

MORAL REASONING, MARKETS AND ORGANIZATIONS

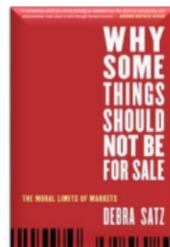
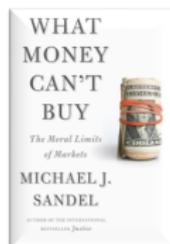
ISEO Summer School

June 21, 2017

Jean Tirole

I. INTRODUCTION MARKETS AND MORALITY

- Most economists spend much of their professional life analyzing market failures, but view markets as essential for a good functioning of society.
- Other social scientists, part of civil society, most religions see it differently, arguing that
 - economists fail to draw a clear line between what has a price and what has dignity (Kant), between the profane and the sacred (Durkheim)
 - markets are a threat to social cohesion.
- Philosophers' different views; e.g.



1. THE MARKET FOR VIRTUOUS INDIGNATION

Both sides of the political spectrum

1.1 Negative sentiments towards economics

- Noxious/ repugnant markets and complaints about economics' Weltanschauung
Sandel: A wide range of goods and services, including **babies for adoption, surrogate motherhood, sexuality, drugs, military service, votes, and organs for transplantation**, are not to be commoditized through markets, no more than **friendship, admissions to elite universities or Nobel Prizes** are to be bought, or **genes and other life forms** to be patented.
- Pope's encyclical Laudato Si (2015) and Sandel about climate change
Pope: "**the environment cannot be "safeguarded or promoted by market forces."** Confronting the climate crisis will require a deeper, spiritual transformation of society, replacing "consumption with sacrifice, greed with generosity, wastefulness with a spirit of sharing."

- Rationing vs. market clearing

Common preference among non-economists for allocations made by lotteries, age, non-elected boards, on first-come-first-served basis; for non-tradable vouchers for food or housing

[Allocating water bottles to hikers at a hilltop on a hot day; Hurricane Katrina and gas, food, and hotel rates; regulators of electricity everywhere feel a strong pressure to resist peak-load pricing and to cap prices.]

Market power/ price gouging argument or dislike of markets?

1.2 The equation economics = markets = laissez-faire

Not exactly...

1.3 The equation economists = selfish, calculating individuals

- Reacting to incentives (financial, career concerns, status...)
- Is economics “performative”, does it create its own reality?
- Always considering trade-offs \Rightarrow losing perspective
 - Safety decisions
[Compare a few child deaths associated with airbags with a larger number of adult lives saved.]
 - Health care and life
[Trade-offs in health-related decisions clashes with generally held views about the sacred character of every human life. Life, as we know, is as priceless as are family, marriage, friendship, loyalty to one's country, democracy, equality or graveyards.]

1.4 Economists: utilitarianism and consequentialism

- Emphasis on (potential) consequences of act (Bentham, Mill tradition)
Little appetite for deontology / duty-based approach (Kant's liar example)
- Willingness to entertain trade-offs (e.g. trolley dilemma; current form: driverless car's software's decision making)

1.5 The economists' difficult communication

Difficulties shared with other sciences

- Current distrust of experts
- Grand claims vs. trade-offs: media communication is not conducive to development of background knowledge
- Economics complex yet familiar: lay persons have opinions.

But also...

- Economists as bearers of bad news (shooting the messenger): economic analysis exposes our deep values.
 - Society wants to believe in its humanity, in its members' ability to behave altruistically, to experience empathy, in the sustainability of the welfare state, in green growth... Such beliefs are reassuring.
 - In the moral domain, motivating law and policy by overarching ethical goals such as fairness and equity avoids a confrontation with tell-tale signs that our morality is not necessarily what we strive to believe it is.
- Economists are bad at predicting, better at identifying factors [Imperfect theories, shortage of data, self-fulfilling phenomena, behavioral factors...]
- Direct vs. indirect effects, visible vs. invisible victims.

2. MODERN ECONOMICS AND MARKETS

Limits to free exchange between consenting adults: Market failures

2.1 Externalities

Repugnant markets applications

- Environmental externalities
- Babies for adoption (correlation WTP and future love for child?)
- Child labor
- Market for diamonds and civil wars
- Market for votes: vote seller and buyer exert an externality on others.

Image externalities

- Dwarf-tossing: Fully consenting and remunerated dwarfs wearing special padded clothing are thrown upon a mattress or a coated wall.
- Market for women's reproductive labor, prostitution.

Linked with demand for dignity: Hard to compensate disabled with large amounts of money if otherwise marginalized. Social inclusion.

2.2 Imperfect markets

- **Market power** (motivates antitrust/public utility regulation)
Repugnant markets applications: Price gouging, contracts written under duress.
- **Asymmetric information**: banking regulation, shrouded attributes. **Asymmetric vulnerability**: professional vs. layman.
Repugnant markets applications: Lack of understanding of LT consequences (organ sales, contract pregnancy and bonding with child...)
- **Asymmetric information**: signaling (adverse selection traps), certification, etc.
Repugnant markets applications: No commodification of friendship, admission in universities, Nobel prize, love.

2.3 Breakdown of insurance: unequal societies

- Insurance market, genetic testing
- Inequality (behind veil of ignorance standard?)
“Across situations”: from tolerance (better hospitals, better schools, safer cars for the rich) to negative attitudes towards wealth-based consumption (limited water, allocation of resources in earthquake, lifeboats on Titanic. . .)
- Wishful thinking: don't want to see we are in an unequal society: organ sales, prostitution, etc. Desperation.
- Titanic choices: would we worry about choice of not having a lifeboat if equal purchasing power?

