Revisiting Incentives:

Values, Laws and Norms

Roland Bénabou

Princeton University

Toulouse Lectures - Lecture III

December 2009

Based on joint work with Jean Tirole (TSE) and Nageeb Ali (UCSD)

Road map to the lectures

L1 Extrinsic, Intrinsic and Attributional Motivation

1. Introduction, evidence
2. The general framework
3. Intrinsic vs. extrinsic motivation

Road map to the lectures

L1 Extrinsic, Intrinsic and Attributional Motivation

1. Introduction, evidence
2. The general framework
3. Intrinsic vs. extrinsic motivation

L2 Laws, Norms and Information

1. Honor, stigma and social norms
2. Welfare and optimal incentives
3. Persuasion and norms-based interventions

Road map to the lectures

L1 Extrinsic, Intrinsic and Attributional Motivation

1. Introduction, evidence
2. The general framework
3. Intrinsic vs. extrinsic motivation

L2 Laws, Norms and Information

1. Honor, stigma and social norms
2. Welfare and optimal incentives
3. Persuasion and norms-based interventions

L3 Social Values and Social Responsibility

1. The expressive content of law
2. Incentives, attributions and crowding out
3. Publicity, privacy and evolving societal values

Lecture III
Social Values and Social Responsibility

1. The expressive content of law
2. Incentives, attributions and crowding out
3. Publicity and the overjustification effect
4. Publicity, privacy and evolving societal values

- Main refs. Bénabou-Tirole AER (2006), (2010), Ali–Bénabou (2010)

The Expressive Function of Law

1. Concept and uses
2. Societal values and incentives and with symmetric information
3. Societal values and incentives with asymmetric Information
4. What forms should punishments (not) take?

The expressive function of law

- Large literature, mostly outside economics, arguing that laws have a dual role:
  - Not just a menu with "prices" for good or bad behaviors
  - Also express society's values: what it approves of or chooses to punish, how it chooses to punish; this expressive function is important

The expressive function of law

- Large literature, mostly outside economics, arguing that laws have a dual role:
  - Not just a menu with "prices" for good or bad behaviors
  - Also express society's values: what it approves of or chooses to punish, how it chooses to punish; this expressive function is important
- Sometimes, expressive considerations used to argue for tougher laws (even inefficiently so), e.g. prison vs. fines or community service.
- Sometimes, used to argue for a gentler hand, e.g. limiting severity of sanctions: corporal punishments, torture, shaming, death penalty

Examples

- Prohibition / legalization of "soft" drugs
- Gay marriage vs. formally equivalent civil union
  Earlier: Georgia's anti-sodomy was, unenforced but still on the books; antimiscegenation laws
- "Symbolic" fines, e.g. for not voting
- Prohibition / legalization of flag burning
- France: contaminated growth hormones trial: prosecutor asking for suspended jail sentence "pour marquer la réprobation de la société"
- Madoff's 150-year sentence

"Is There An Expressive Function of Law? An Empirical Analysis of Voting Laws with Symbolic Fines", Funk, Am. Law & Econ. Rev. (2007):

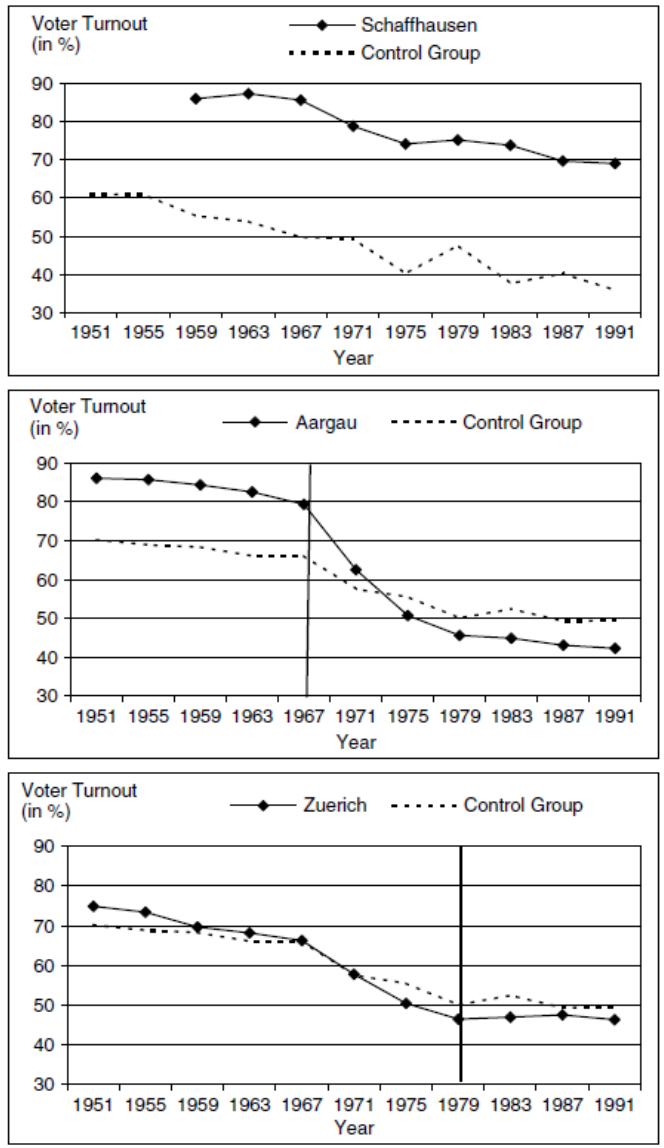


Figure 1. The impact of the legal voting duty on voter turnout.

Table 3. The effect of the legal voting duty on turnout

	(1)	(2)	(3)	(4)
Legal duty with minimal fine	10.20	10.22	6.08	6.19
	(3.48)***	(3.52)***	(3.26)*	(3.47)*
Legal duty without fine		0.24		0.50
		(2.20)		(1.59)
Dummy postal	1.28	1.28	−0.08	−0.04
	(1.83)	(1.84)	(1.44)	(1.47)
Age 0 to 19 (%)	2.03	2.03	0.32	0.24
	(1.89)	(1.90)	(1.19)	(1.28)
Age 20 to 39 (%)	0.84	0.83	−1.51	−1.60
	(1.87)	(1.88)	(1.37)	(1.46)
Age 40 to 59 (%)	1.27	1.26	−0.90	−0.97
	(1.67)	(1.69)	(1.37)	(1.45)
Age 60 to 64 (%)	1.32	1.30	1.32	1.20
	(2.99)	(3.01)	(3.33)	(3.41)
Age 65 to 74 (%)	0.53	0.52	−2.70	−2.78
	(2.97)	(2.99)	(2.40)	(2.47)
Population (in Mio.)	−19.56	−19.49	−16.98	−16.84
	(16.95)	(16.76)	(9.51)*	(9.71)*
Unemployment	−0.38	−0.37	−0.21	−0.20
	(1.05)	(1.05)	(0.72)	(0.72)
Education	0.50	0.49	1.14	1.07
	(0.90)	(0.93)	(0.58)*	(0.70)
Canton-fixed effects	Yes	Yes	Yes	Yes
Time-fixed effects	Yes	Yes	Yes	Yes
Observations	316	316	316	316
R-squared	0.99	0.99	0.99	0.99
Econometric method	OLS	OLS	WLS	WLS

Among law scholars,

- "Consequentialists": argue that expressive role of law can or should ultimately be valued for the consequences it produces.

  Cass Sunstein, Richard Posner (closest to economists)

- "Expressivists": insist more on "social meaning" entering welfare calculus per se. Dan Kahan, Robert Anderson

- Idea starting to appear in scholarly law / law-and-economics discussions, among people interested in social norms (the "new Chicago School" of law)

- There is no consistent framework to analyze these issues, the channels through which they operate, when laws and norms are complements or substitutes, etc.

Modeling expressive law

- Provide analytical framework and results, bringing together elements from Lectures I and II
  1. Individuals care about social / self approval $\rightsquigarrow$ norms
  2. Imperfectly informed about "community standards", aggregate distribution of preferences in society, θ; or about e
  3. Legislator / planner / principal may have information about it. Law, incentives, convey message about it

Modeling expressive law

- Provide analytical framework and results, bringing together elements from Lectures I and II
  1. Individuals care about social / self approval $\rightsquigarrow$ norms
  2. Imperfectly informed about "community standards", aggregate distribution of preferences in society, θ; or about e
  3. Legislator / planner / principal may have information about it. Law, incentives, convey message about it

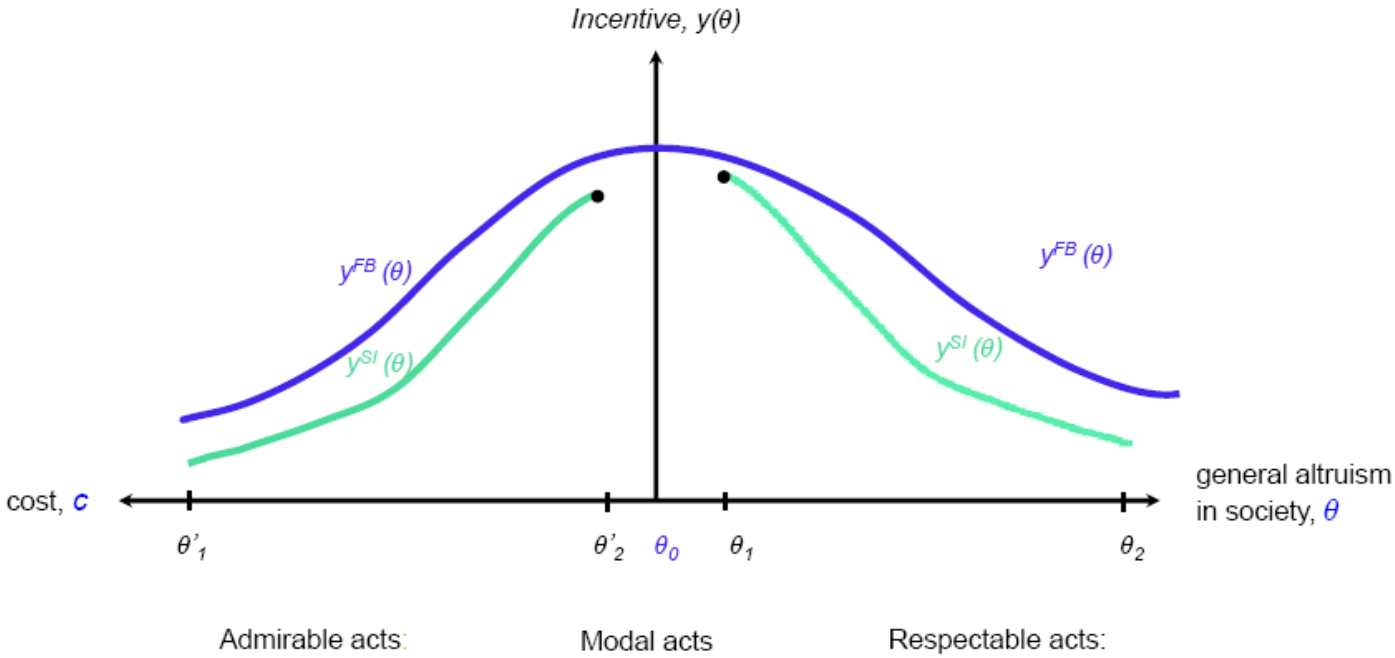
- Two-sided signaling: agents signal their idiosyncratic types, principal signals aggregate state of societal preferences
- Feature of the "task" on which principal's private information bears is now endogenous: social or moral pressure $\mu\Delta$ embodies the equilibrium actions and inferences of all agents

Modeling expressive law

- Provide analytical framework and results, bringing together elements from Lectures I and II
  1. Individuals care about social / self approval $\rightsquigarrow$ norms
  2. Imperfectly informed about "community standards", aggregate distribution of preferences in society, θ; or about e
  3. Legislator / planner / principal may have information about it. Law, incentives, convey message about it

- Two-sided signaling: agents signal their idiosyncratic types, principal signals aggregate state of societal preferences
- Feature of the "task" on which principal's private information bears is now endogenous: social or moral pressure $\mu\Delta$ embodies the equilibrium actions and inferences of all agents
- Can also be situations where legislator or principal is trying to learn about societal preferences, θ. See later.

Reminder: symmetric information case



Optimal incentives with asymmetric information

- Social planner knows aggregate preference θ, hence $G_\theta(v)$
  - For instance, has observed behavior of a representative sample
  - Could instead have private information on c : equivalent

Optimal incentives with asymmetric information

- Social planner knows aggregate preference θ, hence $G_\theta(v)$
  - For instance, has observed behavior of a representative sample
  - Could instead have private information on c : equivalent

- Individuals in society only know that:

  (i) θ lies in some interval (θ_1, θ_2) to the left of $\theta_0 \equiv c - e$
  Alternatively, that θ lies in (θ_1, θ_2) to the right θ_0.

