Revisiting Incentives:

Values, Laws and Norms

Roland Bénabou

Princeton University

Toulouse Lectures - Lecture II

December 2009

Based on joint work with Jean Tirole (TSE)

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

Lecture II
Laws, Norms and Information

1. Honor, Stigma and Social Norms

2. Welfare and Optimal Incentives

3. Persuasion and Norms-Based Interventions

- Main refs: Bénabou-Tirole, AER (2006), (2010)

Honor, Stigma and Social Norms

1. The calculus of reputation
2. Crowding in and (partial) crowding out
3. Welfare and optimal incentives
4. Persuasion and norms-based interventions

Norms

- What makes a behavior socially or morally unacceptable is often the very fact that “it is just not done”. But in other times, other places: “everyone does it”.
    - Choosing surrender vs. death, not going to church, not voting, divorce, welfare dependency, tax evasion, conspicuous consumption...

Norms

- What makes a behavior socially or morally unacceptable is often the very fact that "it is just not done". But in other times, other places: "everyone does it".
  - Choosing surrender vs. death, not going to church, not voting, divorce, welfare dependency, tax evasion, conspicuous consumption...
- Sometimes explained and modeled by some general form of untargeted "reciprocity", or desire for "conformity" : v_a depends on $\bar{a}$.
- Somewhat ad-hoc, plus does not really correspond to a norm:
  - Norms: *"Shared perceptions of appropriate behavior that possess the power to induce people to act publicly in ways that deviate from their private inclinations"* (Miller-Prentice 1996)

Main questions

- Model: social or personal norms will arise endogenously from the interplay of honor and stigma
- When does the fact that more people contribute, behave well increase or decrease the pressure (social, moral) on me to do so? Complements vs. substitutes
- When do incentives crowd out or crowd in social norms?

Revisiting incentives, in three steps

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

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

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, $v_y \equiv 1, v_a = v \sim G(v)$; $\alpha = 0$, $\lambda = 0$
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: $e > 0$, $\mu_a > 0 = \mu_y, v_y \equiv 1$, $v_a = v \sim G(v)$; $\alpha = 1$, $\lambda \geq 0$
3. Incentives and attributional motivation – the “meaning of acts”
   - Signal-extraction by agents and / or principal
     Full model with multidimensional uncertainty (idiosyncratic, aggregate) about the v's, μ's, e

Honor and Stigma

Preliminaries

- Still discrete decisions: $a = 1, 0$: contribute, participate vs. free ride
- $G(v)$: cdf of individuals' intrinsic values. Density $g(v) > 0$ with:
  - finite support, continuously differentiable, mean $\bar{v}$
  - hazard rate $h(v) \equiv g(v) / (1 - G(v))$
  - unimodal: strictly quasiconcave or monotonic
- Key moments

$$\mathcal{M}^{+}(v) \equiv \frac{\int_{v}^{+\infty} \tilde{v} g(\tilde{v}) d\tilde{v}}{1 - G(v)}, \quad \mathcal{M}^{-}(v) \equiv \frac{\int_{-\infty}^{v} \tilde{v} g(\tilde{v}) d\tilde{v}}{G(v)}$$

Preliminaries

- Still discrete decisions: $a = 1, 0$: contribute, participate vs. free ride
- $G(v)$: cdf of individuals' intrinsic values. Density $g(v) > 0$ with:
  - finite support, continuously differentiable, mean $\bar{v}$
  - hazard rate $h(v) \equiv g(v) / (1 - G(v))$
  - unimodal: strictly quasiconcave or monotonic
- Key moments

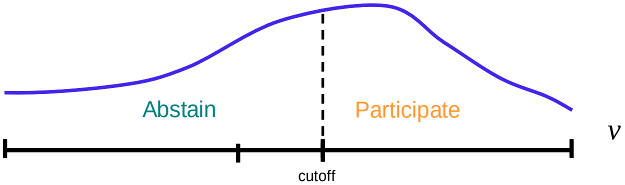
$$\mathcal{M}^+(v) \equiv \frac{\int_v^{+\infty} \tilde{v} g(\tilde{v}) d\tilde{v}}{1 - G(v)}, \quad \mathcal{M}^-(v) \equiv \frac{\int_{-\infty}^{v} \tilde{v} g(\tilde{v}) d\tilde{v}}{G(v)}$$

- When it is known that those choosing $a = 1$ are agents with $v \geq v^*$,
  - $\mathcal{M}^+(v^*)$ governs the "honor" conferred by participation
  - $\mathcal{M}^-(v^*)$ governs the "stigma"from abstention
- Given image concerns μ, net reputational incentive to participate is

$$\Delta(v^*) \equiv \mathcal{M}^+(v^*) - \mathcal{M}^-(v^*) = \text{Honor - Stigma}$$

Intuition

- Fix incentive. Who participates?



$$v \geq c - y - \mu \left[\text{Honor - Stigma}\right] \equiv v^*$$

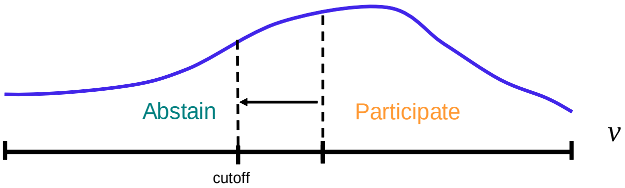
- Participation determined by a cutoff, v^*
  - Honor $=$ average altruism of those above cutoff, $\mathcal{M}^+(v^*)$
  - Stigma $=$ average altruism of those below cutoff, $\mathcal{M}^-(v^*)$

- When interior, cutoff solution to:

$$v^* - c + y + \mu \left[\mathcal{M}^+(v^*) - \mathcal{M}^-(v^*)\right] = 0$$

Intuition

- When more people participate, honor declines, stigma worsens



- Social / moral pressure $\mathcal{M}^{+}\left(v^{*}\right)-\mathcal{M}^{-}\left(v^{*}\right)$, may $\searrow$ or $\nearrow$

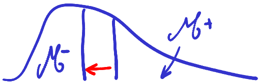
- Same for marginal agent's total non-monetary return to contributing

$$\Psi\left(v^{*}\right) \equiv v^{*}+\mu\left[\mathcal{M}^{+}\left(v^{*}\right)-\mathcal{M}^{-}\left(v^{*}\right)\right] \equiv v^{*}+\mu \Delta\left(v^{*}\right)$$

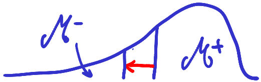
- Key difference between behaviors in which quest for honor versus avoidance of stigma is (endogenously) the main driver of behavior.

Role of the distribution of individual preferences

- Expect honor to dominate when there are only a few heroic or saintly types, whom the mass of more ordinary individuals would like to be identified with



- Expect stigma considerations to dominate when the population includes only a few "bad apples" with very low intrinsic values, which most agents will be eager to differentiate themselves from



- Actions should be
  - Strategic substitutes in first case; unique equilibrium
  - Strategic complements in the second; multiple norms possible

Jewitt's lemma

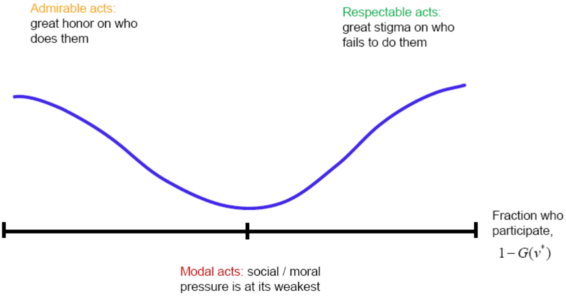
Lemma

The shape of $\Delta(v) = \mathcal{M}^{+}(v^{*}) - \mathcal{M}^{-}(v^{*})$ *mirrors that of density* $g(v)$:

1. *If* g *is everywhere decreasing (increasing), then* Δ *is everywhere increasing (decreasing)*
2. *If* g *has a unique interior maximum, then* Δ *has a unique interior minimum*

- Remarks:
  - Minimum of Δ is not the mode of g.
  - Will sometimes normalize the v's so that $\Delta(v)$ minimized at $v = 0$
- Figure