2.4 Internalities

- **Failure to pursue self-interest:** Self-control

[Mandated cooling-off periods; public interventions in the matter of cigarettes, drugs, addictions, gambling, savings or excessive interest rates (usury) are in part motivated by self-control concerns]

Repugnant markets applications

- Voluntary slavery
 - Organ sales: poor people overweigh ST benefit; long-lasting cost (different from blood sale)
- **Internalities and externalities:** doping in sports.

2.5 Perverse effects of incentives

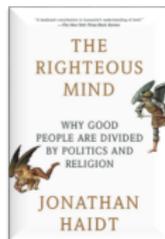
- Not only standard theories on limits to incentives
[Noisy measurement, teamwork, multitasking, collusion, capture, adverse selection...]
- Crowding out theories and empirical tests
[informed principal, overjustification]

3. THINKING ABOUT MORALITY

3.1 What's moral or not?

Haidt's classification

- *“Do no harm”* (externality-based notion)
[Economists' view? Not only: old tradition in moral philosophy: Arendt's moral compass: “Don't do unto others what you don't want done to yourself”; presumably assumes similar preferences...; Turiel 1983: morality = not harming other]
- *Less universal values* (emotion/ belief based, less stable?)
 - Based on disgust, tradition, etiquette, respect for authority, group loyalty;
 - Condemn victimless acts: life insurance, gay marriage,...
 - Banfield (1958)'s distinction between limited morality (clan-based: narrow circle of friends and family) and generalized morality. Gemeinschaft (community) vs. Gesellschaft (civil society)



3.2 Moral postures: feelings of revulsion are an unreliable source of ethical inspiration

Heuristics/ warning signal, but should not go beyond

- **Moral assertions can override freedom of others:** sexual acts between persons of same sex or different races.
- **Markets exist** whether we want them or not: prostitution, organ markets, surrogate motherhood. . .

Issue is to regulate/ prevent them if we so decide, not to assume them away:

- Ban on public executions in France from 1939 through 1981, on corporal punishment.
- Moving prostitution elsewhere
- **Expressive law.**

Markets, morality and social cohesion

Are markets a threat to social cohesion?

- Markets make relationships anonymous (that is the purpose! Emancipation from monopoly and political power); they encourage citizens to distance themselves from traditional institutions.
Yet (a) need for trust; (b) *doux commerce* (Montesquieu): trade as a factor of peace.
- Other factors: urbanization, increased mobility, online communications.

Counterfactual to market?

- Migration toward untaxed, unregulated underground markets?
- Administered: corruption/favoritism/ queueing

Counterfactual to globalization?

- Protectionism

4. EPILOGUE

- Yes, economists are concerned about morality
- Indignation is cheap; only in-depth analysis will do
- Current empirical work sheds much light on moral wriggle room, the role of narratives, determinants of social norms. . . And theory helps organize our thinking.
- Still a long way to go.

II. NARRATIVES, IMPERATIVES AND MORAL REASONING

Joint work with Armin Falk and Roland Bénabou.

- Goal: Understand determinants of moral behavior, and its key feature of malleability.
Drivers: Institutions, social context. Availability of narratives.
- This paper:
 - Starts with (utilitarian) workhorse model that accounts for many regularities on role of self and social image in moral behavior.
 - Reflects upon deontological (Kantian) behaviors, with an application to experimental methods.
 - Analyzes role of narratives, and their horse race with imperatives as means of moral influence.
 - Studies virality of narratives.

WHAT IS A NARRATIVE?

We are agnostic: *any story, life experience or heuristic that has the potential to alter an agent's beliefs, and therefore actions.*

“Beliefs”: reflect whatever surfaces in the agent's mind when acting.

- Retrieval process is selective \Rightarrow vivid life experiences, simple and striking arguments, opinions of others and contextual cues particularly likely to influence beliefs.
- Story may not have any truth, just a *perceived* grain of truth
 - some of the most successful narratives are wrong
 - narrative can be correct, but incomplete and therefore potentially misleading if one jumps to the conclusion
Example: confusion correlation-causality.
 - “believability is the hallmark of the well-formed narrative” (J. Bruner).

Veracity does not matter for *positive* analysis.

- “Story, life experience or heuristic” can be fortuitous, provided by a narrative entrepreneur or someone one is connected to, or else thought of by the agent herself.
 - Tension between immoral behavior and negative personal (self and social) image
 - ⇒ negative narratives serve as excuses, reduce conflict between immoral action and personal image.
 - Narratives play a role in interpersonal everyday interaction...
 - attribution of one’s immorality to external situation (time constraints, missing information).
- ...but also in long-term societal developments
- cruel persecution based on stigmatization of ethnical or political minorities.

NARROW VIEW OF NARRATIVES

(1) Narratives here are *rationales* or *arguments* for behavior.

[We formalize them as signals about the importance of the externality, or cost of behaving morally.]

Narratives may also supply *meaning* (need to have a purpose in one's life, understanding of one's environment).

- Karlsson-Loewenstein-McCafferty (2004) "The Economics of Meaning", Chater-Loewenstein 2016: Narratives, "life stories" provide sense-making: one's preferences & events & life, social and temporal identity.
- Presumably serves some desire for predictability (hedonic/anxiety reducing, or instrumental/helpful for planning; probably both), or reflects evolutionarily wired-in program for pattern seeking.

(2) We are interested only in those narratives that have a moral dimension (affect an agent's feeling of duty and personal image). Not in "They are not making any more land" (Shiller's RE bubble).

1. BASIC MODEL

Builds on Bénabou-Tirole (*AER* 2006, *QJE* 2011, *Laws and Norms*)

- An individual may engage in moral/pro-social behavior ($a = 1$) or not ($a = 0$).

Cost c , perceived as c/β where $\beta \leq 1$ is self-control (hyperbolicity) parameter.

