Extreme Incentives

Andrew Clausen[1]

[1]University of Edinburgh

Motivation.

- Contract theory predicts drastic rewards and punishments when agents are risk neutral. For example,
    - Becker (1968) proposes a small police force with harsh prison terms,
    - Cremer and McLean (1988) propose large punishments when agents report conflicting signals.
- Mirrlees (1975) studies moral hazard contracts with risk averse agents.
    - Punishment now has a cost, i.e. compensation for accepting risk.

Question. Does Mirrlees' approach actually rule out harsh punishments?

Answer. No, Mirrlees predicts that harsh punishments should be typical!

Example 1 – Overview

- The worker's utility from consuming transfers $t \in \mathbb{R}_+$ and exerting effort $e \in \{0, 1\}$:
  - $u(t) - ce$ where u is CRRA(3.5)
  - "Infinite punishment" is possible, i.e. $u(0) = -\infty$.
  - calibrate worker's outside option to $(t, e) = (20000, 0)$.
  - calibrate c so that the worker is indifferent between outside option and $(25000, 1)$.
- The manager is risk neutral, and sees a noisy signal X of e:
  - The signal X is correct with probability $\frac{4}{5}$, i.e. $\mathbb{P}(X = 1|e = 1) = P(X = 0|e = 0) = \frac{4}{5}$.

Example 1 – the manager's problem

$$
\begin{aligned}
\min_{t(0),t(1)} \; & \frac{4}{5}t(1) + \frac{1}{5}t(0) \\
\text{s.t. } \; & \text{(VP)} \; \frac{4}{5}u(t(1)) + \frac{1}{5}u(t(0)) - c = u(20000) \\
& \text{(IC)} \; \underbrace{\frac{4}{5}u(t(1)) + \frac{1}{5}u(t(0)) - c}_{\text{honour payoff}} \geq \underbrace{\frac{1}{5}u(t(1)) + \frac{4}{5}u(t(0)) - 0}_{\text{breach payoff}}.
\end{aligned}
$$

Example 1 – classic solution method.
First-order condition for $t(x)$:

$$\underbrace{\frac{1}{u'(t(x))} = \lambda}_{\text{Optimal risk sharing}} \quad + \quad \mu\left(1 - \underbrace{\frac{\mathbb{P}(X = x|e = 0)}{\mathbb{P}(X = x|e = 1)}}_{\text{Likelihood ratio of breach}}\right).$$

Optimal contract (boring).

- The optimal transfers are $(t^1(1), t^1(0)) = (28034, 18962)$, or 26219 on average.
- The Lagrange multipliers are $(\lambda^1, \mu^1) = (3139, 733)$, so extreme incentives would involve a likelihood ratio of 5.3.
- The bad signal's likelihood ratio is only 4.

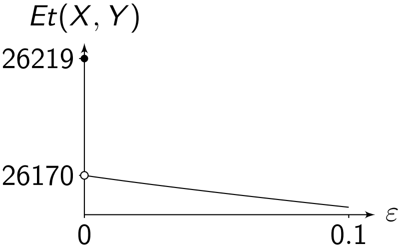
Example 2 – two signals

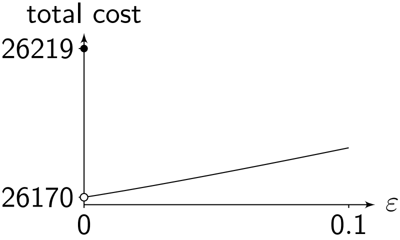
- Suppose the manager can buy a second signal Y for 3000.
    - Y is conditionally independent from X, and
    - Y is correct with probability 2/3, i.e. $\mathbb{P}(Y = 1|e = 1) = \mathbb{P}(Y = 0|e = 0) = 2/3$.
    - The likelihood ratio of $(X, Y) = (0, 0)$ is 6.
- Conditionally on buying Y, the optimal contract is boring with:
    - The optimal transfers are $(t^2(1,1), t^2(1,0), t^2(0,1), t^2(0,0))$ $= (27043, 26822, 24006, 17179)$, or 25819 on average.
    - The Lagrange multipliers are $(\lambda^2, \mu^2) = (2883, 444)$, so extreme incentives would involve a likelihood ratio of 7.5.
- Y is only worth $26219 - 25819 = 400$, so they adopt Example 1's contract.

