cost-types of the seller.
That is, the buyer's optimal incentive-compatible plan could be either to bid $\theta_L = 20$, so that $q_L = 1$, $w_L = 20$, $q_H = 0$, $w_H = 0$ (separating); or to bid $\theta_H = 40$, so that $q_L = 1 = q_H$, $w_L = 40 = w_H$ (pooling).
The pooling offer θ_H is better than the separating offer θ_L for the buyer when
$p_H\pi_H + p_L\pi_L - \theta_H \ge p_L(\pi_L - \theta_L)$,
but this inequality is equivalent to $p_H \ge (\theta_H - \theta_L)/(\pi_H - \theta_L) = 2/3$, which defines the high-p_H case.
To verify optimality for the buyer of the plan to θ_H bid in the high-p_H case, notice that trading at the pooling bid θ_H satisfies the conditions for incentive efficiency for this case (pooling and $q_L = 1$), and it

achieves the maximum of $U_1(w,q|H) + U_2(w,q)$ with $U_1(w,q)|H) = 0$. So it maximizes $U_2(w,q)$ over all feasible plans that satisfy the participation constraints (even while allowing $U_1(L)>0$).
To verify optimalize for the buyer of the plan to bid θ_L in the low-p_H case, notice that trading at the separating bid θ_L satisfies the conditions for optimality in this case ($|$L-incentive binding and $q_L=1$), and it maximizes $(p_H+\alpha_{|L})U_1(w,q|H) + (p_L-\alpha_{|L})U_1(w,q|L) + U_2(w,q)$ with $U_1(w,q|H) = 0 = U_1(w,q|L)$. So it maximizes $U_2(w,q)$ over all feasible plans that satisfy the participation constraints.

But what trading plan might be used when the informed sellers have market power? This would happen when there are many buyers and each buyer could trade with as many sellers as are available. So let us now assume such a market, with many buyers, each of whom could buy from as many sellers as are available.

We might start by considering Walras's concept of market equilibrium, defined for a model in which trade can only occur at a uniform price for all trades that is posted by a neutral Walrasian auctioneer. With uninformed buyers who can trade with as many sellers as available, there would be excess demand if the uninformed buyers could get positive expected profits at the announced prices.
So in equilibrium, the buyers must just get zero expected profit from trading with the types of sellers who are willing to trade at the given price.
In this example (with $\pi_L<\theta_H$), for any p_H there can be such a Walrasian equilibrium with a positive supply at the price $w/q = \pi_L = 30$, where supply comes only from L-types.
This Walrasian equilibrium has no trade for H types and so is not incentive efficient.
When $p_H \geq (\theta_H-\pi_L)/(\pi_H-\pi_L) = (40-30)/(50-30) = 0.5$, there is another Walrasian posted-price equilibrum at the price $w/q = p_H\pi_H+(1-p_H)\pi_L = p_H50+p_L30$, where both types sell.
When it exists, the higher price equilibrium is better for sellers and yields the same expected payoff for buyers. So we may find a Pareto ranking among Walrasian equilibria with adverse selection.

We might also ask: what trading plan maximizes the seller's ex-ante expected payoff?
That is, let us choose (w,q) to maximize $p_HU_1(w,q|H)+p_LU_1(w,q|L)$ subject to feasibility plus interim participation constraints $U_1(w,q|H)\geq 0$, $U_1(w,q|L)\geq 0$, and $U_2(w,q)\geq 0$?
This ex-ante seller's optimum depends on p_H.
If $p_H \geq 0.5$ then it is $q_L = 1 = q_H$, $w_L = w_H = p_H\pi_H+p_L\pi_L = p_H50+p_L30 \geq 40$; but
if $p_H \leq 0.5$ then it is $q_L= 1$, $w_H = 40q_H$, $w_L = 30+10q_Hp_H/p_L$, $q_H = 1/(2-p_H/p_L)$,
to satisfy the equations $w_H = \theta_Hq_H$, $w_L-\theta_L = q_H(\theta_H-\theta_L)$, and $p_L(\pi_L-w_L)+p_Hq_H(\pi_H-\theta_H) = 0$.
For example, with $p_H = 0.2$, we get $q_L=1$, $q_H = 4/7$, $w_H=40q_H$, $q_L = 31\,{}^3/_7 = 31.43$.
($\mathcal{L}$-multipliers with $\lambda_H=0.2$, $\lambda_L=0.8$: $U_2\geq 0$ has $\gamma=8/7$, $U_1(H)\geq 0$ has $\beta_H=1/7$, $\alpha_{|L} = 8/70$, $\alpha_{|H}=0$.)
When p_H is small, this ex-ante optimal plan is bad for an H-type seller (as $U_1(w,q|H)=0$),
but it would involve losses for the buyer ($30q_L-w_L < 0$) if the seller's type were not H.
So if the seller, knowing his type, were to select this trading plan, then the buyer might reasonably infer that the seller's type is L and thus reject this plan!

In such problems (where sellers have private information, and where buyers have no private information but will participate in the market only if expected gains from trade are nonnegative), we may say that a trading plan is separable iff the plan is incentive-compatible and yields a nonnegative expected profit for the buyer with each type of seller.
So a separable trading plan (w,q) is incentive-compatible and satisfies the separate break-even constraints: $\pi_H q_H - w_H \ge 0$, $\pi_L q_L - w_L \ge 0$.

Fact. All types of the informed seller can agree on which separable plan is best.
Proof Suppose to the contrary that the best separable plan for type H was (w,q) but the best separable plan for L was $(\hat{w},\hat{q})$. Then let the plan $(\bar{w},\bar{q})$ coincide with (w,q) for type H but coincide with $(\hat{w},\hat{q})$ for type L. That is, let $(\bar{w}_H,\bar{q}_H,\bar{w}_L,\bar{q}_L) = (w_H,q_H,\hat{w}_L,\hat{q}_L)$. Then
$\bar{w}_H - \theta_H\bar{q}_H = w_H - \theta_H q_H$ (by definition of the $(\bar{w},\bar{q})$ plan when the type is H)
$\ge \hat{w}_H - \theta_H\hat{q}_H$ (because H prefers (w,q) over $(\hat{w},\hat{q})$)
$\ge \hat{w}_L - \theta_H\hat{q}_L$ (because $(\hat{w},\hat{q})$ is incentive compatible)
$= \bar{w}_L - \theta_H\bar{q}_L$ (by definition of the $(\bar{w},\bar{q})$ plan when the type is L)
Also, $\bar{w}_L - \theta_L\bar{q}_L = \hat{w}_L - \theta_L\hat{q}_L \ge w_L - \theta_L q_L \ge w_H - \theta_L q_H = \bar{w}_H - \theta_L\bar{q}_L$. So $(\bar{w},\bar{q})$ is incentive compatible.
$(\bar{w},\bar{q})$ is separable because $\pi_H\bar{q}_H - \bar{w}_H = \pi_H q_H - w_H \ge 0$ and $\pi_L\bar{q}_L - \bar{w}_L = \pi_L\hat{q}_L - \hat{w}_L \ge 0$.
But each type of seller gets the same expected payoff from $(\bar{w},\bar{q})$ as the other plan that was assumed best for it, and so $(\bar{w},\bar{q})$ is best for both types of seller among all separable plans.

So we can talk about the best separable plan for the sellers.

In our example, the best separable plan is $q_L=1$, $w_L = \pi_L = 30$, $q_H = (\pi_L - \theta_L)/(\pi_H - \theta_L) = 1/3$, $w_H = \pi_H q_H = 50/3$, with $U_1(w,q|L) = 30-20 = 10$, $U_1(w,q|H) = (50-40)/3 = 3.333$, $U_2(w,q) = 0$.
Having $q_L=1$ and $w_L=30$ gives type L the highest expected profits subject to the constraint that the buyer cannot pay more than $\pi_L = 30$. Then the best separable plan for H must have (w_H,q_H) that maximizes $w_H - 40q_H$ subject to $50q_H - w_H \ge 0$ and $w_H - 20q_H \le 30-20$.
The solution has $w_H=50q_H$ and $(50-20)q_H = 30-20$ so $q_H = 1/3$.
Notice that the best separable plan does not depend on the probability of the high type p_H.

In the low-p_H case when $p_H < 2/3$, the best separable plan here is interim incentive efficient, because the |L-incentive constraint is binding and $q_L=1$ in this best separable plan.
When the best separable plan is incentive efficient, it can be arguably considered an equilibrium for the adverse-selection market with competitive uninformed buyers, because of the following fact.
Fact. If the best separable plan is incentive efficient then, for any other incentive-compatible plan, the buyers must expect losses from the seller-types that would prefer this plan to the best separable plan. If not, then letting each type of seller choose its preference among this alternative plan and the best separable plan would also be an incentive-compatible plan that would interim Pareto-dominate the best separable plan, contradicting our assumption that the best separable plan is incentive-efficient.

In the high-p_H case when $p_H > 2/3$, however, the best separable plan for this example can be interim dominated by a pooling plan such as $\hat{q}_L = \hat{q}_H = 1$, $\hat{w}_L = \hat{w}_H = 50p_H + 30(1-p_H) = 30+20p_H > 43.33$. For example, when p_H=0.8, a pooling price 46 yields $U_1(L) = 26$ and $U_1(H) = 6$ with $U_2 = 0$.

In this market with many competitive buyers, any one of whom could use all the sellers' supply, we may anticipate that buyers' competition should drive their expected profits to 0 in equilibrium.
If the buyers' expected profits were positive, then one buyer could deviate and increase his price offers (w_H,w_L) slightly, to make greater profit by attracting all the sellers.
In the low-p_H case when $p_H \leq 2/3$, the seller's best separable plan (q_L=1, w_L=30, q_H=1/3, w_H=50/3) is incentive-efficient and we have seen that it may then be considered a competitive equilibrium here.

In high-p_H cases with p_H>2/3, however, it is harder to say what a market equilibrium should be.
In these cases, the only incentive-efficient plans are pooling plans in which the buyers profit from trading with H-sellers but suffer losses from trading with L-sellers.
If the trading plan that is offered by the buyers in the market is not incentive-efficient, then a buyer could gain by a dominating deviation, that is, by deviating to offer an interim Pareto-dominating plan that attracts all types of sellers and gives the buyer greater expected profits, when we average over all types according to their given (p_H,p_L) distribution.