- Individual has deep value $\begin{cases} v_H & \text{with prob. } \rho \text{ (high/moral type)} \\ v_L & \text{with prob. } 1 - \rho \text{ (low/immoral type)} \end{cases}$

Mean: $\bar{v} \equiv \rho v_H + (1 - \rho)v_L$.

- Perceived magnitude of externality: e . Externality is 0 or 1; $e =$ probability that externality is 1 (alternatively could have $e \in \mathbb{R}^+$).
- Consequentialism: intrinsic motivation ve .

For the moment, we will pick parameters such that low type never contributes.

► Parameters

Question: Does high type contribute?

Self or social reputation motive: $\mu\hat{v}$

where \hat{v} = perception of deep value

μ = intensity of (self or social) image concerns.

Payoffs

$$\left(v_{He} - \frac{c}{\beta}\right)a + \mu\hat{v}(a) \quad \text{for moral/high type}$$

and

$$\left(v_{Le} - \frac{c}{\beta}\right)a + \mu\hat{v}(a) \quad \text{for immoral/low type}$$

Equilibrium: Unique or if not, select Pareto-dominant one.

$$a_H = 1 \Leftrightarrow e > e^* \quad \text{where} \quad v_H e^* - \left(\frac{c}{\beta} \right) + \mu(v_H - \bar{v}) = 0.$$

Unethical behavior is encouraged by

- low image concern (low μ)
- low self-control (low β)
- high cost of moral behavior (high c)
- low perceived social benefit from pro-social behavior (low e)
- good initial reputation (high \bar{v}): moral licensing

[this conclusion is less robust: theoretically reputational incentive is in general hump-shaped in \bar{v} : BT on identity *QJE* 2011; empirically: moral licensing vs cleansing]

► Evidence

SEEKING OR AVOIDING MORAL CHOICES

Commitment to $a = 0$ or take a test of morality?

Ex-post self

$$a_H = 1 \quad \text{iff} \quad v_{He} - \left(\frac{c}{\beta}\right) + \mu(v_H - \bar{v}) > 0$$

Ex-ante self (assume that type is unknown)

$$a_H = 1 \quad \text{iff} \quad v_{He} - c > 0 \quad (\text{since expected payoff is } E[(ve - c)a] + \mu\bar{v}).$$

Two wedges:

- self-control problem acts against moral behavior
- image concern (a “zero-sum game”, a “positional good”) promotes moral behavior, perhaps too much from point of view of individual.

If $(c/\beta) - c < \mu(v_H - \bar{v})$, avoids the ask, over a range of parameters.
Conversely, may seek out the ask.

2. THE POPULARITY OF DEONTOLOGICAL KANTIANS

- Our agents are assumed to be utilitarians/consequentialists.
- Evidence shows that in practice they may refuse outright to consider moral trade-offs. Does it mean that they are deontologists?
- More complex than it appears: it may be that some “deontologists” are just posturing. This is related to the idea that morality is malleable/not immutable (we cannot divide the world into “utilitarians”, “deontologists”...)
- Implications for the measurement of prosocial propensities.

▶ Taboos and posturing

Reflexions about the multiple price list/strategy method

[Becker-DeGroot-Marschak BDM]

Minimal level of reward γ that agent is willing to take for picking $a = 0$, knowing that actual reward c will be drawn from $f(c)$?

Low type's loss:

$$L(\gamma) \equiv \int_{\beta v_{Le}}^{\gamma} \left(\frac{c}{\beta} - v_{Le} \right) f(c) dc$$

BDM may lead to pure Kantian behavior by otherwise fully consequentialist agents.

If $L(+\infty) \leq \mu(\bar{v} - v_L)$, then $\gamma_H = \gamma_L = +\infty$

[Everett et al 2016: people who make deontological judgments in moral dilemmas preferred as social partners.]

COMPARISON BDM-DE (direct elicitation)

Direct elicitation

Agent is offered a reward or bribe c to take immoral action.

Threshold

$$v_{He} - \frac{\gamma_H^{DE}}{\beta} + \mu(v_H - \bar{v}) = 0 \quad \text{for high type.}$$

Extend basic analysis to entire range $c \in [0, +\infty)$. For small c , low type may also choose $a = 1$.

Results:

- For $\mu = 0$, $a^{BDM} = a^{DE}$ (γ corresponds to type's true WTA)
- For low μ , $a^{DE} > a^{BDM}$ (more moral behavior under DE)
- For high μ , $a^{BDM} > a^{DE}$ (more moral behavior under BDM).

[Intuition: For low μ , costless separation under BDM, not under DE. For high μ , low-cost pooling under BDM.]

Chen-Schonger (DP 2013)

- multiple price list and direct elicitation equivalent in non-moral domain
- not so in moral domain.]

3. ADDING A NARRATIVE

Negative narrative (excuse, absolving): pushes toward $a = 0$

- *Downplay externality:* Dehumanizing language: “Third party is undeserving”, Nazi propaganda and language degrading Jews.
- *Low level of being pivotal:* Bystander effect: “If I don’t do it, someone else will”. Consequentialist/utilitarian reasoning!
- *Magnification of the cost:* “I only followed orders”.
- *Omission vs commission:* different availability of narratives.

▶ Evidence

Positive narrative (responsibilizing): pushes toward $a = 1$

fairytale, role model, moral precepts, religion,
nation/loyalty/brotherhood...

