
MOTIVATION INTRINSEQUE, INCITATIONS ET NORMES SOCIALES

Conférence Jean-Jacques Laffont

57^{ème} congrès annuel de l'Association Française de Science Economique

La Sorbonne, 18 septembre 2008

Jean TIROLE

Entièrement basé sur des travaux en commun avec Roland Bénabou (Princeton)



I. INTRODUCTION

REVISIT OLD DEBATE ABOUT “DO INCENTIVES WORK?”

Intrinsic motivation

- ✓ “One is said to be intrinsically motivated to perform an activity when one receives no apparent reward except the activity itself” [Deci 1971].
- ✓ Economics: rewards reinforce intrinsic motivation.
- Psychology, sociology: Without negating incentive properties, less confidence in price effect; “crowding out”.

ECONOMICS: LOW-POWERED INCENTIVES DESIRABLE WHEN:

✓ *Single task*

- Noisy performance measurement/lack of alignment between true and measured performance
[team worker].
- Collusion with monitors, capture
[noncontingent wages, cost of service regulation].
- Adverse selection: rent extraction-incentives tradeoff
[cost of service regulation].
- Repeated interaction.

✓ *Multiple tasks*

Incentive on one task crowds out performance on other task:

- Conflict on allocation of time and effort:
 - crowding out of quality by profit or other performance-based incentives
[school teacher; deregulation of airlines, railroads, electricity],
 - helping classmates or coworkers,
 - short termism [educational setting; bonus without stock options].
- Conflict of interest on output side

[advocate/prosecutor; energy conservation programs/output expansion.]

- ✓ Global picture: incentives work, although they may be misemployed/poorly structured, or they may be seriously limited by informational constraints.

MOTIVATION FOR THE WORK

- ✓ People engage in costly, low-individual-benefit activities: refrain from polluting, vote, volunteer, give to charitable organizations, help strangers, join rescue squad, risk life,...

[e.g., Fehr-Schmidt 03 survey of field and lab experiments]

- ✓ While theory and experiments to be discussed usually cast in a context of citizen behavior and public policy, behaviors are highly relevant for organizations as well:
 - cooperation and communication within firms,
 - rules,
 - changing norms.

✓ Some phenomena cannot be explained by sole presence of individuals with other-regarding preferences:

(a) *Crowding out effects*

[Festinger-Carlsmith 59, Deci-Ryan 85, Gneezy-Rustichini 00a,b, many papers by Frey, Fehr and co-authors.]

(b) *Social glory and shame attached to good and selfish deeds* [Codes of honor, shame; conspicuous donations; pressure (Batson 98, Freeman 97, Bandiera et al 05). People contribute more when observed (Funk 05, List 05).]

(c) *Self-image concerns* [Adam Smith 1776's "impartial spectator"; Dana et al 03a,b, Murnighan et al. 01; Kahneman-Knetsch 92 on "purchase of moral satisfaction"] "I am what I do"

Two (informational) crowding-out mechanisms

	<i>"Intrinsic and Extrinsic Motivation"</i> (BT 2003)	<i>"Incentives and Prosocial Behavior"</i> (BT 2006)
Mechanism	Conveys bad news about nature of task, its payoffs, or individual's ability (principal's trust of the agent) <i>(informed principal)</i>	Sullies the meaning of good deeds by creating doubt as to true motivation (overjustification effect) <i>(multidimensional signaling)</i>
Examples	Rewarding child for task, reading. Closely monitoring agent...	Help, blood donation,...
Impact of rewards	<ul style="list-style-type: none"> • Limit immediate reinforcement. • Crowd out future re-engagement 	Immediate
Related work	Souvorov (2004), Herold (2004), Souvorov - van de Ven (2005), Ellingsen-Johannesson (2008)...	[Bernheim (1994), Corneo(1997), Denrell (1998)], Seabright (2005)

Implications for fostering prosocial behaviors/internalization of externalities

- ✓ Economics & law: incentives, deterrence
 - Taxes, subsidies
 - Laws, regulations

- ✓ Psychology, sociology, political science: persuasion and social norms
 - Educating people to consequences of their actions
 - Public appeals to good citizenship
 - Changing the “social meaning” of the act
 - changing/correcting people’s perception of the social norm
 - shifting the norm itself.

- ✓ Unifying approach: two-way role of information
 - Individuals: signaling/reputational/status concerns.
 - State: law and other policies as a signal of social values, preferences, (dis)approval.