On the other hand, if the trading plan in the market is a pooling plan or any other plan in which buyers lose from trading with L-types, then a deviating buyer could do better by making an offer that is slightly better for H-sellers and slightly worse for L-sellers, thus only attracting the profitable H-types. That is, if the other buyers are offering an incentive-compatible plan with $50q_H-w_H > 0$ and $30q_L-w_L<0$, then a deviating buyer could do better by offering to buy some quantity $\hat{q} = q_H-\varepsilon$ for a payment $\hat{w} = w_H-\delta$ such that ε is small and positive and $40\varepsilon>\delta>20\varepsilon$.
Then with $\hat{w}-40\hat{q} > w_H-40q_H$, the H-types will prefer the deviator's offer; and with ε small, the H-types will still yield positive profits $50\hat{q}-\hat{w}>0$. But with $\hat{w}-20\hat{q} < w_H-20q_H \leq w_L-20q_L$, the L-types will prefer to trade with other buyers, and the deviator will gain by avoiding the unprofitable L-trades.
This tactic of leaving unprofitable types to trade with others in the market is called cream-skimming.

Another approach to this problem of defining equilibrium with competitive uninformed buyers is to model it as a game where each of two buyers independently announces an incentive-compatible plan $(\tilde{q}_H,\tilde{w}_H,\tilde{q}_L,\tilde{w}_L)$, and then each informed seller chooses the plan that he prefers (given his type).
The above argument shows that this game has no pure-strategy equilibrium when p_H>2/3.
A symmetric randomized equilibrium can be found. Let us consider the case of $p_H = 0.8$.
In this equilibrium, each buyer independently chooses his $\tilde{q}_H$ from a uniform distribution on [1/3,1], and then lets $\tilde{w}_H = 2+44\tilde{q}_H$, $\tilde{q}_L = 1$, and $\tilde{w}_L = 20\tilde{q}_L + \tilde{w}_H - 20\tilde{q}_H = 22 + 24\tilde{q}_H$.
The low end $\tilde{q}_H$=1/3 is best separable plan where L-sellers get $U_1(L)$=10, H-sellers get $U_1(H)$=10/3.
The high end $\tilde{q}_H$=1 is the pooling plan with $\tilde{w} = p_H\pi_H+p_L\pi_L = 46$, $U_1(L)$=26, $U_1(H)$=6.

All sellers choose whichever buyer offers the higher $\tilde{q}_H$, and this buyer gets zero expected profits, because these plans all satisfy $0.2(30\tilde{q}_L - \tilde{w}_L) + 0.8(50\tilde{q}_H - \tilde{w}_H) = 0.2(8-24\tilde{q}_H) + 0.8(6\tilde{q}_H - 2) = 0$.
The winning buyer gains from H-sellers but loses from L-sellers. When the other buyer's offer is uncertain, no alternative offer could expect to earn positive profits from the types that would prefer it. But ex post, for any plan that buyer 1 might offer, when buyer 2 observes it then buyer 2 could do better either by deviating either to a profitable dominating plan that attracts all types, or to a cream-skimming plan that attracts only H-types and leaves buyer 1 with the unprofitable L-types.
So this approach seems unsatisfactory, as it relies too heavily on any assumption that buyers must commit themselves simultaneously to offers that they then cannot revise.

To formulate a concept of market equilibrium for which general existence can be proven, we must find some way for equilibria to deter some of these dominating or cream-skimming deviations.
Two general ways have been suggested.
One is to allow buyers some ability to react to each others' deviation from the anticipated equilibrium.
The other is to allow that the self-selected sellers who accept a deviator's offer might not be representative of the overall population of sellers. We consider each in turn.

Threats of competitive reactions could make any market uncompetitive. Buyers could sustain monopsonistic collusion if they could threaten to increase their offers against any buyer's deviation. But such collusion could not be sustained if buyers were only allowed to withdraw their offers in response to a deviation. In a standard market without adverse selection, a buyer's profit could never be reduced by other buyers withdrawing some competitive offers.
But in a market with adverse selection, seemingly profitable cream-skimming deviations could be made unprofitable by withdrawing competitive offers that the deviation itself has made unprofitable.
For example, with p_H=0.8, the efficient pooling plan with zero expected profit for buyers has q=1 and $w = p_H\pi_H + p_L\pi_L = 46$. A cream-skimming challenge that attracts the profitable H-types but not the unprofitable L-types must have $\hat{q}=1-\varepsilon$ and $\hat{w}=46-\delta$ where $40\varepsilon > \delta > 20\varepsilon$. But then the average price per unit must be $\hat{w}/\hat{q} > (46-40\varepsilon)/(1-\varepsilon) > 46$, and so the deviation would become unprofitable if the other buyers withdrew their offers that were supposed to attract the unprofitable L-types.
The problem is that, by treating the type-L sellers so well, the efficient pooling plans actually make it easier for a competitive challenger to design terms of trade that appeal only to profitable H types.
Thus, the possibility of defensive withdrawal of offers could deter cream-skimming deviations and thus could sustain the efficient 0-profit pooling plan as a market equilibrium.

The other way of sustaining an equilibrium is by dropping the assumption that a deviating buyer can confidently expect to attract all sellers of any type that prefer the deviator's offer.
We normally think of a competitive equilibrium as being characterized by many small competitors, each of whom serves only a small fraction of the market. (If the buyers were few, each with a large share of the market, it would be hard to justify an assumption that they could not react to each other.)

For a small buyer, a deviation could be worthwhile if it profitably increased the buyer's share of the market, even if it could not hope to suddenly recruit all sellers in the market.
After such a deviation, a small buyer would recruit a self-selected sample of sellers who are attracted by the new offer, and their types might not be representative of the overall type distribution.
Of course, types that prefer the market equilibrium will not do business with the deviator; but other types could also be under- or over-represented in the sample drawn by the deviator.
For a deviation that would attract two or more types, if any of them would be unprofitable for the deviator, then the deviation could be deterred by fear of over-sampling the unprofitable types.
So a dominating deviation by a pooling plan against an inefficient separable plan (in the high-p_H case) could be deterred by fear of L-types being adversely over-represented in the deviator's sample.
Thus, the possibility of such adverse selection against a deviator could deter dominating deviations and thus could sustain the best separable plan at an equilibrium even when it is inefficient.

Consider a dynamic version of this market into which a flow of informed sellers continuously enter, search for a suitable buyer, and then exit when they find an offer that seems competitively optimal.
In any time interval, suppose that 80% of the entering sellers are H-types, 20% are L-types.
Suppose that all buyers' offers are very close to some part of the best separable plan, but most offer only $(q_H,w_H) = (1/3-\varepsilon,50(1/3-\varepsilon))$ for H-types, and L types must search longer to find $(q_L,w_L)=(1,30)$.
So at any point in time, the stock of searching sellers, from whom a deviating buyer must sample, might consist of mostly L-types, and then a short-term deviator could not gain by a pooling plan.
Thus, the best separable plan could be sustained as an equilibrium by a version of Gresham's law: that the bad (L) types circulate more than the good (H) types.

Fact. If both defensive withdrawals and adverse selection are possible against deviators, then any feasible plan that (weakly) interim dominates the best separable plan is sustainable.
For such a sustainable equilibrium, buyers could also offer the best separable plan, which no sellers would be expected to accept (because the given feasible plan interim Pareto-dominates it).
But if any buyer deviated to another offer, then the other buyers could withdraw to the best separable plan, and then the deviator could not earn profits from its self-selected sellers if the fraction of H-types among them were in the low range (below 2/3, where this separable plan cannot be dominated).

[**Some parametric variants**. If the parameters were $\theta_L=30$, $\pi_L=20$, $\theta_H=40$, $\pi_H=50$, then the best separable plan for the sellers would have no trade at all, even with high p_H: $0=q_L=q_H$, $0=w_L=w_H$.
Here buyers fear to trade with L sellers, with whom they cannot profitable trade.
If the parameters were $\theta_L=20$, $\pi_L=50$, $\theta_H=40$, $\pi_H=30$, then the best separable plan for the sellers would be $q_L=1$, $w_L=40$, $0=q_H=w_H$, which lets buyers earn strictly positive profits. Here fear of trading with H-sellers could deter buyers from offering more than 40 for L-sellers who are worth 50.]

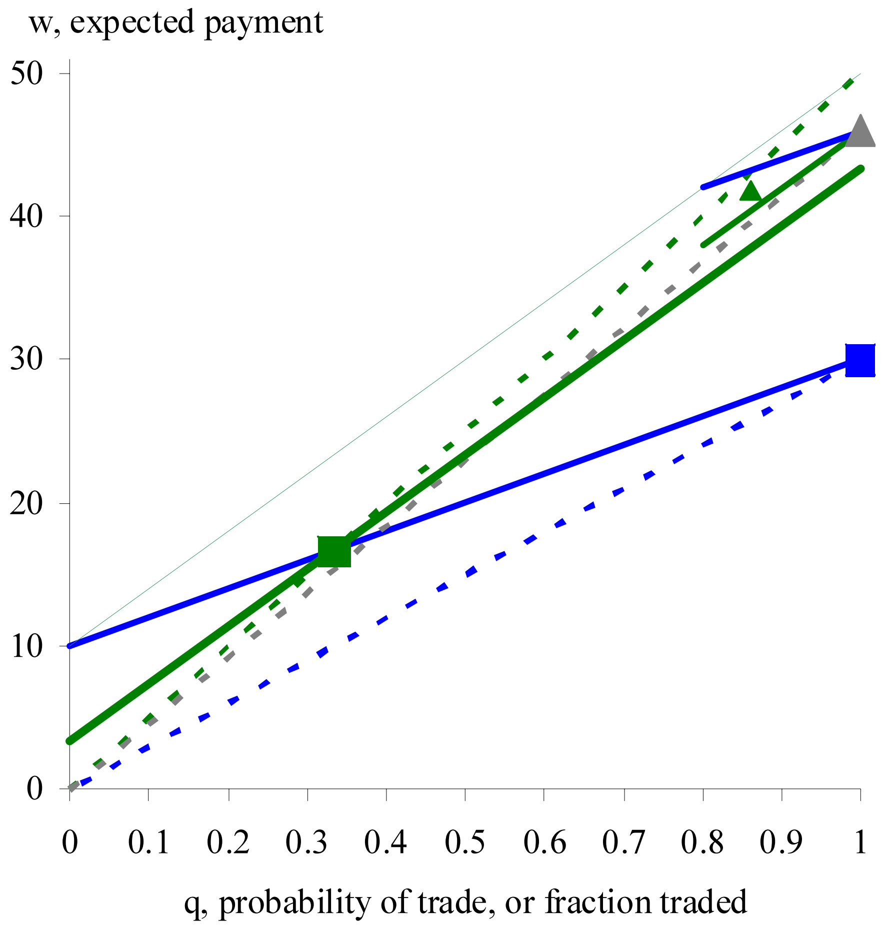


Diagram of the MWG example 23.F.3, with θ_L=20, π_L=30, θ_H=40, π_H=50, p_H=0.8 .

The lowest dashed line (blue) is where the buyer's expected profit is zero with low type, the highest dashed line (green) is where the buyer's expected profit is zero with high type, and the middle dashed line (grey) is where the buyer's expected profit is zero with both types pooled, given fractions p_H and p_L=1−p_H. The solid lines with lower slope (blue) are lines of constant profit to low-type sellers, and the solid lines with higher slope (green) are lines of constant profit to high-type sellers.

