Revisiting Incentives:

Values, Laws and Norms

Roland Bénabou

Princeton University

Toulouse Lectures - December 2009

Based on joint work with Jean Tirole (TSE)

Revisiting Incentives:

Values, Laws and Norms

Roland Bénabou

Princeton University

Toulouse Lectures - December 2009

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

The view from home

- Economics: incentives are key. Effective, when properly applied
- Aware of a number of caveats - carry around a “checklist”.

The view from home

- Economics: incentives are key. Effective, when properly applied
- Aware of a number of caveats - carry around a “checklist”.
- Low-powered incentives desirable when
  - Single task : Noisy performance measurement, teams, collusion with monitors or capture, repeated interactions, adverse selection
  - Multiple tasks: Over-allocate resources (time, effort) to one task at expense of another, crowding out of quality by profit-based incentives, underinvest in helping coworkers or classmates, short-termism

The view from home

- Economics: incentives are key. Effective, when properly applied
- Aware of a number of caveats - carry around a “checklist”.
- Low-powered incentives desirable when
  - Single task : Noisy performance measurement, teams, collusion with monitors or capture, repeated interactions, adverse selection
  - Multiple tasks: Over-allocate resources (time, effort) to one task at expense of another, crowding out of quality by profit-based incentives, underinvest in helping coworkers or classmates, short-termism
- Nonetheless, premise remains that incentives work / can be made to work, albeit limited by informational constraints
- Main focus accordingly remains on achieving compliance through contractual incentives, mechanism design

The view from there

- Psychology, sociology: material incentives are often of limited effectiveness, or even counterproductive
    - “Undermine intrinsic motivation”
    - “Change, sully the meaning of actions”
    - Crowd out valuable social norms, institutions ⏩

The view from there

- Psychology, sociology: material incentives are often of limited effectiveness, or even counterproductive
  - "Undermine intrinsic motivation"
  - "Change, sully the meaning of actions"
  - Crowd out valuable social norms, institutions ⏩

What does it mean? Evidence, then want to understand

when such concerns are more relevant, and when less

- Even when incentives work, other methods of achieving compliance may work as well and be much cheaper
  - Public appeals, "norms-based interventions", social sanctions
  - How do they work, and what are their own pitfalls?

The view from there

- Sometimes, small incentives work surprisingly well, e.g., “symbolic” fines with significant effects Something else is going on.

- Contractual and norms-based incentives often used together, e.g., legal penalties and social esteem / sanctions.

  How do they interact: crowding out or crowding in?

The view from there

- Sometimes, small incentives work surprisingly well, e.g., “symbolic” fines with significant effects Something else is going on.

- Contractual and norms-based incentives often used together, e.g., legal penalties and social esteem / sanctions.

  How do they interact: crowding out or crowding in?

- Societies, electorates: whether or not incentives would work, large fractions of society resist / object to them in certain contexts:
  - Organs, blood, votes, carbon taxes
  - Monetary incentives for students
  - Idea that not everything should have a price, “taboo tradeoffs”

The view from there

- Law: also a somewhat different view.

  Laws = incentives (fines, jail sentences), but also "express" the values of a society (or those it aspires to have)

  - Some punishments that economists like (fines, home monitoring) often rejected as too soft by society, not stigmatizing enough
  - Some very tough (and cheap) incentives have substantial popular support and are still used in many countries: corporal punishments, shaming, torture, death
  - But increasingly renounced by developed societies as not "civilized", contrary to their values. Real issue is not effectiveness or ineffectiveness, but "what kind of a society we are"

- Try to understand this concept, incorporate into economic analysis

  When does expressive role of law call for less strict incentives, or for tougher ones?

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 standards

Lecture I

Intrinsic, Extrinsic and Attributional Motivations

1. Introductory evidence
2. General framework
3. Intrinsic and extrinsic motivation: the trust effect
4. Self-confidence, trust profitability effects

Evidence and Puzzles

Lettre de Didier Chatenay, DR1 CNRS en Physique

- Madame la Présidente, Monsieur le Directeur Général [du CNRS],

  J'ai appris récemment que vous alliez mettre en place une prime d'excellence scientifique (PES) destinée aux chercheurs du CNRS. Parmi les voies explorées... vous songez à une attribution automatique aux médailles du CNRS. Ayant eu l'honneur d'être distingué par mes pairs en recevant la Médaille d'Argent du CNRS en 1999, je vous fais part de mon refus à priori de me voir verser une telle prime

  Par principe je suis en effet totalement opposé à l'existence meme d'un quelconque système de primes, considérant qu'elles ne constituent en aucun cas un mécanisme acceptable d'amélioration des revenus des agents de la fonction publique.

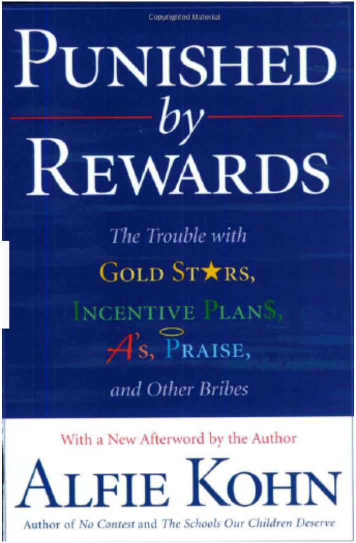
Lettre de Didier Chatenay, DR1 CNRS en Physique

- Madame la Présidente, Monsieur le Directeur Général [du CNRS],

  J'ai appris récemment que vous alliez mettre en place une prime d'excellence scientifique (PES) destinée aux chercheurs du CNRS. Parmi les voies explorées... vous songez à une attribution automatique aux médailles du CNRS. Ayant eu l'honneur d'être distingué par mes pairs en recevant la Médaille d'Argent du CNRS en 1999, je vous fais part de mon refus à priori de me voir verser une telle prime

  Par principe je suis en effet totalement opposé à l'existence meme d'un quelconque système de primes, considérant qu'elles ne constituent en aucun cas un mécanisme acceptable d'amélioration des revenus des agents de la fonction publique.

  Jusqu'ici, une des caracteristiques du monde académique et savant (et tout particulierement du CNRS à travers l'attribution de médailles) résidait dans sa capacite à décerner à certains de ses membres une reconnaissance symbolique dépourvue de tout avantage matériel. L'instauration d'un système de primes va à l'encontre de cette tradition qu'il me semble necessaire de maintenir, les considérations d'ordre matériel ne devant en aucun cas interférer avec des arguments purement scientifiques...



Copyrighted Material

PUNISHED *by* REWARDS

The Trouble with

GOLD ST★RS,

INCENTIVE PLAN$,

A's, PRAISE,

and Other Bribes

With a New Afterword by the Author

ALFIE KOHN

Author of *No Contest* and *The Schools Our Children Deserve*

“Torre Says Yankees’ Offer Showed Lack of Trust” (NYT 2007)

- Torre rejected the Yankees’ [baseball team] contract offer on Thursday, but in a news conference... he said he felt rejected by them. Torre acknowledged that the $5 million the Yankees offered him was generous. But he said he felt insulted by... the deal, which was heavily tied to incentives and not open to negotiation.

“Torre Says Yankees’ Offer Showed Lack of Trust” (NYT 2007)

- Torre rejected the Yankees’ [baseball team] contract offer on Thursday, but in a news conference... he said he felt rejected by them. Torre acknowledged that the $5 million the Yankees offered him was generous. But he said he felt insulted by... the deal, which was heavily tied to incentives and not open to negotiation.

- The structure of the Yankees’ proposal rewarded Torre with a $1 million bonus for each postseason round the team would achieve in 2008. But if he did not reach the World Series, he would not have exceeded this season’s $7.5 million salary.

- The new offer would have kept Torre as the majors’ highest-paid manager, and while Torre bristled at the idea of incentives, ownership saw them as a way for him to exceed his 2007 salary. The [Yankee’s General Manager] said the incentive package would probably not be part of a new manager’s contract.

“Torre Says Yankees’ Offer Showed Lack of Trust” (NYT 2007)

- Torre rejected the Yankees’ [baseball team] contract offer on Thursday, but in a news conference... he said he felt rejected by them. Torre acknowledged that the $5 million the Yankees offered him was generous. But he said he felt insulted by... the deal, which was heavily tied to incentives and not open to negotiation.

- The structure of the Yankees’ proposal rewarded Torre with a $1 million bonus for each postseason round the team would achieve in 2008. But if he did not reach the World Series, he would not have exceeded this season’s $7.5 million salary.

- The new offer would have kept Torre as the majors’ highest-paid manager, and while Torre bristled at the idea of incentives, ownership saw them as a way for him to exceed his 2007 salary. The [Yankee’s General Manager] said the incentive package would probably not be part of a new manager’s contract.

- “I’ve been there for 12 years and I didn’t think motivation was needed,” Torre said.