Psychologist's view of information too narrow; besides it neglects strategic and general equilibrium aspects.

Some good intuitions in law and economics literature, but no consistent model.

- ✓ Methodology: bring together
 - social (or self) signaling models; build here on Bénabou-Tirole (*AER* 2006),
 - informed principal model, link to ideas in Bénabou-Tirole (*RES* 2003).

[Unwritten paper: Bénabou-Tirole “Laws and Norms”.]

II. CROWDING OUT: THE LOOKING-GLASS SELF/INFORMED PRINCIPAL CHANNEL

- AGENT :
- undertakes task if self-confident in his efficacy/finds the task attractive;
 - gain θV if effort, cost c of effort.

- PRINCIPAL:
- has vested interest in agent's undertaking task and succeeding: parent, spouse, friend, teacher, boss, colleague...
 - gain θW if effort.

- AGENT : has more factual knowledge about
- his previous performances,
 - his past efforts,
 - past situational factors (facilitating/inhibiting).

- PRINCIPAL: has more information about
- difficulty of current task,
 - attractiveness of this task,
 - interpretation of agent's past performance.

AGENT EXTRACTS INFORMATION FROM PRINCIPAL'S BEHAVIOR: THE LOOKING-GLASS SELF

✓ *Timing*

- Stage 1: (knowledgeable about θ , V or c) P selects “policy” (reward y ; or help, delegation...).
- Stage 2: A , who has received private signal σ , selects effort/no effort.

✓ *Hypothesis*: A tries to see through P 's ulterior motivation

[Perfect Bayesian equilibrium.]

✓ Two points:

- Intrinsic motivation is endogenous.
- Two reasons why rewards may reduce effort: trust and profitability effects.

Symmetric information. Agent exerts effort iff

$$\theta(V + y) \geq c.$$

Reward is a positive reinforcer.

Intrinsic motivation: $\theta V - c$.

Extrinsic motivation: θy .

Asymmetric information about cost c , say. Agent exerts effort iff

$$\theta(V + y) \geq E(c|\sigma, y).$$

or

$$\sigma \geq \sigma^*(y).$$

No longer a clean separation between intrinsic and extrinsic motivation: $E(c)$ depends on y .

Trust effect

- ✓ How confident is P as to A 's intrinsic motivation?
 P 's view as to how A perceives task and his suitability to it.
- ✓ P pessimistic about A 's motivation
 - ⇒ needs to give stronger incentives
 - ⇒ but bad news.
- ✓ Lower-powered incentives than under symmetric information.

Profitability effect

- ✓ When, keeping A 's effort constant, A 's type (on which P has private information) enters P 's objective function in a way that would lead her to offer different policies to different types of A .
E.g., conditionally on A 's exerting effort:

$$\theta(W - y)$$

⇒ reward more costly if high probability of success θ .

- Result

Reward = bad news

- (1) positive reinforcer (but weakened due to inference process) for current task,

$$[\text{equilibrium } y_1 < y_2 \implies \sigma^*(y_1) > \sigma^*(y_2)]$$

- (2) reduces intrinsic motivation for future (similar) tasks (regardless of σ and outcome).

✓ *When can there be a hidden cost of rewards?*

- principal has information (educational setting vs workplace)
- sorting condition: principal must be more tempted to reward agent when agent has limited ability or task is boring (condition is not satisfied in all contexts!).

- ✓ **Delegation:** may signal trust

- ✓ **Help**
 - depression (dependent personality pattern),
 - overhelping

- ✓ **Disclosure of information and coaching**
 - Encouragement (build self-esteem).
 - Excuses.
 - Undermining other's ego.
 - rivalry ($W < 0$)
 - risk of coasting (effort and talent substitutes)
 - principal's reputation
 - battles for dominance

OTHER APPLICATIONS OF INFORMED PRINCIPAL THEORY

✓ *To bonuses and control*

- *Ellingsen-Johannesson* (*AER* forthcoming)

Players care about being esteemed. Applications to trust game, Falk-Kosfeld (2006), gift exchange.

- *Herold* (2005)

Multitask environment: lack of incentives on one task may signal P 's trust and boost incentives in another, uncontrolled task.

- *Suvorov-Van de Ven* (2006)

Ex post bonuses can be good news (“ P liked my work”) and boost intrinsic motivation.

✓ *To contextual inferences*

- *Kamenica* (*AER* forthcoming)

Product line is a signal, leads to compromise effects.