The seller's best separable plan is indicated by the two large squares, the low type at the blue square getting q_L=1 and w_L=30, the high type at the green square getting q_H=1/3, w_H=50/3.

In this case, because p_H > 2/3, the best separable plan can be interim pareto-dominated by pooling plans. The large grey triangle at q=1, w=46 denotes the best pooling plan for the sellers. But if sellers expected get this pooling plan with buyers, then a buyers who offered the smaller green triangle (at q = 0.86, w = 41.8 = 48.60q) could hope to make a profit by attracting high sellers but not low sellers. Notice that this small green triangle is wedged in between two short lines that are parts of the indifference curves through the pooling plan (q=1,w=46) for the two types of seller, above the green indifference curve of the high type, below the blue indifference curve of the low type, and it is profitable for the buyer with high-type sellers, as it is below the green (highest) dotted line.

But it is above the grey dotted line, so it would become unprofitable if it attracted both types, as would happen if the grey-triangle offer were withdrawn.

Signalling. In the above example, suppose now that there is a signal that sellers could make which would cost $c_L>0$ for L-types but would cost $c_H>0$ for H-types.
Let us apply sequential equilibrium to a market game in which the informed sellers must make their choices about whether to get the costly signal before the buyers compete for them.
Suppose there are excess buyers in the market, and the terms of trade that these competitive buyers offer to any seller will depend only on what the buyers observe about the seller's signal.
So for each observable signal-value, the competitive buyers should just break even, paying exactly the value of the seller's asset conditional on his signal.

For what values of c_H and c_L could there be an equilibrium where the seller would choose to signal when his type is H but not when his type is L?
When the low type is separated, its competitive bids from buyers will be $\pi_L = 30$, and so type-L sellers will get expected gain $\pi_L - \theta_L = 30-20 = 10$ from their nonsignaling separation.
But signaling in this separating scenario becomes evidence of type H, and so those sellers who signal can get competitive bids of $\pi_H = 50$ from the buyers.
So for the low types to not signal, we must have $\pi_H - \theta_L - c_L \leq \pi_L - \theta_L$.
For the high types to signal, we must have $\pi_H - \theta_H - c_H \geq \max\{0, \pi_L - \theta_H\}$.
So this separating scenario is an equilibrium when $50-20-c_L \leq 30-20$ and $50-40-c_H \geq 0$,
that is, when $c_L \geq 20$ and $c_H \leq 10$.
For this equilibrium to be better for the H-sellers than the best separable plan that we found previously, we also need $50-40-c_H > (50-40)/3$, and so $c_H < 6.67$ (with $c_L \geq 20$).

Lagrangean analysis explains why the costly signal can change the sellers' best separable plan only when $c_L/c_H > 3$. The best separable plan here maximizes a Lagrangean (from the low-p_H case with $\alpha_{|L} = 0.5p_H$) in which the H-type's value of his asset is transformed from θ_H to the virtual value $\theta_H+(\theta_H-\theta_L)\alpha_{|L}/p_H = 1.5\theta_H-0.5\theta_L$. That is, in this Lagrangean, an H-type's virtual utility is 1.5 times the actual H utility minus 0.5 times the utility that an L type would get.
So the signal's virtual cost for H is $1.5c_H-0.5c_L$, which becomes a positive virtual benefit (a negative virtual cost) when $c_L > 3c_H > 0$.

For what values of c_H and c_L could there be an equilibrium where both types choose to signal?
If both types are expected to signal, then at best they could both sell the asset for $p_H\pi_H+p_L\pi_L$.
The worst that buyers could infer about a seller is that his type is L, in which case they would pay up to π_L=30 for his object. So a seller always has two alternatives to signaling: he can sell it for π_L=30, or he can keep the asset himself and get zero gains from trade.
Thus, for both types of seller to actually want to use this signal in an equilibrium, we must have $(p_H\pi_H+p_L\pi_L) - \theta_H - c_H \geq \max\{\pi_L-\theta_H, 0\}$ and $(p_H\pi_H+p_L\pi_L) - \theta_L - c_L \geq \max\{\pi_L-\theta_L, 0\}$.
In this example, this becomes $p_H 50+(1-p_H)30 - 40 - c_H \geq 0$, $p_H 50+(1-p_H)30 - 20 - c_L \geq 30-20$,
and so we must have $20p_H - 10 \geq c_H$ and $20p_H \geq c_L$.
The former condition cannot be satisfied by any positive c_H unless $p_H > 1/2$.
But when $p_H = 0.8$, for example, this pooling scenario is an equilibrium if $c_H \leq 6$ and $c_L \leq 16$.
This may be called a "rat-race equilibrium," as the costly signal does nobody any good.

A general model of markets with adverse selection

Consider a large market where sellers have private information, and each seller must be matched with one buyer, but each buyer can be matched with any number of sellers.
(We may think of the sellers as workers, and buyers as employers.)
Each seller privately knows his own type t, which is in some finite set of possible types T,
but a seller can misrepresent his type if he not given an incentive for honesty.
Let p(t) be the fraction of type-t sellers among new arrivals into the market over any interval of time.
Let M denote the set of possible terms of trade or transactions between a matched seller and buyer.
Each transaction m in M includes a full specification of the quantity sold and the price paid (and any required signals by the seller), but for simplicity we treat M here as a finite set (possibly very large).
Let $u_1(m,t)$ and $u_2(m,t)$ denote the utility values to the seller and buyer respectively of transaction m when the seller's type is t.
Suppose that M includes the null transaction m_0 for which $u_1(m_0,t) = 0 = u_2(m_0,t)$ $\forall t \in T$.

A <u>transaction plan</u> is a function from types to probability distributions over transactions, $\mu: T \rightarrow \Delta(M)$.
That is, to describe the anticipated plan of transactions in such a market, we may let $\mu(m|t)$ denote the probability that a seller of type t will get a transaction m in the market.
So transaction plan μ must satisfy the <u>probability constraints</u>:

[1] $\mu(m|t) \geq 0, \forall m \in M, \ \forall t \in T;$ and $\sum_{m \in M} \mu(m|t) = 1, \ \forall t \in T.$

The expected utility for a type-t seller in the plan μ may be written

$$U_1(\mu,t) = \sum_{m \in M} \mu(m|t) u_1(m,t).$$

Because customers can misrepresent their types, such a market plan must be <u>incentive compatible</u>, that is, it must satisfy the informational incentive constraints:

[2] $\sum_{m \in M} \mu(m|t) u_1(m,t) \geq \sum_{m \in M} \mu(m|s) u_1(m,t), \ \forall t \in T, \forall s \in T.$

Buyers should not expect to lose from these transactions. In the overall market, the <u>p-pooled break-even constraint</u> says that aggregate net profit of all buyers must be nonnegative:

[3] $U_2(\mu,p) = \sum_{t \in T} \sum_{m \in M} p(t) \mu(m|t) u_2(m,t) \geq 0.$

We say that a plan μ is a <u>feasible</u> with the type-distribution p (or <u>p-feasible</u>) iff μ satisfies [1]-[3].

Such a trading plan could be derived from informed sellers' choosing among a menu of buyers' offers.
In a competitive market, the buyers could announce offers specifying the transactions that they are willing to make with sellers, and then each seller could choose the offer that is best for his type.
For convexity, let us assume buyers can offer lotteries that randomly assign sellers to transactions in M according to any specified probability distribution in $\Delta(M) = \{\omega \in \mathbb{R}^M | \ \omega(m) \geq 0, \ \sum_{m \in M} \omega(m) = 1\}$.
Then the <u>menu</u> of offers available in the market would be some a subset of $\Delta(M)$, say $\Omega \subseteq \Delta(M)$.
The market trading plan would then be derived from the sellers' choices among the buyers' offers.
Sellers' type-dependent selections in a menu Ω would yield an incentive-compatible plan μ such that $\mu(\bullet|t) \in \operatorname{argmax}_{\omega \in \Omega} \sum_{m \in M} \omega(m) u_1(m,t), \ \forall t \in T$. Here $\mu(\bullet|t)$ denotes the vector $\mu(\bullet|t) = (\mu(m|t))_{m \in M}$.

We will consider the possibility that buyers might withdraw some offers, in response to a deviation.
If buyers were to withdraw some offers from the menu Ω, then the smaller menu $\Omega' \subseteq \Omega$ would yield a

new incentive-compatible plan μ' that cannot be better for any type: $U_1(\mu',t) \leq U_1(\mu,t)\ \forall t\in T$.
So a fallback from μ could be any incentive-compatible plan μ' such that $U_1(\mu',t) \leq U_1(\mu,t)\ \forall t\in T$.
Any such fallback could be implemented with menus $\Omega' = \{\mu'(\bullet|t)|\ t\in T\}$ and $\Omega = \Omega'\cup\{\mu(\bullet|t)|\ t\in T\}$.

If buyers lose money trading with some types, then individual buyers would want to find ways of avoiding those types that yield negative profits for buyers. Such incentives to manipulate the selection process will exist unless the market plan satisfies the following separate break-even constraints, which assert that buyers do not expect to lose with any type:

[4] $\qquad U_2(\mu,t) = \sum_{m\in M} \mu(m|t)u_2(m,t) \geq 0,\ \forall t\in T.$

We may say that a transaction plan μ is separable iff μ satisfies [1], [2], and [4].
Condition [4] implies that the p-pooled break-even constraint [3] is satisfied for any distribution p.

A separable plan $\bar{\mu}$ is best for the sellers iff, for every other separable plan μ, $U_1(\bar{\mu},t) \geq U_1(\mu,t)\ \forall t\in T$.
Fact 1. There exists a separable plan $\bar{\mu}$ that is best for the sellers of all types.

Proof. With M and T being finite sets, the set of separable plans that satisfy [1], [2], and [4] is closed, bounded, and convex. It is also nonempty because the plan μ_0 that always puts probability one on m_0 ($\mu_0(m_0|t)=1\ \forall t$) satisfies all the constraints.
Now pick any vector of positive weights $\lambda(t)>0$ for all types t. (We could let $\lambda(t)=p(t)$, for example.)
The plan $\bar{\mu}$ will be an optimal solution to the problem

[5] $\qquad$ choose μ to maximize $\sum_{t\in T} \lambda(t)\, U_1(\mu,t)$ subject to the constraints [1], [2], and [4].