“Bonus Babies”: NYT Op-Ed by Barry Schwartz, October 24, 2007

- “Torre was right. It is insulting to be offered incentives like these. What, after all, are the incentives for? They’re for doing his job as well as he can. The offer of a bonus implies that without it, the employee would just be mailing it in.

“Bonus Babies”: NYT Op-Ed by Barry Schwartz, October 24, 2007

- “Torre was right. It is insulting to be offered incentives like these. What, after all, are the incentives for? They’re for doing his job as well as he can. The offer of a bonus implies that without it, the employee would just be mailing it in.

  It is true, of course, that people work for money, and if they weren’t getting paid, they wouldn’t work at all. But people aren’t working only for money. They are also working because they think their work serves a purpose, or they are devoted to excellence, or they love what they do. When you offer people bonuses for doing their jobs, you are telling them that money is not just one of many reasons to work, but the only reason.

“Bonus Babies”: NYT Op-Ed by Barry Schwartz, October 24, 2007

- “Torre was right. It is insulting to be offered incentives like these. What, after all, are the incentives for? They’re for doing his job as well as he can. The offer of a bonus implies that without it, the employee would just be mailing it in.

  It is true, of course, that people work for money, and if they weren’t getting paid, they wouldn’t work at all. But people aren’t working only for money. They are also working because they think their work serves a purpose, or they are devoted to excellence, or they love what they do. When you offer people bonuses for doing their jobs, you are telling them that money is not just one of many reasons to work, but the only reason.

  But the insult Torre feels for being offered a bonus for doing something few baseball managers can do is nothing compared with the insult that New York City teachers should be feeling right now... The city announced that it will start offering bonuses to teachers whose students perform well on standardized tests. In other words, teachers can’t be trusted to do their jobs without bonuses. How insulting can you get?

- There are settings in which bonuses may make sense – if the work offers no opportunity to find satisfaction, for instance, or if it really is all about the money. And yes, there should be public acknowledgment of extraordinary performance. But that acknowledgment needn't be financial, and it certainly shouldn't be contractual.The more society embraces the idea that nobody will do anything right unless it pays, the more true it will become that nobody does anything right unless it pays. And this is no way to run a ballclub, a school system, or a country.

Barry Schwartz, a professor of psychology at Swarthmore College, is the author of "The Costs of Living: How Market Freedom Erodes the Best Things in Life."

Letter to the Editor, New York Times, October 25, 2007

- "Bonuses for Teachers?"

Barry Schwartz [a psychologist] has stated the issue very succinctly. Incentives are insulting to teachers, as they imply that they will not do their jobs without bonuses. This is diverting precious resources away from the structural problems that plague the school system. The money would be better spent on reducing class size, providing necessary resources and fixing overt structural problems.

In the teaching profession, money is one of the many reasons we work; however, the incentive or motivation for the work must be to make a difference.

Signed: Larry Hoffner. The writer is a high school teacher.

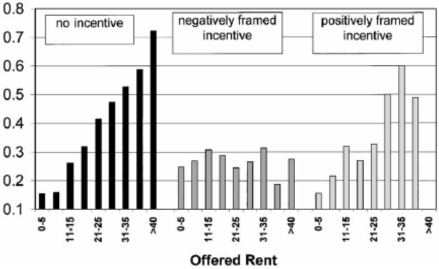
Incentives and effort: Fehr and Gächter (2002)

- No Incentives - baseline:
  - "Employer" makes contract offer: $p =$ non-contingent payment, $\hat{a} =$ desired effort or quality, non-binding. Offered rent: $\hat{U}_A = p - C(\hat{a})$

    Payoff $U_P = Wa - p$ if contract accepted
  - Agent chooses effort a, at some convex cost $C(a)$. Payoff $U_A = p - C(a)$
- Incentives: P can choose a "wage deduction" (fine) $0 \leq f \leq \bar{f}$ that will be imposed if A found to be shirking, $a < \hat{a}$; verification occurs with prob. $1/3$
- Incentives, positively frame: same, but contingent payment framed as "bonus" $0 \leq b \leq f$, to be paid only if verification shows $a \geq \hat{a}$

Incentives and effort: Fehr and Gächter (2002)

- No Incentives - baseline:
  - "Employer" makes contract offer: $p =$ non-contingent payment, $\hat{a} =$ desired effort or quality, non-binding. Offered rent: $\hat{U}_A = p - C(\hat{a})$
    Payoff $U_P = Wa - p$ if contract accepted
  - Agent chooses effort a, at some convex cost $C(a)$. Payoff $U_A = p - C(a)$
- Incentives: P can choose a "wage deduction" (fine) $0 \leq f \leq \bar{f}$ that will be imposed if A found to be shirking, $a < \hat{a}$; verification occurs with prob. $1/3$
- Incentives, positively frame: same, but contingent payment framed as "bonus" $0 \leq b \leq f$, to be paid only if verification shows $a \geq \hat{a}$



Incentives and trust: Fehr and Rockenbach (2001)

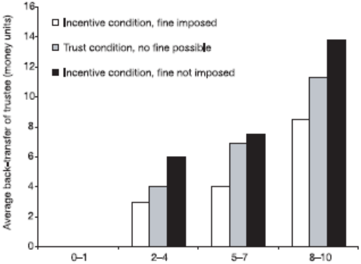
- Standard trust game
  - Investor (I) endowed with 10. Can send $0 \leq x \leq 10$ to Responder (R), which Experimenter triples to $3x$.
  - Responder chooses back-transfer to Investor, $0 \leq y \leq 3x$
- Two variants / conditions:
  - "Trust": when choosing x, I specifies non-binding "desired" back transfer $\hat{y}$
  - "Incentives": same, but I can also (need not) impose fine (pure loss) of $f = 4$, to be levied on R if $y < \hat{y}$
- Standard predictions: "Trust" $\rightsquigarrow x = y = 0$, "Incentives" $\rightsquigarrow I$ makes use of fine, R returns $y = 4$ and so I sends $x =$ 2 or 3

Incentives and trust: Fehr and Rockenbach (2001)

- Standard trust game
  - Investor (I) endowed with 10. Can send $0 \leq x \leq 10$ to Responder (R), which Experimenter triples to $3x$.
  - Responder chooses back-transfer to Investor, $0 \leq y \leq 3x$

- Two variants / conditions:
  - "Trust": when choosing x, I specifies non-binding "desired" back transfer $\hat{y}$
  - "Incentives": same, but I can also (need not) impose fine (pure loss) of $f = 4$, to be levied on R if $y < \hat{y}$

- Standard predictions: "Trust" $\rightsquigarrow$ $x = y = 0$, "Incentives" $\rightsquigarrow$ I makes use of fine, R returns $y = 4$ and so I sends $x =$ 2 or 3



Average Behaviour and payoffs of investors and trustees

	Trust Condition	Incentive Condition, fine chosen	Incentive Condition, no fine chosen
Investment	6.5	6.8	8.7
Desired back-transfer as a percentage of tripled investment	59.9	67.4	63.7
Actual back-transfer	7.8	6.0	12.5
Actual back-transfer as a percentage of tripled investment	40.6	30.3	47.6
Actual back-transfer as a percentage of desired back-transfer	74.4	54.5	74.1
Investor's payoff	11.3	9.2	13.8
Trustee's payoff	21.8	22.4	23.5
Number of Observations	24 pairs	30 pairs	15 pairs

- Responses to vignettes
  - Read “work-situation” like vignettes of an employer who does or does not control employees’ opportunities for cheating / stealing
  - Asked about what their motivation would be as employees
  - 403 subjects, ≈ 2,000 work motivation responses for 10 conditions

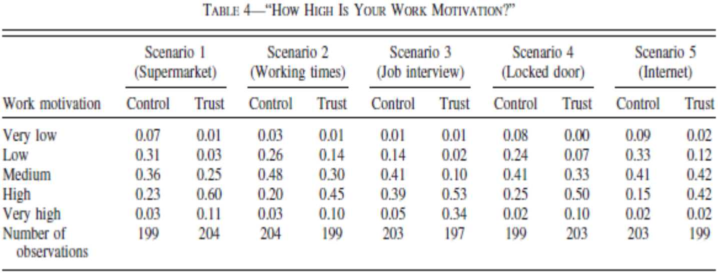


TABLE 4—“HOW HIGH IS YOUR WORK MOTIVATION?”