III. CROWDING OUT: SIGNAL EXTRACTION/ OVERJUSTIFICATION EFFECT

(1) Actions

- People choose their participation level $a \in [0, \infty)$ in some prosocial activity: giving blood, volunteering, recycling.
- If $a \implies$ incurs cost ca : effort, time, resources.
- Incentive: gets ya . Reward, subsidy, tax, etc (\longleftarrow policy, law).

(2) Motivations/preferences

- *First part: direct costs and benefits* from engaging in pro (or anti) social activity

$$[v_a + v_y y - c]a$$

v_a : valuation of extra public good which you provide + “joy of giving”

\implies prosocial orientation

v_y : valuation of money or private consumption.

Individual's true preference type $v = (v_a, v_y)$ is not directly observable by others (sometimes not even accessible to himself).

Private information, must be inferred from actions.

- *Second part: social esteem/self-image* (reputational concerns)
 - Desirable (pleasant, useful) to be perceived as generous, reciprocal, public minded, ...
... and, if heterogenous v_y , undesirable to be perceived as greedy, interested in money, or as poor.
 - What goes for social perceptions goes for self-perception. Judging oneself by actions.

⇒ to people's basic motivations, we add:

concern for *perceived prosociality*,
appearing prosocial in light of behavior

$$\underbrace{\mu_a} \quad \underbrace{E(v_a|a, y)} - \mu_y E(v_y|a, y)$$

- E is for “*expectation*”: what one can expect your true degree of intrinsic social orientation (or greed) to be, given that you did action a for reward y .
- μ is for “*image*”: how much you care about image/self image concerning altruism and greed. Depends in particular on how public or salient behavior is.
- People generally differ in their image concerns (μ) as well as their preferences over public and private goods v . “Type” = (v, μ) .

Three motives for prosocial behavior:
intrinsic + extrinsic + (self) reputational

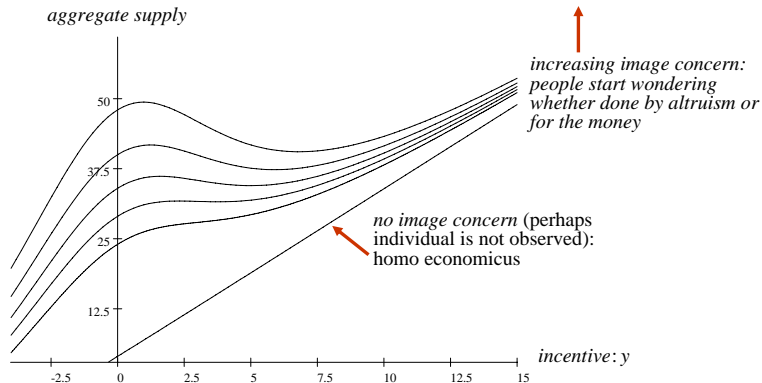
$$U = (v_a + v_y y - c)a + \mu_a E(v_a | a, y) - \mu_y E(v_y | a, y)$$

Policy parameters: material reward y and publicity (amplifying μ).

CROWDING OUT

✓ “Overjustification effect” of rewards

Suppose image concerns are the same for all individuals. When y increases, pro-social behavior is increasingly ascribed to greed. Supply curve downward sloping over a range for large enough μ (normal distribution):



Theory predicts that

- ✓ people give more when observed by others
- ✓ impact of y definitely smaller when act of giving is observed.

Supporting experimental evidence:

Ariely-Bracha-Meier (forthcoming *AER*)

[“Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially”]

✓ “Overjustification effect” of publicity/praise and shame

- Policies based on publicity, prominence, memorability:
Medals, titles, named buildings, public praise and shame, televised arrests, e-registry, pillory,...
- When people also differ in their image concerns μ , one wonders whether prosocial behavior is done for appearances \implies self-limiting impact of publicity.
- Other interesting aspects: self-disclosure of good deeds. Why it is so difficult to turn down rewards.

IV. HONOR, STIGMA, AND SOCIAL 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 over death, not going to church, not voting, divorce, welfare dependency, minor tax evasion, conspicuous modes of consumption,...]

- ✓ Social norms arise *endogenously* from the interplay of **honor** and **stigma**.
- ✓ More generally: when does the fact that others contribute more *increase* or *decrease* the pressure (social, moral) on me to do so? (strategic complements vs. substitutes).