The calculus of reputation



- Will define behavior $a = 1$ as
    - Respectable if "all but the worst types do it": v^* in the lower tail. Thus $\Delta'(v^*) < 0 \Rightarrow SC$. Not beating your spouse and children
    - Admirable if "only the best do it": v^* in the lower tail. Thus $\Delta'(v^*) > 0 \Rightarrow SS$. Donating a kidney to a stranger
    - Modal if both behaviors are prevalent: v^* in middle range

- Participation pool $\left[v^{-}, v^{*}\right]$, abstention $\left[v^{*}, v^{-}\right]$. Incentive at the margin:

$$\Psi(v) \equiv v + \mu\left[\mathcal{M}^{+}(v) - \mathcal{M}^{+}(v)\right] = v + \mu\Delta(v) \gtrless c - y$$

- Participation pool $\left[v^-, v^*\right]$, abstention $\left[v^*, v^-\right]$. Incentive at the margin:

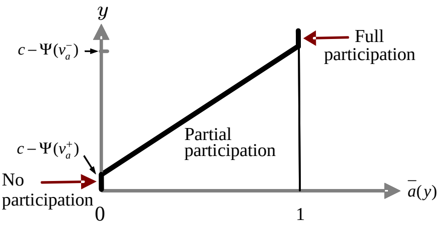
$$\Psi(v) \equiv v + \mu\left[\mathcal{M}^+(v) - \mathcal{M}^+(v)\right] = v + \mu\Delta(v) \gtrless c - y$$

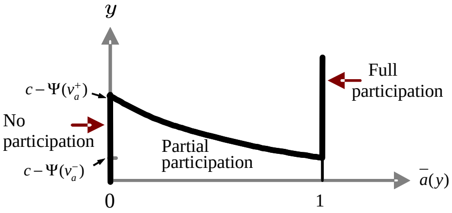
For $1+\mu\Delta' > 0$,
interior eqbm with:

$$\begin{cases} v^* - c + y + \mu\Delta(v^*) = 0 \\ \bar{a}'(y) = \left[1 - G(v^*(y))\right]' = \underbrace{\dfrac{g(v^*(y))}{1+\mu\Delta'(v^*(y))}}_{\text{norms multiplier}} \end{cases}$$

For $1+\mu\Delta' < 0$,
multiple eqba:

$$\bar{a}(y) = 0 \quad \text{and} \quad \bar{a}(y) = 1$$

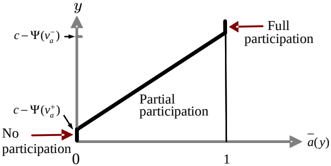


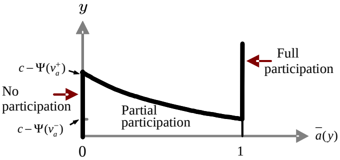


Proposition (honor, stigma, and social norms)

Let $\Psi(v) \equiv v + \mu\Delta(v)$.

1. *When* $\Psi \nearrow$, *there is a unique equilibrium: participation increasing in* y *for* $y \in (c - \Psi(v^+), c - \Psi(v^-))$, *no participation for* $y < c - \Psi(v^+)$, *full participation for* $y > c - \Psi(v^-)$
2. *When* $\Psi \searrow$, *there is / are:*
   - *For* $y \notin (c - \Psi(v^-), c - \Psi(v^+))$, *a unique, corner equilibrium*
   - *For* $y \in (c - \Psi(v^-), c - \Psi(v^+))$, *three equilibria: full participation, no participation, and unstable interior equilibrium*
3. *When* Ψ *is non-monotonic, there is range of values of* y *for which there are at least two stable equilibria, with one at least interior.*





Implications

1. Material incentives (prizes, law) not very effective to spur "admirable", honor- driven behaviors: y weakens social pressure Δ when v^* is high. Pay to vote (Panagopoulos)

   Multiplier $-\mu\Delta' < 0 \rightsquigarrow$ Partial crowding out

Implications

1. Material incentives (prizes, law) not very effective to spur "admirable", honor- driven behaviors: y weakens social pressure Δ when v^* is high. Pay to vote (Panagopoulos)

   Multiplier $-\mu\Delta' < 0 \rightsquigarrow$ Partial crowding out

2. Incentives much more effective to strengthen "respectable", stigma-driven ones: y strengthens social pressure Δ when v^* is low

   Multiplier $-\mu\Delta' > 0 \rightsquigarrow$ Crowding in, partial or complete

3. Small changes in incentives can have large effects, shift social norms, when costs are low and actions observable

   - Continental Airlines $50 bonus program based on company-wide performance for the month (Knez-Simester 2001)
   - Small tax on plastic bags in Ireland

- Ireland: 33¢ tax on plastic shopping bags + awareness campaign

- Ireland: 33¢ tax on plastic shopping bags + awareness campaign

  "Within weeks, plastic bag use dropped 94%. Within a year, nearly everyone had bought reusable cloth bags, keeping them in offices and in the backs of cars."

- Ireland: 33¢ tax on plastic shopping bags + awareness campaign

  "Within weeks, plastic bag use dropped 94%. Within a year, nearly everyone had bought reusable cloth bags, keeping them in offices and in the backs of cars."

- How did it work?

- Ireland: 33¢ tax on plastic shopping bags + awareness campaign

  "Within weeks, plastic bag use dropped 94%. Within a year, nearly everyone had bought reusable cloth bags, keeping them in offices and in the backs of cars."

- How did it work?

  "Plastic bags were not outlawed, but carrying them became socially unacceptable –on a par with wearing a fur coat or not cleaning up after one's dog."

Extensions

1. Crime: a policy of zero tolerance of minor offences can have a "double dividend" (Dur 2006)
   - Increases their signaling value for being "tough" (drives out wimps) ⇒ can result in a decrease in serious crimes as well
2. Symbolic fines, e.g., for not voting: see later

Which situations lead to SS or SC?

1. Shape of the density $g(v)$ $\rightsquigarrow$ stigma vs. honor
   - Related: if reputational payoffs are $\mu_a E\left[\pi(v_a) \mid a, y\right]$, curvature of π
2. Size of cost c
   - Determines which tail of $g(v)$ the cutoff v^* lies in

     Low cost $\rightsquigarrow$ respectable behaviors, high cost $\rightsquigarrow$ admirable behaviors
3. Magnitude of visibility x in $x\mu$
   - Amplifies (1), via $\mu\Delta$
4. Differential visibility / detectability of good and bad deeds
   - Type I vs. type II errors

Allowing for excuses

- Excuses: with probability $\delta \in [0,1]$, an individual faces (unverifiable) circumstances that preclude participation: not being informed, having to deal with some emergency, etc.

- For any potential cutoff v, honor is unchanged, stigma is lessened

$$\mathcal{M}^{P}(v) = \mathcal{M}^{+}(v),$$

$$\mathcal{M}^{NP}(v;\delta) = \frac{\delta\bar{v} + (1-\delta)\,G(v)\,\mathcal{M}^{-}(v)}{\delta + (1-\delta)\,G(v)}$$

- Same if abstention never gives rise to signal that individual contributed, but a contribution may go unnoticed with probability δ

Allowing for skepticism on motives

- Uninformative participation: with probability $\delta' \in [0,1]$, individual is forced or strongly incentivized to contribute, or faces temporarily low c.