Optimal incentives with asymmetric information

- Social planner knows aggregate preference θ, hence $G_\theta(v)$
  - For instance, has observed behavior of a representative sample
  - Could instead have private information on c : equivalent

- Individuals in society only know that:

  (i) θ lies in some interval (θ_1, θ_2) to the left of $\theta_0 \equiv c - e$
  Alternatively, that θ lies in (θ_1, θ_2) to the right θ_0. Thus:

  - uncertainty about societal values is local / not too global, e.g. θ fluctuating around long-run $\bar{\theta} \neq \theta_0$

  - agents have broad sense of whether some behavior is rare and admirable or prevalent and merely respectable

Optimal incentives with asymmetric information

- Social planner knows aggregate preference θ, hence $G_\theta(v)$
  - For instance, has observed behavior of a representative sample
  - Could instead have private information on c : equivalent

- Individuals in society only know that:

  (i) θ lies in some interval (θ_1, θ_2) to the left of $\theta_0 \equiv c - e$
  Alternatively, that θ lies in (θ_1, θ_2) to the right θ_0. Thus:

  - uncertainty about societal values is local / not too global, e.g. θ fluctuating around long-run $\bar{\theta} \neq \theta_0$

  - agents have broad sense of whether some behavior is rare and admirable or prevalent and merely respectable

  (ii) Planner sets incentive $y^{AI}(\theta)$ to maximize social welfare

- Condition (i) required by non-monotonicity of $\Delta \rightsquigarrow y^{FB} \rightsquigarrow y^{FI}$.
Separating equilibrium cannot exist around θ_0

Equilibrium

- Look for separating equilibrium where $y^{AI}(\theta) \nearrow$ on (θ_1, θ_2) if lies to the left of θ_0, $\searrow$ if lies to the right
- Agents invert the policy and infer θ as solution $\hat{\theta}(y)$ to

$$y^{AI}(\hat{\theta}(y)) \equiv y.$$

- Resulting cutoff for participation: $v^*(y, \hat{\theta}(y)) \Rightarrow$ planner maximizes

$$W_\theta^{AI}(y) = \int_{v^*(y,\hat{\theta}(y))}^{+\infty} (e + v - c - \lambda y)\, g_\theta(v) dv + \mu(\bar{v} + \theta)$$

Equilibrium

- Look for separating equilibrium where $y^{AI}(\theta) \nearrow$ on (θ_1, θ_2) if lies to the left of θ_0, $\searrow$ if lies to the right
- Agents invert the policy and infer θ as solution $\hat{\theta}(y)$ to

$$y^{AI}(\hat{\theta}(y)) \equiv y.$$

- Resulting cutoff for participation: $v^*(y, \hat{\theta}(y)) \Rightarrow$ planner maximizes

$$W_\theta^{AI}(y) = \int_{v^*(y,\hat{\theta}(y))}^{+\infty} (e + v - c - \lambda y)\, g_\theta(v) dv + \mu(\bar{v} + \theta)$$

- Assume $W_\theta^{AI}(\cdot)$ quasiconcave, true for λ small enough. FOC:

$$\left(\frac{e - c - \lambda y + v^*(y, \hat{\theta}(y))}{1 + \mu \Delta_\theta'(v^*(y, \hat{\theta}(y)))} \right) \left(\underbrace{1 - \mu \Delta_\theta'(v^*(y, \hat{\theta}(y)))\ \hat{\theta}'(y)}_{\text{informational multiplier}} \right)$$

$$= \frac{\lambda}{h_\theta(v^*(y, \hat{\theta}(y)))}$$

- FOC = implicit DE in $\hat{\theta}(y)$, or its inverse, $y^{AI}(\theta)$

$$\left(\underbrace{\frac{e - c - \lambda y^{AI}(\theta) + v^*(y^{AI}(\theta), \theta)}{1 + \mu \Delta'_\theta(v^*(y^{AI}(\theta), \theta))}}_{\text{norms multiplier}} \right) \left(\underbrace{1 - \frac{\mu \Delta'_\theta(v^*(y^{AI}(\theta), \theta)))}{(y^{AI})'(\theta)}}_{\text{informational multiplier}} \right)$$

$$= \frac{\lambda}{h_\theta(v^*(y^{AI}(\theta), \theta))}$$

- Difference with FI case: reflects planner's taking into account that agents will make inferences from chosen policy, about:
  - Where societal values lie: $\hat{\theta}' = 1/(y^{AI})'$
  - Social norms / sanctions will face as a result: $\mu \Delta'_\theta(v^*(y^{AI}(\theta), \theta))$
- This is the "expressive content of the law" $\rightsquigarrow$ new multiplier
- Note that it becomes irrelevant when $\lambda = 0$. Intuitive.

Existence, uniqueness, properties of AI solution

- No deadweight loss = again a useful benchmark

Proposition

For $\lambda = 0$, the first-best solution, $y^{FB}(\theta) = e - \mu\Delta(c - e - \theta)$, remains an equilibrium, which is separating, on $(-\infty, \theta_0)$ and $(\theta_0, +\infty)$.

Existence, uniqueness, properties of AI solution

- No deadweight loss = again a useful benchmark

Proposition

For $\lambda = 0$, the first-best solution, $y^{FB}(\theta) = e - \mu\Delta(c - e - \theta)$, remains an equilibrium, which is separating, on $(-\infty, \theta_0)$ and $(\theta_0, +\infty)$.

- Solve DE with bound. cond. $y^{AI} = y^{FI}$, for λ relatively small

Existence, uniqueness, properties of AI solution

- No deadweight loss = again a useful benchmark

Proposition

For $\lambda = 0$, the first-best solution, $y^{FB}(\theta) = e - \mu\Delta(c - e - \theta)$, remains an equilibrium, which is separating, on $(-\infty, \theta_0)$ and $(\theta_0, +\infty)$.

- Solve DE with bound. cond. $y^{AI} = y^{FI}$, for λ relatively small

Lemma

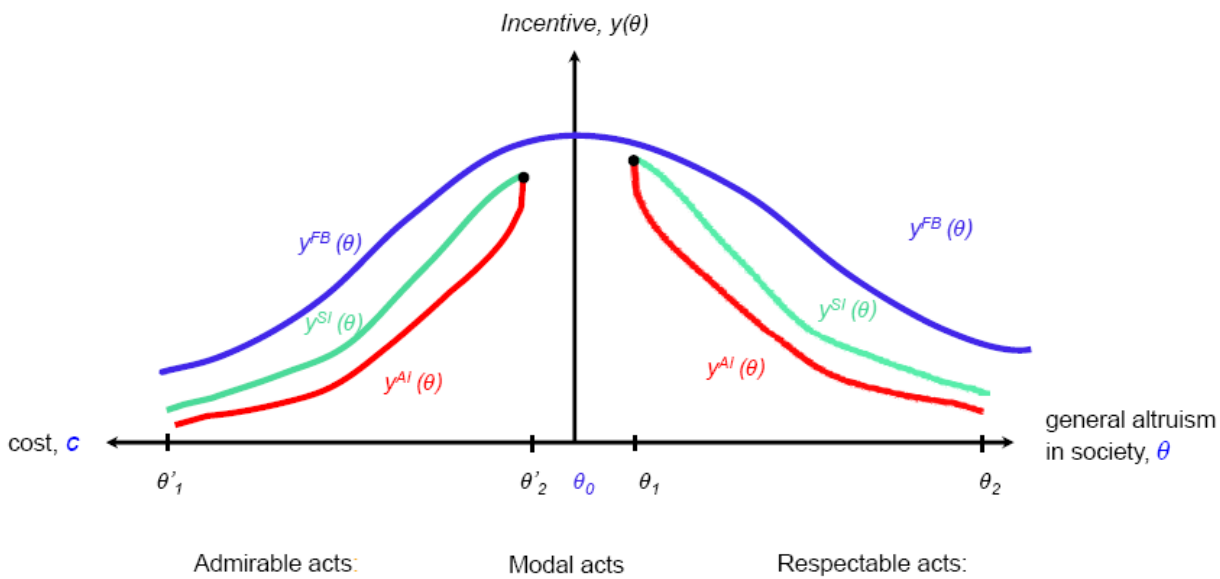
Let (θ_1, θ_2) be any interval with $\theta_0 < \theta_1$ (resp., $\theta_2 < \theta_0$)

For λ small enough, the differential equation (FOC) with boundary cond. $y^{AI}(\theta_1) = y^{FI}(\theta_1)$ (resp., $y^{AI}(\theta_2) = y^{FI}(\theta_2)$) has unique solution on (θ_1, θ_2), with $y^{AI}(\theta) > 0$ and decreasing (resp., increasing) in θ.

Proposition (expressive law: I)

Whether the prosocial action is of a respectable or admirable nature ($\theta_0 < \theta_1$ *or* $\theta_2 < \theta_0$)*, for all* $\lambda > 0$ *low enough:*

1. *Principal always sets lower-powered incentives under asymmetric information:* $y^{AI}(\theta) < y^{FI}(\theta)$ *for all* $\theta \in (\theta_1, \theta_2)$.
2. *Participation is lower than under full information:* $v^*(y^{AI}(\theta), \theta) > v^*(y^{FI}(\theta), \theta)$.



Intuitions

- Respectable activities / SC: lower y credibly conveys the message: "everyone does it, except the most disreputable people who suffer great stigma This is why we need not provide strong extra incentives" $(\hat{\theta}\uparrow,\ \hat{v}\downarrow,\ \Delta' < 0 \Rightarrow \hat{\Delta}\uparrow)$

- Admirable activities / SS: lower y credibly conveys the message "the glory suffices: contributors are rare heroes, who reap such glory and social esteem that no additional incentives are necessary" $(\hat{\theta}\downarrow,\ \hat{v}\uparrow,\ \Delta' > 0 \Rightarrow \hat{\Delta}\uparrow)$

Intuitions

- Respectable activities / SC: lower y credibly conveys the message: "everyone does it, except the most disreputable people who suffer great stigma This is why we need not provide strong extra incentives" $(\hat{\theta}\uparrow,\ \hat{v}\downarrow,\ \Delta' < 0 \Rightarrow \hat{\Delta}\uparrow)$

- Admirable activities / SS: lower y credibly conveys the message "the glory suffices: contributors are rare heroes, who reap such glory and social esteem that no additional incentives are necessary" $(\hat{\theta}\downarrow,\ \hat{v}\uparrow,\ \Delta' > 0 \Rightarrow \hat{\Delta}\uparrow)$

- Expressive law is more responsive to changes in societal values than "standard" law. On both sides of θ_0,
  - Level: $y^{AI}(\theta) < y^{FI}(\theta)$, but
  - Sensitivity: average slope over (θ_1, θ_2) and (especially) local slope near θ_1 (resp., θ_2) is steeper for $y^{AI}(\theta)$

Spillovers across spheres of behavior

- What people learn or perceive concerning others' general degree of prosociality or selfishness carries over between activities

Spillovers across spheres of behavior

- What people learn or perceive concerning others' general degree of prosociality or selfishness carries over between activities

  - "Society is rotten, corrupt": damaging in the "respectable" case $\Delta' < 0$, i.e. when $(\theta_1, \theta_2) > \theta_0$
  - "Everyone is OK": damaging in the "admirable" case $\Delta' > 0$, i.e. when $(\theta_1, \theta_2) < \theta_0$

- Government may refrain from giving too strong incentives in one activity she can monitor closely, because this would adversely affect
  - People's views of societal norms, and thus
  - Behavior in other activities it cannot monitor or incentivize as well

Keizer et al. *Science* (2008). "The Spreading of Disorder"

- Post flyers (advertisements) on parked bicycles.
  One-third of 77 cyclists tossed them on the ground


GRAFFITI

- Same, with graffiti: more than two-thirds littered.

- Same, with graffiti: more than two-thirds littered.
- Leave € 5 note sticking out of mailbox: 13% of subjects pocketed it when in a clean environment, 23% when there was trash around

A simple case:

- Two activities, a and b, both $0, 1$
- Individual's $a-$behavior: observed by other private citizens, but not by principal / gvt.
  - Informational costs, activity done privately, observable not verifiable
  - Cooperating, helping, public goods contributions, not rent-seeking

$$y_a = 0, \quad \mu_a = \mu > 0$$

- Individual's $b-$behavior: observed by principal / gvt.,but not by other private citizens
  - Transactions involving principal: paying / evading taxes, bureaucrats' honesty or corruption; employee productivity.

    Or, other agents less able than principal to sort through excuses

$$y_b = y > 0, \quad \mu_b = 0$$

- For simplicity, same $v_a = v_b = v$ in both activities: general degree of prosociality. More generally, just need correlation

- Two cutoffs:

  $b = 1$ iff

$$v - c_b + y \geq 0, \quad \text{or}$$

$$v \geq c_b - y \equiv v_b^*(y)$$

  $a = 1$ iff

$$v \geq v_a^*(y) \equiv v_a^*(0, \hat{\theta}(y))$$

$$\text{defined by : } v_a^* - c_a + \mu \Delta_{\hat{\theta}(y)}(v_a^*) = 0$$

- Note that v_a^* depends on y, through the inferences about θ

The expressive spillovers of law

- Gvt. or other principal maximizes

$$W_\theta^{AI}(y) = \int_{v_b^*(y)}^{+\infty} (e_b + v - c_b - \lambda y)\, g_\theta(v)\, dv$$

$$+ \int_{v_a^*(y)}^{+\infty} (e_a + v - c_a)\, g_\theta(v)\, dv + \mu(\bar{v} + \theta),$$

$$\frac{\partial W_\theta^{AI}(y)}{\partial y} = (e_b + v_b^*(y) - c_b - \lambda y)\, g_\theta(v_b^*(y)) - \lambda \left[1 - G_\theta(v_b^*(y)\right]$$

$$- (e_a - c_a + v_a^*(y))\, g_\theta(v_a^*(y)) \left(\frac{\partial v_a^*(y)}{\partial y}\right)$$

- Social cost of marginal rise in y now includes:
  - Rents to inframarginal agents choosing $b = 1$
  - Reduction in $\bar{a}$, due to agents' inferring that they face weaker social enforcement in other pro- or anti-social behaviors

- FOC:

$$\frac{\lambda}{h_\theta(v_b^*(y))} = (e_b - (1+\lambda)y)\, g_\theta(v_b^*(y)) - (e_a - c_a + v_a^*(y)) \left(\frac{g_\theta(v_a^*(y))}{g_\theta(v_b^*(y)}\right) \left(\frac{\hat{\theta}'(y) \cdot \mu\Delta_\theta'(v^*(y,\hat{\theta}(y)))}{1+\mu\Delta_\theta'(v^*(y,\hat{\theta}(y)))}\right)$$

- Once again, Δ_θ' and $dy^{AI}/d\theta > 0$ will have same sign, as under FI

$$\Rightarrow \quad \hat{\theta}'(y)\ \Delta_\theta' > 0 \quad \Rightarrow \quad (e_b - (1+\lambda)y)\, g_\theta(v_b^*(y)) > \frac{\lambda}{h_\theta(v_b^*(y))}$$

Proposition

Whether the socially-enforced behavior a is of a respectable or admirable nature ($i.e.$, $\theta_0 < \theta_1$ or $\theta_2 < \theta_0$), for all λ low enough:

1. *Principal sets lower-powered incentives for the privately monitored action b under asymmetric information:for all $\theta \in (\theta_1, \theta_2)$,*

$$y^{AI}(\theta) < y^{FI}(\theta)$$

2. *Participation in b is lower than under full information, participation in a is unchanged (since θ is revealed)*

Two further questions

1. (When) can expressive content make law / incentives more strict rather than more lenient, i.e. $y^{AI} > y^{FI}$?