Now to verify that an optimal solution of [5] is best for all types of seller, suppose to the contrary that there is some other separable plan $\hat{\mu}$ that some types prefer.
Let $S(\hat{\mu})$ denote the set of types that prefer this plan, so $S(\hat{\mu}) = \{s\in T: U_1(\hat{\mu}|s) > U_1(\bar{\mu}|s)\}$.
Then let μ be the plan that gives each type its preference among $\hat{\mu}$ and $\bar{\mu}$; that is:
$\mu(m|t) = \hat{\mu}(m|t)$ if $t \in S(\hat{\mu})$, and $\mu(m|t) = \bar{\mu}(m|t)$ otherwise.
It is straightforward to check that this μ is also a separable plan, and so it would be a strictly better solution to problem [5] if $S(\hat{\mu})$ were nonempty. QED

Let $\Delta(T)$ be the set of probability distributions over T, $\Delta(T) = \{r\in\mathbb{R}^T|r(t)\geq 0\ \forall t\in T, \sum_{t\in T} r(t)=1\}$.
For any r in $\Delta(T)$, a transaction plan μ is r-feasible iff it satisfies [1]-[3] with the type-distribution r replacing p in the buyers' break-even constraint [3]: $U_2(\mu,r) = \sum_{t\in T} \sum_{m\in M} r(t)\mu(m|t)u_2(m,t) \geq 0$.
Being separable, the plan $\bar{\mu}$ is r-feasible with any other type-distribution r in $\Delta(T)$.
An r-feasible plan μ is r-efficient for the informed sellers iff there exists no r-feasible plan μ^* such that $U_1(\mu^*,t) \geq U_1(\mu.t)\ \forall t\in T$ with a strict inequality $U_1(\mu^*,t) > U_1(\mu,t)$ for some type t.

Fact 2. There exists a type-distribution $\bar{r}$ in $\Delta(T)$ such that the best separable plan $\bar{\mu}$ is $\bar{r}$-efficient for the informed sellers.

Proof: The above problem [5] (with λ) that characterizes $\bar{\mu}$ is a linear programming problem.
So the optimal solution $\bar{\mu}$ of [5] is supported by nonnegative Lagrange multipliers (α,γ) such that $\bar{\mu}$ maximizes the following Lagrangean subject to only the probability constraints [1]

$$\mathcal{L}(\mu;\alpha,\gamma) = \textstyle\sum_{t\in T} \lambda(t)U_1(\mu,t) + \sum_{t\in T} \sum_{s\in T} \alpha(s|t) \sum_{m\in M} [\mu(m|t)-\mu(m|s)]u_1(m,t)$$
$$+ \textstyle\sum_{t\in T} \gamma(t) \sum_{m\in M} \mu(m|t)u_2(m,t),$$

and we have the following complementary slackness conditions

$\alpha(s|t) \geq 0$ and $\alpha(s|t) \sum_{m\in M} [\bar{\mu}(m|t) - \bar{\mu}(m|s)]u_1(m,t) = 0,\ \forall t\in T,\ \forall s\in T,$

$\gamma(t) \geq 0$ and $\gamma(t) [\sum_{m\in M} \bar{\mu}(m|t)u_2(m,t)] = 0,\ \forall t\in T.$

The complementary slackness conditions imply that the maximized value of the Lagrangean is just

$\mathcal{L}(\bar{\mu};\alpha,\gamma) = \sum_{t\in T} \lambda(t)\, U_1(\bar{\mu},t).$

Let $\Gamma = \sum_{t\in T} \gamma(t) \geq 0$. Choose $\bar{r}$ in $\Delta(T)$ so that $r(t)\Gamma = \gamma(t), \forall t\in T$.

(If Γ is 0 then let $\bar{r}$ be any distribution in $\Delta(T)$; otherwise, with $\Gamma\neq 0$, let $\bar{r}(t) = \gamma(t)/\Gamma\ \forall t$.)

If μ is feasible with the type-distribution $\bar{r}$, then we have $\sum_{m\in M} [\mu(m|t)-\mu(m|s)]u_1(m,t) \geq 0\ \forall t\ \forall s$,

and $\sum_{t\in T} \gamma(t) \sum_{m\in M} \mu(m|t)u_2(m,t) = \Gamma \sum_{t\in T} \bar{r}(t)\sum_{m\in M} \mu(m|t)u_2(m,t) \geq 0$.

So Lagrangean-optimality of $\bar{\mu}$ implies $\sum_{t\in T} \lambda(t)U_1(\bar{\mu},t) = \mathcal{L}(\bar{\mu};\alpha,\gamma) \geq \mathcal{L}(\mu;\alpha,\gamma) \geq \sum_{t\in T} \lambda(t)U_1(\mu,t)$.

With all $\lambda(t)>0$, if any s has $U_1(\mu,s)>U_1(\bar{\mu},t)$, then another type t must have $U_1(\mu,t) <U_1(\bar{\mu},t)$. QED.

Rothschild-Stiglitz (1976) showed that simple concepts of competitive equilibrium in markets with adverse selection may yield no equilibria in many cases. Wilson (1977) and Hellwig (1987) suggested that, to get equilibrium existence, we should admit that buyers could react against another buyer's deviation by withdrawing some offers that the deviation may make unprofitable.
We should not allow buyers to add new offers in reaction to a deviation, however, because such reactions could make any market uncompetitive. Buyers could sustain monopsonistic collusion if they could threaten to increase their offers against another buyer's deviation. But in a conventional market without adverse selection, a buyer's profit cannot be reduced by his competitors withdrawing offers, and so theats of offer-withdrawals could not be used to sustain collusion among the buyers.

In a market where buyers' offers yield a plan μ, suppose a buyer is considering a deviation that offers another incentive-compatible plan η that can attract at least one type t such that $U_1(\eta,t) > U_1(\mu,t)$.
In response to this deviation, the other buyers could withdraw some of their offers so that their reduced menu would yield a fallback plan μ' such that $U_1(\mu',t) \leq U_1(\mu,t)\ \forall t\in T$.
Seller types t with $U_1(\eta,t) \leq U_1(\mu',t)$ would have no incentive to select into the deviation η, and the deviating buyer could send any such types away to trade with the other buyers who offer μ'.
So the types selecting η could have a distribution r such that $r(t)=0$ unless $U_1(\eta,t) > U_1(\mu',t)$.

A plan μ is sustainable with adverse selection iff μ is feasible with the given type-distribution p, and, for every incentive-compatible deviation η such that $\{t\in T|\ U_1(\eta,t) > U_1(\mu,t)\} \neq \emptyset$, there exists a p-feasible fallback plan μ', satisfying $U_1(\mu'|t) \leq U_1(\mu|t)\ \forall t\in T$, and there exists a type-distribution r such that $\{t\in T|\ r(t)>0\} \subseteq \{t\in T|\ U_1(\eta,t) > U_1(\mu',t)\}$ and η is not r-feasible.
When η is incentive compatible, not being r-feasible means $U_2(\eta,r) < 0$.

Fact 3. A p-feasible plan μ is sustainable with adverse selection iff $U_1(\mu,t) \geq U_1(\bar{\mu},t)\ \forall t\in T$, where $\bar{\mu}$ is the best separable plan. Any such μ can be sustained against a deviation η by the fallback $\mu' = \bar{\mu}$ and distribution r such that any t in $S = \{s\in T|\ U_1(\eta,s)>U_1(\bar{\mu},s)\}$ has probability $r(t) = \bar{r}(t)/\sum_{s\in S} \bar{r}(s)$.

Proof. If there is any type t such that $U_1(\mu,t) < U_1(\bar{\mu},t)$, then μ cannot be sustainable, because the best separable plan $\bar{\mu}$ is a deviation that would be feasible with any type-distribution r.
Now consider any μ such that $U_1(\mu,t) \geq U_1(\bar{\mu},t)$ $\forall t \in T$, so that $\bar{\mu}$ can be a fallback from μ.
Consider any incentive-compatible deviation η such that $\{s \in T | U_1(\eta,s) > U_1(\mu,s)\} \neq \emptyset$.
Let $S = \{s \in T | U_1(\eta,s) > U_1(\bar{\mu},s)\}$.
Define the plan $\tilde{\eta}$ such that $\tilde{\eta}(m|t) = \eta(m|t)$ if $t \in S$, and otherwise $\tilde{\eta}(m|t) = \bar{\mu}(m|t)$.
It is straightforward to verify that $\tilde{\eta}$ must be incentive compatible for the sellers, because it could be implemented by letting each type of seller choose between the incentive-compatible plans η and $\bar{\mu}$.
We have $U_1(\tilde{\eta},t) > U_1(\mu,t) \geq U_1(\bar{\mu},t)$ for all t in S, $U_1(\tilde{\eta},t) = U_1(\bar{\mu},t)$ for all other t.
So $\tilde{\eta}$ is Pareto-superior to the best separable plan $\bar{\mu}$ for the informed sellers.
Thus, by Fact 2, $\tilde{\eta}$ cannot be feasible with the type distribution $\bar{r}$.
That is, we must have $\sum_{t \in T} \bar{r}(t) \sum_{m \in M} \tilde{\eta}(m|t)u_2(m,t) < 0$.
In this negative sum, the terms for types that choose $\bar{\mu}$ are all nonnegative, and the other types are all in S where $\tilde{\eta}$ is the same as η. Thus, $\sum_{t \in S} \bar{r}(t) \sum_{m \in M} \eta(m|t)u_2(m,t) < 0$.
So some t in S must have $\bar{r}(t) > 0$. Then we can let $r(t) = \bar{r}(t)/\sum_{s \in S} \bar{r}(s)$ for all t in S, and let $r(t) = 0$ for all other t. So $U_2(\eta,r)<0$ and η is not r-feasible, as required for μ to be sustained against η. QED.

Reaction to deviations may not actually be needed to sustain a plan μ such that $U_1(\mu,t) \geq U_1(\bar{\mu},t)$ $\forall t$.
We are assuming that sellers are continuously flowing into the market with the type-distribution p.
But when sellers apply for a transaction in μ, a small (infinitesimal) fraction of them could be put into a special wait list from which they will be matched according to the best separable plan $\bar{\mu}$.
Because $U_1(\mu,t) \geq U_1(\bar{\mu},t)$ for all t, sellers in this wait list would be more eager than others of the same type to trade with any buyer who deviates from the μ system.
Different types could have different probabilities of entering the wait list (after revealing types while expecting the incentive-compatible μ); so the type-distribution in the wait list could be $\bar{r}$ from Fact 2.
So if all sellers are matched with buyers quickly except for those going on the wait list, then the stock of sellers who are available in the market at any point in time may consist almost entirely of the wait list with plan $\bar{\mu}$ and type-distribution $\bar{r}$, from which no deviation η can profitable sample (by Fact 2).
The key idea here, that the stock of traders who are actually available at any point in time may differ adversely from the flow of traders who enter the market over any interval of time, is really a version of Gresham's law: that bad types may circulate more than good types.
See also R. Myerson, "Sustainable matching plans with adverse selection," *Games and Economic Behavior* 9:35-65 (1995); or *Game Theory* section 10.9.

Finding the distribution $\bar{r}$ for an example.