Assume for simplicity that :

- $v_y = 1$ (call $v_a = v$ and $\mu_a = \mu$ to simplify notation)
- $a = 0$ or 1

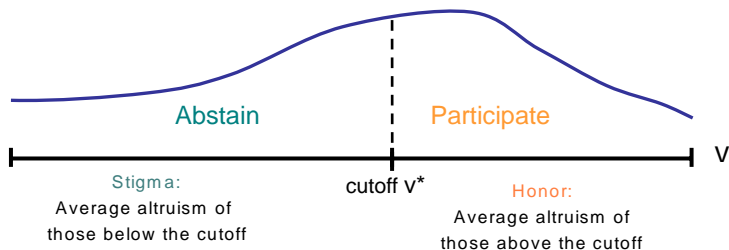
Utility:

$$\left\{ \begin{array}{ll} v + y - c + \mu \overbrace{E(v|a = 1, y)}^{\text{honor}} & \text{if contributes} \\ \mu \underbrace{E(v|a = 0, y)}_{\text{stigma}} & \text{if does not contribute} \end{array} \right.$$

- ✓ People who contribute are those with

$$v + y - c + \mu \times [\text{honor} - \text{stigma}] \geq 0$$

- ✓ Fix incentive. Who participates?



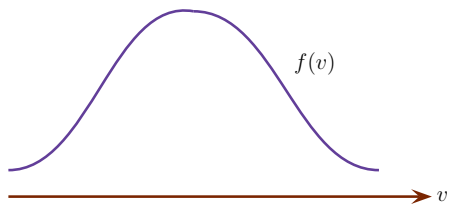
- ✓ When more people participate, *honor declines* but *stigma worsens!*
 \implies social/moral pressure to participate may decrease or increase (“multiplier”).
- ✓ Key difference between behaviors in which quest for *honor* versus avoidance of *stigma* is main driver of behavior.

Cutoff: $v^* + y - c + \Delta(v^*) = 0$, where Δ is the reputational concern:

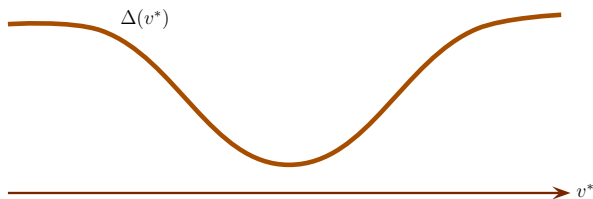
$$\Delta(v^*) \equiv \underbrace{\mu[E(v|v \geq v^*)]}_{\text{honor}} - \underbrace{E(v|v < v^*)}_{\text{stigma}}$$

Note: participation very elastic to incentive when $\Delta' < 0$.

We assume that the density is unimodal

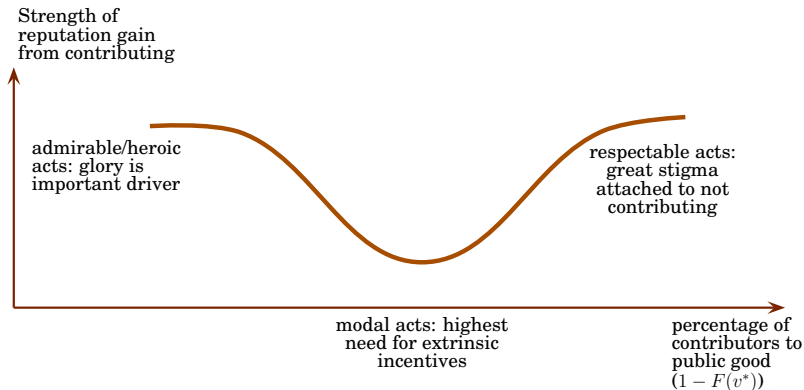


⇒ Reputational incentive as a function of the cutoff v^* :



Basics of reputational incentives

Assume that v has a unimodal distribution, and rule out multiple equilibria ($\Delta' > -1$):



Note that incentives powerful when non-participation of a minority only ($\Delta' < 0$).

Consistent with e.g., Posner (*J. Legal Studies* 1998) :

“ Making an activity subject to a significant risk of nontrivial sanctions changes the mixture of people in it by driving away people who have decently remunerative lawful alternatives... We need to pay more attention to selection phenomena in general.” (p562)

Key point: pursuit of esteem is a zero-sum game: the average reputation in society remains fixed, because the distribution of types is given.

In other words, esteem, and even self-esteem, is by its very nature a *positional good*.