- Stigma from abstention now unchanged, but the honor is dulled

$$\mathcal{M}^{NP}(v) = \mathcal{M}^{-}(v),$$

$$\mathcal{M}^{P}\left(v;\delta'\right) = \frac{\delta'\bar{v} + (1-\delta')\left[1 - G(v)\right]\mathcal{M}^{+}(v)}{\delta' + (1-\delta')\left[1 - G(v)\right]}$$

- Same if participation always detected, but non-participation can go undetected with probability δ'

Excuses, constrained participation, and observability

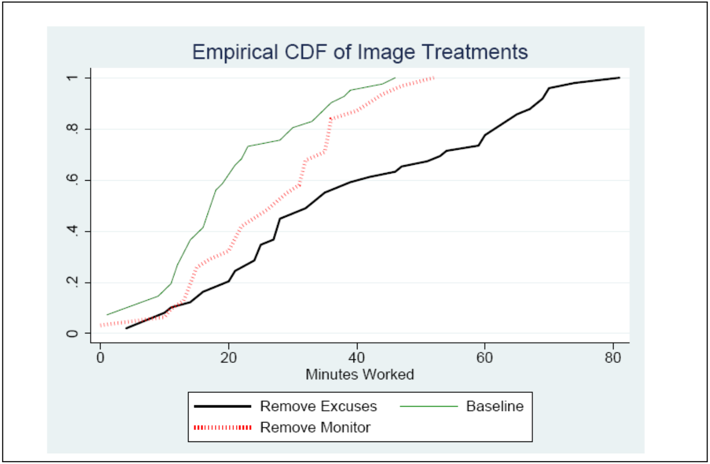
Proposition (differential observability)

1. *An increase in the probability of unobserved constrained participation δ' facilitates the emergence of strategic complementarities and multiple social norms. An increase in the probability of (unobserved) involuntary non-participation δ inhibits it*
2. *Same for, respectively, the probability δ' that abstention may escape detection and the probability δ that a good deed goes unnoticed*

- Experimental test: Linardi-McConnell (2009)

Linardi-McConnell (2009)

- Minutes of volunteering time (database work for a nonprofit)

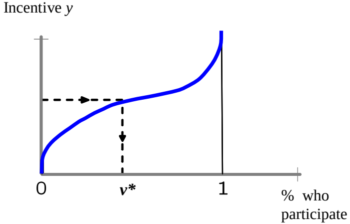


- Baseline: possible excuses (stochastic termination time), experimenter present to monitor
- Removing excuses increases provision significantly
- Removing monitor also increases provision

Summary

When honor motive is dominant:

- Individuals' decisions are substitutes
- Incentives → partial crowding out
  (still work, but weakened)

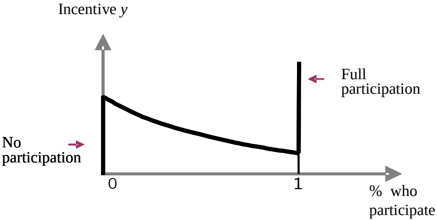


This occurs when:

- Most people are "mediocre", only rare "saintly" types with v well above most others (heroism, organ donation)
- Action is very costly
- There are possible "excuses" for not contributing, and / or one can do it without being noticed (⇒ weak stigma)

When stigma motive is dominant:

- Individual's decisions are complements
- Multiple norms may coexist
- Small incentives can have large effects: shift norms, crowding in



This occurs when

- Most people are "OK", only a few "rotten apples" with v well below most others (crime, child neglect)
- Action is relatively cheap
- There are possible non-glorious reasons for contributing (e.g., fear of the law), and/or it may go unnoticed (⇒ weak honor)

Welfare and Optimal Incentives

Preliminaries

- From now on, assume that v has a unimodal distribution
- To avoid multiplicity, $1 + \mu\Delta'(v) > 0$ everywhere; μ not too large
  Still allows both SS and SC $(\Delta' \gtrless 0)$

Preliminaries

- From now on, assume that v has a unimodal distribution
- To avoid multiplicity, $1 + \mu\Delta'(v) > 0$ everywhere; μ not too large
  Still allows both SS and SC $(\Delta' \gtrless 0)$

Proposition

Aggregate supply, $a(y) = 1 - G(v^*(y))$, is upward-sloping and S-shaped, with a unique inflection point at $\tilde{y}$, defined by

$$\frac{g'(\tilde{v})}{g(\tilde{v})} = \frac{\mu\Delta'(\tilde{v})}{1 + \mu\Delta'(\tilde{v})}, \quad \textit{where } \tilde{v} \equiv v^*(\tilde{y})$$

- Will denote elasticity as

$$\varepsilon(y) \equiv \frac{y\ a'(y)}{a(y)} = \frac{h(v^*(y))}{1 + \mu\Delta'(v^*(y))} \qquad \leftarrow\!\!-\,- \text{ norms multiplier}$$

- Will focus on P maximizing social welfare, but will see that this also covers selfish P and all cases in-between

Welfare calculus

- Net social value of an individual contribution, e.g., buying a Prius?
- Agent gets
  - Cost to individual: $-c$
  - Intrinsic value v: how much he values the improvement in public good (air quality) that his action brings about + pure "joy or giving"
  - Extrinsic reward: y. Subsidy, tax rebate, penalty avoided, etc.
  - Improved (self) image: $\mu \times$ (Honor $-$ Stigma)

Welfare calculus

- Net social value of an individual contribution, e.g., buying a Prius?
- Agent gets
  - Cost to individual: $-c$
  - Intrinsic value v: how much he values the improvement in public good (air quality) that his action brings about + pure "joy or giving"
  - Extrinsic reward: y. Subsidy, tax rebate, penalty avoided, etc.
  - Improved (self) image: $\mu \times$ (Honor $-$ Stigma)
- Others get
  - Benefit e created by unit increment to the public good, $\bar{a}$
  - Incentive payments: $-y(1+\lambda)$, from taxes or private sources
  - Loss of self image: stigma of non-contributors rises, honor of contributors falls (SUV owners, but also Prius owners)

Welfare calculus

- Net social value of an individual contribution, e.g., buying a Prius?
- Agent gets
  - Cost to individual: $-c$
  - Intrinsic value v: how much he values the improvement in public good (air quality) that his action brings about + pure “joy or giving”
  - Extrinsic reward: y. Subsidy, tax rebate, penalty avoided, etc.
  - Improved (self) image: $\mu \times$ (Honor $-$ Stigma)
- Others get
  - Benefit e created by unit increment to the public good, $\bar{a}$
  - Incentive payments: $-y(1+\lambda)$, from taxes or private sources
  - Loss of self image: stigma of non-contributors rises, honor of contributors falls (SUV owners, but also Prius owners)
- Key point: pursuit of esteem is a zero-sum game: average reputation in society remains fixed, since distribution of types is fixed.
- Esteem, or even self-esteem is, by its very nature, a positional good

"When I was making money, I made the most money, and now that I'm spiritual I'm the most spiritual."

Agents' welfare

- Given y, behavior of private agents characterized by cutoff v^* (v^* affected by y differently under different info. conditions)

Agents' welfare

- Given y, behavior of private agents characterized by cutoff v^* (v^* affected by y differently under different info. conditions)

- Agents' average utility

$$\begin{aligned}\bar{U}\left(v^*; y\right) &= \int_{v^*}^{+\infty} \left(e + v - c + y + \mu E\left[\tilde{v} \mid \tilde{v} \geq v^*\right]\right) \; g(v) \; dv \\ &+ \int_{-\infty}^{v^*} \mu E\left[\tilde{v} \mid \tilde{v} \leq v^*\right] g(v) dv \\ &= \int_{v^*}^{+\infty} \left[e + v - c + y\right] g(v) dv + \mu \bar{v}\end{aligned}$$

- Shows reputation as zero-sum game, positional good

Agents' welfare

- Given y, behavior of private agents characterized by cutoff v^* (v^* affected by y differently under different info. conditions)

- Agents' average utility

$$\begin{aligned}\bar{U}\left(v^*;y\right) &= \int_{v^*}^{+\infty} \left(e+v-c+y+\mu E\left[\tilde{v} \mid \tilde{v} \geq v^*\right]\right) \, g(v)\, dv \\ &+ \int_{-\infty}^{v^*} \mu E\left[\tilde{v} \mid \tilde{v} \leq v^*\right] g(v) dv \\ &= \int_{v^*}^{+\infty} \left[e+v-c+y\right] g(v) dv + \mu\bar{v}\end{aligned}$$

- Shows reputation as zero-sum game, positional good

- Not zero-sum iff reputational payoff nonlinear in probabilities, or if μ varies with v