Recall again the example where the seller's cost-types are θ_L=20 and θ_H=40, the corresponding buyer's values are π_L=30 and π_H=50.

The set of possible transactions is the set of all (q,w) in M = [0,1]×ℝ. This is an infinite set, but the linearity of utility functions will make the problem [5] that characterizes the best separable plan a linear programming problem as follows: Choose (q_H,w_H,q_L,w_L) to

maximize $\lambda_H(w_H-40q_H) + \lambda_L(w_L-20q_L)$ subject to

$$\begin{array}{ll} w_H-40q_H - (w_L-40q_L) \geq 0 & [\alpha_{|H}] \\ w_L-20q_L - (w_H-20q_H) \geq 0 & [\alpha_{|L}] \\ 50q_H-w_H \geq 0 & [\gamma_H] \\ 30q_L-w_L \geq 0 & [\gamma_L] \\ 1\geq q_H \geq 0,\ \ 1\geq q_L \geq 0. & \end{array}$$

Lagrangean conditions for w_H and w_L yield $0 = \lambda_H + \alpha_{|H} - \alpha_{|L} - \gamma_H$, $0 = \lambda_L + \alpha_{|L} - \alpha_{|H} - \gamma_L$.

To support a separable plan with $1 > q_H > 0$ and slack in the |H-incentive constraint, we must have $0 = 50\gamma_H + 20\alpha_{|L} - 40(\lambda_H+\alpha_{|H})$ and $\alpha_{|H} = 0$.

These four equations give us

$\alpha_{|L} = (1/3)\lambda_H$, $\gamma_H = \lambda_H-\alpha_{|L} = (2/3)\lambda_H$, $\gamma_L = \lambda_L + \alpha_{|L} = \lambda_L+(1/3)\lambda_H$.

Then q_L has Lagrangean cocfficient $30\gamma_L + 40\alpha_{|H} - 20(\lambda_L+\alpha_{|L}) = 10\lambda_L + (10/3)\lambda_H > 0$, for q_L=1.

From above, these Lagrange multipliers yield an $\bar{r}$ distribution that sustains the best separable plan:

$\bar{r}_H = \gamma_H/(\gamma_H+\gamma_L) = (2/3)/(\lambda_L/\lambda_H + 1)$, $\bar{r}_L = \gamma_L/(\gamma_H+\gamma_L) = 1-\bar{r}_H$.

As the given λ_L/λ_H can range from 0 to infinity, this can yield any $\bar{r}_H$ between 0 and 2/3.

So we confirm that the seller's best separable plan becomes incentive-efficient when the fraction of H-sellers is less than 2/3.

Motivating an agent with a linear type drawn from a continuous distribution on an interval.

Suppose that the agent's type $\mathbf{t}$ is a random variable drawn from an interval [A,B].

The agent's type t is his cost of effort, and his utility for income w and effort q is $w - t\,q$.

Consider any contract $(w(\bullet),q(\bullet))$ where the terms of trade for each type θ would be $(w(t),q(t))$.

Let $U(w,q|t) = w(t) - tq(t)$ denote the expected utility of type t under this contract.

For any pair of possible types t and s in [A,B], the (s|t)-informational incentive constraint says

$U(w,q|t) = w(t) - t\,q(t) \geq w(s) - t\,q(s) = U(w,q|s) + (s-t)q(s)$.

Similarly, the (t|s)-incentive constraint implies $U(w,q|s) \geq U(w,q|t) + (t-s)q(t)$

So the (t|s) and (s|t) constraints together imply $(s-t)q(t) \geq U(w,q|t)-U(w,q|s) \geq (s-t)q(s)$.

So when $s > t$ we must have $q(t) \geq q(s)$, and so q(t) is a decreasing function of the cost-type t.

These inequalities over many small steps from t up to B yield the information-rent equation:

$U(w,q|t) = U(w,q|B) + \int_t^B q(r)\,dr$.

(Seller type B has the least motivation to trade.) The expected income of any type t is then

$w(t) = U(w,q|t) + tq(t) = U(w,q|B) + \int_t^B q(r)\,dr \ + tq(t) = U(w,q|B) + q(B)B + \int_{q(B)}^{q(t)} q^{-1}(\gamma)\,d\gamma$.

With this $w(\bullet)$, we get $U(w,q|t)-[w(s)-tq(s)] = U(w,q|t)-U(w,q|s)-(s-t)q(s) = \int_t^s [q(r)-q(s)]dr$,

which is always nonnegative (verifying incentive compatibility) if q(s) is weakly decreasing in s.

Suppose the principal's beliefs about the agent's type are described by the cumulative distribution

$F(t) = P(\tilde{t} \leq t)$, and $f(t) = F'(t)$ is the continuous probability density of this distribution,

with $f(t)>0$ for all t in [A,B]. Here $F(B)=1$, $F(A)=0$, and $P(a\leq\mathbf{t}\leq b) = F(b)-F(a)=\int_a^b f(t)\,dt$

whenever $a \leq b$. Then the expected wage bill is

$\int_A^B w(t)\,f(t)\,dt = \int_A^B [U(w,q|B) + tq(t) + \int_t^B q(r)\,dr]f(t)\,dt$

$\quad = U(w,q|B) + \int_A^B t\,q(t)\,f(t)\,dt + \int_A^B \int_A^r f(t)\,dt\,q(r)\,dr$

$\quad = U(w,q|B) + \int_A^B t\,q(t)\,f(t)\,dt + \int_A^B F(r)\,q(r)\,dr = U(w,q|B) + \int_A^B q(t)\,[t + F(t)/f(t)]\,f(t)\,dt$.

So the incentive-compatible expected wage $E(w(\mathbf{t}))$ looks like what the principal would have to pay without incentive constraints if the cost of each type t were increased to a virtual cost $t+F(t)/f(t)$.

This virtual-cost formula expresses the fact that, when we ask more effort from any type t, we increase the amount that we must pay all types below t, because of incentive constraints.

Example: Akerlof's Lemons. The "agent" is the seller of a unique object, of which the "principal" is the only potential buyer. The seller's type is the value of the object to him, which depends on his unverifiable private information about its quality. Then q(t) can be reinterpreted as the probability of his selling the good if he acts like type t, which must satisfy $0 \leq q(t) \leq 1$, and w(t) is his expected revenue from selling if he acts like type t.

Suppose $\mathbf{t}$ is drawn from a Uniform distribution on the interval from 0 to 100, but the value of the object to the buyer also depends on the quality (which the buyer would learn only after the transaction) and would be $1.5\mathbf{t}$. So the object would always be worth 50% more to the buyer.

If (w,q) satisfies the incentive constraints and $U(w,q|t)\geq 0$, the buyer's expected gain from trade is

$\int_0^{100} [1.5tq(t)-w(t)]f(t)dt = \int_0^{100} [1.5t-t-F(t)/f(t)]q(t)f(t)dt-U(w,q|100)$

$= \int_0^{100} [1.5t-2t]q(t)dt/100 - U(w,q|100) \leq 0$. The buyer can only expect to lose if any $q(t)>0$.

<u>A continuous example:</u> Now suppose that θ is drawn from a Uniform distribution on [1,2], keeping everything else the same as in the previous example, with $\pi(q|\theta) = (1+\theta)q^{0.5}$.

So $F(\theta) = (\theta-1)/(2-1)$, $f(\theta) = 1/(2-1) = 1$, $F(\theta)/f(\theta) = \theta-1$, for any θ in [1,2].

Then the principal's expected gains from trade are

$\int_1^2 [\pi(q(\theta)|\theta) - w(\theta)]\, f(\theta)\, d\theta = \int_1^2 \{\pi(q(\theta)|\theta) - q(\theta)[\theta+F(\theta)/f(\theta)]\}f(\theta)d\theta - U(w,q|2)$

$= \int_1^2 [(\theta+1)q(\theta)^{0.5} - q(\theta)(2\theta-1)]\, d\theta - [w(2)-2q(2)]$.

To maximize the integrand at every θ, we want

$0 = 0.5(\theta+1)q^{-0.5} - (2\theta-1)$, which yields $q(\theta) = [0.5(\theta+1)/(2\theta-1)]^2$.

Because this $q(\theta)$ is monotone decreasing in θ, we know that it is actually feasible.

[If this $q(\theta)$ were increasing over any part of the interval, then it would not be feasible, and we would have an "irregular" case which is more complicated to solve (e.g: Myerson, 1981).]

At the extreme types, we get $q(1) = 1$, $q(2) = 0.25$.

To get $U(w,q|2) = 0$, the optimal solution must have $w(2) = 2q(2) = 2\times0.25 = 0.5$.

Then the information-rent equations give us

$U(w,q|\theta) = \int_\theta^2 q(t)\, dt = 0.25[0.25t + 0.75\mathrm{LN}(2t-1) - 2.25/(4t-2)]\big|_\theta^2$,

and $w(\theta) = \theta q(\theta) + \int_\theta^2 q(t)\, dt$.

In particular, we get $U(w,q|1) = 0.456$ and $w(1) = 1.456$.

(*Recall:* In the discrete case where types $\theta_L=1$ and $\theta_H=2$ each had probability 1/2, the optimal solution also had $q(1)=1$, $q(2)=0.25$, and $w(2)=0.5$, but it had $w(1)=1.25$, so type 1 got a smaller information rent.)

Now let's do the analogous result for the case where **the agent is a buyer** of some object.

Here the agent's type t is interpreted as his valuation of the object.

Now let $q(t)$ denote the probability of the agent buying the object if her type is t, and let $w(t)$ denote expected amount that the agent will have to pay if her type is t.

(If $\hat{w}(t)$ denotes the price that the type-t agent will pay if she buys, and if she would pay nothing if she does not buy the object, then our $w(t)$ is equal to $q(t)\, \hat{w}(t)$.)

So the expected gains from trade for a type-t buyer are $U(w,q|t) = t\, q(t) - w(t)$.

An incentive compatible trading plan must satisfy, for all types s and t in the interval [A,B],

$U(w,q|t) = t\, q(t) - w(t) \geq t\, q(s) - w(s) = U(w,q|s) + (t-s)q(s)$.

Assuming differentiability, $0 = \partial/\partial s\, [tq(s)-w(s)]\big|_{s=t}$, so $w'(t) = tq'(t)$, and $w''(t) = tq''(t)+q'(t)$.

Then $0 \geq \partial^2/\partial s^2\, [tq(s)-w(s)]\big|_{s=t} = tq''(t)-w''(t) = -q'(t)$, and so a buyer's $q(\bullet)$ must be increasing.

Notice $\partial/\partial s\, [tq(s)-w(s)] = tq'(s)-w'(s) = (t-s)q'(s)$. With $q'\geq0$, this is ≤0 if $s>t$, ≥0 if $s<t$.