Two further questions

1. (When) can expressive content make law / incentives more strict rather than more lenient, i.e. $y^{AI} > y^{FI}$?
   - "Lock them up and throw away the key. We need to send a message"

Two further questions

1. (When) can expressive content make law / incentives more strict rather than more lenient, i.e. $y^{AI} > y^{FI}$?
   - "Lock them up and throw away the key. We need to send a message"

Two further questions

1. (When) can expressive content make law / incentives more strict rather than more lenient, i.e. $y^{AI} > y^{FI}$?
   - "Lock them up and throw away the key. We need to send a message"

2. People's intrinsic motivation "should" be linked to how useful their action is for others: making one's contribution to the firm, to public goods that others enjoy, to social welfare. Thus: $e \rightsquigarrow v$
   - With small numbers, consistent with "pure" altruism: an individual correctly values the difference that he makes to $\bar{a}$. Consequentialist.
   - With large numbers, it is not: negligible individual effect on $\bar{a}$.

     Intrinsic motivation must then be pure preference for / "joy of" giving.

Two further questions

1. (When) can expressive content make law / incentives more strict rather than more lenient, i.e. $y^{AI} > y^{FI}$?
   - "Lock them up and throw away the key. We need to send a message"

2. People's intrinsic motivation "should" be linked to how useful their action is for others: making one's contribution to the firm, to public goods that others enjoy, to social welfare. Thus: $e \rightsquigarrow v$
   - With small numbers, consistent with "pure" altruism: an individual correctly values the difference that he makes to $\bar{a}$. Consequentialist.
   - With large numbers, it is not: negligible individual effect on $\bar{a}$.

     Intrinsic motivation must then be pure preference for / "joy of" giving.
   - But, sensibly, should derive more intrinsic utility from giving to more useful causes, rather than unimportant ones
     - Unlike what happens with (self-) image motivations
     - May also reflect a Kantian or similar rule-based reasoning

A version with "ConseKantialism"

- Intrinsic motivation is now *ve*, with $v \sim G(v)$
- Reputation / self-image still bears on v = degree of social concern
  - μ could also vary with e
- Principal knows e : how damaging are CO_2 emissions, how much good \$1 can do in poor countries, negative externalities from drunk driving, drugs, how important to firm is quality / customer service...
- Participation cutoff $v^* = v^*(y)$ under *FI* is

$$ev^*(y) - c + y + \mu\Delta(v^*(y)) \equiv 0$$

Unique, provided $e + \mu\Delta' > 0$

Modified Pigou again

- Samuelson condition: $e + ev^*(y^{FB}) = c$, or

$$v^*(y^{FB}) = (c - e)/e$$

$v^* > 0$ requires $c > e$

- Optimality condition + cutoff rule $\Rightarrow$

$$y^{FB} = e - \mu\Delta\left(\frac{c - e}{e}\right)$$

- Could, in, general, have $y^{FB} < 0$: people demonstrate great social concern by paying significant costs for trivially small social benefits
- Reputation gain can increase more than 1 for 1 with e, hence reputation tax also: $dy^{FB}/de < 0$
- Will abstract from this case; interested in relatively large e's

Asymmetric information about e

- Look for separating eqbm. where $y^{AI}(e)$ is $\nearrow$ on $\mathbb{R}$, like $y^{FI}(e)$
- Agents infer e as solution $\hat{e}(y)$ to $y^{AI}(\hat{e}(y)) \equiv y$
- Participation cutoff $\hat{v}(y)$ given by

$$\hat{e}(y)\hat{v}(y) - c + y + \mu\Delta(\hat{v}(y)) \equiv 0 \Rightarrow$$

$$\frac{d\hat{v}(y)}{dy} = -\frac{1 + \hat{v}(y)\hat{e}'(y)}{\hat{e}(y) + \mu\Delta'(\hat{v}(y))}$$

- Knowing this, planner maximizes

$$W_e^{AI}(y) = \int_{\hat{v}(y)}^{+\infty} (e + ev - c - \lambda y)\, g(v)dv + \mu\bar{v},$$

Expressive role makes law tougher

- FOC:

$$\left(\frac{\hat{e}(y)\left[1+\hat{v}(y)\right]-c-\lambda y}{\hat{e}(y)+\mu\Delta'(\hat{v}(y))}\right)\left(1+\hat{v}(y)\frac{d\hat{e}(y)}{dy}\right)=\frac{\lambda}{h_\theta(\hat{v}(y))}$$

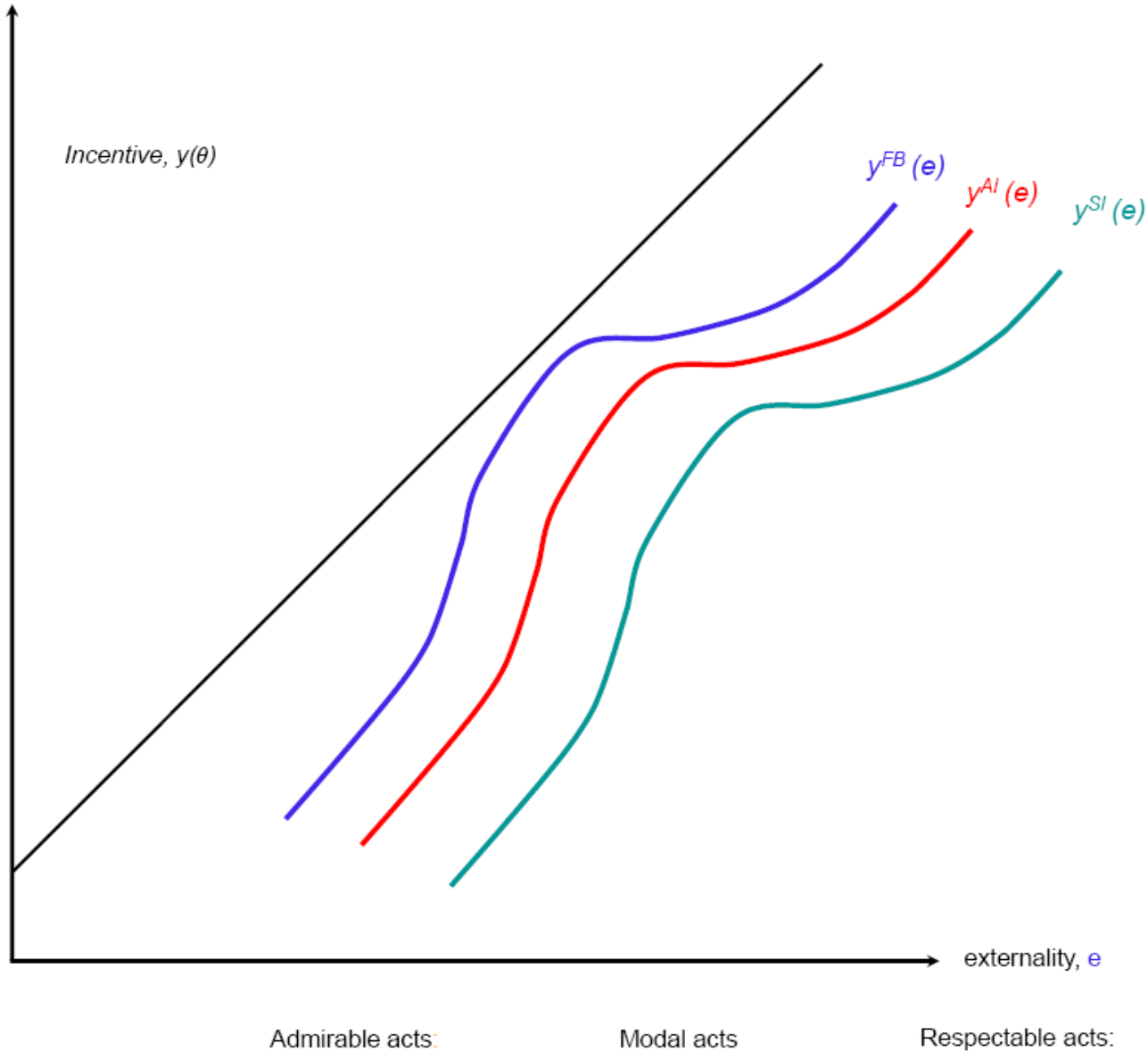
- DE in $\hat{e}(y)$, or in its inverse, $y^{AI}(e)$.
- Since $y^{AI}(e) \nearrow$ in e, expect for λ small, $\hat{e}(y) \nearrow$ in y

Proposition (expressive law: II)

Whether the prosocial action is of a respectable or admirable nature, for all $\lambda > 0$ low enough:

1. *Principal always sets higher-powered incentives under asymmetric information: $y^{AI}(e) > y^{FI}(e)$ for all e.*
2. *Participation is higher than under full information: $v^*(y^{AI}(e)) < v^*(y^{FI}(e))$.*

Expressiveness about externalities makes law tougher



Affective beliefs about societal preferences, human nature

- In previous cases, planner / principal had instrumental reason for wanting to alter agents' beliefs about the social (dis)approval θ or the public-welfare implications e of their behavior

- Can also be hedonic, comfort value for people to think that their neighbors, other citizens, etc., are generally good, generous people rather than, deep down, Hobbesian selfish brutes

  - Scary to think people will act badly when law enforcement fails (blackouts, Katrina, war...) or when one is in need of help (accident, heart attack...)

- Similar in spirit to spillovers across activities via $\hat{\theta}$. Formalize as

$$W_\theta^{AI}(y) = \int_{v^*(y,\hat{\theta}(y))}^{+\infty} (e + v - c - \lambda y)\, g_\theta(v) dv - \beta\hat{\theta}(y) + \mu(\bar{v} + \theta)$$

- FOC:

$$\left(\frac{e-c-\lambda y+v^*(y,\hat{\theta}(y))}{1+\mu\Delta'_\theta(v^*(y,\hat{\theta}(y)))}\right)\left(1-\mu\Delta'_\theta(v^*(y,\hat{\theta}(y)))\ \hat{\theta}'(y)\right)$$
$$=\frac{\lambda}{h_\theta(v^*(y,\hat{\theta}(y)))}+\frac{\beta\hat{\theta}'(y)}{g_\theta(v^*(y,\hat{\theta}(y)))}$$

- New "expressive" term in β leads to lower y, compared to FI
- This effect no longer disappears for $\lambda = 0$
- Remark: equilibrium separating $\Rightarrow$ in fine, no one is fooled or feels better than under FI. Discuss in context of next application

Alternative sanctions, cruel and unusual punishments

- Economists typically favor fines, community service, compensation, etc., over prison
  - Politically unpopular: explained (e.g., Kahan) as due to not carrying enough appropriate symbolism: insufficient stigma on condemned, devalue victims
  - Electorates often favor death penalty, corporal punishments, torture, shaming
  - Many countries still use them

Alternative sanctions, cruel and unusual punishments

- Economists typically favor fines, community service, compensation, etc., over prison
  - Politically unpopular: explained (e.g., Kahan) as due to not carrying enough appropriate symbolism: insufficient stigma on condemned, devalue victims
  - Electorates often favor death penalty, corporal punishments, torture, shaming
  - Many countries still use them
  - Others (increasingly) forbid themselves to use them. This is done not really based on considerations of (in)effectiveness, but on "what it makes us", what "civilized" peoples do or don't.

Strong incentives in Singapore

- Singapore: GDP/capita = US \$51,226, 3^d or 4^{th}in the world, a few places ahead of USA(\$ 47,440). Literacy rate 95%.

Strong incentives in Singapore

- Singapore: GDP/capita = US \$51,226, 3^d or 4^{th}in the world, a few places ahead of USA(\$ 47,440). Literacy rate 95%.

- Singaporean law allows caning to be ordered for over 30 offences, including robbery, gang robbery with murder, drug use, vandalism, and rioting. Caning is mandatory punishment for certain offences such as rape, drug trafficking and for visiting foreigners who overstay their visa by more than 90 days

- For male criminals between the ages of 18 and 50, certified medically fit by medical officer; maximum of 24 strokes on any one occasion. Under 18, maximum 10 strokes, with lighter cane. Males under 16 may be sentenced to caning only by the High Court, not district courts.

- In 1993: 3,244 criminals sentenced to caning . By 2007: doubled, to 6,404. About 95% of sentences were implemented.