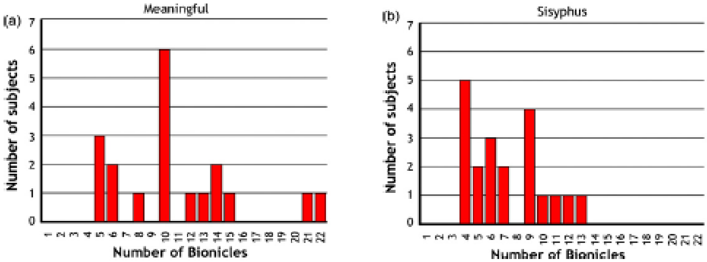
	Scenario 1 (Supermarket)		Scenario 2 (Working times)		Scenario 3 (Job interview)		Scenario 4 (Locked door)		Scenario 5 (Internet)	
Work motivation	Control	Trust	Control	Trust	Control	Trust	Control	Trust	Control	Trust
Very low	0.07	0.01	0.03	0.01	0.01	0.01	0.08	0.00	0.09	0.02
Low	0.31	0.03	0.26	0.14	0.14	0.02	0.24	0.07	0.33	0.12
Medium	0.36	0.25	0.48	0.30	0.41	0.10	0.41	0.33	0.41	0.42
High	0.23	0.60	0.20	0.45	0.39	0.53	0.25	0.50	0.15	0.42
Very high	0.03	0.11	0.03	0.10	0.05	0.34	0.02	0.10	0.02	0.02
Number of observations	199	204	204	199	203	197	199	203	203	199

"Man's Search for Meaning... " Ariely et al. (JEBO 2007)

- Assembling Lego "bionicles" (robot-like figurines), made of 40 pieces each
- Mean time ≈ 10 min. Paid \$2.00 for the first \$1.89 (11¢ less) for the second one, and so on linearly. For the 20th+, \$0.02.
- Two conditions (20 subjects in each):
  - "Meaningful": each figurine from new box, assembled ones lined up on shelf
  - "Sysiphus": just 2 boxes, experimenter disassembles previous bionicle while subject is working on the next

“Man’s Search for Meaning... ” Ariely et al. (JEBO 2007)

- Assembling Lego “bionicles” (robot-like figurines), made of 40 pieces each
- Mean time ≈ 10 min. Paid \$2.00 for the first \$1.89 (11¢ less) for the second one, and so on linearly. For the 20th+, \$0.02.
- Two conditions (20 subjects in each):
  - “Meaningful”: each figurine from new box, assembled ones lined up on shelf
  - “Sysiphus”: just 2 boxes, experimenter disassembles previous bionicle while subject is working on the next



Second experiment, with possibility of cheating

- Boring task: subjects given sheet of paper with seemingly random sequence of letters, paid \$0.55 for finding 10 instances of two consecutive letters 's.'

- After completing first page, asked if want to complete second one for \$0.50 (5¢ less), etc., with wages declining by 5¢ per sheet, until decides to stop

Second experiment, with possibility of cheating

- Boring task: subjects given sheet of paper with seemingly random sequence of letters, paid $0.55 for finding 10 instances of two consecutive letters 's.'

- After completing first page, asked if want to complete second one for $0.50 (5¢ less), etc., with wages declining by 5¢ per sheet, until decides to stop

- Three conditions:

  - Acknowledged: asked to write their name on each sheet. Instructions explained that, after completing task, would hand it over to the experimenter who would examine it and file it in a folder

  - Ignored: not instructed to write their name. Instructions explained that, after completing task, experimenter would place the sheet on a high stack of papers. Experimenter did so without examining the completed sheets

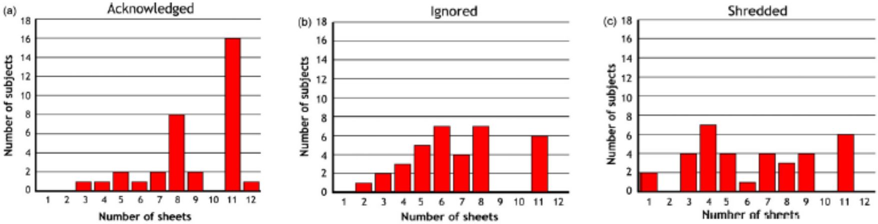
  - Shredded: same as Ignored, except instructions explained that completed sheets would be immediately put through a shredder. As the subjects turned in sheets, experimenter shredded them without a glance.

- Subjects could cheat in all the conditions, given the absence of monitoring. Incentives to cheat arguably higher in "Ignored" and even higher in "Shredded", where cheating not only impossible to detect, but also of no consequence since sheets were immediately destroyed.

- Standard theory: highest reservation wage / stop earlier in "Acknowledged" which requires more conscientious attention to fastidious task, and lowest in "Shredded"

- Subjects could cheat in all the conditions, given the absence of monitoring. Incentives to cheat arguably higher in “Ignored” and even higher in “Shredded”, where cheating not only impossible to detect, but also of no consequence since sheets were immediately destroyed.

- Standard theory: highest reservation wage / stop earlier in “Acknowledged” which requires more conscientious attention to fastidious task, and lowest in “Shredded”



Roland Fryer's large-scale educational experiments

- Student incentives: experiments in 261 schools in 4 major US cities
  - 19,264 students, \$10,000,000+ distributed
  - Large / expert management team
- Focus on very poor, heavily minority, inner-city schools

  Main goal = close the achievement gap in educational attainment, esp. for most disadvantaged students
- Also another project on (school-level) teachers incentives

  Very preliminary results

Incentives for students: Fryer (2009)

- Dallas: "Earning by Learning"
  - 43 schools opted in, 22 randomly chosen for treatment

    3788, 2nd grade students
  - Paid \$2 / book read, short test to check Rewards given 3 times / year
  - \$360,000 total cost, 80% consent rate. 1 dedicated project manager
- New York: "Spark"
  - 143 schools opted in, 63 schools randomly chosen for treatment.
  - 8,176, 4th and 7th grade students, select 8th grade students
  - Rewards: 4th graders can earn up to \$25 per test and \$250 per year ; 7th graders up to \$50 per test and \$500 per year. Rewards paid 5 times / year
  - \$6,000,000 distributed. 66% opened bank accounts. 82% consent rate; 3 dedicated project managers

- Washington, DC: “Capital Gains”
  - 17 schools randomly chosen to participate from set of all DC middle schools. 3,269 middle school students
  - Students paid for attendance, discipline, wearing uniform. Can earn up to \$100 / week, \$1500 / year. Paid every two weeks
  - \$2,300,000 distributed., 99.9% consent rate, 2 dedicated project managers.

- Chicago: PaPer
  - 70 schools opted in, 20 randomly chosen for treatment. 4,120, 9th graders
  - Paid for grades: A = \$50, B=\$35, C=\$20, D=\$0, F=\$0 in any classes. Can earn up to \$250 per report card and \$2000 per year. Paid every 5 weeks / report card. Half of the rewards given immediately, the other at graduation
  - \$3,000,000 distributed, 2 dedicated project managers.

- Evaluation of programs: relative gain performance on state tests

  Comparable metrics
  - $+$ 0.08 stand. dev. $\approx +$ 1 month of schooling
  - $+$ 0.17 stand. dev. $\approx$ reducing class size from 24 to 16 (Krueger)

Main findings

- Paying to read books (Dallas) had sizeable and significant effect on English test scores (only, e.g., not math); about 1.6 stand. dev.'s
    - Gains = mostly for "stronger" students
    - Poorest Spanish-speaking students showed instead some losses
- Paying for test scores, attendance, discipline, etc., improved attendance and discipline had no significant effect on test scores
- Paying for grades had no significant impact on test scores
- Fryer's hypothesis:
    - Students, esp. those with most disadvantaged backgrounds, do not know the " production function" for educational achievement
    - Therefore, better to subsidize inputs rather than (as standard theory would suggest) output

Main findings

- Paying to read books (Dallas) had sizeable and significant effect on English test scores (only, e.g., not math); about 1.6 stand. dev.'s
  - Gains = mostly for "stronger" students
  - Poorest Spanish-speaking students showed instead some losses
- Paying for test scores, attendance, discipline, etc., improved attendance and discipline had no significant effect on test scores
- Paying for grades had no significant impact on test scores
- Fryer's hypothesis:
  - Students, esp. those with most disadvantaged backgrounds, do not know the " production function" for educational achievement
  - Therefore, better to subsidize inputs rather than (as standard theory would suggest) output
- Teacher incentives (NYC) for student achievement
  - Thousands of $ / teacher at stake
  - Preliminary results = 0.0

Panagopoulos (2009): “Turning Out, Cashing In”