Example 3 – divisible signal

- The manager pays 3000ε to observe Y with probability ε.
- There is no optimal contract! Limit as $\varepsilon \to 0$:

$$
\begin{array}{ll}
t^3(1,1) = 27722 & t^3(0,1) = 23625 \\
t^3(1,\emptyset) = 27608 & t^3(0,\emptyset) = 20417 \\
t^3(1,0) = 27435 & t^3(0,0) = \color{red}{0} \quad \color{red}{\leftarrow \quad \text{extreme incentive}}
\end{array}
$$



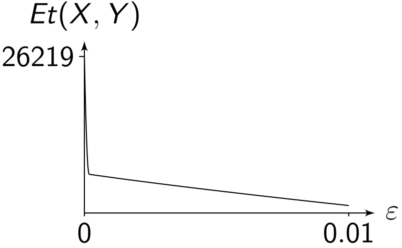


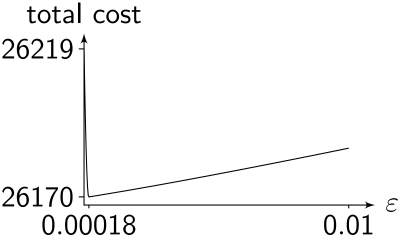
Example 3 – Challenge to moderate incentives

- Suppose a moderate contract based on a signal X is optimal. (Example 1)
- Isn't there **always** another signal Y that provides slightly more information, albeit at a prohibit cost? (Example 2)
- If so, then extreme incentives are **always** optimal! (Example 3)

Example 4 – limited liability

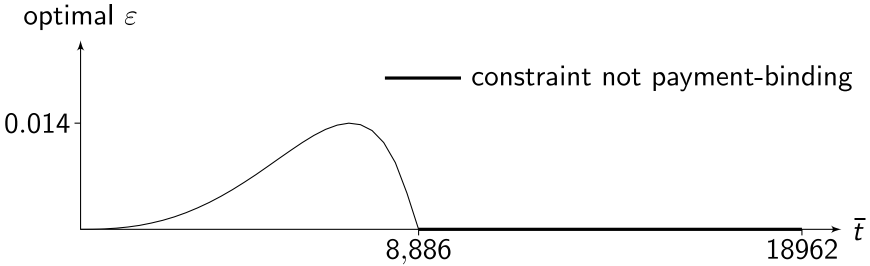
- As before, but the worker must be paid at least 1000.





Example 4 – limited liability (continued)

- More generally, when the worker must be paid at least $\bar{t}$:



- Innocent workers receive the maximum punishment by mistake with probability up to 0.1%.

Likelihood ratio representation

- In the examples, we saw that the likelihood ratio

$$L = \frac{\mathbb{P}(X = x|e = 0)}{\mathbb{P}(X = x|e = 1)}$$

  is a sufficient statistic for optimal transfers.
- Without loss of generality, assume
  - the principal only sees L, and
  - L is distributed according to $f(\ell) = \mathbb{P}(L = \ell|e = 1)$.
- Useful likelihood ratio properties:
  - f also encodes the distribution for shirking as $\mathbb{P}(L = \ell|e = 0) = \ell f(\ell)$.
  - If f and g are two different signals, then $h(\ell) = (1 - \varepsilon)f(\ell) + \varepsilon g(\ell)$ is the distribution of a randomisation between f and g.

Principal's problem.

$$C(f,g,P) = \min_{\varepsilon, t(\cdot)} \sum_{\ell} h(\ell;\varepsilon)t(\ell) + \varepsilon P$$

$$\text{s.t. (VP)} \;\; \sum_{\ell} h(\ell;\varepsilon)u(t(\ell)) - c_1 \geq 0$$

$$\text{(IC)} \;\; \sum_{\ell} h(\ell;\varepsilon)u(t(\ell)) - c_1 \geq \sum_{\ell} \ell h(\ell;\varepsilon)u(t(\ell)) - c_0$$

where $h(\ell;\varepsilon) = (1-\varepsilon)f(\ell) + \varepsilon g(\ell)$.

Maintained assumptions.

- u is concave and differentiable.
- Inada conditions: $\lim_{t\to 0} u(t) = -\infty$ and $\lim_{t\to 0} u'(t) = \infty$.
- f and g have finite supports.