Social planner and other principals

- Benevolent social planner with shadow cost of funds λ would maximize over y

$$W(v^*; y) = \bar{U}(v^*; y) - \int_{v^*}^{+\infty} (1+\lambda) y g(v) dv$$
$$= \int_{v^*}^{+\infty} \left[e + v - c - \lambda y \right] g(v) dv + \mu \bar{v}$$

subject to:$v^* = v^*(y; \ldots)$

Social planner and other principals

- Benevolent social planner with shadow cost of funds λ would maximize over y

$$\begin{aligned} W\left(v^*; y\right) &= \bar{U}\left(v^*; y\right) - \int_{v^*}^{+\infty} (1+\lambda) y g(v) dv \\ &= \int_{v^*}^{+\infty} \left[e + v - c - \lambda y\right] g(v) dv + \mu \bar{v} \end{aligned}$$

subject to:$v^* = v^*(y; \ldots)$

- Other principal: NGO, gvt. agency, church, etc.
  - May derive private benefits B from agents' participation / effort
  - May put weight $0 \leq \alpha \leq 1$ on agents' welfare

$$W(y) \equiv \alpha \bar{U}(v^*; y) + (B - y)\left[1 - G(v^*)\right]$$

- Polar case: selfish principal, e.g. employer: $B > 0,\ \alpha = 0$

Renormalization as planner's problem

- For arbitrary principal with preferences (B, α) :

$$\begin{aligned} W(y) = \alpha \int_{v^*}^{+\infty} & \left[e + v - c + y\right] g(v) dv \\ & + (B - y) \int_{v^*}^{+\infty} g(v) dv + \alpha \mu \bar{v} \\ & = \int_{v^*}^{+\infty} \left[e^{'} + v' - c' - \lambda^{'} y\right] g(v) dv + \alpha \mu \bar{v}, \end{aligned}$$

  - Social externality from participation $e^{'} \equiv \alpha e + B$
  - Private costs and values $v' \equiv \alpha v, c' \equiv \alpha c$
  - Shadow cost of funds $\lambda' \equiv 1 - \alpha > 0$

$\Rightarrow$ New planner's problem, with renormalized e, λ, v, c.

- Selfish principal: $\alpha = 0,\ e' = B,\ \lambda' = 1$

Shifts in societal values

- Will study situations with changes in / aggregate uncertainty about preferences of society: v distributed according to

  $$G_\theta(v) \equiv G(v - \theta),$$

  i.e. G shifted right by $\theta \in \mathbb{R}$: known or uncertain

Shifts in societal values

- Will study situations with changes in / aggregate uncertainty about preferences of society: v distributed according to

$$G_\theta(v) \equiv G(v-\theta),$$

  i.e. G shifted right by $\theta \in \mathbb{R}$: known or uncertain

- Density $g_\theta(v) = g(v-\theta)$, hazard rate $h_\theta = h(v-\theta)$, mean $\bar{v}+\theta$

- Given θ, reputational return is

$$\Delta_\theta(v) = E_G\left[v' \mid v'+\theta \geq v\right] - E_G\left[v' \mid v'+\theta < v\right] = \Delta(v-\theta)$$

- If normalize $\Delta'(0) = 0 \Rightarrow$ point of minimum reputation under G_θ is located at $v = \theta$

- For known θ, all results so far unchanged, with $g \rightsquigarrow g_\theta$, $\Delta \rightsquigarrow \Delta_\theta$ ....

Individual decisions

- Given y, μ, θ, agent with valuation v contributes iff

$$v - c + y + \mu\Delta_\theta(v^*) \geq 0$$

- Participation cutoff $v^* = v^*(y, \theta)$, given (if interior) by

$$v^*(y,\theta) - c + y + \mu\Delta_\theta(v^*(y,\theta)) \equiv 0$$

$$\frac{\partial v^*}{\partial y} = \frac{-1}{1 + \mu\Delta'_\theta(v^*)} < 0, \quad \frac{\partial v^*}{\partial \theta} = \frac{\mu\Delta'_\theta(v^*)}{1 + \mu\Delta' v^*)} \gtrless 0 \text{ as } v^* \lessgtr \theta.$$

- On net, incentives always increase compliance, but
  - If v^* above θ, $\Delta'_\theta(v^*) > 0$: actions are SS , $y \rightsquigarrow$ crowding out
  - If v^* below θ, $\Delta'_\theta(v^*) < 0$: actions are SC, $y \rightsquigarrow$ crowding in

Effects of known shifts in societal preferences

- Cutoff $v^*(y, \theta)$ defined by

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

$$\Rightarrow \; v^*(y, \theta) - \theta = v^*\left(y + \theta, 0\right).$$

- A known shift in societal preferences θ has same effect on aggregate behavior $a_\theta(y)$ and social norms $\Delta_\theta(v^*(y, \theta))$ as an increase in material incentive y, or a decrease in cost c, of the same magnitude

Effects of known shifts in societal preferences

- Cutoff $v^*(y,\theta)$ defined by

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

$$\Rightarrow \; v^*(y,\theta) - \theta = v^*\left(y + \theta, 0\right).$$

- A known shift in societal preferences θ has same effect on aggregate behavior $a_\theta(y)$ and social norms $\Delta_\theta(v^*(y,\theta))$ as an increase in material incentive y, or a decrease in cost c, of the same magnitude

- Peeking ahead: when Principal has private information about θ, she may want to substitute messages / signals about what "community standards" are, instead of costly incentives

Optimal incentives under symmetric information

- Planner sets y to maximize

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

Optimal incentives under symmetric information

- Planner sets y to maximize

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

- W_θ^{FI} strictly quasiconcave for λ small enough. FOC:

$$\left[e + v^*(y,\theta) - c - \lambda y\right] g_\theta(v^*(y,\theta)) \left(\frac{-\partial v^*(y,\theta)}{\partial y}\right)$$
$$= \lambda \left[1 - G_\theta(v^*(y,\theta))\right]$$

- Ramsey taxation
  - LHS = Net social marginal benefit of raising y by \$1, inducing $da_\theta = \left(-\partial v^* / \partial y\right) g_\theta$ new agents to participate
  - RHS = deadweight loss incurred by paying \$1 more to all inframarginal agents (informational rents)

Optimal incentives under symmetric information

- Planner sets y to maximize

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

- W_θ^{FI} strictly quasiconcave for λ small enough. FOC:

$$[e + v^*(y,\theta) - c - \lambda y]\, g_\theta(v^*(y,\theta)) \left(\frac{-\partial v^*(y,\theta)}{\partial y} \right)$$
$$= \lambda \left[1 - G_\theta(v^*(y,\theta))\right]$$

- Ramsey taxation
  - LHS = Net social marginal benefit of raising y by \$1, inducing $da_\theta = (-\partial v^* / \partial y)\, g_\theta$ new agents to participate
  - RHS = deadweight loss incurred by paying \$1 more to all inframarginal agents (informational rents)
- Equivalently:

$$y = \frac{e + v^*(y,\theta) - c}{\lambda \left[1 + 1/\varepsilon_\theta(y)\right]}$$

Solving

- Equilibrium cutoff: $y \rightsquigarrow v^*$

$$v^* - c + y + \mu \Delta_\theta(v^*) \equiv 0$$

- FOC: $v^* \rightsquigarrow y$:

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

- System of two implicit equations in y and $v^* = v^*$

First-best case

- No-distortion, $\lambda = 0$: important benchmark under both symmetric and asymmetric information. FOC simplifies to:

$$e + v^*(y^{FB}(\theta), \theta) = c$$

- Standard Samuelson condition: equating total social benefit and cost of the marginal contribution.

First-best case

- No-distortion, $\lambda = 0$: important benchmark under both symmetric and asymmetric information. FOC simplifies to:

$$e + v^*(y^{FB}(\theta), \theta) = c$$

- Standard Samuelson condition: equating total social benefit and cost of the marginal contribution. Recall also

$$v^*(y, \theta) - c + y + \mu\Delta_\theta(v^*(y, \theta)) = 0.$$

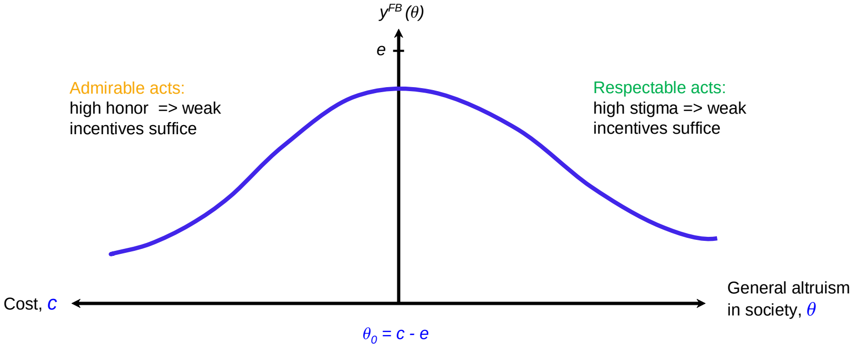
Proposition (modified Pigou)

The first-best subsidy $y^{FB}(\theta)$ under symmetric information and no tax distortion is

$$y^{FB}(\theta) = e - \mu\Delta_\theta(c - e) = e - \mu\Delta(c - e - \theta)$$

It is unimodal with respect to θ and c, and maximized at $\theta_0 \equiv c - e$.