- Caning also in prison, military, schools

Versailles, June 17, 1939

Execution of Eugene Weidmann, six-time murderer, June 1939

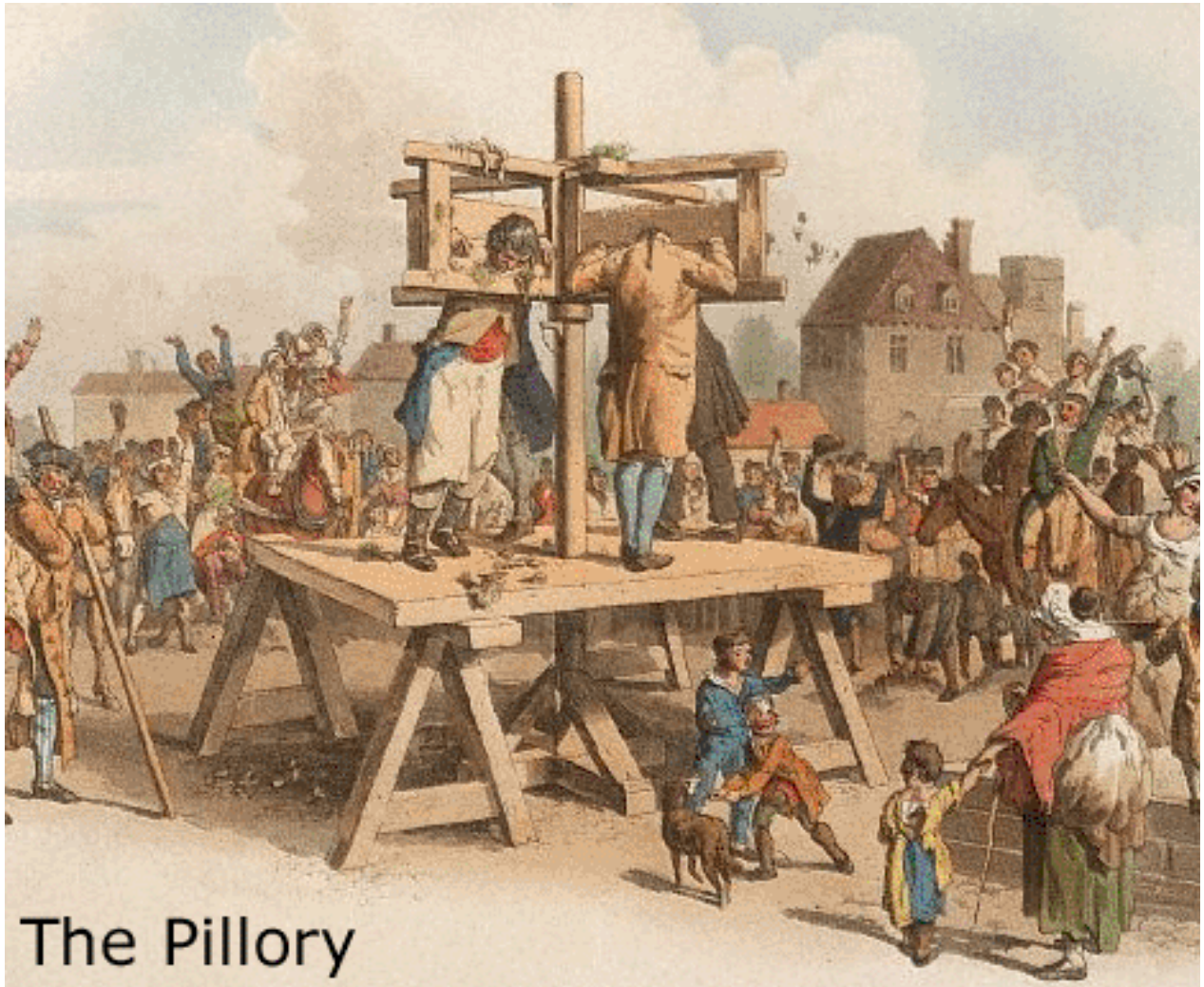
"A huge crowd gathered the night before, but was kept out of the street by a police barrier so the view of the execution scene shows only a half circle of a few hundred spectators, the ones with official passes, allowing them through the police blockade.

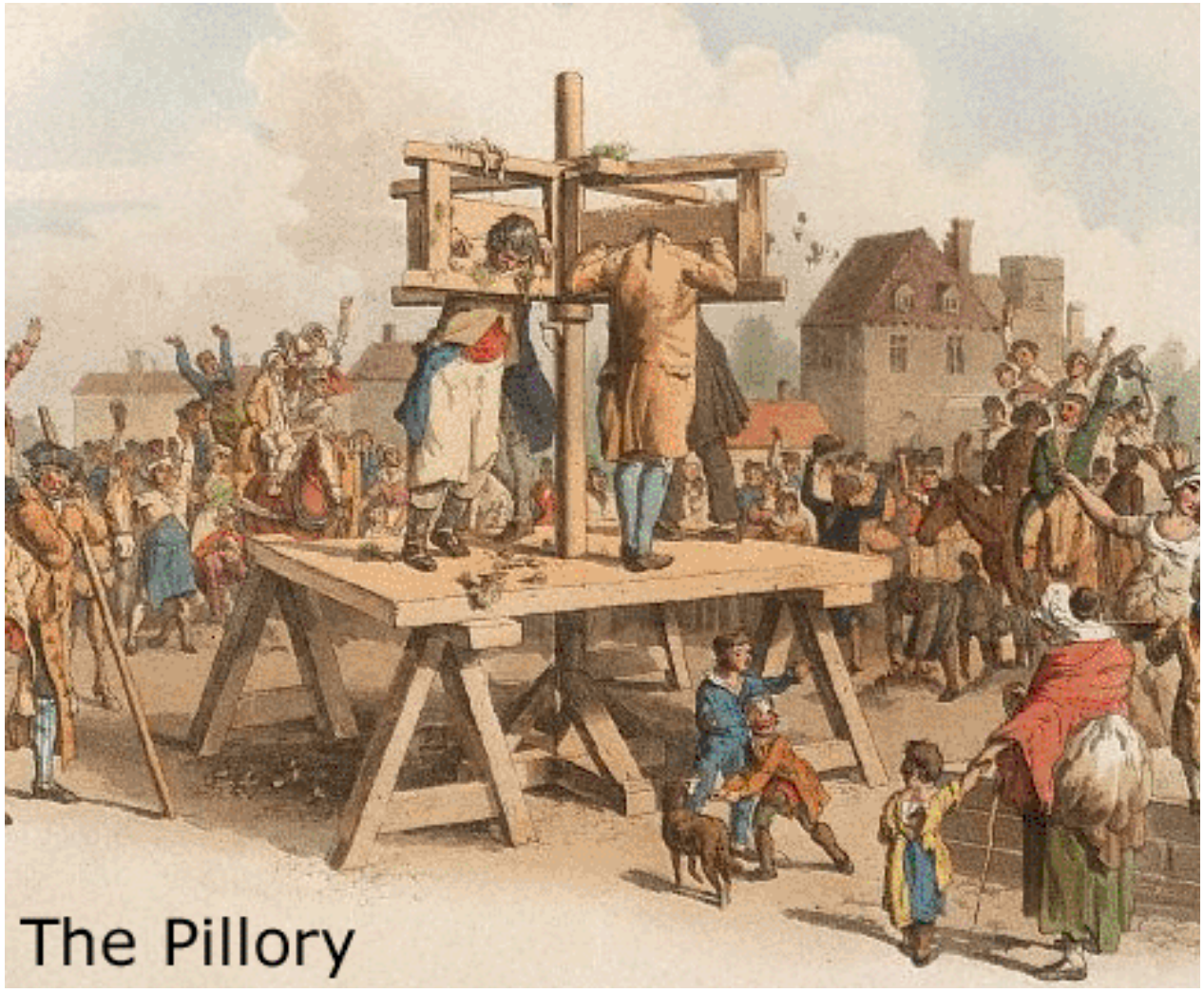
The government downplayed the story and to this day the picture with the small crowd is still used to dispel the "myth" of the near-riot situation that occurred that morning. The reality was that around 30-40,000 rowdy, drunken, screaming and singing "would-be" spectators spent the night partying in the surrounding streets.

After the execution was over and the guillotine had been dismantled, this bloodthirsty crowd invaded the area. Reports of women dipping handkerchiefs in the bloody water on the sidewalk were, in fact, true.

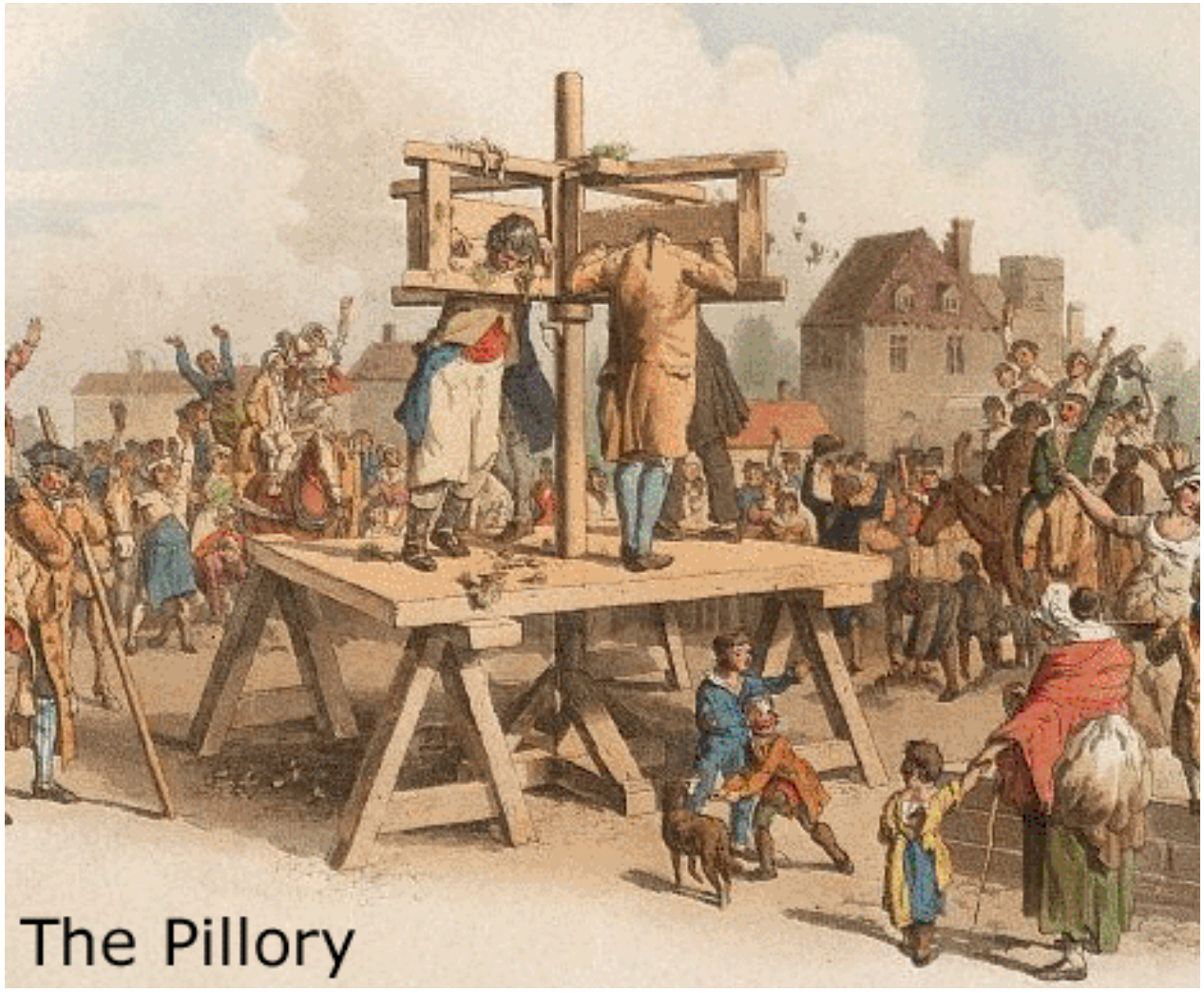
It is not known if the crowd's undignified behavior, the illegal photography and filming, the flashy press coverage or the new executioner's apparent incompetence prompted it, but the government put an end to public executions by the following month."

Source: http://boisdejustice.com/History/History.html


The Pillory



- Also abolished



- Also abolished.. but coming back (see later)

Shaming sanctions (modern pillory)

- Fast-increasing trend in the U.S.

Shaming sanctions (modern pillory)

- Fast-increasing trend in the U.S.
- A number of legal scholars (e.g., Kahan; later recanted) have argued for more recourse to them:
  have both good efficiency (very cheap) and expressive properties.

Shaming sanctions (modern pillory)

- Fast-increasing trend in the U.S.
- A number of legal scholars (e.g., Kahan; later recanted) have argued for more recourse to them:
  have both good efficiency (very cheap) and expressive properties.
- Arguments against?
  - Ineffective, many offenders have no shame. Or, more permanent stigma than prison, hence bad incentives ex post. But highly debatable (middle class offenders), not based on any evidence
  - Cruel, enlists the public in punishment; activates preferences for debasement. Habit formation?

Shaming sanctions (modern pillory)

- Fast-increasing trend in the U.S.
- A number of legal scholars (e.g., Kahan; later recanted) have argued for more recourse to them:
  have both good efficiency (very cheap) and expressive properties.
- Arguments against?
  - Ineffective, many offenders have no shame. Or, more permanent stigma than prison, hence bad incentives ex post. But highly debatable (middle class offenders), not based on any evidence
  - Cruel, enlists the public in punishment; activates preferences for debasement. Habit formation?
  - Variance: severity hard to predict / control, as relies on socially enforced sanctions such as ostracism from community $\rightsquigarrow$ issues of coordination, emotions. May over or underpunish, erratic
    Variability in μ, Lecture III.
  - Richard Posner: inflicting not just shame but humiliation, thus neglecting expressive content of other important societal values (respect for human dignity of all).

Wall of shame

- Let us go to Arizona.

Wall of shame

- Let us go to Arizona. Many other states, counties
- A federal judge in March 2003 ordered X... to stand for eight hours outside a San Francisco post office wearing a two-sided "sandwich board" bearing the words: "I stole mail. This is my punishment."
- Shoplifters have been required to stand outside stores with signs announcing their crimes.
- In Escambia County, Fla., and in Ohio, drunken drivers are issued special license plates that identify them to fellow motorists.
- In Houston and Corpus Christi, Texas, convicted sex offenders ordered to place signs on their front lawns that warn away children.
- In Pennsylvania,... the driver of a car that caused a fatal accident was forced to carry a picture of the victim.

Wall of shame

- "Sick of johns [customers for prostitution] cruising International Boulevard at all hours, driving down property values and creating a market for the sex trade,... Oakland is about to strike back with a new weapon: shame.
- Those caught by surveillance cameras and convicted of solicitation will be at risk of having their faces plastered on bus stop signs or even 10-foot by 22-foot billboards. Clear Channel is providing the advertising space. [CSR?]
- As elected officials, state judges know that few things please the public as much as hoisting a wretch in public. One Texas state judge, Ted Poe, was known as "The King of Shame" for his signature use of punishments like shoveling manure. Poe said that he liked to humiliate people because "[t]he people I see have too good a self-esteem."