Definitions.

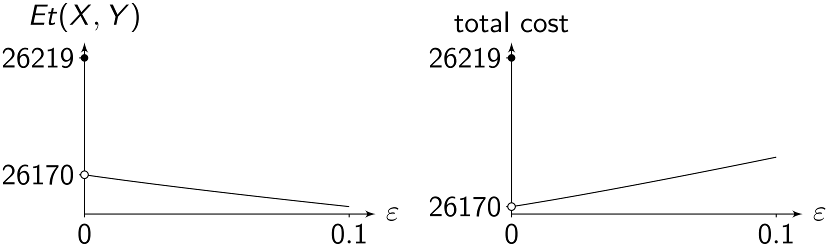
- g's **most incriminating signal** is the largest likelihood ratio in the support of g, denoted $\lceil g \rceil$.
- Definitions arising from non-existence:
  - (ε_k, t_k) are **asymptotically optimal contracts** if costs converges to $C(f, g, P)$.
  - g **is used sparingly** if all asymptotically optimal contracts involve $\varepsilon_k \to 0$.
  - **Infinite punishment is threatened** if all asymptotically optimal contracts involve $t_k(\lceil g \rceil) \to 0$.

Theorem 1 (unlimited liability). If

- $\lceil g \rceil$ is sufficiently incriminating, and
- the price P of g is sufficiently high given g

then

1. there is no optimal contract,
2. g is used sparingly,
3. infinite punishment is threatened,
4. the benefit of using g sparingly ranges from none to first-best, depending continuously on $\lceil g \rceil$.



Principal's problem decomposed.

$$C(f,g,P) = \min_{\varepsilon} T((1-\varepsilon)f + \varepsilon g) + \varepsilon P$$

where

$$\begin{aligned} T(h) = \min_{t(\cdot)} & \sum_{\ell} h(\ell)t(\ell) \\ \text{s.t. (VP)} & \sum_{\ell} h(\ell)u(t(\ell)) = c_1 \\ \text{(IC)} & \sum_{\ell} (1-\ell)h(\ell)u(t(\ell)) = \Delta c, \end{aligned}$$

and $\Delta c = c_1 - c_0$.

Jewitt Duality.

Consider a firm that produces and sells "utility" at price p. Its profit function is

$$\pi(p) = \max_t pu(t) - t.$$

Lemma (Jewitt). $T(h) = \max_{\lambda,\mu} \lambda c_1 + \mu \Delta c - \sum_{\ell} h(\ell)\pi(\lambda + \mu(1-\ell)).$

Proof.

- Let $\mathcal{L}(t, \lambda, \mu)$ be the Lagrangian for the optimal transfer problem.
- By Strong Lagrangian Duality,

$$T(h) = \max_{\lambda,\mu} \min_t \mathcal{L}(t, \lambda, \mu).$$

- This can be re-arranged into the specified form. $\square$

Lemma (discontinuity). If the agent has unlimited liability, then

$$\lim_{\varepsilon \to 0} T((1-\varepsilon)f + \varepsilon g) = \max_{\lambda,\mu} \lambda c_1 + \mu \Delta c - \sum_{\ell} f(\ell)\pi(\lambda + \mu(1-\ell))$$

$$\text{s.t. } \lambda + \mu(1 - \lceil g \rceil) \geq 0.$$

Proof of discontinuity lemma.

Let $\phi(\varepsilon,\lambda,\mu) = \lambda c_1 + \mu \Delta c - \sum_\ell h(\ell;\varepsilon)\pi(\lambda + \mu(1-\ell))$. If g is a mean-preserving spread of f, then

$$
\begin{aligned}
&\lim_{\varepsilon\to 0} T((1-\varepsilon)f + \varepsilon g) && (1)\\
&= \lim_{\varepsilon\to 0} \max_{\lambda,\mu} \phi(\varepsilon,\lambda,\mu) && (2)\\
&= \sup_{\varepsilon\in(0,1]} \max_{\lambda,\mu} \phi(\varepsilon,\lambda,\mu) && (3)\\
&= \sup_{\varepsilon\in(0,1]} [\max_{\lambda,\mu} \phi(\varepsilon,\lambda,\mu) \text{ s.t. } \lambda + \mu(1-\lceil g\rceil) \geq 0] && (4)\\
&= \max_{\lambda,\mu} \left[\max_{\varepsilon\in[0,1]} \phi(\varepsilon,\lambda,\mu) \right] \text{ s.t. } \lambda + \mu(1-\lceil g\rceil) \geq 0 && (5)\\
&= \max_{\lambda,\mu} \phi(0,\lambda,\mu) \text{ s.t. } \lambda + \mu(1-\lceil g\rceil) \geq 0. && (6)
\end{aligned}
$$