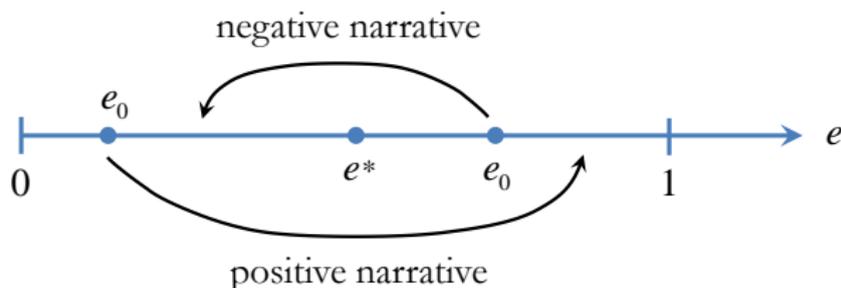
▶ Narratives can help define moral duties

EXOGENOUS NARRATIVES

Prior mean e_0 .

Possible narrative gives rise to posterior belief e distributed according to cdf $F(e)$ and density $f(e)$ on $[0, 1]$, with $E_F(e) = e_0$.

- Suppose $e_0 > e^*$. Then default action is $a = 1$; *negative narrative* = “reason not to act” (leads to $a = 0$ instead of default action $a = 1$ for H type; is an excuse for L type).
- Suppose $e_0 \leq e^*$. Then default action is $a = 0$; *positive narrative* = “reason to act” (leads to $a = 1$ instead of default action $a = 0$ for H type).



SELF-SUPPLIED NARRATIVES

Basic insights

Search for narrative prior to acting

Prior to acting, agent

- can learn e with probability x at cost $\psi(x)$
- can disclose e to audience if learns it.

Discloses and picks $a = 0$ if $e \leq \hat{e}$ where \hat{e} may differ from e^* , since being informed says something about one's type. Two opposite forces:

- H type has decision-making incentives to search, which L type does not have;
- L type particularly eager to find an excuse.

Searching for reason	Who uses narratives?
not to act ($e_0 > e^*$)	$x_H > x_L$ iff $e^* - M^-(e^*) > \mu\rho(v_H - \bar{v})/v_H$ (F "bottom-heavy"). Avoid false positives.
to act ($e_0 \leq e^*$)	$x_H > x_L$ always (L type less eager to find positive narrative)

Ex-post rationalization of behavior

- Pure ex-post rationalization useless. Must be the case that narrative may have been obtained prior to acting.
- Low type has more incentive to rationalize ex post

4. VIRAL NARRATIVES

Why do narratives spread? Reputation concern vs. influence concern

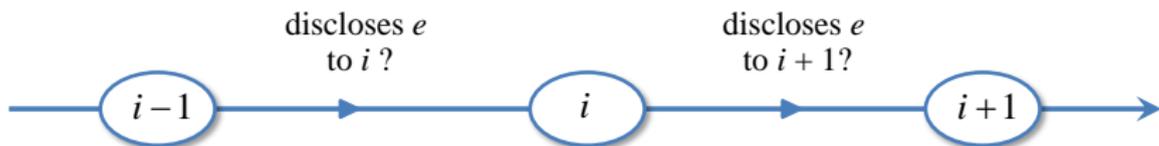
By changing audience's perception, sharing of *excuse* has two effects:

- enhances reputation of agent who has behaved immorally
($a = 0$)
- impacts negatively audience's behavior if it confronts similar choice

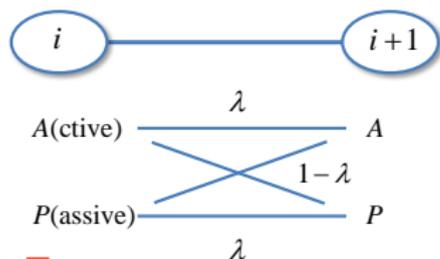
And conversely for a positive narrative.

When reputation concerns are paramount, negatives narrative are more likely to be shared than positive narratives.

When influence concerns are paramount, positive narrative are more likely to be shared than negatives narratives.



- learns e exogenously (prob. x)?
- observes a_{i-1} (if $i-1 \in A$)
- picks a_i (if $i \in A$), which is observed by $i+1$, her audience.



► Model

Interesting trade-offs concerning disclosure

(1) *Who is most eager to spread positive narratives / refrain from spreading negative narratives?*

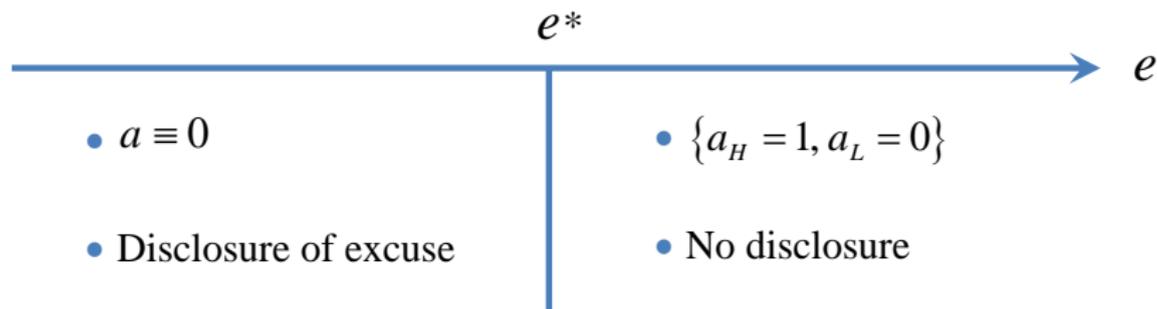
High type

- has higher influence concerns
- has lower need for excuses (less likely to choose $a = 0$).

(2) Look at *strategic substitutabilities (SS)/complementarities (SC)*:
Influence effect stronger if sharing by subsequent agents.

- Negative narrative: SS.
- Positive narrative: SC.

Example: equilibrium in which $a_H(\emptyset) = 1$



Negative narratives then spread, positive ones don't. More likely if:

- large image concerns μ (incentive to provide excuse)
 - heterophily (λ small)
 - bad society (ρ small)
- } reduces virality.