- The *net social return* (what I do not internalize in my decision) to contributing is thus

$$S = \text{benefit created for others from increased public good } (e) - \text{others' loss in self or social image}$$
$$= \text{benefit created for others from increased public good } (e) - \text{gain in own self or social image.}$$

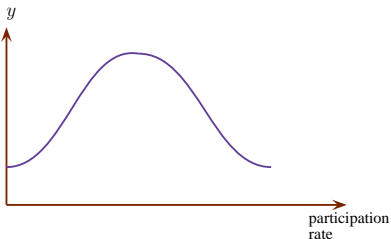
Difference between *free-riding* effect and *reputation-stealing* effect \implies

Proposition (modified Pigovian policy)

The socially optimal incentive rate is strictly less than the standard subsidy that leads agents to internalize the full public-good value of their contribution. It subtracts the value of image “bought”, and is equal to S .

Costless transfers case:

$$y = e - \Delta$$



V. NORM-BASED INTERVENTIONS AND EXPRESSIVE CONTENT OF LAW

Recall emphasis in other social sciences on persuasion and changing social norms and the perception thereof as *one* approach to encourage the internalization of externalities.

- ✓ Policy signal
 - direct: disclosure of information (“norm-based interventions”)
 - indirect: choice of incentive y (“expressive content of the law”).

- ✓ Policy-maker may try to express:
 - externality e
 - distribution of preferences/participation ($F(v|\theta)$)
 - visibility or others’ attention to one’s behavior μ .

What we do:

- ✓ Stage 1: (knowledgeable) state selects a “policy” (incentive, disclosure of information)
- ✓ Stage 2: Public reacts (reputational concerns).

Law influences norm, as it changes reputational incentives.

Norm influences law (for example, substitute in encouraging compliance).

- ✓ *Descriptive norms*: what most other people (in your community) do.
- ✓ *Prescriptive norms*: what most other people (in your community) approve of.

Descriptive norms are prescriptive if high v individuals pay more attention to behavior (care more; self-serving; ...). Then

$$\begin{aligned}\mu &= \mu_0 + \mu_1 v \\ \implies \mu(\theta) &= \mu_0 + \mu_1 \theta \quad \text{if } v = \theta + \varepsilon.\end{aligned}$$

EXPRESSIVE CONTENT OF THE LAW

Incentives (y) as signals of behavior, values, or externalities (θ , μ or e).

[Cooter: “The expression of social values is possibly the most important function of the courts”.]

- ✓ Large literature (mostly outside economics) arguing that laws have dual role: not just “price” good or bad behavior, but also “express” society’s values (what it approves of, or chooses to punish).
- ✓ Ancillary debate: *how* to express moral condemnation?
 - Kahan on shaming sanctions vs Posner/Whitman,
 - alternative vs inefficient sanctions.

Choice of reward (shadow cost $\lambda \geq 0$). Policy-maker knows θ .

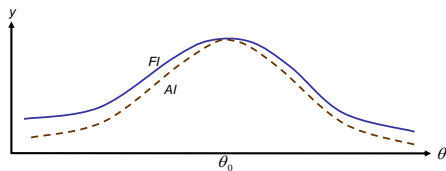
Again, we look for a separating equilibrium.

$$W(y, \theta) \equiv \int_{\hat{v}(y)}^{\infty} (e + v - c - \lambda y) g(v - \theta) dv$$

where

$$\hat{v}(y) = v^*(y, \hat{\theta}(y)):$$

$$\hat{v}(y) + y - c + \Delta(\hat{v}(y) - \hat{\theta}(y)) = 0$$



(For small λ at least) $y^{AI}(\theta) < y^{FI}(\theta)$.

EXAMPLES OF NORM-BASED INTERVENTIONS

- (1) **Communicating on e** (“scarcity situations” experiments)
- ✓ Reiss-White (2008): during 2000-2001 California energy crisis, San Diego, facing rolling blackout, launched \$65 million public campaign to promote energy conservation . Reduced consumption continuously over few months, to about -7% .
 - ✓ De Janvry et al. (2006): shortage of flu vaccines in Fall 2004; Center for Disease Control recommended that people in non-priority groups delay vaccination. Call on self restraint reduced it by 37.5% (esp. among target group).
 - ✓ Monin-Norton (2003): Water shortage at Princeton U. following Hurricane Floyd in 1999. Three-day ban on showering. On average, 41% students in dormitories reported not taking showers, and (perhaps more reliably) students on average estimated that 40% of others abstained.