Then the envelope theorem yields $U'(w,q|t) = d/dt\, [tq(t)-w(t)] = q(t)$,

and so we get the buyer's <u>information-rent equation</u>: $U(w,q|t) = U(w,q|A) + \int_A^t q(s)\, ds$.

(Buyer type A has the least motivation to trade.) So the expected payment from any type t is

$w(t) = tq(t) - U(w,q|t) = t\, q(t) - \int_A^t q(s)\, ds - U(w,q|A) = q(A)A + \int_{q(A)}^{q(t)} q^{-1}(\gamma)d\gamma - U(w,q|A)$.

The overall expected payment from the buyer, before her type is known, is

$\int_A^B w(t)\, f(t)\, dt = \int_A^B [t\, q(t) - \int_A^t q(s)\, ds - U(w,q|A)]f(t)\, dt$

$= \int_A^B tq(t)f(t)dt - \int_A^B \int_s^B f(t)dtq(s)ds - U(w,q|A) = \int_A^B q(t)\, [t - (1-F(t))/f(t)]\, f(t)\, dt - U(w,q|A)$.

Facts about Uniform distributions. Suppose that $\mathbf{X}$ is a random variable drawn from a Uniform distribution on the interval from A to B, where $A < B$. Then $E(\mathbf{X}) = (A+B)/2$, and $\forall\theta\in[A,B]$:
$F(\theta) = P(\mathbf{X}\leq\theta) = P(\mathbf{X}<\theta) = (\theta-A)/(B-A)$, $f(\theta) = F'(\theta) = 1/(B-A)$, $1-F(\theta) = (B-\theta)/(B-A)$,
$E(\mathbf{X}|\mathbf{X}\leq\theta) = E(\mathbf{X}|\mathbf{X}<\theta) = (A+\theta)/2$, $E(\mathbf{X}|\mathbf{X}\geq\theta) = E(\mathbf{X}|\mathbf{X}>\theta) = (\theta+B)/2$,
$\theta + F(\theta)/f(\theta) = \theta + (\theta-A) = 2\theta-A$, $\theta - (1-F(\theta))/f(\theta) = \theta - (B-\theta) = 2\theta-B$.

Revenue equivalence in auctions Let $\{1,2,...,n\}$ be n bidders to buy an object, and assume that each bidder i has an independent private value for this object $\mathbf{t}_i$ that is drawn from a Uniform distribution on an interval A_i to B_i ($A_i < B_i$). Let unsubscripted $\mathbf{t} = (\mathbf{t}_1,...,\mathbf{t}_n)$ denote the profile of n bidders' types. Let $Q_i(t) = Q_i(t_1,...,t_n)$ denote the conditional probability of i winning the object given that the type-profile $\mathbf{t}$ is equal to t, and let $W_i(t)$ denote the conditional expected payment from i given the same type-profile t. Let $q_i(t_i) = E(Q_i(\mathbf{t}_{-i},t_i)) = E(Q_i(\mathbf{t})|\mathbf{t}_i=t_i)$ and $w_i(t_i) = E(W_i(\mathbf{t}_{-i},t_i)) = E(W_i(\mathbf{t})|\mathbf{t}_i=t_i)$ denote the conditional probability of i winning and i's conditional expected payment given i's own type t_i, but not knowing the others' independent types. With the Uniform $[A_i,B_i]$ distribution, $t_i - (1-F_i(t_i))/f_i(t_i) = 2t_i-B_i$. The seller's total expected revenue from the auction is:
$\sum_i E(W_i(\mathbf{t})) = \sum_i E(w_i(\mathbf{t}_i)) = \sum_i E\{q_i(\mathbf{t}_i)[\mathbf{t}_i-(1-F_i(\mathbf{t}_i))/f_i(\mathbf{t}_i)]\} - \sum_i U_i(W,Q|A_i)$
$= \sum_i E[q_i(\mathbf{t}_i)\,(2\mathbf{t}_i-B_i)] - \sum_i U_i(W,Q|A_i) = E\,[\sum_i Q_i(\mathbf{t})\,(2\mathbf{t}_i-B_i)] - \sum_i U_i(W,Q|A_i)$.
Consider a symmetric case where all A_i=A and all B_i=B. Any auction that always delivers the object to the bidder who actually values it most would yield $q_i(t_i) = ((t_i-A)/(B-A))^{n-1}$ for all t_i in [A,B]. If it allows no profit to the minimal-value type-A bidders $U_i(W,Q|A)=0$, then these facts determine the seller's expected revenue, regardless of the other details of the auction.
But the formula for expected revenue that we have derived above would be maximized, subject to the constraints that $\sum_i Q_i(t) = 1$ $\forall t$ and all $Q_i(t)\geq 0$, by the seller keeping the object when $\max_i 2\mathbf{t}_i-B < 0$, but otherwise giving it to the bidder with highest $\mathbf{t}_i$.

Bilateral bargaining between a seller and a buyer for a single object
The seller is i=1, the buyer is i=2. Each individual i has an independent private value $\mathbf{t}_i$ for the object that is drawn from some probability distribution with cumulative probability distribution $F_i(t_i)$ and positive density $f_i(t_i)$ on the interval $[A_i,B_i]$.
Let $Q(t) = Q(t_1,t_2)$ denote the conditional probability of trade given their types t_1 and t_2
Let $q_i(t_i)$ denote the conditional probability of trade given that i's type is t_i.
Here $q_1(t_1) = E(Q(t_1,\mathbf{t}_2))$ must be decreasing in t_1, and $q_2(t_2) = E(Q(\mathbf{t}_1,t_2))$ must be increasing in t_2.
Let $w_1(t_1)$ be the conditional expected payment to the seller given that his type is t_1.
Let $w_2(t_2)$ be the conditional expected payment from the buyer given that her type is t_2.
Incentive compatibility implies that $E(w_1(\mathbf{t}_1)) = E(q_1(\mathbf{t}_1)[\mathbf{t}_1+F_1(t_1)/f_1(\mathbf{t}_1)]) + U_1(B_1)$,
and $E(w_2(\mathbf{t}_2)) = E(q_2(\mathbf{t}_1)[\mathbf{t}_2-(1-F_2(\mathbf{t}_2))/f_2(\mathbf{t}_2)]) - U_2(A_2)$.
The difference $\Delta = E[w_1(\mathbf{t}_1)) - E(w_2(\mathbf{t}_2))]$ is the expected net subsidy (if any).
So interim participation constraints yield the following inequality, from Myerson-Satterthwaite:
$0 \leq U_1(B_1) + U_2(A_2) = \Delta + E\{Q(\mathbf{t}_1,\mathbf{t}_2)([\mathbf{t}_2-(1-F_2(\mathbf{t}_2))/f_2(\mathbf{t}_2)] - [\mathbf{t}_1+F_1(\mathbf{t}_1)/f_1(\mathbf{t}_1)]\}$
When each $\mathbf{t}_i$ is Uniform $[A_i,B_i]$, for no subsidy (Δ=0) we need $0 \leq E\{Q(\mathbf{t}_1,\mathbf{t}_2)[2(\mathbf{t}_2-\mathbf{t}_1)-(B_2-A_1)]\}$

A Uniform bilateral trading problem like Akerlof's

Let's consider a bilateral trading problem where agent 1 is the seller of some unique object which he owns, and agent 2 is the only possible buyer of this object.

Depending on the object's quality, it may be worth as little as \$40 to agent 1 and \$60 to agent 2 (if its quality is low) or as much as \$100 to agent 1 and \$120 to agent 2 (if its quality is high).

Player 1 knows the quality of the object. Let 1's cost type $\tilde{t}_1$ is his value of keeping the object.

With any quality, the object would be worth \$20 more to player 2 than to player 1.

That is, given 1's type $\tilde{t}_1$, the value of the object to player 2 would be $g(t_1) = \tilde{t}_1+20$.

Player 2's belief about $\tilde{t}_1$ is described by a Uniform distribution on the interval \$40 to \$100.

Game where buyer bids Suppose first that agent 2 can offer to buy for any positive price r, and then agent 1 will accept or reject the offer. If the offer is rejected then they each get profit 0.

If the offer is accepted then 1's profit is $r-\tilde{t}_1$ and 2's profit is $g(\tilde{t}_1)-r$.

In a subgame-perfect equilibrium, agent 1 will accept if $\tilde{t}_1 < r$, but agent 1 will reject if $\tilde{t}_1 > r$.

Agent 2's expected profit from offering any price r is $Y(r) = P(\tilde{t}_1<r)\ [E(g(\tilde{t}_1)|\ \tilde{t}_1<r) - r]$.

For any number r between 40 and 100, this expected profit is

$Y(r) = P(\tilde{t}_1<r)\ [E(\tilde{t}_1+20|\ \tilde{t}_1<r) - r] = P(\tilde{t}_1<r)\ [(E(\tilde{t}_1|\ \tilde{t}_1<r) + 20 - r] =$

$= [(r-40)/(100-40)][(40+r)/2 + 20 - r] = (r-40)(80-r)/120 = (-3200 +120r - r^2)/120.$

This quadratic formula is maximized by letting $r = 60$.

(The buyer cannot gain by bidding less than 40 or more than 100, because a bid below 40 would be surely rejected, and a bid above 100 would be worse than the surely-accepted bid of 100.)

So in the unique subgame-perfect equilibrium of this game, agent 2 offers to buy for \$60, and agent 1 accepts if $\tilde{t}_1 < 60$. The probability of trade is $P(\text{trade}) = (60-40)/(100-40) = 1/3$.

Game where seller bids Suppose now that agent 1 can offer to buy for any positive price r, and then agent 2 will accept or reject the offer. If the offer is rejected then they each get profit 0.

If the offer is accepted, then 1's profit is $r-\tilde{t}_1$ and 2's profit is $g(\tilde{t}_1)-r$.

In this game, the price is named by the agent who has private information, and so signaling effects give us many equilibria.

We may also reinterpret this as a market, with many sellers (1) and enough identical buyers (2) to buy the entire supply, where each seller 1, knowing his own type, publicly commits to a price that he chooses. Then he must sell at this price (if any buyer accepts his bid) or keep the object.

Let's look first for an equilibrium where there is some price r such that agent 2 would surely accept an offer to sell for r but would surely reject an offer to sell for any price higher than r.

In this equilibrium, agent 1 will offer r if $\tilde{t}_1 < r$.

For agent 2 to accept the offer r, 2's expected profit from accepting r must not be negative,

so $0 \le E(g(\tilde{t}_1)|\ \tilde{t}_1<r) - r = E(\tilde{t}_1+20|\tilde{t}_1<r) - r = (40+r)/2 + 20 - r$, which implies $r \le 80$.

For agent 2 to reject any offer to sell at a price higher than r, such a trade must be unprofitable for agent 2 when she makes the worst inference about agent 1, which is that his type is 40, in which case the object would be worth $40+20 = \$60$ to agent 2.