5. NARRATIVES VS. IMPERATIVES

What is an imperative?

- Located at the opposite end of action-conditioning messages
- Its impact is entirely determined by **who** issues them
 - soft information, i.e. uninformative by themselves
 - do not live in a vacuum (by contrast, narrative can by itself alter the individual's beliefs and actions);
may be followed, be ineffective or backfire depending on whether their author is regarded as trustworthy, neutral or adversary.
- Mere opinions in this respect share some of the characteristics of imperatives.

View imperatives as broad recommendations for *actions* (“ $a = 1$ ” say): soft rather than (semi-)hard information.

Benefits of imperatives:

- less fragile (interpretation uncertainty, cheaper to communicate)
- may allow to pool states in which agent would be reluctant with other states where he would be eager.

Costs of imperatives:

- require congruence
- more rigid (non contingent).

[Could also be studied: lower communication costs (particularly important if shared information) and easier monitoring (0/1 nature) of imperatives.]

Rationales for moral construction: Who is the “principal”?

- Parents or ex-ante self: Individual may behave in way that is detrimental to self or social image. Self control problem vs. excessive image concerns.
- Society: Wants to promote creation of positive externalities/ discourage negative ones.

Example: Principal may have objective function

$$[\tau w e + (v e - c)]a$$

$w = 0$: parents maximizing child's welfare

$w = 1$: utilitarian social planner

$w = +\infty$: moral entrepreneur with no empathy for agent

$w = -\infty$: immoral entrepreneur with no empathy for agent.

Preferences

Suppose:

- Principal's preferences affine function of e
- Define e^P : $U^P(e^P) \equiv 0$.

Moral entrepreneur assumption: $e^P < e^*$.

Principal's information/beliefs

Prior about externality: e_0

Principal has narrative \iff posterior beliefs e
with distribution $F(e)$ on $[0, 1]$
and $E[e] = e_0$.

Assumption 1: good behavior is not innate

$$e_0 < e^*$$

Convincing positive narrative $e_0 < e^* < e$

where, remember, $v_H e^* - \frac{c}{\beta} + \mu(v_H - \bar{v}) \equiv 0$.

Assumption 2: Suppose that principal is a *moral entrepreneur*: wants to promote pro-social behavior whenever $e > e^P$ where $e^P < e^*$.

Timing

- (1) Principal chooses between disclosing narrative e and issuing an imperative (say, “ $a = 1$ ”).
- (2) Agent chooses $a \in \{0, 1\}$.

“Effective imperative”: one that induces probability of pro-social behavior $a > 0$.

Equilibrium: moral entrepreneur picks imperative if both imperative and narrative are effective ($a_H = 1$).

▶ (Slightly) imperfect communication

WHEN IS IMPERATIVE EFFECTIVE?

For imperative to be an equilibrium, it must be the case that

(a) Anticipating obedience, principal recommends $a = 1$ iff

$$U^P(e) \geq 0 \Leftrightarrow e \geq e^P$$

(b) Obedience: agent picks $a = 1$ when told to pick $a = 1$. That is

$$M^+(e^P) \equiv E[e|e \geq e^P] \geq e^*. \quad (1)$$

If $M^+(e^P) < e^*$, then no imperative. Narrative is effective iff $e > e^*$.

Comparative statics

$$M^+(e^P) \equiv E[e|e \geq e^P] \geq e^*. \quad (1)$$

- As congruence increases, imperative becomes more likely
 e^P increases \Rightarrow condition (1) more likely to be satisfied.

We accept imperatives from principals with high moral standing (parents, religious or wise persons...). Narrative can be spread by anyone.

- *Large expected externality favors imperative*

$$F(e - \theta) \text{ with monotone hazard rate } \left(\frac{f}{1 - F} \right)' < 0.$$

Then if $\theta_1 < \theta_2$

$$M_1^+(e^P) \geq e^* \Rightarrow M_2^+(e^P) > e^*.$$

- *Imperative more likely if P perceived to have sound judgment*

Higher ρ , when $\partial M^+(e^P, \rho) / \partial \rho > 0$

(uniform, Pareto, exponential; or to left of rotation point).

► Value of flexibility

6. CONCLUSION

- (1) Need to understand how narratives are constructed, are shared and spread. Starting point: these respond to incentives
 - to sustain a self- and social reputation,
 - to induce others to adopt a moral or immoral behavior (we are all narrative entrepreneurs).
- (2) Popularity of deontological Kantians.
Implications for the measurement of moral preferences.
- (3) Competition between narratives and imperatives to shape moral behavior.
- (4) Virality of narratives.

Many alleys for future research...

- Ongoing one on organizational design: shared control and the sub-additivity of responsibility: individual vs. collective veto power (decision rule: $a = a_i a_j$ or $1 - (1 - a_i)(1 - a_j)$?); individual vs. collective accountability (observability of a or a_i ?); individual vs. team incentives (cost c incurred when $a = 1$ or $a_i = 1$?).
- What defines a “moral act”? Consequentialist/externality approach in model.
Relatedly different views on morality.
- Act may create positive and negative externalities.