Wall of shame

- "Sick of johns [customers for prostitution] cruising International Boulevard at all hours, driving down property values and creating a market for the sex trade,... Oakland is about to strike back with a new weapon: shame.
- Those caught by surveillance cameras and convicted of solicitation will be at risk of having their faces plastered on bus stop signs or even 10-foot by 22-foot billboards. Clear Channel is providing the advertising space. [CSR?]
- As elected officials, state judges know that few things please the public as much as hoisting a wretch in public. One Texas state judge, Ted Poe, was known as "The King of Shame" for his signature use of punishments like shoveling manure. Poe said that he liked to humiliate people because "[t]he people I see have too good a self-esteem." Poe was so popular for what he called "Poe-tic Justice" that he literally shamed himself right into Congress and is now serving as a member of the House of Representatives"

Cruel and unusual punishments - what kind of society?

- The "banality of evil": a fraction κ (for "cruel") of agents in society actually enjoy the suffering of others. Could also be "vengeful" types
  - Uncorrelated with v_a, for simplicity. Can relax
  - Seeing cheaters, criminals, etc. punished harshly (p) and publicly is an opportunity (and an excuse) to obtain such enjoyment
  - Utility $\kappa \times p\left[1 - G\left(v^*(p)\right)\right]$; normalize intensity to 1

- People do not like to think they are surrounded by many cruel types
  - Could hurt them in other circumstances
  - Low social trust, reducing cooperative investments
  - Will assume cruel types also have such preferred beliefs (inessential)

- Assume here simple, linear affective dislike, $U \rightsquigarrow \ U - \beta E\left[\kappa \mid p\right]$
  - Could endogenize, make functional (as with θ), or nonlinear

Optimal severity of punishments

- Government observes κ : via judicial system / crimes, prison life, how people behave in circumstances of lawlessness, war

- Sets incentive $p = p(\kappa)$: "painful" penalty levied on those who choose $a = 0$ in, say, theft, fraud, drunk driving, child abuse
  - Here, only policy tool. Could also have alternative, "non-painful" incentives y (fines, jail; rewards): do not generate as much enjoyment for cruel types, but more costly (pain is cheap)

- Weight $0 < \zeta \leq 1$ on the utility of cruel types

  Equivalently: their political influence

- Enforcement may be costly: $\lambda \geq 0$ per unit. Or: "empathic" types who dislike seeing pain inflicted even on the guilty

- Agents infer κ as solution $\hat{\kappa}(p)$ to $p(\hat{\kappa}(p)) = p$

- Cutoff for participation:

$$v^*(p) - c + p + \mu \Delta v^*(p) = 0$$

independent of $\kappa \Rightarrow$ planner maximizes

$$W_\kappa^{AI}(p) = \int_{v^*(p)}^{+\infty} (e + v - c)\, g(v) dv + \mu(\bar{v} + \theta)$$

$$- p\,(1 + \lambda - \kappa\zeta)\,[1 - G\,(v^*(p))] - \beta\hat{\kappa}(p)$$

- $W_\kappa^{AI}(\cdot)$ quasiconcave for λ, κ small enough. FOC:

$$\left(\frac{e - c + v^*(p) + p\,(1 + \lambda - \kappa\zeta)}{1 + \mu\Delta'(v^*(p))} \right) g(v^*(p))$$

$$= \frac{1 + \lambda - \kappa\zeta}{h(v^*(p))} + \beta \frac{\hat{\kappa}'(p)}{g(v^*(p))}$$

Implications

- Presence of κ-types reduces deadweight loss from punishment, if society counts their utility: $\lambda \rightsquigarrow \lambda - \kappa\zeta$

  $\Rightarrow$ harsher punishments, closer to or even beyond pure deterrence level for $\lambda = 0$

- Desired belief for not living in an evil world, $\beta > 0$, leads to

  - Restrictions on cruel punishments (physical, psychological) whether or not effective at the margin

    as well as

  - Restrictions on public infliction of harsh punishments

    (what goes on in prisons: out of sight)

Remark

- Equilibrium is separating $\Rightarrow$ ultimately no one fooled, no welfare gain P would like to be able to commit to the truth; standard, but should not put too much weight on it
- Choice of a particular objective function
- Could also look for pooling equilibria, e.g., with discrete types
  - May be welfare gains from pooling if, say indivisibilities in cooperative investments make social payoffs nonlinear
  - Greater losses from increase in distrust $\hat{\kappa}$ than gain from equivalent decrease
- All agents 100% sophisticated Bayesians: unrealistic
  - A fraction may be naifs, esp. children. Take p at face value: $\hat{\kappa}_n(\cdot) = (p^{FI})^{-1}(\cdot)$ instead of $\hat{\kappa}_{Bayes}(\cdot) = (p^{AI})^{-1}(\cdot)$

Summary

1. Analyzed how social esteem and stigma shape behavior
   - Admirable behaviors: few people do, SS, multiplier $-\mu\Delta' < 0$ incentives y partially dampened by crowding out
   - Respectable behaviors: most people do, SC, positive multiplier $-\mu\Delta' > 0$, incentives y amplified by crowding in
2. Social or self esteem is by its very nature a positional good
   - Prosocial actions inefficiently distorted toward the most visible
3. Optimal incentives under symmetric info: Pigou- Ramsey, adjusted by reputation tax
4. Norms based interventions
5. Optimal incentives under asymmetric info: expressive role of law
   - Weakens optimal incentives when informative about society's general "goodness" θ or "cruelty" κ. Strengthens them when informative about importance of externalities e
   - What is expressed concerning θ by law or incentives bearing on one activity carries over to people's attitudes and behavior in others

Incentives, Attributions, and Crowding Out

1. Multiple motives and signal extraction
2. Material incentives
3. Image incentives

Back to crowding out

- Saw in L1, a first mechanism, where y crowds out intrinsic motivation v_a. Incentives convey "bad news", lack of trust.
  - Requires informed principal and appropriate sorting condition

Back to crowding out

- Saw in L1, a first mechanism, where y crowds out intrinsic motivation v_a. Incentives convey "bad news", lack of trust.
  - Requires informed principal and appropriate sorting condition
- Saw in L2 a second mechanism, that gives only partial crowding out: for honor-driven behaviors (only), social multiplier $-\mu\Delta' < 0 \quad \Rightarrow$ the more people do it (e.g., because of higher y), the lower the reputational incentive to do it.
  - Not specific to y, and cannot generate a net negative response

Back to crowding out

- Saw in L1, a first mechanism, where y crowds out intrinsic motivation v_a. Incentives convey "bad news", lack of trust.
  - Requires informed principal and appropriate sorting condition
- Saw in L2 a second mechanism, that gives only partial crowding out: for honor-driven behaviors (only), social multiplier $-\mu\Delta' < 0 \quad \Rightarrow$ the more people do it (e.g., because of higher y), the lower the reputational incentive to do it.
  - Not specific to y, and cannot generate a net negative response
- Other idea: incentives (esp. money) "sully the meaning" of good actions: no longer clear, to others and oneself, whether done for the "right" reason (virtue) or the wrong one ("greed").

Back to crowding out

- Saw in L1, a first mechanism, where y crowds out intrinsic motivation v_a. Incentives convey "bad news", lack of trust.
  - Requires informed principal and appropriate sorting condition
- Saw in L2 a second mechanism, that gives only partial crowding out: for honor-driven behaviors (only), social multiplier $-\mu\Delta' < 0 \quad \Rightarrow$ the more people do it (e.g., because of higher y), the lower the reputational incentive to do it.
  - Not specific to y, and cannot generate a net negative response
- Other idea: incentives (esp. money) "sully the meaning" of good actions: no longer clear, to others and oneself, whether done for the "right" reason (virtue) or the wrong one ("greed").
- Formalize this idea. Show can generate full crowding out and many other interesting results.

Revisiting incentives, in three steps

$$U = (v_a + v_y y)a - C(a) + x\mu_a E(v_a|a, y, x) - x\mu_y E(v_y|a, y, x) + e\bar{a}$$

$$W = \alpha \bar{U}(x, y) + [B - (1 + \lambda)y]\,\bar{a}(x, y) - \varphi(x)$$

1. Incentives and intrinsic motivation: y affects perceived v_a or $C(a)$
   - Focus on private P-A setup: $e = 0$, $\mu_a = \mu_y \equiv 0$, x irrelevant, $v_y \equiv 1$; $\alpha = 0$, $\lambda = 0$
2. Incentives and attributional motivation – social norms: y affects $x\mu_a E(v_a|a, y, x)$; also role of x
   - Focus on basic public-goods setup with unidimensional uncertainty: $e > 0$, $\mu_a > 0 = \mu_y$, $v_y \equiv 1$, $\alpha = 1$, $\lambda \geq 0$
3. Incentives and attributional motivation – the "meaning of acts" Signal-extraction by agents and / or principal
   - Key: multidimensional uncertainty (idiosyncratic, aggregate) about the v's, μ's, e

Preferences: attributional motivations

- Desire, instrumental / hedonic, for being seen as having a high v_a :
  - Private-goods context: career concerns $\rightsquigarrow$ valuable to bee seen by employers as motivated for the activity or sector in question; as perfectionist, honest, ethical, etc. In the spirit of Holmström.

    Type signaled = some general "talent", not employer-specific
  - Public-goods context: desirable to be perceived as generous, public minded, reciprocal, good citizen, etc. More likely to be chosen as mate, friend, leader, elected to office, etc.

- May also care about perceptions concerning v_y
  - In most contexts, undesirable to be perceived as greedy, willing to do anything for money, or poor / needy
  - In a rare cases (Wall Street of old) may be good to be seen as "hungry", because then easily controllable by incentives
    - 1980's Gordon Gekko: "Greed is good"

Preferences: attributional motivations

- Desire, instrumental / hedonic, for being seen as having a high v_a :
  - Private-goods context: career concerns $\rightsquigarrow$ valuable to bee seen by employers as motivated for the activity or sector in question; as perfectionist, honest, ethical, etc. In the spirit of Holmström.

    Type signaled = some general "talent", not employer-specific
  - Public-goods context: desirable to be perceived as generous, public minded, reciprocal, good citizen, etc. More likely to be chosen as mate, friend, leader, elected to office, etc.

- May also care about perceptions concerning v_y
  - In most contexts, undesirable to be perceived as greedy, willing to do anything for money, or poor / needy
  - In a rare cases (Wall Street of old) may be good to be seen as "hungry", because then easily controllable by incentives
    - 1980's Gordon Gekko: "Greed is good"
    - 2010's: Lloyd Blankfein, CEO of Goldman Sachs:

      "I am just a banker doing God's work"

Multidimensional heterogeneity and signaling

- Actions a vary continuously over $\mathbb{R}$, cost $C(a) = ka^2/2$. FOC:

$$v_a + v_y y + \underbrace{\mu_a \frac{\partial E\left(v_a|a,y\right)}{\partial a} + \mu_y \frac{\partial E\left(v_y|a,y\right)}{\partial a}}_{\text{analogue to } \Delta} = ka$$

- Agents' valuations $\mathbf{v} \equiv (v_a, v_y)$ are distributed in the population as

$$\begin{pmatrix} v_a \\ v_y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \bar{v}_a \\ \bar{v}_y \end{pmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{ay} \\ \sigma_{ay} & \sigma_y^2 \end{bmatrix} \right), \quad \bar{v}_a \gtrless 0, \quad \bar{v}_y > 0,$$

- Focus first on case where everyone has same reputational concerns, $\boldsymbol{\mu} \equiv (\mu_a, \mu_y) = (\bar{\mu}_a, \bar{\mu}_y) \rightsquigarrow$ study material rewards

- Then, extend analysis to case where $\boldsymbol{\mu}$ is also normally distributed across individuals $\rightsquigarrow$ study image rewards

Material rewards

- Common $\boldsymbol{\mu} = \bar{\mu}$ $\Rightarrow$ same reputational return for all agents

$$\bar{r}(a, y) \equiv \bar{\mu}_a \frac{\partial E\left(v_a | a, y\right)}{\partial a} - \bar{\mu}_y \frac{\partial E\left(v_y | a, y\right)}{\partial a}$$

- So by FOC, an agent's choice of a reveals the combination

$$v_a + y v_y = ka - \bar{r}(a, y).$$

- Signal extraction with normal random variables $\Rightarrow$

$$E\left(v_a | a, y\right) = \bar{v}_a + \rho(y) \cdot \left[ka - \bar{v}_a - \bar{v}_y \; y - \bar{r}(a, y)\right]$$

$$E\left(v_y | a, y\right) = \bar{v}_y + \chi(y) \cdot \left[ka - \bar{v}_a - \bar{v}_y \; y - \bar{r}(a, y)\right]$$

- Assessed prosociality = weighted average of prior $\bar{v}_a$ and marginal cost ka of contribution, net of mean extrinsic and image incentives to contribute at that level

$$\rho(y) \equiv Corr\left(v_a, \; v_a + y v_y\right) \quad \text{and} \quad y\chi(y) \equiv 1 - \rho(y)$$

Solving

- Weights

$$\rho(y) \equiv \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2} \qquad \text{and} \quad y\chi(y) \equiv 1 - \rho(y)$$

- Combine signal-extraction rules + marginal return to signaling $\Rightarrow$ equilibrium defined by a function $\bar{r}(a, y)$ solving linear DE in a (note: y is fixed here)

$$\bar{r}(a, y) = [\bar{\mu}_a \rho(y) - \bar{\mu}_y \chi(y)] \cdot \left(k - \bar{r}'(a, y)\right)$$

- Unique solution that is non-explosive, consistent with global maximum in agent's objective function is constant one

$$\bar{r}(a, y) = [\bar{\mu}_a \rho(y) - \bar{\mu}_y \chi(y)] \cdot k$$

Proposition

Let all agents have the same image concern $(\bar{\mu}_a, \bar{\mu}_y)$.