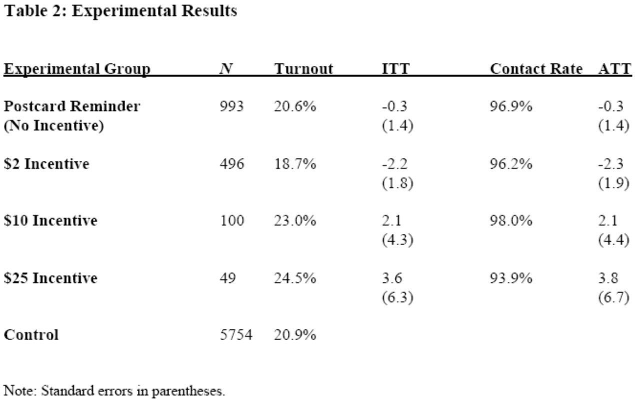
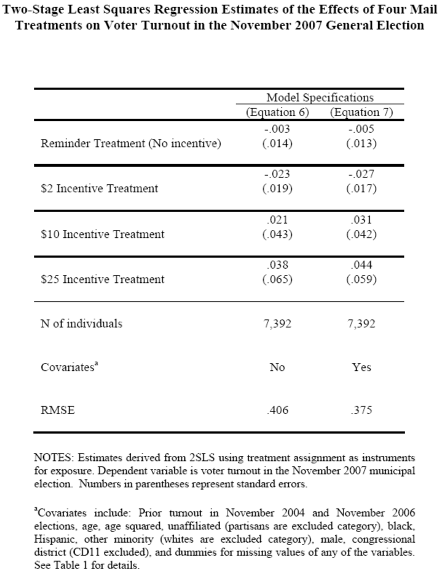


Table 2: Experimental Results

Experimental Group	*N*	Turnout	ITT	Contact Rate	ATT
Postcard Reminder (No Incentive)	993	20.6%	-0.3 (1.4)	96.9%	-0.3 (1.4)
$2 Incentive	496	18.7%	-2.2 (1.8)	96.2%	-2.3 (1.9)
$10 Incentive	100	23.0%	2.1 (4.3)	98.0%	2.1 (4.4)
$25 Incentive	49	24.5%	3.6 (6.3)	93.9%	3.8 (6.7)
Control	5754	20.9%			

Note: Standard errors in parentheses.



Two-Stage Least Squares Regression Estimates of the Effects of Four Mail Treatments on Voter Turnout in the November 2007 General Election

	Model Specifications	
	(Equation 6)	(Equation 7)
Reminder Treatment (No incentive)	-.003 (.014)	-.005 (.013)
$2 Incentive Treatment	-.023 (.019)	-.027 (.017)
$10 Incentive Treatment	.021 (.043)	.031 (.042)
$25 Incentive Treatment	.038 (.065)	.044 (.059)
N of individuals	7,392	7,392
Covariates[a]	No	Yes
RMSE	.406	.375

NOTES: Estimates derived from 2SLS using treatment assignment as instruments for exposure. Dependent variable is voter turnout in the November 2007 municipal election. Numbers in parentheses represent standard errors.

[a] Covariates include: Prior turnout in November 2004 and November 2006 elections, age, age squared, unaffiliated (partisans are excluded category), black, Hispanic, other minority (whites are excluded category), male, congressional district (CD11 excluded), and dummies for missing values of any of the variables. See Table 1 for details.

Gerber et al. “Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment”, APSR (2008)

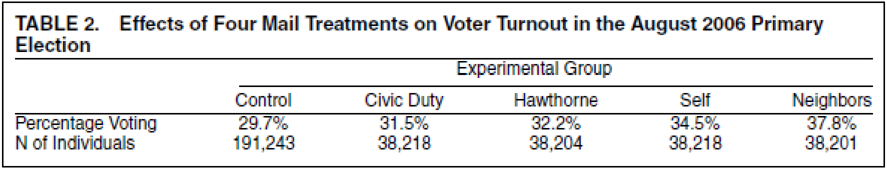


TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

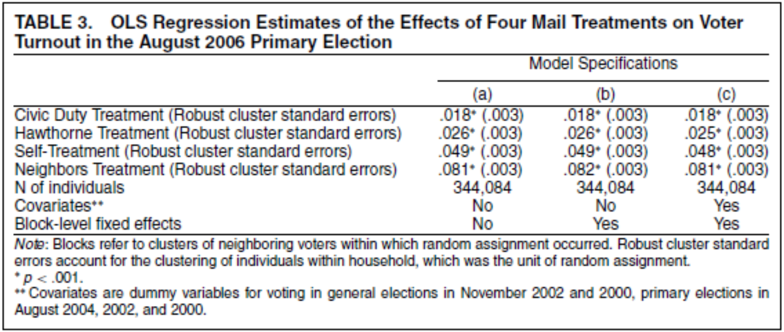


TABLE 3. OLS Regression Estimates of the Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Model Specifications		
	(a)	(b)	(c)
Civic Duty Treatment (Robust cluster standard errors)	.018* (.003)	.018* (.003)	.018* (.003)
Hawthorne Treatment (Robust cluster standard errors)	.026* (.003)	.026* (.003)	.025* (.003)
Self-Treatment (Robust cluster standard errors)	.049* (.003)	.049* (.003)	.048* (.003)
Neighbors Treatment (Robust cluster standard errors)	.081* (.003)	.082* (.003)	.081* (.003)
N of individuals	344,084	344,084	344,084
Covariates**	No	No	Yes
Block-level fixed effects	No	Yes	Yes

Note: Blocks refer to clusters of neighboring voters within which random assignment occurred. Robust cluster standard errors account for the clustering of individuals within household, which was the unit of random assignment.

* $p < .001$.

** Covariates are dummy variables for voting in general elections in November 2002 and 2000, primary elections in August 2004, 2002, and 2000.

The General Framework

Agents: actions

- Agents (one or many) choose action a, at cost $C(a)$: effort, time, resources. May be discrete or continuous
    - Private-goods context: effort in the firm, non-opportunism...
    - Public-goods context: volunteering, voting, giving blood, helping, contributing to a good cause, not polluting...
- Incentive: receive y per unit of a, from some principal
    - Private-goods context: wage for effort, performance-contingent bonus, penalty for failure, etc.
    - Public-goods context: subsidy, tax, fine, prison
- Action may (depending on context) also be observed by others: coworkers, friends, rest of society

Preferences: intrinsic and extrinsic motivation

- Simple linear (or linearized) utility function

$$(v_a + v_y y)\, a - C(a) + e\bar{a}$$

- v_y: valuation for money, consumption, other "'extrinsic" incentives
- v_a: "intrinsic motivation"
  - Private-goods context: liking and motivation for the task (e.g., research), work ethic, perfectionism, company spirit, etc.
  - Public-goods context: degree of altruism / prosocial orientation

    ▷ Pure altruism: valuing others' benefits from increase in $\bar{a}$.
    Large groups: "Kantian"-type reasonings; overscaling one's impact

    ▷ Impure altruism: "joy of giving"

Preferences: intrinsic and extrinsic motivation

- Simple linear (or linearized) utility function

$$(v_a + v_y y)\, a - C(a) + e\bar{a}$$

- v_y: valuation for money, consumption, other "‘extrinsic" incentives

- v_a: "intrinsic motivation"
  - Private-goods context: liking and motivation for the task (e.g., research), work ethic, perfectionism, company spirit, etc.
  - Public-goods context: degree of altruism / prosocial orientation

    ▷ Pure altruism: valuing others' benefits from increase in $\bar{a}$.
    Large groups: "Kantian"-type reasonings; overscaling one's impact

    ▷ Impure altruism: "joy of giving"

- Public-goods case: derives benefit $e\bar{a}$ from supply of public good $\bar{a}$
  - Common e for all agents, but easy to allow heterogeneity
  - e for "externality"; set to 0 in private-goods context

Preferences: attributional motivations

- Individual's true type $v = (v_a\,,\ v_y)$ not observable by others
  Sometimes not even accessible to himself

- People care about how they / their “values” are perceived

Preferences: attributional motivations

- Individual's true type $v = (v_a\,,\ v_y)$ not observable by others
  Sometimes not even accessible to himself

- People care about how they / their "values" are perceived

- Desire, instrumental or / hedonic, for being seen as having a high v_a :
  - Private-goods context: career concerns make it valuable to be seen by employers as motivated for the activity or sector in question; strong work ethic, perfectionist, passionate, honest, etc.

    Applies if type signaled is 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.

Preferences: attributional motivations

- Individual's true type $v = (v_a\,,\ v_y)$ not observable by others
  Sometimes not even accessible to himself

- People care about how they / their "values" are perceived

- Desire, instrumental or / hedonic, for being seen as having a high v_a :
  - Private-goods context: career concerns make it valuable to be seen by employers as motivated for the activity or sector in question; strong work ethic, perfectionist, passionate, honest, etc.