If not, then (3) is an inequality so that (6) is a upper bound.
But (6) is also a lower bound, if g is replaced by a mean-preserving spread $\tilde{g}$ with $\lceil \tilde{g}\rceil = \lceil g\rceil$, by Kim (1995). □

Theorem 1 (unlimited liability). If

- $\lceil g \rceil$ is sufficiently incriminating, and
- the price P of g is sufficiently high given g

then

1. there is no optimal contract,
2. g is used sparingly,
3. infinite punishment is threatened,
4. the benefit of using g sparingly ranges from none to first-best, depending continuously on $\lceil g \rceil$.

Proof.

- The first two items are due to the Discontinuity Lemma. If $\lceil g \rceil$ is large enough for the constraint to bind, then the expected transfer is discontinuous at $\varepsilon = 0$. If P is large, then smaller $\varepsilon > 0$ is better.
- A vanishing probability of g implies infinite punishment.
- Since g is used sparingly, the implementation cost equals expected transfers. By Berge's theorem, this is a continuous value function of $\lceil g \rceil$. The intermediate value theorem applies. □

Accommodating limited liability.

$$C(f, g, P; \bar{t}) = \min_{\varepsilon} T((1-\varepsilon)f + \varepsilon g; \bar{t}) + \varepsilon P$$

where

$$\begin{aligned} T(h; \bar{t}) = \min_{t(\cdot)} \sum_{\ell} h(\ell) t(\ell) & \\ \text{s.t.} \quad (\text{LL}) \quad & t(\ell) \geq \bar{t}, \\ (\text{VP}) \quad & \sum_{\ell} h(\ell) u(t(\ell)) = c_1, \\ (\text{IC}) \quad & \sum_{\ell} (1-\ell) h(\ell) u(t(\ell)) = \Delta c, \\ = \max_{\lambda, \mu} \lambda c_1 + \mu \Delta c & - \sum_{\ell} h(\ell) \pi(\lambda + \mu(1-\ell); \bar{t}) \end{aligned}$$

and

$$\pi(p; \bar{t}) = \max_{t \in [\bar{t}, \infty)} pu(t) - t.$$

Accommodating limited liability.

$$C(f, g, P; \bar{t}) = \min_{\varepsilon} T((1-\varepsilon)f + \varepsilon g; \bar{t}) + \varepsilon P$$

where

$$\begin{aligned} T(h; \bar{t}) = \min_{t(\cdot)} & \sum_{\ell} h(\ell)t(\ell) \\ \text{s.t.} \quad & \text{(LL)} \;\; t(\ell) \geq \bar{t}, \\ & \text{(VP)} \;\; \sum_{\ell} h(\ell)u(t(\ell)) = c_1, \\ & \text{(IC)} \;\; \sum_{\ell} (1-\ell)h(\ell)u(t(\ell)) = \Delta c, \\ = \max_{\lambda,\mu} & \; \lambda c_1 + \mu \Delta c - \sum_{\ell} h(\ell)\pi(\lambda + \mu(1-\ell); \bar{t}) \end{aligned}$$

and

$$\pi(p; \bar{t}) = \max_{t \in [\bar{t}, \infty)} pu(t) - t.$$

Observation: $\lim_{\bar{t} \to 0+} \pi(p, \bar{t}) = -\infty$ for all $p < 0$.

Theorem 2 (limited liability). If

- $\lceil g \rceil$ is sufficiently incriminating,
- the price P of g is sufficiently high given g, and
- the agent's liability limit $\bar{t}$ is sufficiently small given (g, P)

then

1. an optimal contract (ε^*, t^*) exists,
2. g is used with probability $\varepsilon^* > 0$, and
3. the maximum liability is threatened, i.e. $t^*(\lceil g \rceil) = \bar{t}$.