First-best incentives



Admirable acts: high honor => weak incentives suffice

Respectable acts: high stigma => weak incentives suffice

Modal acts: social / moral pressure is at its weakest, strongest incentives needed

$$y^{FB}(\theta) = e - \mu\Delta(c - e - \theta)$$

Intuitions for first-best policy

$$y^{FB}(\theta) = e - \mu\Delta(c - e - \theta)$$

- Reputation tax: participation has both positive spillover e, and negative one, "reputation-stealing" $\mu\Delta$

- Non-monotonicity:
  - Admirable behaviors: when θ is low or c is high, most people do not contribute $\Rightarrow$ being among the "elite few" who do conveys significant honor. Low incentive y required
  - Respectable behaviors: when θ is high or c is low, most people participate $\Rightarrow$ the few "bad apples" who do not are subject to strong stigma. Low incentive y required
  - Modal behaviors: when θ is around $c - e$, social pressure at its weakest: contributing and abstaining are both common behaviors. Higher incentive is required.

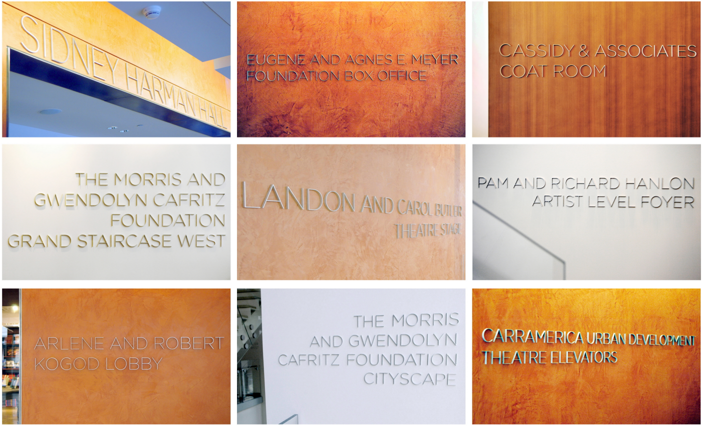
- In general, $y^{FB} \gtrless 0$: may tax, if $\mu\Delta \gg e$. Can ensure $y^{FB} > 0$

Implications

- Optimal tax deduction rate for donations may be lower than thought
- Pattern of contributions distorted toward the most visible (high μ):
  - Alumni giving to wealthy universities rather than high schools, primary schools, preschool programs. Name on building, chair, etc. at Harvard, Princeton, TSE, rather than public school in small town
  - Giving to big hospitals, museums, etc., rather than rural clinics, vaccination programs in LDC's
  - Chinese-American's giving
- Gets worse with sponsor competition, e.g., between nonprofits, NGO's, universities, etc. ⇒ arms race in image seeking

  Can be worse for social welfare than a monopoly, because competing sponsors do not internalize reputation-stealing externality

“The Graffiti of the Philanthropic Class”


SIDNEY HARMAN HALL
EUGENE AND AGNES E. MEYER FOUNDATION BOX OFFICE
CASSIDY & ASSOCIATES COAT ROOM
THE MORRIS AND GWENDOLYN CAFRITZ FOUNDATION GRAND STAIRCASE WEST
LANDON AND CAROL BUTLER THEATRE STAGE
PAM AND RICHARD HANLON ARTIST LEVEL FOYER
ARLENE AND ROBERT KOGOD LOBBY
THE MORRIS AND GWENDOLYN CAFRITZ FOUNDATION CITYSCAPE
CARRAMERICA URBAN DEVELOPMENT THEATRE ELEVATORS

Source: Isherwood (2007)

- Tax deductibility of donations should vary inversely with publicity / image value inherent to them
  - Not easy, but not impossible: implicit market value for a named building, plaque, chair, etc.
  - Rating agencies: should aim to incorporate a “discreteness” premium or “publicity discount” in their scores

- Same for ethical funds, fair trade “green” products: premium also buys you social and self image. . . and confers stigma / bad conscience on others

- Other uses of same money can do more social good but don’t have those image private benefits and social externalities: food kitchens, orphanages, etc. ⇒ too little of them

Optimal incentives with cost of public funds

- Realistically, gvt. faces $\lambda > 0$.

  Or, for non-benevolent principal, $\lambda = 1 - \alpha$

Optimal incentives with cost of public funds

- Realistically, gvt. faces $\lambda > 0$.
  Or, for non-benevolent principal, $\lambda = 1 - \alpha$

- Recall FOC:

$$\underbrace{\frac{e + v^*(y,\theta) - c - \lambda y}{1 + \mu \Delta'_\theta(v^*(y,\theta))}}_{\text{net social benefit of marginal contribution}} = \underbrace{\frac{\lambda}{h_\theta(v^*(y,\theta))}}_{\text{net resource cost of marginal contribution}}$$

$$\Rightarrow \quad e + v^*(y,\theta) > c \quad \Rightarrow \quad y^{FI}(\theta) < y^{FB}(\theta).$$

- Second-best: net social benefit from marginal contribution exceeds its social cost. Too costly to subsidize further

Optimal incentives with cost of public funds

- Realistically, gvt. faces $\lambda > 0$.
  Or, for non-benevolent principal, $\lambda = 1 - \alpha$

- Recall FOC:

$$\underbrace{\frac{e + v^*(y,\theta) - c - \lambda y}{1 + \mu \Delta'_\theta(v^*(y,\theta))}}_{\text{net social benefit of marginal contribution}} = \underbrace{\frac{\lambda}{h_\theta(v^*(y,\theta))}}_{\text{net resource cost of marginal contribution}}$$

$$\Rightarrow \quad e + v^*(y,\theta) > c \quad \Rightarrow \quad y^{FI}(\theta) < y^{FB}(\theta).$$

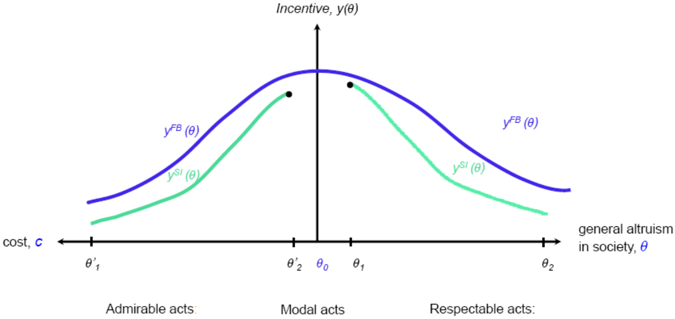
- Second-best: net social benefit from marginal contribution exceeds its social cost. Too costly to subsidize further

- Expect shape of $y^{FI}(\theta)$ close to that of $y^{FB}(\theta)$, for λ small enough

Proposition (second best - symmetric information)

Let (θ_1, θ_2) *be any interval not containing* $\theta_{0.}$ *For* $\lambda > 0$ *low enough,*

1. *The symmetric-information policy* $y^{FI}(\theta)$ *is uniquely defined on* (θ_1, θ_2), *with* $0 < y^{FI}(\theta) < y^{FB}(\theta)$
2. *The incentive* $y^{FI}(\theta)$ *strictly increasing in* θ *when* $\theta_2 < \theta_0$ *and strictly decreasing when* $\theta_0 < \theta_1$.



Persuasion and Norms-Based Interventions

1. Public appeals
2. Norms-based interventions and pluralistic ignorance
3. Formalization: communicating on θ, μ, e

Public appeals: communicating on e

- Reiss-White (2008): 2000-2001 California energy crisis, San Diego
  - Uncapped electricity prices, resulting in large spike (×2.3 on average) within three months ⇒ quick decline in energy consumption (-13%)
  - Under political pressure, re-capped at approximately previous level (even gave rebates); demand went right back up.

    Model: y constrained
  - Then, facing rolling blackout, launched $65 million public campaign to promote energy conservation ⇒ Reduced consumption continuously over few months, to about -7%.

Figure 6. Average Within-Household Consumption Changes, 2000 to 2002. Changes are relative to the same months during pre-crisis years, with weather and trend removed.