    Applies if type signaled is 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 as poor / needy
  - More rarely, good to be seen as "hungry": easily controllable by monetary incentives

Reputational payoffs

- To people's "direct" motivations, we add

  $$\mu_a E(v_a|a,y) - \mu_y E(v_y|a,y)$$

  with $\mu_a > 0$ and μ_y usually ≥ 0, but can be < 0; will be 0 or irrelevant in many applications

- Simple linear (reduced) form here: will capture key effects

Reputational payoffs

- To people's "direct" motivations, we add

  $$\mu_a E(v_a|a, y) - \mu_y E(v_y|a, y)$$

  with $\mu_a > 0$ and μ_y usually ≥ 0, but can be < 0; will be 0 or irrelevant in many applications

- Simple linear (reduced) form here: will capture key effects

- When reputational payoffs are endogenized:
  - May involve nonlinear moments, e.g. $E(\pi(v_a)|a, y)$.
    Can just transform the distribution of v_a's
  - May depend nonlinearly on ex-post beliefs
  - Weights μ may depend on type v
  - Last two cases: will affect welfare analysis, not positive results

Social perceptions and self-perception

- We like to think of ourself as self-directed, honest, generous, good citizen, not greedy or venal, etc. May just be pleasant, or have instrumental value (help overcome temptations)

- We judge ourselves by our own actions, which define "who we are" Adam Smith's "impartial spectator within the breast", Montesquieu
  - Psychology, cognitive dissonance (Festinger 1957), self-perception theory (Bem 1972)
  - Deep-down, requires non-standard imperfection in self-knowledge: imperfect recall of / insights into our own motives
    Bénabou-Tirole (2005, 2007); Bodner-Prelec (2003)

Social perceptions and self-perception

- We like to think of ourself as self-directed, honest, generous, good citizen, not greedy or venal, etc. May just be pleasant, or have instrumental value (help overcome temptations)

- We judge ourselves by our own actions, which define "who we are" Adam Smith's "impartial spectator within the breast", Montesquieu
  - Psychology, cognitive dissonance (Festinger 1957), self-perception theory (Bem 1972)
  - Deep-down, requires non-standard imperfection in self-knowledge: imperfect recall of / insights into our own motives
    Bénabou-Tirole (2005, 2007); Bodner-Prelec (2003)

- But basic message is simple:
  - Self-image works just like social image: inferring (with some probability) one's "true values" or "identity" from one's conduct
  - Similarly, self-signaling works much like social signaling:

$$\mu_a E(v_a|a, y) - \mu_y E(v_y|a, y)$$

Summarizing: agents' preferences

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

intrinsic + extrinsic + (self) reputational motivations

- E is for "*expectation*", μ is for "*image*"
- People differ in image concerns as well as preferences over public and private goods. Most general type is $(v_a, v_y; \mu_a, \mu_y)$

Summarizing: agents' preferences

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

intrinsic + extrinsic + (self) reputational motivations

- E is for "*expectation*", μ is for "*image*"
- People differ in image concerns as well as preferences over public and private goods. Most general type is $(v_a, v_y; \mu_a, \mu_y)$
- Policy parameters of principal (employer, government, NGO...):
  - Material incentive y : reward, punishment, other extrinsic incentives
  - Publicity x : making actions more visible, memorable, etc...
    Amplifying $\mu \rightarrow x\mu$

Social planner and other principals

- Self-interested principal: most relevant in private-goods context, e.g. firm. Maximizes over y and / or x :

$$W(x, y) = (B - y)\bar{a}(x, y) - \varphi(x)$$

  - $\bar{a}(x, y)$: aggregate supply by agents, in equilibrium under policy (x, y)
  - B : principal's private benefit from agents' supply of a (e.g., effort)

- Benevolent social planner: most relevant in public-goods context, e.g. law. Given shadow cost of funds λ, maximizes over y and / or x :

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

  - $\bar{U}(x, y)$: agents' aggregate welfare, in equilibrium under policy (x, y)

Social planner and other principals

- Self-interested principal: most relevant in private-goods context, e.g. firm. Maximizes over y and / or x :

$$W(x,y) = (B-y)\bar{a}(x,y) - \varphi(x)$$

  - $\bar{a}(x,y)$: aggregate supply by agents, in equilibrium under policy (x,y)
  - B : principal's private benefit from agents' supply of a (e.g., effort)

- Benevolent social planner: most relevant in public-goods context, e.g. law. Given shadow cost of funds λ, maximizes over y and / or x :

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

  - $\bar{U}(x,y)$: agents' aggregate welfare, in equilibrium under policy (x,y)

- General case: weight $0 \leq \alpha \leq 1$ on agents' welfare; private benefit B

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

NGO, government agency. Can be reduced to planner's case

Information: key roles

- Idiosyncratic uncertainty:
  - Individual preferences $(v_a, v_y; \mu_a; \mu_y)$ privately known
  - May have noisy signals about the costs / benefits of their actions
  - Principal may have private information about some features of task (cost, returns), agent(s)' ability, or match to the job

Information: key roles

- Idiosyncratic uncertainty:
  - Individual preferences $(v_a, v_y; \mu_a; \mu_y)$ privately known
  - May have noisy signals about the costs / benefits of their actions
  - Principal may have private information about some features of task (cost, returns), agent(s)' ability, or match to the job
- Aggregate uncertainty:
  - Distribution of preferences in society $(v_a, v_y; \mu_a; \mu_y; e)$ may also be subject to aggregate shocks: changing values and norms, variable importance of reputational concerns, technology...
  - Principal may be better informed: observing a representative subsample's behavior, opinion surveys, measuring spillovers...
  - Agents may be better informed about recent societal shifts in preferences, norms

Revisiting incentives, in three steps

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

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

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: $e > 0$, $\mu_a = \mu > 0 = \mu_y$, $v_y \equiv 1$, $v_a = v \sim G(v)$; $\alpha = 1$, $\lambda \geq 0$
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

Intrinsic and Extrinsic Motivation

1. Evidence
2. The Game
3. Performance- incentives and the trust effect
4. Performance-incentives, transfers, and the profitability effect

- Main ref: Bénabou-Tirole RES (2003)

Payoffs

- Agent
  - Binary effort or contribution decision, $a = 0, 1$
  - Probability of success : ability θ
  - Expected gain $v = \theta V$ if effort, cost c of effort
  - Exerts effort if self-confident in his efficacy / finds task attractive
- Principal
  - Has vested interest W in agent's undertaking task and succeeding: parent, teacher, boss, colleague...
  - Expected gain $B = \theta W$ if effort
- No externalities on others / public goods here, nor reputational concerns: $e = 0 = \mu_a = \mu_y$

Information

In general,

- Principal may have more information about
  - Difficulty / attractiveness of current task
  - Long-run return, agent's ability
  - Interpretation of agent's past performances
- Agent may have more knowledge of
  - His previous efforts and performances
  - Past situational factors (facilitating / inhibiting)
- Emphasize here P's informational advantage, but A's information will also play a role

The looking-glass self

- Timing
  - Stage 1: P selects "policy": contingent reward or bonus b; flat payment m; delegation vs. monitoring; help...
  - Stage 2: A selects effort / no effort.

The looking-glass self

- Timing
  - Stage 1: P selects "policy": contingent reward or bonus b; flat payment m; delegation vs. monitoring; help...
  - Stage 2: A selects effort / no effort.
- Key: A tries to see through P's ulterior motivation.
- How can performance incentives reduce current or future effort? Conveying discouraging information to the agent, via either:
  - Trust effect: A's perception of his own incentives for effort
  - Profitability effect: value to P of policy differs across A types

Trust and profitability effects

- Trust effect: how confident is P in A's intrinsic motivation?
  - P's view of how A perceives task and his suitability to it:

    $$E_P\left[E_A\left[\theta V - c\right]\right]$$

  - If P pessimistic about A's motivation $\Rightarrow$ needs to give stronger incentives $\Rightarrow$ bad news
- Equilibrium: lower-powered incentives than under symmetric information, or even completely non-contingent pay
- Profitability effect: standard sorting condition / cross derivative
  - When, keeping A's effort constant, A's type t enters P's objective function in a way that would lead her to offer different policies p to different types of $A's$ under FI :

    $$\partial^2 E_P\left[\theta W - cost(p;\theta)\right]/\partial p \partial t$$

    e.g., conditionally on A's exerting effort, more profitable / less risky to delegate to, and monitor less, a smart agent (high θ).
  - Can also (need not) lead to weaker incentives.