1. *There is a unique (differentiable-reputation) equilibrium, in which an agent with preferences* (v_a, v_y) *contributes*

$$a = \frac{v_a + v_y\ y}{k} + \bar{\mu}_a \rho(y) - \bar{\mu}_y \chi(y),$$

   with $\rho(y)$ *and* $\chi(y)$ *defined earlier.*

2. *Reputational returns: are* $\partial E(v_a)/\partial a = \rho(y)k$, $\partial E(v_y)/\partial a = \chi(y)k$, *with net value*

$$\bar{r}(y) = k\left[\bar{\mu}_a \rho(y) - \bar{\mu}_y \chi(y)\right].$$

Proposition

Let all agents have the same image concern $(\bar{\mu}_a, \bar{\mu}_y)$.

1. *There is a unique (differentiable-reputation) equilibrium, in which an agent with preferences* (v_a, v_y) *contributes*

$$a = \frac{v_a + v_y \ y}{k} + \bar{\mu}_a \rho(y) - \bar{\mu}_y \chi(y),$$

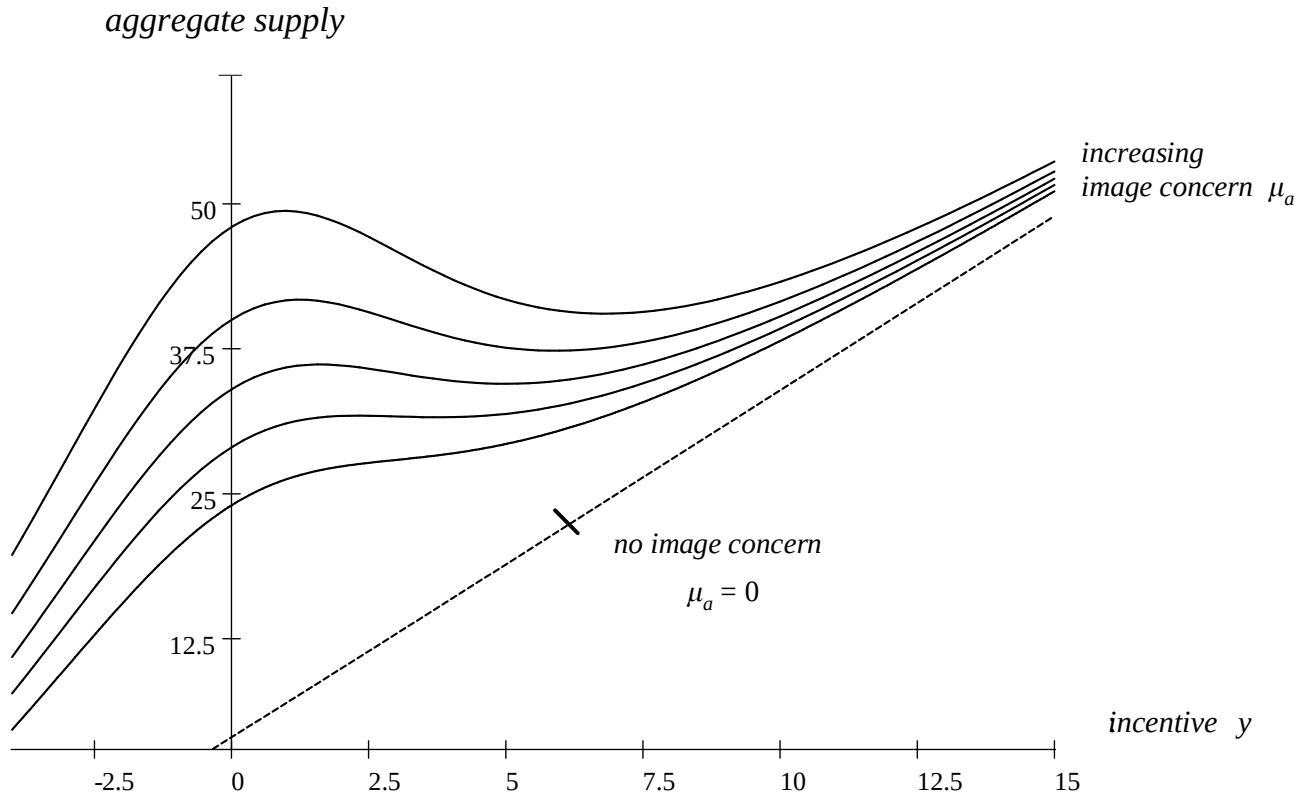
   with $\rho(y)$ *and* $\chi(y)$ *defined earlier.*

2. *Reputational returns: are* $\partial E(v_a)/\partial a = \rho(y)k$, $\partial E(v_y)/\partial a = \chi(y)k$, *with net value*

$$\bar{r}(y) = k\left[\bar{\mu}_a \rho(y) - \bar{\mu}_y \chi(y)\right].$$

- Effects of extrinsic incentives on inferences and behaviors:
  - Higher y increases direct payoff from contributing, $v_a + v_y \ y$
  - But also tends to impair signaling value, along both dimensions

- With $\sigma_{ay} = 0$: $\quad \bar{a}(y) = \frac{\bar{v}_a + \bar{v}_y \ y}{k} + \frac{\bar{\mu}_a - \bar{\mu}_y y \sigma_y^2 / \sigma_a^2}{1 + y^2 \sigma_y^2 / \sigma_a^2}$



- Drawn for $\mu_a \nearrow$, with $\bar{\mu}_y = 0$: no stigma on greed / neediness
- When y increases, pro-social behavior is increasingly ascribed to greed, and less to genuine altruism

Proposition (overjustification and crowding out)

Let $\sigma_{ay} = 0$.

1. *Incentives are counterproductive,* $\bar{a}'(y) < 0$, *at all levels such that*

$$\frac{\bar{v}_y}{k} < \bar{\mu}_a \cdot \frac{2y\sigma_y^2/\sigma_a^2}{\left(1 + y^2\sigma_y^2/\sigma_a^2\right)^2} + \bar{\mu}_y \cdot \frac{\sigma_y^2/\sigma_a^2\left(1 - y^2\sigma_y^2/\sigma_a^2\right)}{\left(1 + y^2\sigma_y^2/\sigma_a^2\right)^2}$$

2. *For all* $\bar{\mu}_a$ *above some threshold* $\mu_a^* \geq 0$ *there is a range* $[y_1, y_2]$ *such that* $\bar{a}(y)$ *is decreasing on* $[y_1, y_2]$ *and increasing elsewhere on* $\mathbb{R}$.

Implications and empirical tests

- Recall Chatenay letter
- People contribute more when observed by others: $\partial \bar{a}/\partial \mu > 0$ (standard), but also
  - This should attenuate when they are rewarded for it: $\partial^2 \bar{a}/\partial y \partial \mu < 0$
  - Equivalently, effectiveness of incentives y smaller, or even reversed when contribution and reward are observed is observed
  - Experimental test: Ariely-Bracha-Meier (2007)
- Goeschl-Perino (2009): experimental evidences that pollution taxes crowd out intrinsic motivation to purchase and retire (actual) CO_2 emission permits from the European Union Emission Trading System

Ariely, Bracha, Meier (2007): "Click for charity"

- 161 Princeton undergraduates
- Task: sequentially pressing keys X and Z on the keyboard for up to 5 minutes.
- For every X-Z pair, pay money in participant's name to an assigned charity: 1 cent for each of first 200 pairs, 0.5 cents for each of next 200 pairs, 0.25 cents for each of next 200 pairs,... 0.01 cents for each above 1,200.
- Design: $2 \times 2 \times 2$:
  - "Good" or "Bad" Charity: American Red Cross, National Rifle Association
  - Incentives: either no payment to self, or same schedule as for charity,. Implemented with random draw
  - Private vs. public condition: anonymous, vs. at the end, must tell other participants which charity was assigned to, $ earned for it and for oneself

Ariely et al. (2007)

Figure 1: Effect of Private Incentive for "Good" Charity

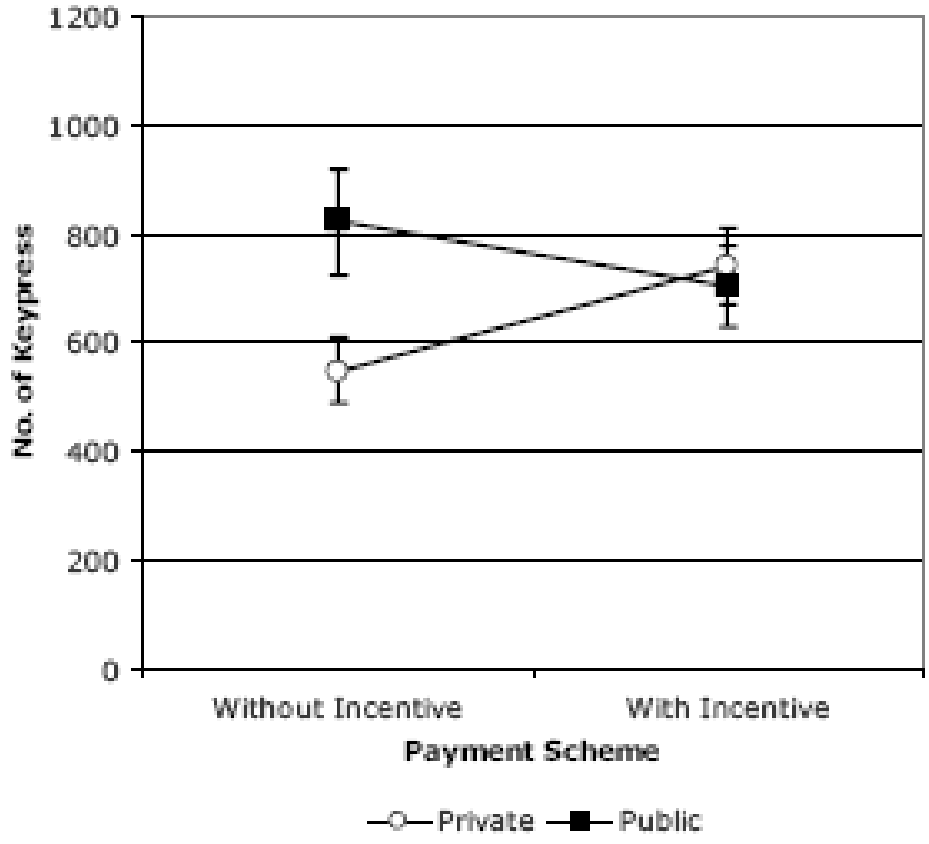
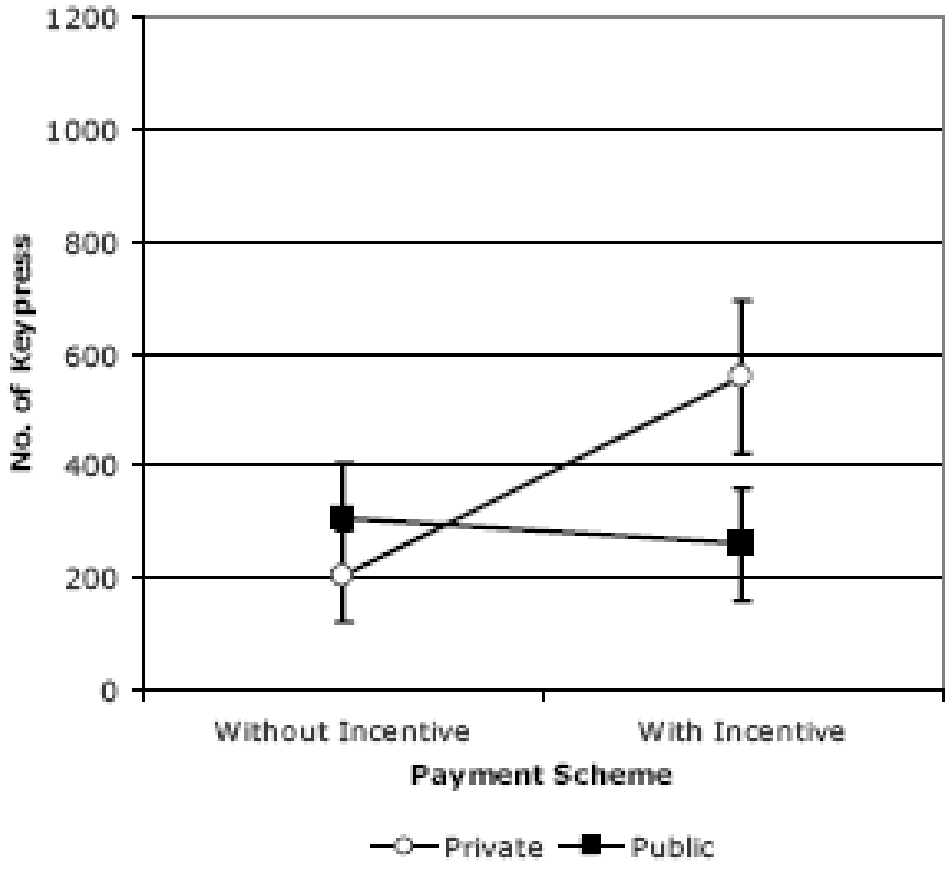