Disclosure of information about externality

- ✓ • Policy-maker knows true externality e .
- $v = ze$
(valuations reflect some consequentialism).
- Cost of signal \hat{e} : $C(\hat{e}|e)$ (with $C_1 > 0$, $C_{12} < 0$)

Separating equilibrium

$$\hat{e} > e.$$

Want to convey notion that externality sizeable to put reputational incentives in motion.

(2) Communicating on θ and μ (descriptive and prescriptive norms)

- ✓ Schultz et al. (2007): monitored electricity meters/consumption of 290 households in a California town.

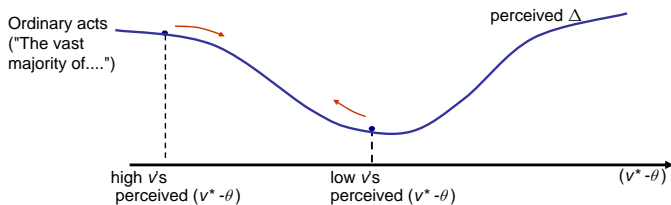
Each week, hung on their door a visible feedback form with (randomized):

- *Descriptive condition*: Own electricity consumption + average consumption of households in their neighborhood (+ tips on conservation) \Rightarrow convergence toward mean.

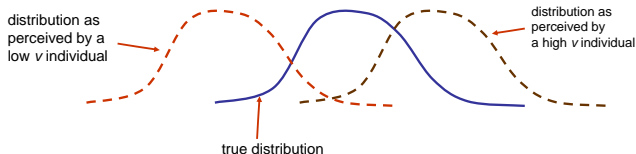
[See also evidence on social norm marketing campaigns against binge drinking.]

- *Prescriptive condition*: Same, plus smily face if below average, frowning face if above \Rightarrow high consumers still decrease, low consumers no longer increase consumption.

Possible interpretation of descriptive condition



(Bayesian version of) false consensus effect



[Example: $v = \theta + \varepsilon$

$$\hat{\theta} = \frac{h_{\theta}}{h_{\theta} + h_{\varepsilon}} \bar{\theta} + \frac{h_{\varepsilon}}{h_{\theta} + h_{\varepsilon}} v]$$

- ✓ *Descriptive condition*: Some evidence that highly publicized crackdowns on tax cheating increases frequency of tax evasion the next year.
- ✓ *Prescriptive condition*: add communication on μ
 - “people care”
 - salience
- ✓ Cialdini’s policy recommendation:
 - if most people behave well, use descriptive norm (or both);
 - if most people behave badly, use prescriptive, avoid descriptive.

Fits well with the theory developed here.

(3) “Pluralistic Ignorance”

- ✓ Why do such “verbal” interventions work? Psychologists’ view:
 - People care about being seen/seeing themselves as moral, prosocial, etc.
 - Judge what “one should do” in that respect by what they see or believe others do and/or approve of
 - Often misperceive what most others do, and especially what they value/approve of, because of:
 - limited information, greater visibility of certain actions
 - *pluralistic ignorance*: because people see or perceive that most others do *X*, they take it to mean that everyone values/approves of *X*. Do not realize/properly account for the fact that others are also conforming to a common perceived norm.

Dispelling pluralistic ignorance

- ✓ Vast problem of excess drinking by college undergraduates, and the young more generally.

Efforts at individual education (to risks, etc.) and public campaigns very limited.

Role of peer influences widely recognized. But why/how?

- ✓ Prentice-Miller (1993): students asked about their level of comfort with drinking on campus, and about their perception of the general attitude of other students about it. Consistently find that:
 - Students significantly overestimate the extent to which others are comfortable with drinking. “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.

VI. CONCLUDING REMARKS

We need to know *when* incentives work and when they don't.

- ✓ Internalization of externalities depends on
 - (1) material incentives,
 - (2) intrinsic motivation,
 - (3) social or self-image concerns.
- ✓ These three incentives interact with each other.
- ✓ Organizational choices affect not only (1) (economics), but also (2) and (3), for two basic reasons:
 - information conveyed by the choice of policy,
 - altering incentives or visibility changes the meaning attached by observers to prosocial or antisocial behavior.
- ✓ Richer view of what incentives achieve, forces us to think about sorting conditions, strategic complementarities/substitutabilities, etc. Results are rather intuitive, though.
- ✓ Managerial implications by and large unexplored.