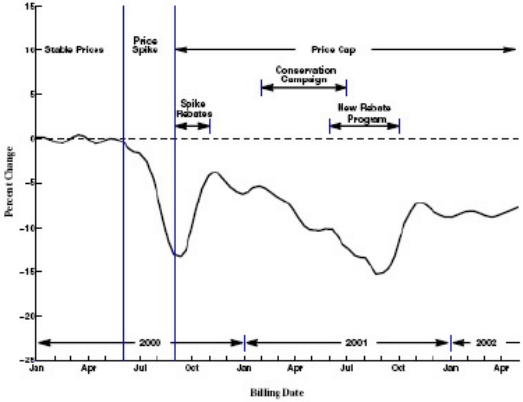
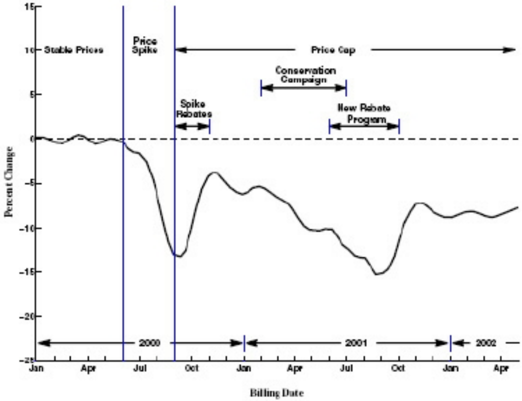


FIGURE 6. Average Within-Household Consumption Changes, 2000 to 2002. Changes are relative to the same months during pre-crisis years, with weather and trend removed.



- Prices matter, but words too

De Janvry et al. (2006): shortage of flu vaccines, Fall 2004

- Center for Disease Control recommended that people in non-priority groups delay vaccination

- Randomized US campus experiment. Send emails to different departments, reminding them of either
  - Scarcity: reduction in number of flu clinics on campus, times open
  - Scarcity + reiterating CDC appeal for non-priority groups to defer.

De Janvry et al. (2006): shortage of flu vaccines, Fall 2004

- Center for Disease Control recommended that people in non-priority groups delay vaccination
- Randomized US campus experiment. Send emails to different departments, reminding them of either
  - Scarcity: reduction in number of flu clinics on campus, times open
  - Scarcity + reiterating CDC appeal for non-priority groups to defer.
- Scarcity information led to 110% increase in demand (from non-target group, with fair amount of cheating)
- Call on self restraint reduced it by 37.5% (esp. in target group)

Norms-based interventions

- Widely used concepts (Cialdini, “Influence” 1984)
  - Descriptive norms: what most other people (in your community) do. The norm of “is”
  - Prescriptive / injunctive norms: what most other people (in your community) approve of. The norm of “ought”

Norms-based interventions

- Widely used concepts (Cialdini, "Influence" 1984)
  - Descriptive norms: what most other people (in your community) do. The norm of "is"
  - Prescriptive / injunctive norms: what most other people (in your community) approve of. The norm of "ought"
- Idea of NBI's: change people's perceptions of what is "normal" behavior or "normal" values.
- Model: communicating on θ, e or μ

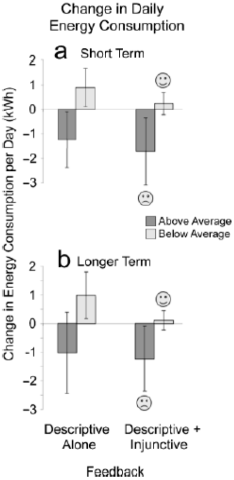
Example: Schultz, Nolan, Cialdini et al. (2007)

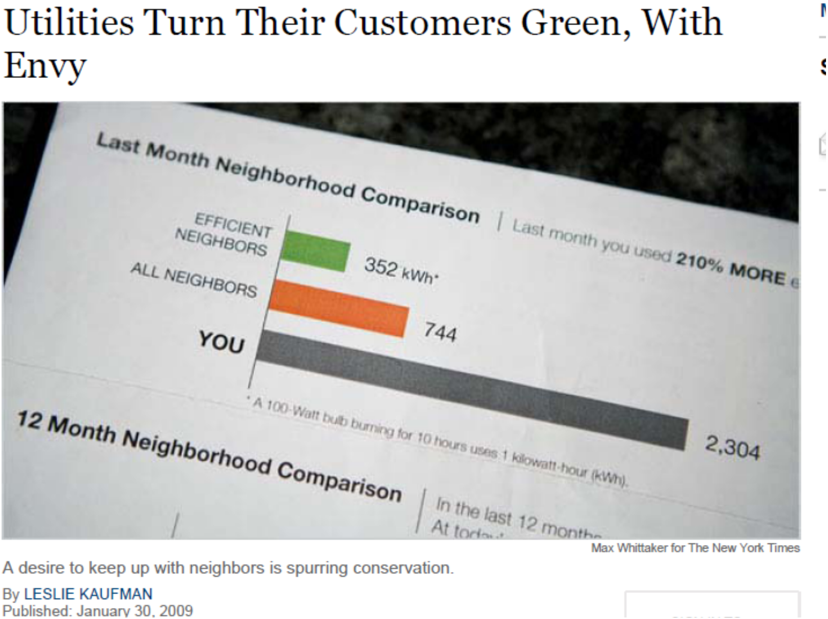
- Monitored electricity meters of 290 households in a California town.
- Each week, hung on their door a visible feedback form with (randomized):

Example: Schultz, Nolan, Cialdini et al. (2007)

- Monitored electricity meters of 290 households in a California town.
- Each week, hung on their door a visible feedback form with (randomized):
  - Descriptive condition: own electricity consumption + average consumption of households in their neighborhood + tips on conservation) ⇒ convergence toward mean
  - Prescriptive condition: same, plus smiley face if below average, frowning face if above ⇒ high consumers still decrease, low consumers no longer increase consumption.

Schultz et al. “The Constructive, Destructive, and Reconstructive Power of Social Norms”, Psy. Science (2007),





Last Month Neighborhood Comparison | Last month you used 210% MORE e
EFFICIENT NEIGHBORS
352 kWh*
ALL NEIGHBORS
744
YOU
2,304
* A 100-Watt bulb burning for 10 hours uses 1 kilowatt-hour (kWh).
12 Month Neighborhood Comparison | In the last 12 month
At toda

Utilities Turn Their Customers Green, With Envy

Max Whittaker for The New York Times

A desire to keep up with neighbors is spurring conservation.

By LESLIE KAUFMAN
Published: January 30, 2009

- A frowny face is not what most electric customers expect to see on their utility statements, but Greg Dyer got one. He earned it, the utility said, by using a lot more energy than his neighbors. Two other Sacramento residents, however, ... were feeling good. They got one smiley face on their statement for energy efficiency and saw the promise of getting another.

- A frowny face is not what most electric customers expect to see on their utility statements, but Greg Dyer got one. He earned it, the utility said, by using a lot more energy than his neighbors. Two other Sacramento residents, however, ... were feeling good. They got one smiley face on their statement for energy efficiency and saw the promise of getting another.

- The district had been trying for years to prod customers into using less energy with tactics like rebates for energy-saving appliances. But the traditional approaches were not meeting the energy reduction goals set by the nonprofit utility's board. So, in a move that has proved surprisingly effective, the district decided to tap into a time-honored American passion: keeping up with the neighbors.

- A frowny face is not what most electric customers expect to see on their utility statements, but Greg Dyer got one. He earned it, the utility said, by using a lot more energy than his neighbors. Two other Sacramento residents, however, ... were feeling good. They got one smiley face on their statement for energy efficiency and saw the promise of getting another.

- The district had been trying for years to prod customers into using less energy with tactics like rebates for energy-saving appliances. But the traditional approaches were not meeting the energy reduction goals set by the nonprofit utility's board. So, in a move that has proved surprisingly effective, the district decided to tap into a time-honored American passion: keeping up with the neighbors.

- Sent out statements to 35,000 randomly selected customers, rating them on their energy use compared with that of neighbors in 100 homes of similar size that used the same heating fuel. The customers were also compared with the 20 neighbors who were especially efficient in saving energy. Customers who scored high earned two smiley faces on their statements. "Good" conservation got a single smiley face. Customers ... in the "below average" category, got frowns, but the utility stopped using them after a few customers got upset.