THANK YOU VERY MUCH FOR YOUR ATTENTION

PARAMETER RANGE

Assume that the high type contributes when certain that there is an externality ($e = 1$), but not when there is none ($e = 0$):

$$v_H - \left(\frac{c}{\beta}\right) + \mu(v_H - \bar{v}) > 0 > v_L - \left(\frac{c}{\beta}\right) + \mu(v_H - v_L).$$

benefit of contributing

cost of contributing

dominant strategy for immoral type not to contribute

maximal reputation gain

suppose no contribution by high type; then $a = 0$ delivers reputation \bar{v} , while contribution would indicate v_H .

return

EVIDENCE: 1. SELF-CONTROL (β)

Evidence somewhat mixed (ego depletion, response times, cognitive load)

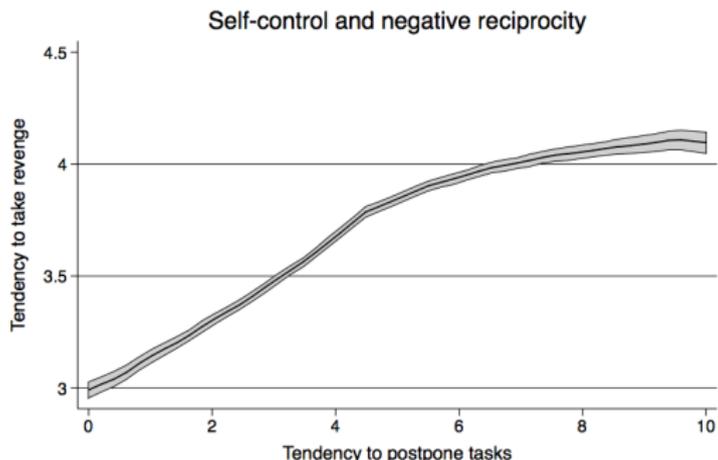
- *Altruism*: Higher ability to use self-control strategies (self-reported: measured on the Rosenbaum psychometric scale*) enhances willingness to share in dictator game (charity) (Martinsson et al., 2012).
- *Cooperation*: Higher self-control (measured in lab) enhances cooperative behavior in sequential two-player prisoner's dilemma. Both for first and second movers, cooperation is higher, the higher is β (Burks et al., 2009).
- *Tragedy of the commons*: Patience (measured in lab) decreases exploitation of fish population (common pool resource) by fishermen in Brazil (Fehr and Leibbrandt, 2011).

* Questions about getting distracted, work-pleasure behavior, inappropriate speech. . .

EVIDENCE: 1. SELF-CONTROL (β)

Revenge: Higher degree of self-control problems increases tendency to take revenge (Source: Falk et al., 2015; representative assessment of preferences in 76 countries, n=80,000).

Revenge: If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so (Scale: 0-10). Self-control: I tend to postpone tasks even if I know it would be better to do them right away. (Scale: 0-10)



EVIDENCE: 2. MORAL LICENSING (\bar{v})

- *Discrimination:* Monin and Miller (2001) let participants demonstrate their lack of prejudice before asking them to play the role of an employer. Compare with “after” condition.
Participants who had been able to demonstrate non-prejudiced attitudes said that a (police) job was better suited for a White person.
Similarly, after having the opportunity to endorse Obama, people are more likely to favor whites than blacks (Effron et al., 2009).
- *Altruism:* After being asked to write a self-relevant story including positive (negative) traits people donate less (more) to a charity (Sachdeva et. al, 2009).
Opposite evidence: foot in the door experiments.

EVIDENCE: 3. EXTERNALITY (e) AND COSTS (c)

- *Externalities (e)*: Higher external return (gain for each other person) and larger group size holding internal costs (cost for subject) constant increases contribution in public good experiments (Goeree et al., 2002).

People are to some extent *consequentialist* / utilitarians (consider trade-offs).

- *Cost of contribution (c)*: The same study shows that higher internal costs reduce contribution.

EVIDENCE: 4. BEING PIVOTAL (e)

Application Falk and Szech (2015) study whether diffusion of being pivotal lowers moral behavior using the mouse paradigm.

n-member committee: $a = 0$ if $a_i = 0$ for at least one member.

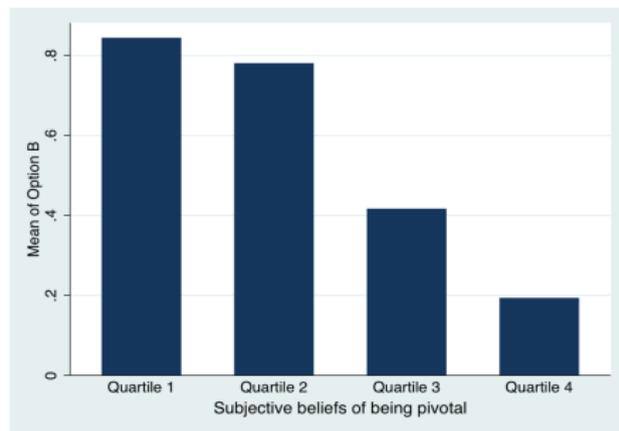
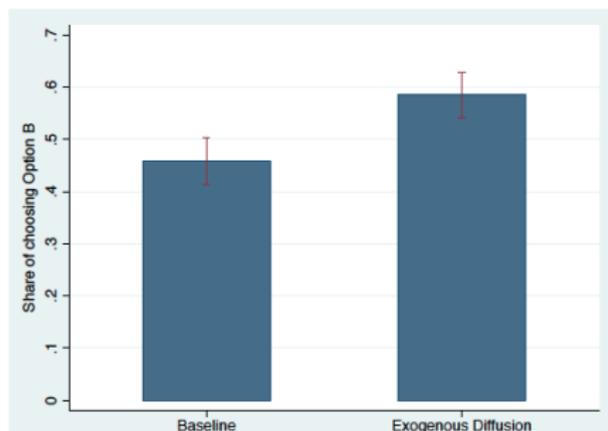
[Different from *unethical bilateral trade*: there, $a = 0$ iff $a_i = 0$ for all i .]

- Delegating responsibility to such committees makes more people become willing to support immoral activity.
- Individual incentives: cost c paid regardless of a_j .
- Moral conceptions must be utilitarian instead of Kantian to make diffusion of being pivotal work.

EVIDENCE: 4. BEING PIVOTAL (e)

Mouse killing paradigm. Source: Falk and Szech, 2015.