Uncertain motivation: the trust effect

- Symmetric information. Agent exerts effort iff

$$\theta(V + b) \geq c$$

  - Intrinsic motivation (net): $v - c = \theta V - c$
  - Extrinsic motivation: $y = \theta b$, or just b
  - Reward is a positive reinforcer

Uncertain motivation: the trust effect

- Symmetric information. Agent exerts effort iff

$$\theta(V+b) \geq c$$

  - Intrinsic motivation (net): $v - c = \theta V - c$
  - Extrinsic motivation: $y = \theta b$, or just b
  - Reward is a positive reinforcer

- Asymmetric information about cost c (could also be V)
  - Principal knows c (perfectly, for simplicity)
  - Offers incentive b conditional on success;
    or: wage $y = \theta b$ conditional on effort
  - Agent has only a noisy signal $\sigma \in [0,1]$ of c; e.g., from talking to others, or own experience as he starts doing the task
  - Higher σ is "good news": MLRP

$$\forall\ \sigma_1 \text{ and } \sigma_2 \text{ with } \sigma_1 > \sigma_2, \quad \frac{g(\sigma_1|c)}{g(\sigma_2|c)} \text{ is decreasing in } c \qquad .$$

Strategies

- Agent's (interim) assessment of task difficulty,

$$\widehat{c}\left(\sigma, b\right) \equiv \mathrm{E}\left[c|\sigma, b\right]$$

is weakly decreasing in signal $\sigma \Rightarrow$

exerts effort, $a = 1$, iff signal exceeds threshold $\sigma^{*}(b)$ defined by:

$$\widehat{c}\left(\sigma^{*}(b), b\right) = \theta\left(V + b\right)$$

- No longer clean separation between intrinsic and extrinsic motivation!
- Principal, observing c, chooses b or $y = \theta b$ to

$$\max_{\mathrm{b}\geq 0}\{U_P \equiv E\left[\left(B - y\right)a\right]$$

$$= \theta\left[1 - G\left(\sigma^{*}(b)|c\right)\right]\left[W - b\right]\}$$

Proposition (hidden costs of incentives)

In any equilibrium,

1. *Offered rewards are positive (but weakened) short-term reinforcers:*

$$\text{if} \quad b_1 < b_2, \Rightarrow \sigma^*(b_1) > \sigma^*(b_2).$$

2. *Rewards are bad news: if b_1 is offered when task difficulty is c_1 and b_2 when it is c_2,*

$$\text{if} \quad c_1 < c_2, \Rightarrow b_1 \leq b_2.$$

3. *Rewards undermine agent's motivation for the task:* $\forall \ (\sigma_1, \sigma_2)$ *and all equilibrium rewards* $b_1 < b_2$,

$$\mathrm{E}\left[c|\sigma_1, b_1\right] < \mathrm{E}\left[c|\sigma_2, b_2\right].$$

Future motivation is also reduced by higher incentives, even after agent's action $a = 0, 1$ and outcome $\omega = S, F$ are realized:

$$\forall(\sigma, a, \omega), \quad \mathrm{E}\left[c|\sigma, b, a, \omega\right] \quad \text{is} \searrow \text{in } b.$$

Proof

1. Otherwise, principal could get more effort by offering less

   NB: in experimental outcomes, no optimization

2. Revealed preference argument.

   Let b_i be an optimal bonus when principal has information c_i, $i = 1, 2$. Denote $\sigma_i = \sigma^*(b_i)$. Since b_i is optimal given c_i,

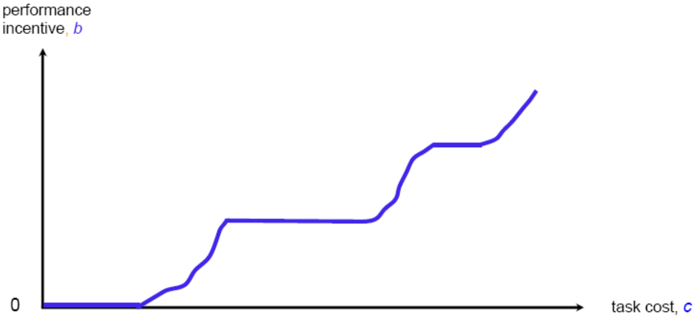
$$\theta \left[1 - G\left(\sigma_i \mid c_i\right)\right] \left[W - b_i\right] \geq \theta \left[1 - G\left(\sigma_j \mid c_i\right)\right] \left[W - b_j\right] \Rightarrow$$

$$\frac{1 - G\left(\sigma_1 \mid c_1\right)}{1 - G\left(\sigma_2 \mid c_1\right)} \geq \frac{W - b_2}{W - b_1} \geq \frac{1 - G\left(\sigma_1 \mid c_2\right)}{1 - G\left(\sigma_2 \mid c_2\right)}$$

   - Since $c_2 > c_1$, MLRP $\Rightarrow$ $\sigma_1 \geq \sigma_2$: higher σ's more likely under c_1
   - Hence $b_1 \leq b_2$, since $\sigma^*(\cdot)$ is decreasing.

   $\Rightarrow$ Pooling thus occur only over intervals. Graph.

3. If principal offers b_1 to types $[\underline{c}_1, \overline{c}_1]$ and $b_2 > b_1$ to types $[\underline{c}_2, \overline{c}_2]$, it must be that $\overline{c}_1 \leq \underline{c}_2$.


performance incentive, b
0
task cost, c

Relation to the literature

- Many classical references on the hidden costs of rewards (Lepper et al. 1973, Deci 1975, Deci-Ryan 1985) emphasize their informational impact
  - "Every reward (including feedback) has two aspects, a controlling aspect and an informational aspect which provides the recipient with information about his competence and self-determination."

    (Deci 1975)

- Also stress distinction between engagement and re-engagement effects
  - "Reinforcement has two effects. First, predictably it gains control of [an] activity, increasing its frequency. Second,...when reinforcement is later withdrawn, people engage in the activity even less than they did before reinforcement was introduced."

    Lepper et al. (1973)

Conditions for incentives to reduce I.M.

1. Principal must have private information. Here, about the agent, the task, or the match between the two
    - May explain why performance incentives are more controversial in educational settings than in the workplace

Conditions for incentives to reduce I.M.

1. Principal must have private information. Here, about the agent, the task, or the match between the two
   - May explain why performance incentives are more controversial in educational settings than in the workplace
2. Principal's sorting condition must go "in the right direction": here, make her more inclined to offer performance incentives to less able agent, or for less rewarding task.
   - Seen cases where it does. Examples going the other way:
   - "Empowerment". Manager promoted from fixed-salary job, given leadership of new project or division + pay-for-performance scheme. Contingent reward associated here with high level of trust.
   - Task subject to learning: trying will reveal c or V prior to completion or repetition. Offering reward says "I am confident you will like this / are good at it, and will come to agree. Encouraging you to try would make no sense for me otherwise"

Retrospective justification and self-perception.

- “Why am I doing this?”

Retrospective justification and self-perception.

- “Why am I doing this?” Someone engaged in writing a book, proving a theorem, running a marathon, etc. may, at times, be seized by doubt as to whether the intellectual and ego benefits from successful completion will, ultimately, justify current efforts

- Psychological literature on ‘insufficient justification” effect and “escalating commitments” suggest more likely to persevere if undertook task under low extrinsic incentives (or high costs)

  (Festinger and Carlsmith, 1959, Bem 1967, Staw 1977)

Retrospective justification and self-perception

- Combine earlier model with imperfect memory. Suppose that
  - Agent faces choice of whether to undertake, or persevere in, a long-term project similar to others he completed previously
  - Knows that he chose to engage in it previously, and any extrinsic incentives that apply, but does not recall his intrinsic interest in the task / overall enjoyment of its completion V

Retrospective justification and self-perception

- Combine earlier model with imperfect memory. Suppose that
  - Agent faces choice of whether to undertake, or persevere in, a long-term project similar to others he completed previously
  - Knows that he chose to engage in it previously, and any extrinsic incentives that apply, but does not recall his intrinsic interest in the task / overall enjoyment of its completion V

- Earlier result on c applies equally to agent's long-term payoff V

$$\mathrm{E}\left[V \mid b\right] > \mathrm{E}\left[V \mid b'\right] \qquad \text{for} \quad b < b'$$

- Agent will reflect that since he embarked on the project in spite of low extrinsic (financial, career) incentives, the personal enjoyment from previous completions (which, at this later and stressful stage, he cannot quite recall) must have been high.

- Hence, it is likely to be worth persevering on the chosen path.

Other Implications

- Forbidden fruits
  - Saw that high-powered incentives can reduce intrinsic motivation
  - Conversely, “forbidden fruits” are the most appealing. Optimal bonus can be zero, or even negative. “Tom Sawyer” effect.
- Improper causal attributions
  - Probability of effort, $1 - G(\sigma^*(b) \mid c)$ and probability of success, $\theta\left[1 - G(\sigma^*(b)| c)\right]$ both decreasing in c, known to P only
  - c covaries positively with b in equilibrium $\Rightarrow$ observer who just correlates b with outcomes may incorrectly conclude that incentives are negative reinforcers even in the short run

Immediate adverse effects

- Information conveyed by incentives spills over to correlated tasks (e.g., math & science homework). Especially if bears on θ
- Spillovers across states when performance measurement is imperfect
  - Same model, principal uses incentives, but effectiveness of monitoring fluctuates randomly
  - Agent learns, before making his decision, whether he is likely to be caught if he misbehaves, or to escape detection

    $\Rightarrow$ Threat of punishment has positive (short term) reinforcement effect in instances when agent knows that monitoring is effective, but only a negative one when thinks he can "get away with it"
  - Teenager's heightened temptation to violate parents' strict prohibition on drinking, smoking, etc., in situations where they cannot catch him

Profitability effect

- Principal now has private information about agent's ability θ, which affects U_P, rather than task cost c, which does not
- Agent has imperfect signal $\sigma \in [0, 1]$ of θ, MLRP;
  c and V are common knowledge
- Agent's effort unobservable to $P \Rightarrow$ conditions bonus b on successful performance
- Two steps:
  - Restrict attention to contracts without unconditional transfers (lump-sum payments) in either direction. Just trust effect again
  - Allow lump-sum payments. Profitability effect

Profitability effect

- Principal now has private information about agent's ability θ, which affects U_P, rather than task cost c, which does not
- Agent has imperfect signal $\sigma \in [0,1]$ of θ, MLRP;

  c and V are common knowledge
- Agent's effort unobservable to $P \Rightarrow$ conditions bonus b on successful performance
- Two steps:
  - Restrict attention to contracts without unconditional transfers (lump-sum payments) in either direction. Just trust effect again
  - Allow lump-sum payments. Profitability effect

Proposition (unkown ability, no lump-sums)

All results obtained for unknown task difficulty c apply (with appropriate changes in notation / terminology) when the P's private information and A's noisy signal bear instead on the agent's probability of success θ.