- After six months, customers who received the personalized report reduced energy use by 2% more than those who got standard statements .
- The approach has now been picked up by utilities in 10 major metropolitan areas... including Chicago and Seattle, according to Positive Energy, the software company that conceived of the reports and contracts to produce them. Following Sacramento's lead, they award smiley faces only.

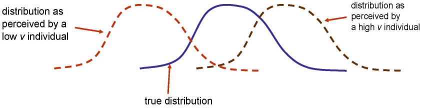
- After six months, customers who received the personalized report reduced energy use by 2% more than those who got standard statements .

- The approach has now been picked up by utilities in 10 major metropolitan areas... including Chicago and Seattle, according to Positive Energy, the software company that conceived of the reports and contracts to produce them. Following Sacramento's lead, they award smiley faces only.

- Robert Cialdini, a social psychologist at Arizona State University, studies how to get Americans –even those who did not care about the environment– to lower energy consumption. And while there are many ways, Dr. Cialdini said, few are as effective as comparing people with their peers. ... "It is fundamental and primitive," said Dr. Cialdini, who owns a stake in Positive Energy. "The mere perception of the normal behavior of those around us is very powerful."

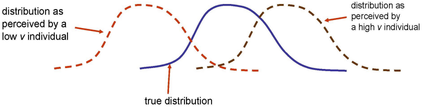
Interpreting the descriptive intervention

1. In addition to aggregate uncertainty about θ, and hence $\bar{a}$, there is idiosyncratic variability in households' perceptions $\hat{\theta}^i$ of it
   - May have received different external signals
   - May be using own v_i as an indicator of / representative of the mean: "false consensus" effect, or Bayesian updating in small sample


distribution as perceived by a low *v* individual
distribution as perceived by a high *v* individual
true distribution

Interpreting the descriptive intervention

1. In addition to aggregate uncertainty about θ, and hence $\bar{a}$, there is idiosyncratic variability in households' perceptions $\hat{\theta}^i$ of it
   - May have received different external signals
   - May be using own v_i as an indicator of / representative of the mean: "false consensus" effect, or Bayesian updating in small sample


distribution as perceived by a low *v* individual
true distribution
distribution as perceived by a high *v* individual

2. Energy conservation: "respectable", not heroic: $\Delta' < 0$, SC $\Rightarrow$
   - Conserve more, the higher is perceived mean $\hat{\theta}^i$
   - Utility revealing $\bar{a}$ reveals (or informs about) $\theta \Rightarrow$ those with $\hat{\theta}^i < \theta$ now feel greater pressure to conserve, those with $\hat{\theta}^i > \theta$ lower pressure

Interpreting prescriptive intervention

- Communication on e or μ
  - “People are strongly affected by this problem”
  - “People care about it and are watching / judging others
  - Salience, both social and to self (Cialdini);
    especially if negative info or frame (Baumeister)

Interpreting prescriptive intervention

- Communication on e or μ
  - "People are strongly affected by this problem"
  - "People care about it and are watching / judging others
  - Salience, both social and to self (Cialdini); especially if negative info or frame (Baumeister)
- Many other such field experiments / NBI's. Usually small, sometimes quite ambitious: Paluck's (2007) randomized media experiments in Rwanda and Sudan

Interpreting prescriptive intervention

- Communication on e or μ
  - "People are strongly affected by this problem"
  - "People care about it and are watching / judging others
  - Salience, both social and to self (Cialdini); especially if negative info or frame (Baumeister)
- Many other such field experiments / NBI's. Usually small, sometimes quite ambitious: Paluck's (2007) randomized media experiments in Rwanda and Sudan
- Cialdini's policy recommendation:
  - If most people behave well, use descriptive norm (or both)
  - If most people behave badly, use prescriptive, avoid descriptive

Interpreting prescriptive intervention

- Communication on e or μ
  - “People are strongly affected by this problem”
  - “People care about it and are watching / judging others
  - Salience, both social and to self (Cialdini); especially if negative info or frame (Baumeister)
- Many other such field experiments / NBI’s. Usually small, sometimes quite ambitious: Paluck’s (2007) randomized media experiments in Rwanda and Sudan
- Cialdini’s policy recommendation:
  - If most people behave well, use descriptive norm (or both)
  - If most people behave badly, use prescriptive, avoid descriptive
- Caveats:
  - Relevant in the “respectable” / SS range only. OK
  - Sensible, but potential credibility / consistency problem

When descriptive and injuctive norms conflict

- "What's Obscene? Google Could Have an Answer" (NYT 06/24/2008)

  Judges and jurors who must decide whether sexually explicit material is obscene are asked to use a local yardstick: does the material violate community standards? That is often a tricky question because there is no simple, concrete way to gauge a community's tastes and values [uncertainty over θ]

When descriptive and injuctive norms conflict

- "What's Obscene? Google Could Have an Answer" (NYT 06/24/2008)

  Judges and jurors who must decide whether sexually explicit material is obscene are asked to use a local yardstick: does the material violate community standards? That is often a tricky question because there is no simple, concrete way to gauge a community's tastes and values [uncertainty over θ]

  In a novel approach, the defense... in the trial of a pornographic Web site operator... plans to show that residents of Pensacola are more likely to use Google to search for terms like "orgy" than for "apple pie" or "watermelon." "Time and time again you'll have jurors sitting on a jury panel who will condemn material that they routinely consume in private," said... the defense lawyer. Using the Internet data, "we can show how people really think and feel and act in their own homes...

When descriptive and injuctive norms conflict

- "What's Obscene? Google Could Have an Answer" (NYT 06/24/2008)

  Judges and jurors who must decide whether sexually explicit material is obscene are asked to use a local yardstick: does the material violate community standards? That is often a tricky question because there is no simple, concrete way to gauge a community's tastes and values [uncertainty over θ]

  In a novel approach, the defense... in the trial of a pornographic Web site operator... plans to show that residents of Pensacola are more likely to use Google to search for terms like "orgy" than for "apple pie" or "watermelon." "Time and time again you'll have jurors sitting on a jury panel who will condemn material that they routinely consume in private," said... the defense lawyer. Using the Internet data, "we can show how people really think and feel and act in their own homes...

  The Florida state prosecutor, said he... would try to block the search data's use in court. He.... said that the popularity of sex-related Web sites had no bearing on whether Mr. McCowen was in violation of community standards. "How many times you do something doesn't necessarily speak to standards and values," he said.

Pluralistic ignorance

- Why do “verbal”, N.B.-interventions work? Psychologists’ view:
  - People care about being seen/ seeing themselves as moral, prosocial
  - Judge what “one should do” by what they see or believe others do and/or approve of

Pluralistic ignorance

- Why do “verbal”, N.B.-interventions work? Psychologists’ view:
  - People care about being seen/ seeing themselves as moral, prosocial
  - Judge what “one should do” by what they see or believe others do and/or approve of But often misperceive what most others do, or, especially, think because of:

Pluralistic ignorance

- Why do “verbal”, N.B.-interventions work? Psychologists’ view:
  - People care about being seen/ seeing themselves as moral, prosocial
  - Judge what “one should do” by what they see or believe others do and/or approve of But often misperceive what most others do, or, especially, think because of:

1. Limited information, differential visibility of actions, cognitive biases (e.g.,false consensus effect, self-serving bias)

Pluralistic ignorance

- Why do “verbal”, N.B.-interventions work? Psychologists’ view:
  - People care about being seen/ seeing themselves as moral, prosocial
  - Judge what “one should do” by what they see or believe others do and/or approve of But often misperceive what most others do, or, especially, think because of:

1. Limited information, differential visibility of actions, cognitive biases (e.g.,false consensus effect, self-serving bias)
2. “Pluralistic ignorance”: because people see or perceive that most others do X, take it to mean that everyone values / approves of X.
   - Do not properly account for fact that others are also conforming to a common perceived norm. Instance of the “fundamental attribution error” (Jones-Harris, Ross): always underestimate the “power of the situation”.
   - “Social proof” (Cialdini 1984), “preference falsification” (Kuran 1995)

- Dispelling *PI* can bring about sudden and large shift in the norm

Dispelling pluralistic ignorance

- Vast problem of excess drinking by the young, e.g. undergraduates
  - Efforts at individual education (to risks, etc.) and public campaigns have had very limited effectiveness
  - Role of peer influences widely recognized.