So we can construct such an equilibrium for any r such that $60 \le r \le 80$.

In such an equilibrium, types higher than r may be expected to make some offer higher than 120, which agent 2 could never profitably accept.

An offer between r and 120 may be rejected by agent 2 because this surprise offer may lead agent 2 to believe that 1's type is 40, in which case the object is only worth 60 to agent 2. Among these almost-pooling equilibria, agent 1 most prefers the equilibrium with r = 80.
In this equilibrium, the probability of trade is $\Pr(\text{trade}) = \Pr(\tilde{t}_1<80) = (80-40)/(100-40) = 2/3$.
<u>Reinterpretation in market</u>: We'd get excess demand if r<80, and the market clears only at r=80.
So an uninformed Walrasian auctioneer who posts a price to clear the market would choose r=80.

There are many other equilibria where different types of seller (1) choose different prices.
Let's look for an equilibrium in which some types of agent 1 would offer to sell for \$70, but all higher types would offer to sell for \$100, and agent 2 would be sure to accept \$70 but her probability of accepting \$100 would be between 0 and 1. To find this equilibrium, we have two unknowns to find: let q denote the probability that agent 2 would accept an offer of \$100, and let θ denote the highest type of agent 1 that would offer \$70.
For agent 2 to be willing to randomize between accepting and rejecting \$100, her expected profit from accepting it must be 0, and so
$0 = E(g(\tilde{t}_1)|\tilde{t}_1>\theta) - 100 = E(\tilde{t}_1+20|\tilde{t}_1>\theta) - 100 = (\theta+100)/2 + 20 - 100$, and so $\theta = 60$.
For agent 1 to offer \$70 below when his type is below θ but \$100 when his type is above θ, we need that $70-t_1 \geq q(100-t_1)$ when $t_1<\theta$, and $70-t_1 \leq q(100-t_1)$ when $t_1>\theta$.
These inequalities imply $70-\theta = q(100-\theta)$, and so $q = (70-60)/(100-60) = 1/4$.
In this equilibrium, $\Pr(\text{trade}) = \Pr(\tilde{t}_1<\theta) + \Pr(\tilde{t}_1>\theta)q = (20/60) + (40/60)(1/4) = 1/2$.
This is also a <u>market equilibrium</u>, with no excess demand at either price: $E(g(\tilde{t}_1)|\tilde{t}_1\leq\theta) = 70$.

There is a separating equilibrium in which each possible type t_1 of agent 1 would offer to sell for $r(t_1) = t_1+20$, and the probability of agent 2 accepting would depend on the offer r according to the formula $Q(r) = e^{-(r-60)/20}$, for any $r \geq 60$. The derivation is as follows:
For $r = t_1+20$ to maximize $Q(r)(r-t_1)$, we need $0 = Q'(r)(r-t_1) + Q(r)$ when $r=t_1+20$, and so $-1/20 = Q'(r)/Q(r) = d/dr\, \mathrm{LN}(Q(r))$.
In our usual notation for incentive-compatible trading plans, this separating equilibrium has, for any seller's type t, the type-conditional probability of trade $q(t) = Q(t+20) = e^{-(t-40)/20}$
and the type-conditional expected payment $w(t) = (t+20)q(t) = (t+20)e^{-(t-40)/20}$.
This function q() could also be derived from the incentive constraints, for all t in [40,100]:
$w(t)-tq(t) = \max_s w(s)-tq(s) = \max_s q(s)(s+20-t)$, which yields $0 = q'(t)(t+20-t) - q(t)1$.
The low-end boundary condition q(40)=1 identifies the best separating equilibrium.

If we compare different types expected gains from trade in these three equilibria, we find that low types ($t_1\leq 73$) prefer the pooling equilibrium where all trade occurs at the price 80, middle types ($74\leq t_1\leq 94$) prefer the two-price equilibrium where trade occurs at 70 or 100, and high type ($t_1\geq 95$) prefer the separating equilibrium. But the separating equilibrium is actually interim pareto-dominated by the following semi-separating equilibrium:

Low types in [40,60] are pooled together and sell for \$70 with probability q=1,
but any higher type t>60 separates and sells with a probability q(t)<1 at price t+20.
For t=60 to be indifferent between selling for price 70 with probability 1 or selling for price 60+20 with a lower probability, that probability must be q(60)=0.5. Then as before, we get $q'(t)/q(t) = -1/20$ and so $q(t) = 0.5e^{-(t-60)/20}$. (Compare: $e^{-(t-40)/20} = 0.368 < 0.5$.)

Mechanism design with informational incentive constraints: a general framework
$i,j \in \{1,...,I\} = \{\text{agents}\}$. $x \in X = \{\text{feasible allocations}\}$. $s_i,\theta_i \in \Theta_i = \{\text{possible types of agent } i\}$.
$\theta = (\theta_1,...,\theta_I) = (\theta_{-i},\theta_i) \in \Theta = \Theta_1\times...\times\Theta_I$. $\theta_{-i} \in \Theta_{-i} = \times_{j\neq i}\,\Theta_j$.
We restrict our attention to the case where agents' types are independent random variables.
Let p_i be the probability distribution of i's type in Θ_i. $p(\theta) = \prod_{i\in N} p_i(\theta_i)$. $p_{-i}(\theta_{-i}) = \prod_{j\neq i} p_j(\theta_j)$.
$u_i{:}X\times\Theta - \mathbb{R}$ is i's utility function.
A <u>direct-revelation mechanism</u> is any mapping from type-profiles to allocations $\mu{:}\Theta\to X$.
$U_i(\mu|\theta_i) = \sum_{\theta_{-i}\in\Theta_{-i}} p_{-i}(\theta_{-i})u_i(\mu(\theta),\theta)$. $\hat{U}_i(\mu,s_i|\theta_i) = \sum_{\theta_{-i}\in\Theta_{-i}} p_{-i}(\theta_{-i})u_i(\mu(\theta_{-i},s_i),\theta)$.
μ is <u>incentive compatible</u> iff $\forall i\in\{1,...,I\}, \forall\theta_i\in\Theta_i, \forall s_i\in\Theta_i,\ U_i(\mu|\theta_i) \geq \hat{U}_i(\mu,s_i|\theta_i)$.

A generalized mechanism is of the form $\gamma{:}S_1\times...\times S_I\to X$.
An equilibrium of such γ is of the form $(\sigma_1{:}\Theta_1\to S_1,...,\sigma_I{:}\Theta_I\to S_I)$.
For any equilibrium σ of any generalized mechanism γ, $\mu(\theta) = \gamma(\sigma_1(\theta_1),...,\sigma_I(\theta_I))$ defines an equivalent incentive-compatible direct-revelation mechanism. So without loss of generality, we may restrict our attention to incentive compatible mechanisms (<u>revelation principle</u>).

A mechanism μ is <u>(strongly) interim (Pareto-)dominated</u> by another mechanism $\hat{\mu}$ iff
$U_i(\hat{\mu}|\theta_i) > U_i(\mu|\theta_i),\ \forall i\in\{1,...,I\}, \forall\theta_i\in\Theta_i$.
$\mu{:}\Theta\to X$ is <u>(weakly) (interim) incentive efficient</u> iff μ is incentive compatible and μ is not interim Pareto-dominated by any other incentive-compatible mechanism $\hat{\mu}{:}\Theta\to X$.

<u>Fact</u> $\bar{\mu}$ is interim incentive-efficient iff we can find nonnegative weights $\lambda_i(\theta_i)$ for all types θ_i of all agents i such that: at least some $\lambda_i(\theta_i)>0$, and μ is an optimal solution to the problem
maximize$_{\mu:\Theta\to X}$ $\sum_{i\in\{1,...,I\}}\sum_{\theta_i\in\Theta_i}\lambda_i(\theta_i)U_i(\mu|\theta_i)$
subject to $U_i(\mu|\theta_i) \geq \hat{U}_i(\mu,s_i|\theta_i)\ \forall i\in\{1,...,I\}, \forall\theta_i\in\Theta_i, \forall s_i\in\Theta_i$. ($s_i|\theta_i$ incentive constraint)

The Lagrangean of this problem (with constraint multipliers $\alpha_i(s_i|\theta_i)\geq 0, \forall i, \forall\theta_i, \forall s_i$) is
$\mathcal{L}(\mu;\alpha) = \sum_{i\in\{1,...,I\}}\sum_{\theta_i\in\Theta_i}\lambda_i(\theta_i)U_i(\mu|\theta_i) + \sum_{i\in\{1,...,I\}}\sum_{\theta_i\in\Theta_i}\sum_{s_i\in\Theta_i}\alpha_i(s_i|\theta_i)(U_i(\mu|\theta_i)-\hat{U}_i(\mu,s_i|\theta_i))$
$= \sum_\theta p(\theta)\sum_i [\lambda_i(\theta_i)\,u_i(\mu(\theta),\theta) + \sum_{si}\alpha_i(s_i|\theta_i)(u_i(\mu(\theta),\theta) - u_i(\mu(\theta_{-i},s_i),\theta)]/p_i(\theta_i)$
$= \sum_\theta p(\theta)\{\sum_i [\lambda_i(\theta_i)+\sum_{si}\alpha_i(s_i|\theta_i)]u_i(\mu(\theta),\theta) - \sum_{si}\alpha_i(s_i|\theta_i)u_i(\mu(\theta_{-i},s_i),\theta)\}/p_i(\theta_i)$
$= \sum_\theta p(\theta)\{\sum_i [\lambda_i(\theta_i)+\sum_{si}\alpha_i(s_i|\theta_i)]u_i(\mu(\theta),\theta) - \sum_{si}\alpha_i(\theta_i|s_i)u_i(\mu(\theta),(\theta_{-i},s_i))\}/p_i(\theta_i)$
$= \sum_\theta p(\theta)\sum_i v_i(\mu(\theta),\theta,\lambda,\alpha)$, where v_i is the <u>(λ,α)-virtual utility function</u> for agent i,
defined to be $v_i(x,\theta,\lambda,\alpha) = [(\lambda_i(\theta_i)+\sum_{si}\alpha_i(s_i|\theta_i))u_i(x,\theta) - \sum_{si}\alpha_i(\theta_i|s_i)\,u_i(x,(\theta_{-i},s_i)]/p_i(\theta_i)$.

Lagrange multipliers can be positive $\alpha_i(\theta_i|s_i)>0$ only for binding constraints $U_i(\mu|s_i)=U_i(\mu,\theta_i|s_i)$.
With such an incentive-efficient μ, we say that type s_i <u>jeopardizes</u> θ_i iff the incentive constraint $U_i(\mu|s_i) \geq \hat{U}_i(\mu,\theta_i|s_i)$ is binding and has positive multiplier $\alpha_i(\theta_i|s_i)>0$ in the Lagrangean for μ.
Virtual utility of any type θ_i of agent i differs from the actual utility of type θ_i by exaggerating the differences from i's other types s_i that jeopardize θ_i.
The incentive-efficient mechanism μ appears ex-post efficient in terms of such virtual utilities.