- To study richer contracts, specialize model:
  - θ takes values θ_H (prob. f_H) or θ_L (prob. f_L), with $\theta_H > \theta$
  - Contingent rewards cannot be negative: $b \geq 0$
- Minimum effort-inducing bonus when agent knows his ability:

$$b_k^* \equiv \max\{0, c/\theta_k - V\}, \quad \text{for} \quad k = L, H$$

Assume $0 = b_H^* < b_L^* < W$; normalize reservation $\bar{U} = 0$

- To study richer contracts, specialize model:
  - θ takes values θ_H (prob. f_H) or θ_L (prob. f_L), with $\theta_H > \theta$
  - Contingent rewards cannot be negative: $b \geq 0$

- Minimum effort-inducing bonus when agent knows his ability:

$$b_k^* \equiv \max\left\{0, c/\theta_k - V\right\}, \quad \text{for} \quad k = L, H$$

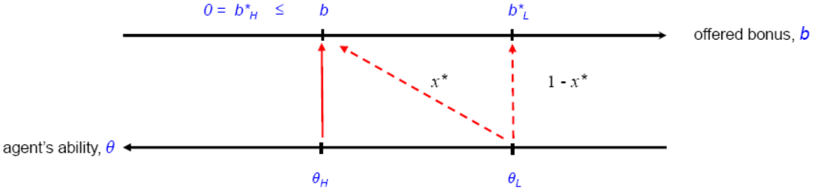
Assume $0 = b_H^* < b_L^* < W$; normalize reservation $\bar{U} = 0$

Proposition (two-type case, no lump-sums)

1. *In any equilibrium, principal offers a low bonus $b < b_L^*$ to a more able agent (θ_H) and randomizes between bonuses b and b_L^* when dealing with a less able agent (θ_L).*

2. *There is a unique D1 / NWBR equilibrium, and it has $b = 0$. The probability $x^* > 0$ of pooling (offering $b = 0$ to θ_L), and the unconditional probability of no reward, $f_H + f_L x^*$, both increase with agent's initial self-confidence, f_H.*

Two-type case without lump-sum payments



- Message:
  - Trust effect forces principal to adopt low-powered incentives
  - The more so, the more self-confident the agent
    (even when he is actually a low type)

Two-type case with lump-sum payments

- So far, ruled out unconditional transfers
  - Any equilibrium outcome absent lump-sum payments is still an equilibrium when they are allowed (sustained by OEB's that transfers convey no information)
  - May not be feasible: one party has no cash, or limited liability
  - Tend to attract undesirable (lazy) types: adverse selection

Two-type case with lump-sum payments

- So far, ruled out unconditional transfers
  - Any equilibrium outcome absent lump-sum payments is still an equilibrium when they are allowed (sustained by OEB's that transfers convey no information)
  - May not be feasible: one party has no cash, or limited liability
  - Tend to attract undesirable (lazy) types: adverse selection

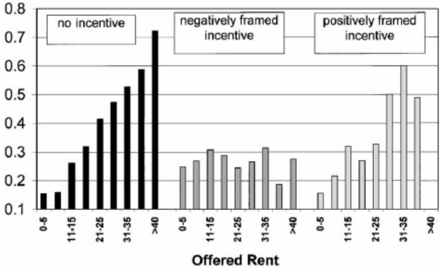
- Sometimes, more general contracts feasible: P can propose up-front payment $m \gtrless 0$, together with bonus $b \geq 0$ for success ⏩

Two-type case with lump-sum payments

- So far, ruled out unconditional transfers
  - Any equilibrium outcome absent lump-sum payments is still an equilibrium when they are allowed (sustained by OEB's that transfers convey no information)
  - May not be feasible: one party has no cash, or limited liability
  - Tend to attract undesirable (lazy) types: adverse selection
- Sometimes, more general contracts feasible: P can propose up-front payment $m \gtrless 0$, together with bonus $b \geq 0$ for success
- Symmetric information: m only enables P to tax (high-ability) agent's rents from the activity
- Private information: P can use m to signal confidence in agent
  - Similar to "burning money" (or time, e.g., pep talks): offer just enough to wipe out expected profits from inducing low-ability agent to work, leaving positive surplus for P when high-ability agent persuaded to work

Incentives and effort: Fehr and Gächter (2002)

- No Incentives - baseline:
  - "Employer" makes contract offer: $p =$ non-contingent payment, $\hat{a} =$ desired effort or quality, non-binding. Offered rent: $\hat{U}_A = p - C(\hat{a})$ Payoff $U_P = Wa - p$ if contract accepted
  - Agent chooses effort a, at some convex cost $C(a)$. Payoff $U_A = p - C(e)$
- Incentives: P can choose a "wage deduction" (fine) $0 \leq f \leq \bar{f}$ that will be imposed if A found to be shirking, $e < \hat{e}$; verification occurs with prob. $1/3$
- Incentives, positively frame: same, but contingent payment framed as "bonus" $0 \leq b \leq f$, to be paid only if verification shows $e \geq \hat{e}$



- Cdf's of agent's signal with ability $\theta_H, \theta_L : G_H(\sigma) < G_L(\sigma)$
- Assume a "limited-informativeness" condition

$$\theta_H G_H(\sigma) > \theta_L G_L(\sigma)\,, \text{ for all } \sigma > 0$$

  - Signal's distribution does not vary too much with underlying state

Proposition (unknown ability, lump-sum transfers)

There is a unique PBE satisfying Cho-Kreps' intuitive criterion, and it is separating.

1. *Agent of ability θ_k, $k = L, H$, is offered contract (m_k, b_k), with*

$$b_L = b_L^* = c/\theta_L - V,\ b_H = b_H^* = 0,$$
$$m_L = 0 < c - \theta_L V = m_H.$$

2. *Principal's and agent's expected utilities are*

$$\theta = \theta_L : \qquad U_P^L \equiv \theta_L(V + W) - c, \qquad U_A^L = 0$$
$$\theta = \theta_H : \qquad U_P^H \equiv \theta_L V + \theta_H W - c, \qquad U_A^L = (\theta_H - \theta_L)V$$

Implications

- More general class of contracts: remains the case that, in equilibrium, a more high-powered incentive scheme (higher b, lower m):
  - Is a positive reinforcer in the short-run: leads θ_L agent to exert effort, would otherwise not have done so
  - Is bad news for the agent, permanently damaging his motivation / self-confidence, even if task succeeds. Also lowers utility.

    ▷ Torre contract: incentives b vs. fixed a.

Implications

- More general class of contracts: remains the case that, in equilibrium, a more high-powered incentive scheme (higher b, lower m):
  - Is a positive reinforcer in the short-run: leads θ_L agent to exert effort, would otherwise not have done so
  - Is bad news for the agent, permanently damaging his motivation / self-confidence, even if task succeeds. Also lowers utility.

    ▷ Torre contract: incentives b vs. fixed a.

- Highlights workings of the profitability effect that comes into play with lump-sum payments:
  - $(\partial U_P / \partial a) / (\partial U_P / \partial b) = (W - b) / a$ independent of θ, due to multiplicative form of expected output, but
  - $-(\partial U_P / \partial a) / (\partial U_P / \partial m) = (W - b) / \theta$ is increasing in θ : lump-sum transfer is an investment (in signaling) that has higher rate of return when agent is talented

- Differences with no-lump-sum-transfer case:
  - Equilibrium fully separating $\Rightarrow$ informed-principal game leads to no distortion of incentives; same slopes
  - Because of the fixed wage $m_H > 0$, agent's utility is higher than under symmetric information. Rents.
- Similarities:
  - The general weakening of performance-based compensation, which is model's main insight, takes the form of a lower share of contingent compensation in total compensation.
  - Distribution of earnings in population of agents is again more equal (Lorenz) under AI, due to the motivation-management problem

Similar aspects of the looking-glass self

- Delegation: may signal trust, though profitability condition: "I put myself in your hands"

- Help: may signal trust ("I believe in you / this project") or lack thereof ("you are in trouble")
  - Depends on shape of principal's payoff (upside or downside more important)
  - Psychology literature on overhelping and dependency

Other applications and extensions

- Once a reward is offered, it will be required –and "expected"– every time task has to be performed –perhaps even in increasing amounts. Ratchet effect.