Figure 2: Effect of Private Incentive for "Bad" Charity

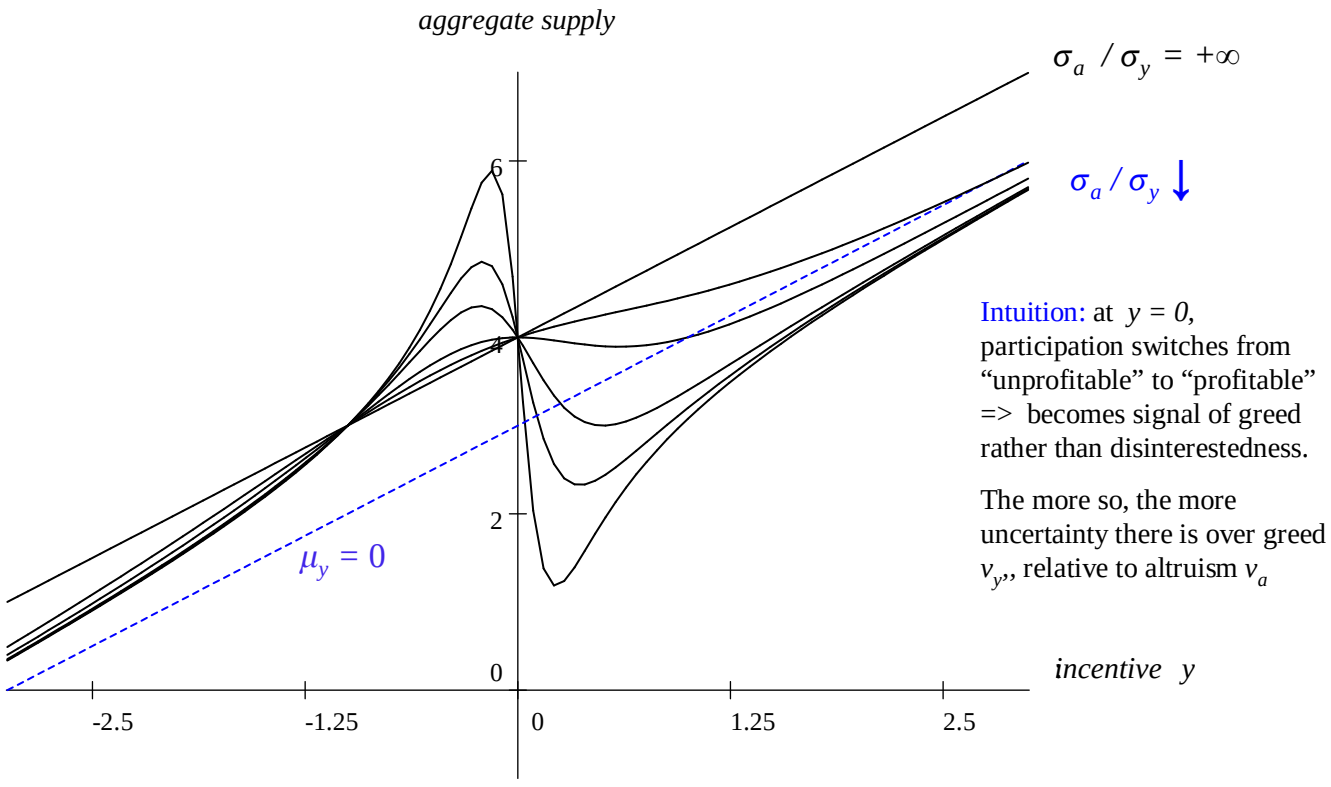


The case of "small rewards"

- Some studies find crowding out ($\bar{a}(y) \searrow$) to occur mostly at low \$ amounts. Then, why relevant?
- Sometimes suggested that the main effect is a discontinuity at zero in subjects' response to incentives. Appeal to framing

  (e.g., Gneezy-Rustichini 2000b, Bowles-Reyes 2009)
- Is there something qualitatively different between "unrewarded" and "rewarded" activities that could cause rational agents to behave in this way?
- Show that there is. But also that relevant notion of "small" rewards is quite different in real-world .vs. lab.

- With $\sigma_{ay} = 0$, $\bar{a}'(0) = \frac{\bar{v}_y}{k} - \bar{\mu}_y \left(\frac{\sigma_y}{\sigma_a}\right)^2$



- Illustrate with $\bar{\mu}_y > 0 = \mu_a$: no concern to appear prosocial
- In situations with much more uncertainty (more to learn) about individuals' desire for money than about their motivation for task at hand, even minimal concern about appearing greedy (small $\bar{\mu}_y > 0$) is sufficient to cause sharply negative response to small incentives $\rightsquigarrow$ downward discontinuity in supply

Small rewards and signal-reversal

Proposition (signal-reversal)

1. *Small incentives are counterproductive,* $\bar{a}'(0) < 0$, *whenever*

$$\frac{\bar{v}_y}{k} < \bar{\mu}_a \left(\frac{\sigma_{ay}}{\sigma_a^2} \right) + \bar{\mu}_y \left(\frac{\sigma_y^2 - 2\sigma_{ay}^2/\sigma_a^2}{\sigma_a^2} \right)$$

2. *Let* v_a *and* v_y *be uncorrelated, or not too correlated. As* $\sigma_a/\sigma_y \to 0$, *the supply function's slope at* $y = 0$ *tends to* $-\infty$.

3. *Let participation entails unit opportunity cost with monetary value* $\tilde{y}$. *Then* $\bar{a}'(\tilde{y}) < 0$ *and* $\bar{a}'(\tilde{y}) \to -\infty$ *under conditions (1) and (2).*

Remarks

- Result applies whether or not the task has prosocial dimension ($\bar{\mu}_a \geq 0$). Explains why adverse effects of small rewards found both in experiments with private, puzzle-solving tasks and others involving public-goods provision (raising money for charity)

- Signal-reversal effect creates, around zero net reward, additional source of crowding out on top of signal-jamming ($\rho(y) \downarrow$), which operates at all y's for acts with $\mu_a > 0$

- If empirical validity of crowding-out / discontinuity was restricted to very small prizes and fines, it would be of limited interest.

- Proposition shows that relevant "tipping point" is not really zero –except in lab, where subjects have no alternative uses of time. It is instead agents' opportunity cost value of time, can be significant.

- Suggests future work should involve situations where opportunity costs are (known to be) non-trivial and vary across subjects.

Image-based incentives

- Public authorities and private sponsors make heavy use of both public displays and private mementos conveying honor or shame
  - Nations award medals and honorific titles, non-profits give bumper stickers and T-shirts with logos, charities send donors pictures of "their" sponsored child, universities award honorary degrees
  - The new pillory: ⏩ televised arrests, internet posting of drunk drivers, parents delinquent on child support,... Publishing licence plates of cars photographed in areas of drug trafficking or prostitution
- Discussed and modelled earlier the expressive value of extreme shaming as cruel "humiliation".
- Now, different issue: effectiveness of image incentives, whether
  - Bilateral / negative, but non-cruel: e.g., revealing who voted or not, (Gerber et al., Funk), how much people contributed to charity, etc.
  - Especially, positive: honors, distinctions.

Image rewards and image concerns

- Idea: when people are heterogeneous in their image concerns, giving greater visibility to behavior will cause good actions to be tainted with suspicion of being image-motivated

- Greater publicity or prominence: increase in (μ_a, μ_y)

- Let agents' image concerns, like their v's, be normally distributed:

$$\begin{pmatrix} \mu_a \\ \mu_y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \bar{\mu}_a \\ \bar{\mu}_y \end{pmatrix}, \begin{bmatrix} \omega_a^2 & \omega_{ay} \\ \omega_{ay} & \omega_y^2 \end{bmatrix} \right), \quad \bar{\mu}_a \geq 0, \quad \bar{\mu}_y \geq 0,$$

  with $\mathbf{v}$ and $\boldsymbol{\mu}$ independent

- FOC now has both exogenous + endogenous "noise" affecting signal

$$v_a + v_y y + \mu_a \frac{\partial E\left(v_a | a, y\right)}{\partial a} + \mu_y \frac{\partial E\left(v_y | a, y\right)}{\partial a} = ka$$

- Reputational return $r(a, y; \boldsymbol{\mu})$ also normal and, conditionally on a, independent of $\mathbf{v}$. Mean $\bar{r}(a, y)$ given by

$$\bar{r}(a, y) \equiv \bar{\mu}_a \frac{\partial E\left(v_a | a, y\right)}{\partial a} - \bar{\mu}_y \frac{\partial E\left(v_y | a, y\right)}{\partial a}$$

and variance

$$\Omega(a, y)^2 \equiv \left(\begin{array}{cc} \frac{\partial E(v_a|a,y)}{\partial a} & -\frac{\partial E(v_y|a,y)}{\partial a} \end{array} \right) \left[\begin{array}{cc} \omega_a^2 & \omega_{ay} \\ \omega_{ay} & \omega_y^2 \end{array} \right] \left(\begin{array}{c} \frac{\partial E(v_a|a,y)}{\partial a} \\ -\frac{\partial E(v_y|a,y)}{\partial a} \end{array} \right)$$

- Signal-extraction rules unchanged, with new updating coefficients

$$\rho(a, y) \equiv \frac{\sigma_a^2 + y\sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2 + \Omega(a, y)^2},$$

$$\chi(a, y) \equiv \frac{y\sigma_y^2 + \sigma_{ay}}{\sigma_a^2 + 2y\sigma_{ay} + y^2\sigma_y^2 + \Omega(a, y)^2}.$$

Solving

- Equilibrium = again a pair of functions $E\left(v_a|a,y\right)$ and $E\left(v_y|a,y\right)$ that solve system of DE
- System is now nonlinear, due to term $\Omega(y)^2 = Var\left(r(y;\boldsymbol{\mu})\right)$ in ρ and χ : greater variability of image motives makes behavior a more noisy measure of underlying values (v_a, v_y), reducing $\rho(y)$ and $\chi(y)$
- This variance is endogenous, however, (depends on $\rho(y)$ and $\chi(y)$) : agents' reputational calculus takes into account how their collective behavior affects observers' signal-extraction-problem.
- This is reflected in the fixed-point nature of Ω

Proposition

1. *A linear-reputation equilibrium corresponds to a fixed-point* $\Omega(y)$,

$$\Omega(y)^2/k^2 \equiv \omega_a^2\ \rho(y)^2 - 2\omega_{ay}\ \rho(y)\chi(y) + \omega_y^2\ \chi(y)^2,$$

   with $\rho(y)$ *and* $\chi(y)$ *given earlier as functions of* $\Omega(y)$.

2. *An agent with type* $(\mathbf{v},\boldsymbol{\mu})$ *contributes*

$$a = \frac{v_a + y \cdot v_y}{k} + \mu_a \rho(y) - \mu_y \chi(y)$$

3. *Reputational returns are* $\partial E\left(v_a\right)/\partial a = \rho(y)k,\ \partial E\left(v_y\right)/\partial a = \chi(y)k$, *with net value for the agent*

$$r(y;\boldsymbol{\mu}) = (\mu_a \rho(y) - \mu_y \chi(y))k$$

4. *There exists such an equilibrium. If* $\omega_{ay} = 0$ *it is unique (linear).*

Publicity and the overjustification effect

- Reputational weights $\boldsymbol{\mu} = (\mu_a, \mu_y)$ scaled up by prominence or memorability factor, x. Material incentive y remains constant.

- Aggregate supply

$$\bar{a}(y, x) = \frac{\bar{v}_a + y \cdot \bar{v}_y}{k} + x\bar{\mu}_a\rho(y, x) - x\bar{\mu}_y\chi(y, x)$$

  - Dependence on x indicates that all $\mu-$covariance terms affecting variance Ω of reputational returns are now multiplied by x^2

Publicity and the overjustification effect

- Reputational weights $\boldsymbol{\mu} = (\mu_a, \mu_y)$ scaled up by prominence or memorability factor, x. Material incentive y remains constant.

- Aggregate supply

$$\bar{a}(y, x) = \frac{\bar{v}_a + y \cdot \bar{v}_y}{k} + x\bar{\mu}_a\rho(y, x) - x\bar{\mu}_y\chi(y, x)$$

  - Dependence on x indicates that all $\mu-$covariance terms affecting variance Ω of reputational returns are now multiplied by x^2

- Greater visibility of actions $\Longrightarrow$ two offsetting effects:

  - Direct amplifying effect: sign is that of $\mu_a\rho(y, x) - \mu_y\chi(y, x)$ for an individual, and $\bar{\mu}_a\rho(y, x) - \bar{\mu}_y\chi(y, x)$ on average
  - Dampening effect: reputation becomes less sensitive to behavior, which observers increasingly ascribe to image-seeking: $\rho,\ \chi \searrow$

Implications

- For instance, when μ_y is known $(\omega_y = 0)$ while μ_a varies across people, and as x becomes large

$$\rho(y,x) \approx \left(\frac{\sigma_a^2 + y\sigma_{ay}}{k^2\omega_a^2}\right)^{1/3} x^{-2/3}.$$

- Effectiveness of publicity has rapidly decreasing returns: $x\rho(y,x)\bar{\mu}_a$ grows only as $x^{1/3} \Rightarrow$ marginal cost $\varphi'(x)$ vs. $x^{-2/3}$

- Message: policies by parents, teachers, governments and other principals that rely on the "currency" of praise and shame are effective up to a point, but eventually self-limiting

Publicity, Privacy and Evolving Societal Values:

Image Versus Information

(work in progress with Nageeb Ali, UCSD)

What's wrong with publicizing everything?

- Saw, both theoretically and empirically, that publicity / visibility x, amplifying honor and (especially) stigma, is a powerful incentive
- It is also very cheap
- So what is wrong with using it extensively?

What's wrong with publicizing everything?

- Saw, both theoretically and empirically, that publicity / visibility x, amplifying honor and (especially) stigma, is a powerful incentive
- It is also very cheap
- So what is wrong with using it extensively?

1. Deep shaming as a cruel humiliation with bad "expressive" properties
   - Let's leave that out, focus on "pure" publicity / transparency

What's wrong with publicizing everything?

- Saw, both theoretically and empirically, that publicity / visibility x, amplifying honor and (especially) stigma, is a powerful incentive
- It is also very cheap
- So what is wrong with using it extensively?

1. Deep shaming as a cruel humiliation with bad "expressive" properties
   - Let's leave that out, focus on "pure" publicity / transparency
2. Overjustification effect if people are heterogeneous in image concerns, value of publicity has decreasing returns
   - But does not negate it.

What's wrong with publicizing everything?

- Saw, both theoretically and empirically, that publicity / visibility x, amplifying honor and (especially) stigma, is a powerful incentive
- It is also very cheap
- So what is wrong with using it extensively?

1. Deep shaming as a cruel humiliation with bad "expressive" properties
   - Let's leave that out, focus on "pure" publicity / transparency
2. Overjustification effect if people are heterogeneous in image concerns, value of publicity has decreasing returns
   - But does not negate it.