- Baseline: If $a = 0$, mouse is killed (individual is fully pivotal).
- Exogenous Diffusion Treatment: Groups of eight; if at least one group member chooses $a_i = 0$, eight mice get killed; private costs, i.e., each subject receives 10 euros for choosing $a_i = 0$ (option B), irrespective of behavior of others.



EVIDENCE: OMISSION VS. COMMISSION

Experiment (Spranca et al, 1991).

Setup:

- Subjects read scenarios concerning pairs of options. One option was an omission, the other, a commission.

Result:

- Subjects often rated harmful omissions as less immoral than harmful commissions.
- Such ratings were associated with judgments that omissions do not cause outcomes.

Omission: easier to find excuses (“I forgot”, “I did not draw the connection”, “I was in a rush”, ... “I did not lie”).

return

WHEN NARRATIVES HELP DEFINE MORAL DUTIES

Ubiquitous *“Imagine that everyone did this”/Kant’s famous categorical imperative “Act only according to that maxim whereby you can, at the same time, will that it should become a universal law”*

- At first sight, rather silly counterfactual
 - Game theory: take other strategies as a given
 - Even if the game is repeated, unlikely that one’s behavior will change others’ if the number of players is large.
- Actually, another (more logical) narrative states *“In a large society one’s action, say a minor pollution, makes no difference”*. Similarly, one’s behavior is never going to become universal law.

Conjecture: does the paper thrown to the ground significantly deteriorate the environment? Is the externality e high enough to “justify” the cost c , “implying” a moral duty?

Narrative magnifies both cost and externality (one then envisions dirty cities), facilitating the comparison.

SELF-SUPPLIED NARRATIVE

- Convex cost $\psi(x)$ of finding narrative with probability x . Otherwise finds no information (\emptyset).
- Search intensities x_H for moral type, x_L for immoral type.

Equilibrium:

Searching for negative narratives

Search: at cost $\psi(x)$, learns e (distribution $F(e)$ on $[0, 1]$, $E_F(e) = e_0$) with prob. x , learns nothing (\emptyset) with prob. $1 - x$.

Who is most eager to find negative narratives (and disclose them)?

$$x_H > x_L \Leftrightarrow \underbrace{v_H[e^* - M^-(e^*)]}_{\substack{H \text{ type stops} \\ \text{picking } a = 1 \text{ when} \\ \text{posterior is below } e^* \\ (M^-(e^*) = \text{truncated} \\ \text{mean } E[e|e \leq e^*])}} > \underbrace{\mu\rho(v_H - \bar{v})}_{\substack{L \text{ type gains more} \\ \text{reputation when} \\ \text{finding an excuse} \\ (H \text{ type loses} \\ \text{reputation})}}$$

Searching for positive narratives

$$x_H > x_L$$

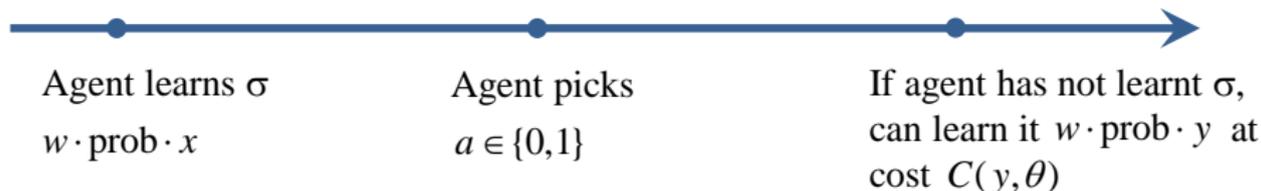
Low type

- will not act when finding a positive narrative
- does not want to increase stigma further.

High type

- decision-making gain
- reputational gain.

Ex-post rationalization of behavior



Searching for negative narratives ($e_0 > e^*$)

- (1) No ex-ante information ($x = 0$) \Rightarrow no ex-post rationalization ($y = 0$)
- (2) Only v_L rationalizes ex post (v_H picks $a = 1$ if ex-ante information \Rightarrow no need to rationalize: $y_L > y_H = 0$).

return

APPLICATION: POSITIVE NARRATIVES AND DEVELOPMENT OF PRO-SOCIALITY

Narratives and development of pro-social behavior:

- Morally relevant narratives (oral, written, and cinematic) are “an essential component of effective moral education” (Vitz 1990, p. 709).
- Allow for identification and self-representation, increase empathy and awareness, and provide positive role models (see, e.g., Bennett 1993, McAdams and Koppensteiner 1992, Tappan and Brown 1989, Mar and Oatley 2008, Johnson 2012).

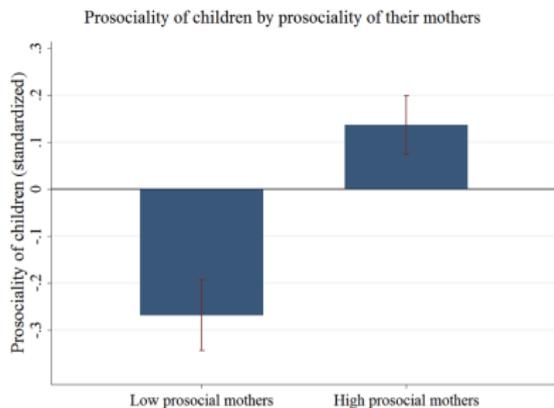
Bandura (1965, 1986): exposure to role model affects pro-social behavior.

Loose connection at this stage: positive role models as positive narratives?