- Suvorov (2002): addiction to rewards
  Two-period extension of model. Additional effect: agent now has a strategic incentive to appear demotivated (having a low σ), in order to be given a higher bonus in the future.

- Three monotonicity results:
  - In each period, low type is offered (weakly) higher bonus
  - For each type, bonus is (weakly) increasing over time
  - For each type, initial bonus is lower when P is same in both periods than with two different principals. Long-lasting principal internalizes the fact that rewards are habit-forming

- Explains people's (e.g., parents') reluctance to offer rewards, even when small price to pay to get the current job done

Other applications and Extensions

- Suvorov-Van de Ven (2006): ex post, non-contingent bonuses can be good news ("*P* liked my work") and boost intrinsic motivation

  ▷ Schwartz letter

- Herold (2005): multitasking ⇒ lack of incentives on one task mays signal *P*'s trust and boost incentives in another, uncontrolled task.

- Ellingsen-Johannesson (2009): Players care about being esteemed. Applications to trust game, Falk-Kosfeld (2006), gift exchange.

"The Hidden Costs of Control": Falk-Kosfeld (AER 2006)

- Agent chooses action $x \in \{0, 1, ...120\}$, resulting in payoffs

  $U_A = 120 - x$ for him and $U_P = 2x$ for Principal

- Prior to choice of x, principal can choose to either
  - constrain agent's choice to $x \geq \underline{x}$,

    $\underline{x}$ a fixed lower bound $= 5, 10$, or 20
  - leave it unconstrained

- 804 subjects. Use strategy method (elicit all conditional choices of A's)

“The Hidden Costs of Control”: Falk-Kosfeld (AER 2006)

- Agent chooses action $x \in \{0, 1, ...120\}$, resulting in payoffs

  $U_A = 120 - x$ for him and $U_P = 2x$ for Principal

- Prior to choice of x, principal can choose to either
  - constrain agent's choice to $x \geq \underline{x}$,

    $\underline{x}$ a fixed lower bound $= 5, 10$, or 20
  - leave it unconstrained

- 804 subjects. Use strategy method (elicit all conditional choices of A's)

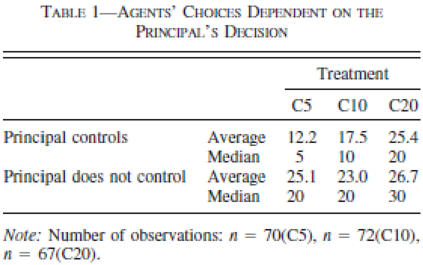


TABLE 1—AGENTS' CHOICES DEPENDENT ON THE PRINCIPAL'S DECISION

		Treatment		
		C5	C10	C20
Principal controls	Average	12.2	17.5	25.4
	Median	5	10	20
Principal does not control	Average	25.1	23.0	26.7
	Median	20	20	30

Note: Number of observations: n = 70(C5), n = 72(C10), n = 67(C20).

- Agents' choices

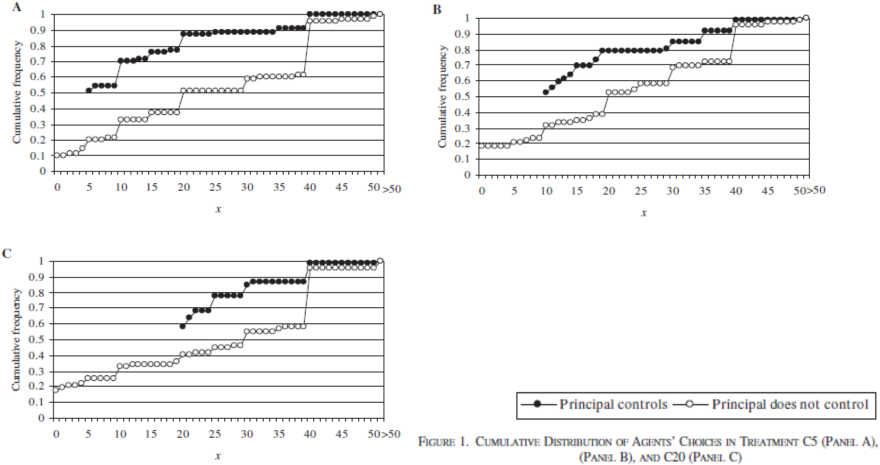


FIGURE 1. CUMULATIVE DISTRIBUTION OF AGENTS' CHOICES IN TREATMENT C5 (PANEL A), C10 (PANEL B), AND C20 (PANEL C)

Other applications and extensions

- Once a reward is offered, it will be required –and "expected"– every time task has to be performed –perhaps even in increasing amounts. Ratchet effect.

- Suvorov (2002): addiction to rewards
  Two-period extension of model. Additional effect: agent now has a strategic incentive to appear demotivated (having a low σ), in order to be given a higher bonus in the future.

- Three monotonicity results:
  - In each period, low type is offered (weakly) higher bonus
  - For each type, bonus is (weakly) increasing over time
  - For each type, initial bonus is lower when P is same in both periods than with two different principals. Long-lasting principal internalizes the fact that rewards are habit-forming

- Explains people's (e.g., parents') reluctance to offer rewards, even when small price to pay to get the current job done

- Principals’ choices

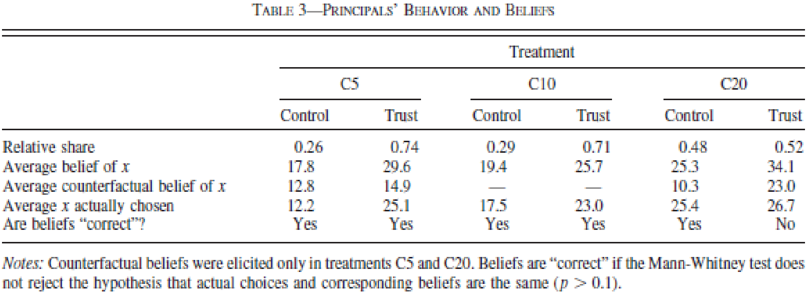


TABLE 3—PRINCIPALS’ BEHAVIOR AND BELIEFS

	Treatment					
	C5		C10		C20	
	Control	Trust	Control	Trust	Control	Trust
Relative share	0.26	0.74	0.29	0.71	0.48	0.52
Average belief of x	17.8	29.6	19.4	25.7	25.3	34.1
Average counterfactual belief of x	12.8	14.9	—	—	10.3	23.0
Average x actually chosen	12.2	25.1	17.5	23.0	25.4	26.7
Are beliefs “correct”?	Yes	Yes	Yes	Yes	Yes	No

Notes: Counterfactual beliefs were elicited only in treatments C5 and C20. Beliefs are “correct” if the Mann-Whitney test does not reject the hypothesis that actual choices and corresponding beliefs are the same ($p > 0.1$).

  - Most choose to forego control
  - For those who do not, “self-fulfilling prophecy of distrust”

Summary of Lecture I

- Gave precise content to "intrinsic motivation" and idea that it may be undermined or "crowded out" by incentives

- Mechanism identified here requires
    - Principal has private information
      Applicable to some settings / instances of crowding out.
      For others, will need a different explanation
    - "Trust effect" or profitability effect" generating sorting condition in the appropriate direction
      Will depend on the structure of payoffs and signals.

- Low-powered incentives and unconditional payments / sacrifices are two ways in which P's confidence-management motive can be reflected in equilibrium contracts
    - Each with its own domain of applicability
    - But similar effects on wage inequality and long-run motivation

Broadening the picture

- Private information of the principal could also be her own preferences or beliefs
  - Altruism toward A
  - Trusting or untrusting priors

- Agent caring about it for non-instrumental, affective reasons
  - Reciprocal altruism
  - Desire to be perceived well (signal to) certain types of principals, or to other agents

    (Ellingsen-Johannesson 2009)

Broadening the picture

- Private information of the principal could also be her own preferences or beliefs
  - Altruism toward *A*
  - Trusting or untrusting priors
- Agent caring about it for non-instrumental, affective reasons
  - Reciprocal altruism
  - Desire to be perceived well (signal to) certain types of principals, or to other agents
    (Ellingsen-Johannesson 2009)
- Unknown feature of an activity when it will be observable to others may be the social norms governing it.
- First, understand norms, and how interact with material incentives
  - Turn (self) image motivation back on
  - Private information now on agents' side. Signaling