Proof. Let $\phi(\varepsilon, \bar{t}) = T((1-\varepsilon)f + \varepsilon g; \bar{t}) + \varepsilon P$.
An optimal contract exists. ϕ is continuous on a compact domain.
Signal g is used with probability $\varepsilon > 0$, since $\phi_\varepsilon(0, \bar{t}) < 0$ for small $\bar{t}$:

$$\phi_\varepsilon(\varepsilon, \bar{t}) = P - \frac{\partial}{\partial \varepsilon}\left[\sum_\ell (g-f)(\ell)\pi(\lambda + \mu(1-\ell); \bar{t})\right]_{\lambda=\lambda(\varepsilon,\bar{t}),\mu=\mu(\varepsilon,\bar{t})}$$

- For $\lceil g \rceil$ sufficiently large:
    - For small $\bar{t}$, $\lambda(0, \bar{t}) + \mu(0, \bar{t})(1 - \lceil g \rceil) < 0$.
    - Recall $\lim_{\bar{t}\to 0+} \pi(p, \bar{t}) = \infty$ for all $p < 0$.
    - Hence $\lim_{\bar{t}\to 0+} \phi_\varepsilon(0, \bar{t}) = -\infty$.

The maximum liability is threatened.

- For large P, ε^* is small.
- By Berge's theorem, $\lambda(\varepsilon, \bar{t})$ and $\mu(\varepsilon, \bar{t})$ are continuous for $\bar{t} > 0$.
- Therefore, for large P and small $\bar{t}$, $\lambda(\varepsilon^*, \bar{t}) + \mu(\varepsilon^*, \bar{t})(1 - \lceil g \rceil) < 0$. $\square$

Implications for Mirrlees' (1975) theory of moral hazard (1/2)

Conventional wisdom:

- Extreme incentives require compensating the agent for risk.
  Therefore, extreme incentives are typically not optimal.
- Exception: when evidence is overwhelmingly incriminating. (Mirrlees 1975)
  - In this case, incentives do not reflect the social costs and benefits of effort.
  - More generally, incentives reflect strength of evidence.
- Extra signals improve welfare by reducing the agent's risk exposure. (Holmstrom 1979 and Kim 1995)

Implications for Mirrlees' (1975) theory of moral hazard (2/2)

Recent development: Moroni and Swinkels (2013) discovered a glitch:

- Jewitt, Kadane and Swinkels (2008) incorrectly claimed that optimal contracts exist under general conditions.
- If the incentive problem is severe (e.g. Δc is sufficiently large), then no optimal contract exists and extreme incentives are optimal.
- Did not suggest that this glitch is the most likely outcome.

What's new:

- Extreme incentives are optimal when the best evidence arises rarely.
- If the best evidence is expensive, then it will be acquired sparingly. **Hence, extreme incentives are typically optimal.**
- Information might have a low marginal value beyond the discontinuous benefit.

Crime literature: Theory

Becker (1968) proposes a small police force with harsh prison terms.

- A tempering force is that it's socially optimal to break rules sometimes, e.g. double parking for a medical emergency.
- Polinsky and Shavell (1979) introduce risk aversion.
- Kaplow and Shavell (1994) introduce self-reporting and detection mistakes.

Crime literature: Evaluating the death penalty

- The literature focuses on deterrence:
  - Ehrlich (1975) focuses on "the tradeoff between the execution of an offender and the lives of potential victims it might have saved" (1:8 in 1933-1967, US)
  - This neglects the benefit of being more lenient on most innocent defendants.
- The literature assumes that crime is continuous in the execution probability:
  - The recent literature has given up (Katz, Levitt and Shustrorovich, 2003; and Donohue and Wolfers, 2005/2009)
  - Problem: murder rates but not executions fluctuate drastically.
  - The discontinuity might give a new identification strategy? New challenges: gangs, extra-judicial punishment, and police brutality, might also be important.

Conclusions

- Becker's (1968) logic is surprisingly robust:
  - Even when agents are risk averse and signals are noisy (as in Mirrlees 1975), my model predicts extreme incentives.
  - Extreme punishments arise as a way to reduce punishments imposed based on (moderately) weaker evidence.
  - The absence of capital punishment for parking violations, arriving late to work, etc., is a puzzle. There must be another powerful force at play.
- The conclusion is robust to limited liability.
- Empirical evaluations of the death penalty could investigate/exploit the theoretical discontinuity in punishment probability.