Now suppose agents have separable linear utility for money w_i, that is, any social choice x can be decomposed into a pair $x = (q_i,w_i)$ such that i's utility is $u_i(x,\theta) = u_i(q_i,w_i,\theta) = u_i(q_i,0,\theta) + w_i$, and these money transfers w_i are not bounded (no liquidity constraints).

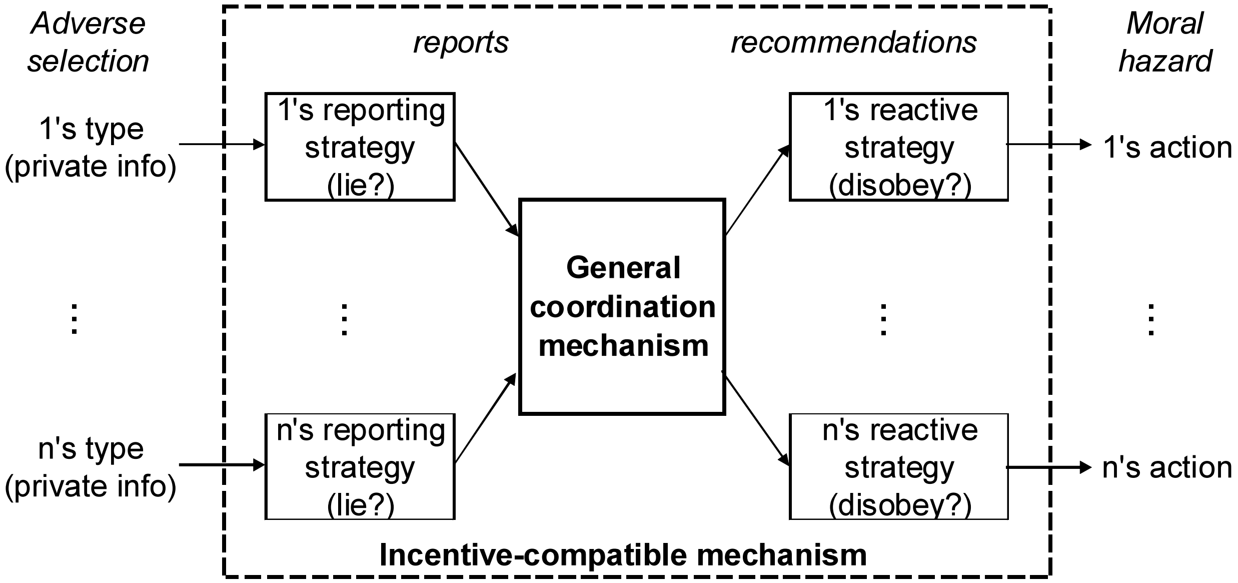
Then all types of all agents must have the same marginal virtual-utility of money (or else Lagrangean optimality would call for infinite transfers from agents with lower marginal utility of money to those with higher marginal utility for money). We may take this constant marginal virtual-utility for money to equal 1, without loss of generality (if not, just rescale all the λ-weights). Then the conditions $\partial v_i((q_i,w_i),\theta,\lambda,\alpha)/\partial w_i = 1$ give us the "balance" equations:
$\lambda_i(\theta_i) + \sum_{si} \alpha_i(s_i|\theta_i) - \sum_{si} \alpha_i(\theta_i|s_i) = p_i(\theta_i)$, $\forall i$, $\forall \theta_i$.
Substituting these equations into the formula for the (λ,α)-virtual utility function, it simplifies to
$v_i(x,\theta,\lambda,\alpha) = v_i(x,\theta,\alpha) = u_i(x,\theta) + \sum_{si} [u_i(x,\theta) - u_i(x,(\theta_{-i},s_i))]\alpha_i(\theta_i|s_i)/p_i(\theta_i)$.

Fact. With independent types and linear utility for money, an incentive-compatible mechanism μ is incentive-efficient iff there exists some vector α of nonnegative Lagrange multipliers
$\alpha_i(s_i|\theta_i) \geq 0$, $\forall s_i$, $\forall \theta_i$, $\forall i$
that have complementary slackness with the corresponding constraints
$\alpha_i(s_i|\theta_i)[U_i(\mu|\theta_i)-\hat{U}_i(\mu,s_i|\theta_i)] = 0$, $\forall s_i$, $\forall \theta_i$, $\forall i$
and yield nonnegative λ weights for all possible types of all individuals
$\lambda_i(\theta_i) = p_i(\theta_i) + \sum_{si} (\alpha_i(\theta_i|s_i) - \alpha_i(s_i|\theta_i)) \geq 0$, $\forall \theta_i$, $\forall i$
such that the mechanism μ maximizes the sum of virtual-utility payoffs for each profile of types
$\sum_i v_i(\mu(\theta),\theta,\alpha) = \max_x \sum_i v_i(x,\theta,\alpha)$, $\forall \theta$.

[There are examples where player i is a seller with an unknown cost type, and the possible cost types are $\{A, A+\delta, A+2\delta,..., B\}$. If the question is how to maximize the buyer's expected payoff, then we find that the participation constraint is binding only for i's highest possible type B, and so only that highest type has positive λ_i; that is $\lambda_i(B)=1$ and all other $\lambda_i(t_i)=0$. In such examples, we often find that the incentive constraints that have positive Lagrange multiplier are $\alpha_i(t_i|t_i-\delta) = \sum_{s<ti} p_i(s)$. Then the cost-type t_i is associated with the virtual cost
$t_i+(t_i-(t_i-\delta))\alpha_i(t_i|t_i-\delta)/p_i(t_i) = t_i + (\sum_{s<ti} p_i(s))/(p_i(t_i)/\delta)$.]

Revelation principle

For any coordination plan, any equilibrium of people's (possibly dishonest) reporting and (possibly disobedient) reactions is equivalent to an incentive-compatible plan that makes it an equilibrium for everyone to be honest and obedient. That is, without loss of generality, a trustworthy mediator can plan to make honesty and obedience the best policy for everyone.


Adverse selection
reports
recommendations
Moral hazard
1's type (private info)
1's reporting strategy (lie?)
1's reactive strategy (disobey?)
1's action
General coordination mechanism
⋮
⋮
⋮
⋮
n's type (private info)
n's reporting strategy (lie?)
n's reactive strategy (disobey?)
n's action
Incentive-compatible mechanism

An optional exercise to verify a randomized equilibrium of the competitive screening game in MWG 13.D, where two firms simultaneously offer menus of employment contracts:

Consider a discrete example of our basic labor-market model. There are two types of workers: low types with reservation wage $\theta_L=2$ and productivity $\pi_L=3$, and high types high types with reservation wage $\theta_H=4$ and productivity $\pi_H=5$. The fraction of high-type workers in the population is $p_H=0.8$, and so the expected productivity of a randomly-sampled worker is $E(\tilde{\pi}) = 0.2\times3+0.8\times5 = 4.6$. Each worker has one unit of labor to sell, but he can accept a contract to sell only some fraction q of it to one firm. So an employment contract specifies the fraction q of the worker's labor that he will sell, and a total salary w that he will be paid.

When a worker of type θ and productivity π accepts contract (q,w), the worker's payoff is $w-\theta q$, and the firm's payoff is $\pi-w$. We may say that a menu of contracts $((q_L,w_L),(q_H,w_H))$ is <u>incentive compatible</u> if, among these two contracts, low types weakly prefer (q_L,w_L) and high types weakly prefer (q_H,w_H). Given any menu $((q_L,w_L),(q_H,w_H))$, applying the first contract to a low-type worker would give the firm payoff $y_L = 3q_L-w_L$ and give the worker payoff $u_L =w_L-2q_L$, and applying the second contract to a high-type worker would give the firm payoff $y_H = 5q_H-w_H$ and give the worker payoff $u_H = w_H-4q_H$.

We may say that the <u>|L incentive constraint is binding</u> in such an incentive compatible menu if low types would be indifferent among the two contracts. (Read "|L" as "given-L.")

(a) Show that, for any incentive-compatible menu $((q_L,w_L),(q_H,w_H))$, there is an amount $\hat{w}_L$ such that $((1,\hat{w}_L),(q_H,w_H))$ is incentive compatible, the low-type workers would be indifferent between $(1,\hat{w}_L)$ and (q_H,w_H), and firms would prefer hiring low-type workers under the $(1,\hat{w}_L)$ contract than the (q_L,w_L) contract. (Thus we can assume without loss of generality that firms only offer such incentive compatible menus where low types work full-time and the |L-constraint is binding.)

(b) Show that, when $((1,w_L),(q_H,w_H))$ is an incentive-compatible menu in which the |L-incentive constraint is binding, the firm's profits (y_L,y_H) from hiring each type of worker (in its intended contract) can be expressed as a linear function of the two worker-types' payoffs (u_L,u_H).

(c) Suppose firm 1 will announce a randomly generated menu $((1,\tilde{w}_L),(\tilde{q}_H,\tilde{w}_H))$ such that:
$\tilde{w}_L$ is drawn from a uniform distribution over the interval from 3 to 4.6 (from π_L to $E(\tilde{\pi})$),
$\tilde{w}_L-2 = \tilde{w}_H-2\tilde{q}_H$ (so the low types would be indifferent between the two contracts), and
$0.2(3-\tilde{w}_L)+0.8(5\tilde{q}_H-\tilde{w}_H) = 0$ (so the menu applied to all workers would yield expected profit 0).
With these two equations, $\tilde{q}_H$ and $\tilde{w}_H$ depend linearly on $\tilde{w}_L$. Show that the resulting menu is incentive compatible. Then show that the two types' expected payoffs under such a random menu $\tilde{u}_L = \tilde{w}_L-2$ and $\tilde{u}_H = \tilde{w}_H-4\tilde{q}_H$ are also uniform random variables over certain intervals.
Show the formulas for the cumulative probabilities $F_L(u) = P(\tilde{u}_L\leq u)$ and $F_H(u) = P(\tilde{u}_H\leq u)$.

(d) Given that firm 1 is behaving according to the randomized strategy in (c), show that firm 2 could get zero expected profit by choosing any incentive-compatible menu $((1,w_L),(q_H,w_H))$ that satisfies the conditions $w_L-2 = w_H-2q_H$, $0.2(3-w_L)+0.8(5q_H-w_H) = 0$, and $3 \leq w_L \leq 4.6$.

(e) Now suppose that firm 2 considers any incentive-compatible menu $((1,w_L),(q_H,w_H))$ in which the |L-incentive constraint is binding. Using your formulas from (b) and (c), show that firm 2's expected profit per worker $0.2\ F_L(u_L)\ y_L + 0.8\ F_H(u_H)\ y_H$ cannot be positive.