- Still, have some unease at the idea of a society with zero privacy and systematic public dissemination of good and bad behaviors

Two new arguments I

1. Unpredictability / variance: the severity of the punishment is hard to control / predict a priori.
    - Real sanction is in the social ostracism of the exposed perpetrator
    - Because this involves both the emotional response of many others and their degree of coordination, it can vary significantly over place, time, groups, offenses, and individuals (Eric Posner)

Two new arguments II

- But of course, purpose back then was precisely to discourage / repress such behaviors, as were the laws of those times.
- Problem is that societal preferences change, due to technology, migration, exposure to other cultures, enlightenment, etc. $\Rightarrow$

  If behavior is too constrained by fear of social shame and associated sanctions, these changes remain hidden from legislator and other decision-makers.
- Model: variability in μ, amplified by x, confronts the principal (e.g., legislator) with own signal-extraction problem in trying to learn or update on θ.

Other applications

- Donations
  - Agents have information about specific public good or charitable cause
  - Principal (church, foundation) motivates them to donate by publicizing who gives what
  - However, Principal (a foundation) may wish to learn how valuable the project actually is, and look to volume of donation as an indicator
- Moral hazard in teams
  - Agents exert effort to sell company's product, and privately observe how well product matches tastes
  - Principal makes investment decisions based on sales (R&D, broaden product line)
  - Publicizing accomplishments incentivizes but crowds out information

- Social norms / political correctness
  - Agents engage in some behavior or speech that is socially approved and refrain from engaging in some that is disapproved
  - Planner / legislator seeks to encourage socially approved behavior or speech using publicity, plus other incentives (e.g., law)
  - Societal values change over time. Principal tries to assess "community standards" by what people do ($\sim$ descriptive norm), but this may be a poor indicator of what people really value and think ($\sim$ prescriptive norm). Google example.
- Corporate social responsibility, green for consumers, goods, etc.
  - Is increasing trend / popularity the result of genuine change in values, or rising visibility concerns?

Agents and principal

- Participation:
  - Each agent i chooses a participation level $a^i \in \mathbb{R} \rightsquigarrow$ aggregate $\bar{a}$
  - Principal chooses own participation a_P (or some other policy)
- Payoffs:
  - Agents derive payoffs from own participation and total participation
  - Principal derives payoffs from total participation
  - All payoffs are increasing in the "quality" of the public good, or other aggregate shift in preferences, denoted θ.
  - Agents obtain private i.i.d. signals θ^i.

Agents' payoffs

- Direct

$$U^i\left(a^i, \bar{a}, \theta\right) = \underbrace{\left(v_a^i + \theta\right)}_{\text{intrinsic motivation}} a^i + \overbrace{\left(e + \theta\right)}^{\text{value from public good}} \left(\bar{a} + a_P\right) - \underbrace{k\frac{a_i^2}{2}}_{\text{cost}}$$

  - Both v_a^i and e increased by higher "quality" of public good or other preference shift θ
  - No price incentive, $y = 0$. Could allow.
  - Normalize $k = 1$

- Reputational
  - Agents have different signals θ^i as well as different v_a^i's $\Rightarrow$ different j's will judge same a differently.
  - Esteem = average assessment of others:

$$R\left(a, \theta^i\right) = \underbrace{\mu_a}_{\text{value of image}} \times E\left[\int_0^1 \underbrace{E\left[v_a^i \mid a, \bar{a}, \theta^j\right] dj}_{\text{what } j \text{ thinks of } i} \mid \theta^i\right]$$

  - No image concerns over v_y : $\mu_y \equiv 0$

Principal

- Cares about total provision of public good, net of costs, and adjusted for quality / utility

$$W\left(a_P, \bar{a}, \theta\right) = \left(e + \theta\right)\left(\bar{a} + a_P\right) - \frac{k_P a_P^2}{2} - \alpha \int_0^1 \frac{a_i^2}{2} di \qquad (1)$$

- Chooses publicity / privacy $x \rightsquigarrow \mu x$ to affects agents behaviors
- Observes resulting aggregate compliance $\bar{a}$, then chooses own a_p
  - Here, direct contribution Could also be updating a matching rate or subsidy y
- $\alpha \leq 1$: extent to which P internalizes agents' costs . Set $\alpha = 1$.
  - Could also internalize other terms, e.g., their intrinsic motivation or overall welfare, $\bar{U}$, as before

Distributional assumptions

- Agent i, with value v_a^i and signal θ^i, maximizes over a

$$E\left[\left(v_a^i+\theta\right)a+\left(e+\theta\right)\left(\bar{a}+a_P\right)-a^2/2 \mid \theta^i\right]+xR\left(a,\theta^i\right)$$

- Intrinsic motivation: $v_a^i \sim N\left(\bar{v}_a,\sigma_a^2\right)$
- Quality or preference shift: $\theta \sim N\left(\bar{\theta},\sigma_\theta^2\right)$
- Idiosyncratic signal: $\theta_i \mid \theta \sim N\left(\theta,s_\theta^2\right)$
- Image concern / intensity of social sanctions: $\mu \sim N\left(\bar{\mu},\sigma_\mu^2\right)$.
  - Abstract from heterogeneity in image-motives for now $\mu_a^i=\mu_a$
  - Can allow, but focus here on aggregate variability: $\mu \sim \mathcal{N}\left(\bar{\mu},\sigma_\mu^2\right)$
- Agents know μ, Principal may or may not.

Agents' behavior

- Solve for (unique) linear equilibrium
  - Tractability, similar tradeoffs for other equilibria
- Agent i, with intrinsic motivation v_a^i and signal θ^i, chooses

$$a^i = a\left(v_a^i, \theta^i; \mu\right) = \underbrace{v_a^i}_{\text{own IM}} + \underbrace{\rho\theta^i + (1-\rho)\,\bar{\theta}}_{\text{expected quality of public good}} + \underbrace{x\mu\xi}_{\text{image effect}},$$

$$\rho = \frac{\sigma_\theta^2}{\sigma_\theta^2 + s_\theta^2}, \quad \xi = \frac{\sigma_a^2}{\sigma_a^2 + \rho^2 s_\theta^2}$$

- ρ = informativeness of agent's signal: extracting θ from θ^i
- ξ = marginal reputational return = informativeness of individual behavior: observers others extracting v_a^i from a^i

$$E\left[v_a^i \middle|\, a^i\right] = (1-\xi)\,\bar{v} + \xi\left(\bar{v} + a^i - \bar{a}\right) = \bar{v} + \xi\left(a^i - \bar{a}\right)$$

- Aggregate contribution from agents

$$\bar{a} = \bar{v}_a + \rho\theta + (1-\rho)\bar{\theta} + x\mu\xi$$

- Ex-post, $\bar{a}$ observed $\Rightarrow$ agents can retrieve the true θ,since they know μ

- Variability in $\mu \Rightarrow$ two important implications
  - Inefficient fluctuations in $\bar{a}$, not reflecting variations in θ
  - If Principal does not know μ, she faces a signal-extraction problem

Principal's problem: symmetric information

- Benchmark: suppose first that P, like the agents, will observe realization of μ, hence will know true θ when choosing a_P
- Can compute $P's$ ex-ante utility from a given x, then take *FOC*

$$\frac{dE[W^{SI}]}{dx} = \underbrace{(\xi\bar{\mu})\left[\left(e+\bar{\theta}\right)-\left(\bar{v}+\bar{\theta}\right)\right]}_{\text{incentive effect}} - \underbrace{x\xi^2\left(\bar{\mu}^2+\sigma_\mu^2\right)}_{\text{variance effect}}.$$

The variance effect

Proposition (incentive and variability effects)

Under symmetric information, the Principal sets publicity level

$$x^{SI} = \frac{\bar{\mu}\left[e - \alpha\bar{v} + \bar{\theta}\left(1-\alpha\right)\right]}{\alpha\xi\left(\bar{\mu}^2 + \sigma_\mu^2\right)}$$

- Thus, with $\alpha = 1$ (P fully internalizes A's costs):
  - If $\sigma_\mu^2 = 0$, he is able to perfectly offset free-riding with publicity, by setting (Pigou-like)
  $$x^{FB} = \frac{e - \bar{v}}{\alpha\xi\bar{\mu}}$$
  - If $\sigma_\mu^2 > 0$, he must trade off the incentive gains and variability costs / distortions of publicity, resulting in a lower optimal level, $x^{SI} < x^{FB}$

Principal's problem: asymmetric information

- When P does not observe μ, high aggregate contributions or compliance may reflect high quality / demand θ, or high visibility concerns / social enforcement, $\not\mu$
- P knows that

$$\bar{a} = \bar{v}_a + \rho\theta + (1-\rho)\bar{\theta} + x\mu\xi$$

so his observation of $\bar{a}$ generates a signal

$$\hat{\theta} = \frac{1}{\rho}\left[\bar{a} - \bar{v}_a - x\xi\bar{\mu} - (1-\rho)\,\bar{\theta}\right] = \theta + (x\xi/\rho)\,(\mu - \bar{\mu})$$

- By magnifying agents' signaling / social compliance motives, publicity increases the noisiness of the signal that P can use to learn θ

$$\hat{\theta} \mid \theta \sim \mathcal{N}\left(\theta, \frac{x^2\xi^2\sigma_\mu^2}{\rho^2}\right)$$

Principal's information quality

- P's optimal forecast his a variance-weighted combination of the signal $\hat{\theta}$ retrieved from $\bar{a}$, and the prior $\bar{\theta}$

$$E\left[\theta \mid \bar{a}\right] = \gamma\left(x\right)\hat{\theta} + \left(1 - \gamma(x)\right)\bar{\theta}$$

- $\gamma\left(x\right)$ = precision of the information that P obtains from $\bar{a}$,

$$\gamma\left(x\right) = \frac{\rho^2\sigma_\theta^2}{\rho^2\sigma_\theta^2 + x^2\xi^2\sigma_\mu^2},$$

  which is clearly decreasing in x

- Conditioning on the true realizations of θ and μ, her error is

$$E\left[\theta|\bar{a}\right] - \theta = \left(1 - \gamma\right)\left(\bar{\theta} - \theta\right) + \frac{\gamma\xi\left(\mu - \bar{\mu}\right)}{\rho}x$$

Principals' optimal choice of publicity

- P's ex-ante utility given x can now be written

$$E\left[W^{AI}\right] = E\left[W^{SI}\right] - \underbrace{\frac{\sigma_\theta^2}{2k_P}\left[1-\gamma(x)\right]}_{\text{information cost}} \Rightarrow$$

$$\frac{dE\left[W^{AI}\right]}{dx} = \underbrace{\left(\xi\bar{\mu}\right)\left(\left(w+\bar{\theta}\right)-\lambda\left(\bar{v}+\bar{\theta}\right)\right)}_{\text{incentive effect}} - \underbrace{\alpha x\xi^2\left(\bar{\mu}^2+\sigma_\mu^2\right)}_{\text{variance effect}}$$

$$- \underbrace{\frac{x\gamma(x)^2\xi^2\sigma_\mu^2}{\rho^2 k_P}}_{\text{information distortion effect}},$$

Shifting societal preferences and benefits of privacy

Proposition

When the Principal faces uncertainty about both θ and μ, she selects a lower degree of visibility, $x^{AI} < x^{FI}$, uniquely given by implicit equation

$$x = \left(\frac{\bar{\mu}}{\xi}\right)\left(\frac{e - \lambda v + \bar{\theta}\left(1-\alpha\right)}{\lambda\left(\bar{\mu}^2 + \sigma_\mu^2\right) + \gamma\left(x\right)^2 \sigma_\mu^2 / \rho^2 k_P}\right)$$

Shifting societal preferences and benefits of privacy

Proposition

When the Principal faces uncertainty about both θ and μ, she selects a lower degree of visibility, $x^{AI} < x^{FI}$, uniquely given by implicit equation

$$x = \left(\frac{\bar{\mu}}{\xi}\right)\left(\frac{e - \lambda v + \bar{\theta}\,(1-\alpha)}{\lambda\left(\bar{\mu}^2 + \sigma_\mu^2\right) + \gamma\,(x)^2\,\sigma_\mu^2/\rho^2 k_P}\right)$$

- Comparative statics
  - Publicity is increasing in k_P and e and decreasing in $\bar{v}$
  - Others need not be monotone (in progress)

The main lessons I

1. Important to incorporate / model multiple human motivations –intrinsic, extrinsic and (self) reputational, which

- Differ unobservably across people
- Interact endogenously with each other
- Respond to the social and economic (strategic, informational) environment

The main lessons I

1. Important to incorporate / model multiple human motivations –intrinsic, extrinsic and (self) reputational, which

- Differ unobservably across people
- Interact endogenously with each other
- Respond to the social and economic (strategic, informational) environment

2. Allowed us to identify (and test) several mechanisms that generate crowding out or crowding in, both complete or partial

- Informed principal
- Norms multiplier: substitutability for admirable, honor-driven behaviors, complementarity for respectable, stigma-driven ones
- Overjustification effect of incentives, whether material or image based,
- under multidimensional uncertainty

The main lessons II

3. Broadened the analysis of incentives to include explicit or implicit communication, norms-based interventions, and publicity

The main lessons II

3. Broadened the analysis of incentives to include explicit or implicit communication, norms-based interventions, and publicity

4. Gave content to the "expressive function of law", leading to an informational multiplier

- Uncertain societal preferences $\rightsquigarrow$ softer optimal incentives
- Uncertain externalities $\rightsquigarrow$ tougher optimal incentives
- Modeled use and renunciation of cruel and unusual punishments

The main lessons II

3. Broadened the analysis of incentives to include explicit or implicit communication, norms-based interventions, and publicity

4. Gave content to the "expressive function of law", leading to an informational multiplier

- Uncertain societal preferences $\rightsquigarrow$ softer optimal incentives
- Uncertain externalities $\rightsquigarrow$ tougher optimal incentives
- Modeled use and renunciation of cruel and unusual punishments

5. Identified the benefits, limits and costs of image and social pressure as an incentive

- Cheap and often powerful, albeit limited by overjustification effect
- Involves unpredictable variations in severity of social sanctions
- Low privacy makes evolutions in societal values less transparent to principal, rigidifying law and other formal incentives