Dispelling pluralistic ignorance

- Vast problem of excess drinking by the young, e.g. undergraduates
  - Efforts at individual education (to risks, etc.) and public campaigns have had very limited effectiveness
  - Role of peer influences widely recognized.
- Prentice-Miller (1993): students asked about own level of comfort with drinking on campus + their perception of the general attitude of other students about it. Find that:
  - Students significantly overestimate the extent to which others are comfortable with drinking. PI = "illusion of universality"
  - Perceived level of tolerance by peers strong predictor of own use
  - Over time, males (mostly) tend to adjust their (reported) attitudes toward what they perceive to be the norm.

Prentice- Schroeder (1998) experiment

- Idea: intervene not at level of individual beliefs and attitudes about own alcohol consumption, but at that of beliefs about the social aspects / others' attitudes about it

Prentice- Schroeder (1998) experiment

- Idea: intervene not at level of individual beliefs and attitudes about own alcohol consumption, but at that of beliefs about the social aspects / others' attitudes about it

- Entering students asked same questions as above, then randomly assigned to two discussion groups / conditions:
  - "Individual"- based discussions: risks, how to make responsible decisions about alcohol
  - "Peer"-based: shown previous evidence of pluralistic ignorance, discuss how it works, social dynamics surrounding drinking

- Follow up-4-6 months later: questions about own and peer attitudes, and about own past alcohol consumption.
  - All showed PI initially. Mostly eliminated six months later, for both conditions. Learning.
  - Students in "peer" condition reported significantly lower levels of consumption over the period
  - Mediated by individual's score on "fear of negative evaluation" questionnaire: for those in the individual condition, higher FNE raised sensitivity of their consumption to their perception of peer's comfort with alcohol. For those in "peer" condition, this effect was cancelled.
- Interpretation: making students aware early on of actual distribution of values dispelled PI and weakened prescriptive effect of the norm

A necessary caveat

- Dispelling *PI* can bring about sudden and large shift in the norm

A necessary caveat

- Dispelling *PI* can bring about sudden and large shift in the norm
- Problem: *PI* need not take the form of excessive pessimism about others' behavior or values, as in the campus-drinking example.
- Can also be excessive optimism, e.g., drugs, pornography / Google example

A necessary caveat

- Dispelling *PI* can bring about sudden and large shift in the norm
- Problem: *PI* need not take the form of excessive pessimism about others' behavior or values, as in the campus-drinking example.
- Can also be excessive optimism, e.g., drugs, pornography / Google example
- In such cases, back to the information revelation / credibility issue
  - Principal (field experimenter, sponsor, government) finds herself in the position of trying to hide "depressing truths" from agents
  - Will appeal instead to injuctive norms: what people say they value (even though they do not do it). Much softer, less credible evidence

A necessary caveat

- Dispelling *PI* can bring about sudden and large shift in the norm

- Problem: *PI* need not take the form of excessive pessimism about others' behavior or values, as in the campus-drinking example.

- Can also be excessive optimism, e.g., drugs, pornography / Google example

- In such cases, back to the information revelation / credibility issue
  - Principal (field experimenter, sponsor, government) finds herself in the position of trying to hide "depressing truths" from agents
  - Will appeal instead to injuctive norms: what people say they value (even though they do not do it). Much softer, less credible evidence

- Problem not really recognized / discussed by psychologists

- Purely injuctive norm probably more credible for behaviors commonly subject to lack of self-control (intrapersonal preference conflict)

Formalizing norms-based interventions (finger exercises)

- Descriptive: information about participation
  - High $\bar{a} \iff$ high θ : strong average taste for prosocial behavior
  - No reward: $y = 0$
  - Convex cost of misrepresentation $C(\widehat{\theta} - \theta)$, minimized at true θ
    Reputational or falsification costs. Alternatives: (non-) disclosure of hard information, burning money, costly signaling; see later

- Focus here on "respectable" actions i.e. $\theta > \theta_0$, multiplier > 0
  - simplicity + relevant case for most existing NBI's

Formalizing norms-based interventions (finger exercises)

- Descriptive: information about participation
  - High $\bar{a} \iff$ high θ : strong average taste for prosocial behavior
  - No reward: $y = 0$
  - Convex cost of misrepresentation $C(\widehat{\theta} - \theta)$, minimized at true θ
    Reputational or falsification costs. Alternatives: (non-) disclosure of hard information, burning money, costly signaling; see later
- Focus here on "respectable" actions i.e. $\theta > \theta_0$, multiplier > 0
  - simplicity + relevant case for most existing NBI's
- Look for separating equilibrium - just sketch argument
  - Principal's strategy: $\widehat{\theta} = S(\theta)$, $\nearrow$
  - Agents' interpretation of announcement $\hat{\theta}$: $T(\hat{\theta})$, with $T = S^{-1}$
- Assume $e > \mu \Delta_\theta(v^*(0, \theta))$ $\forall \theta$: insufficient public good under FI
- Principal will always overstate the extent to which contributing / behaving well "is the norm": $S(\theta) > \theta$ for all $\theta > \theta_0$

- Prescriptive: what most other people approve of

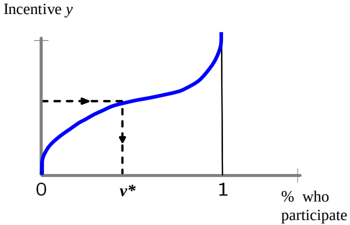
- Info about μ : again, P will always be tempted to overstate, as this boosts $\mu\Delta$ (directly + indirectly under SC), hence compliance
  - How communicated: surveys, votes. Subject to credibility problem
  - Alternative: actually increase μ, by making individual good and bad deeds more public. Will see effectiveness, but also limitations / costs later on
  - Descriptive norms may also be somewhat prescriptive: if high-v individuals pay more attention to other's behavior, announcing a higher θ (or $\bar{a}$) is also announcing a higher μ

- Info about e : if individual values v reflect the importance of e, P will always be tempted to overstate e: boosts both intrinsic and reputational motivation (SC), hence compliance
  - Credibility issue again. Need some form of costly signaling

Summary of Lecture II

When honor motive is dominant:

- Individuals' decisions are substitutes
- Incentives → partial crowding out
  (still work, but weakened)

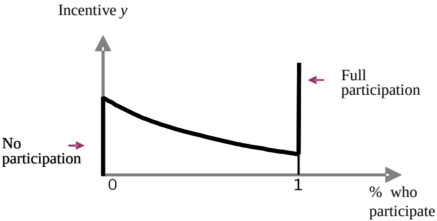


This occurs when:

- Most people are "mediocre", only rare "saintly" types with *v* well above most others (heroism, organ donation)
- Action is very costly
- There are possible "excuses" for not contributing, and / or one can do it without being noticed (⇒ weak stigma)

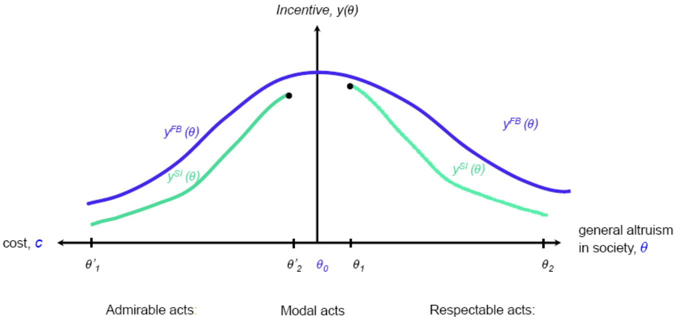
When stigma motive is dominant:

- Individual's decisions are complements
- Multiple norms may coexist
- Small incentives can have large effects: shift norms, crowding in



This occurs when

- Most people are "OK", only a few "rotten apples" with *v* well below most others (crime, child neglect)
- Action is relatively cheap
- There are possible non-glorious reasons for contributing (e.g., fear of the law), and/or it may go unnoticed (⇒ weak honor)



$$y^{FB} = e - \mu\Delta(c - e - \theta)$$

$$y^{SI} = \frac{e - y^{SI} - \mu\Delta(v^*(y^{SI}) - \theta)}{\lambda\left[1 + 1/\varepsilon_\theta(y^{SI})\right]}$$

Norms-based intervention,

descriptive or injunctive: communicating on θ, e, μ