EVIDENCE: POSITIVE NARRATIVES/ROLE MODELS AND DEVELOPMENT OF PRO-SOCIALITY

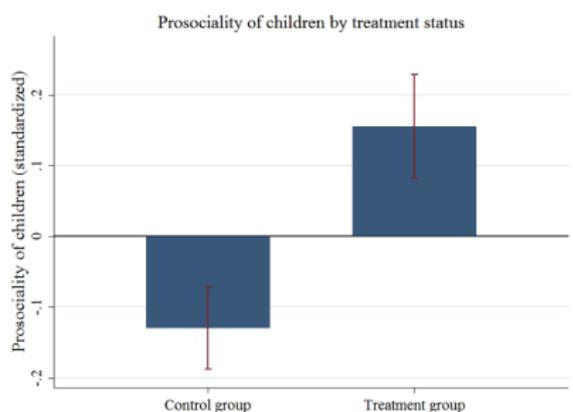
Pro-sociality: sample of young children, standardized measures of altruism, trust, pro-social behavior; source: Kosse et al., 2015.

Impact of mother



N=410

Impact of mentor



N=490

EVIDENCE: INSENSITIVITY TO NUMBERS

Utilitarianism vs. Kantianism

Choice paradigm: reservation prices to kill one, two or three mice, between subject design. Source: Falk and Szech, 2015.

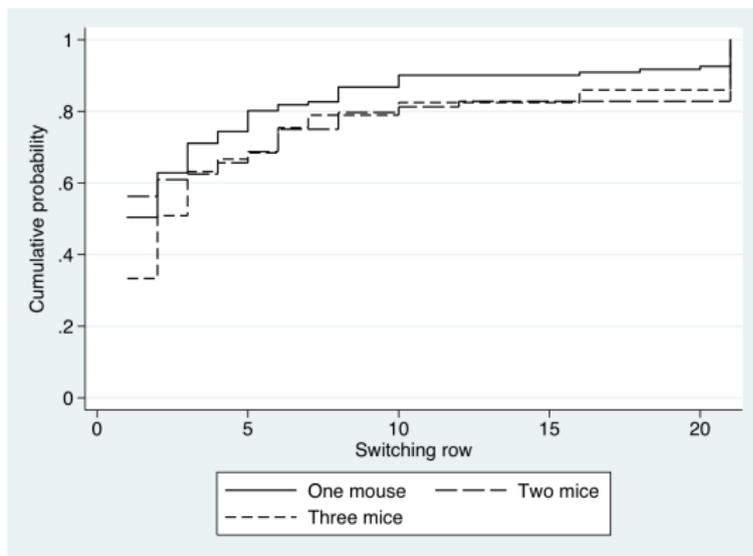


Figure 1: Cumulative probabilities of switching rows in the One Mouse, the Two Mice and the Three Mice treatments.

[return](#)

UNCERTAINTY ABOUT INTERPRETATION/COMMUNICATION

Agent $\left\{ \begin{array}{l} \text{understands argument with probability } x \Rightarrow \text{posterior } e \\ \text{fails to understand } (\emptyset) \text{ with probability } 1 - x \text{ (can be } \simeq 0). \end{array} \right.$

Principal also has “meaningless narrative” (\emptyset), that the agent is unable to distinguish from lack of understanding of true narrative.

return

VALUE OF FLEXIBILITY

Narrative generates thought: Agent may have (existing or context specific) complementary information that refines narrative.

$G(\sigma|e)$ such that $E(\sigma|e) = e$ (example: $G(\sigma|e) = \sigma^{\frac{e}{1-e}}$ (so $E(\sigma|e) = e$)).

Important point: $U^P(\sigma)$ and $U^A(\sigma)$.

Can an imperative exist?

Narrative: $V^P(e) \equiv \int_{e^*}^1 U^P(\sigma) dG(\sigma|e)$.

Imperative: $V^P(e) \equiv U^P(e)$ (provided that $M^+(e^P) \geq e^*$).

Conditions for existence of imperative. Let $I = \{e|e \text{ picks imperative}\}$

Obedience

$$E(e|e \in I) \geq e^*. \quad (2)$$

Optimality $\Delta(e) =$ net benefit of narrative when e .

$$\Delta(e) \equiv \int_0^{e^*} U^P(\sigma) dG(\sigma|e) \leq 0 \quad \text{iff } e \in I. \quad (3)$$

Remarks

- (3) is not satisfied at e^P : $\int_0^1 U^P(\sigma) dG(\sigma|e^P) = 0$,
and $U^P(e) > 0$ for $e \geq e^*$.
- *Impact of self-control.* Lower self-control ($\beta \searrow$) $\Rightarrow e^* \nearrow$
 - imperative more attractive ((3) more likely)
 - but obedience more problematic ((2) less likely).

Then $\Delta(e) \geq 0 \Rightarrow \Delta(e') > 0$ for $e' > 0$, and conversely.

Hence $I = [e^\dagger, 1]$ with $e^\dagger > e^P$ if there exists an equilibrium with imperative. Obedience condition can be rewritten

$$M^+(e^\dagger) \geq e^*$$

e^\dagger is decreasing in e^*

- as self-control problem worsens, more temptation to go for imperative (I expands)
- however, obedience condition at some point is no longer fulfilled. Very serious self-control problem \Rightarrow no imperative.

▶ Evidence

return

Model of viral narratives

- Agents $i \in \mathbb{N}$ on a line from $-\infty$ to $+\infty$.
- Agents each learn the narrative e exogenously with probability x (i.i.d.).
- Agents can be passive (P) or active (A). They don't know their successor's type (v_{i+1} and A/P). Serial correlation $\lambda \geq 1/2$ in A/P dimension.
- Active agents choose $a_i = 0$ or 1 . Agent i 's action is observed by agent $i + 1$.
- Agent i chooses whether to disclose to agent $i + 1$ the narrative, if any.
- Agent i cares about his reputation vis-à-vis agent $i + 1$.

return

Taboos (Fiske-Tetlock on “Taboo tradeoffs”, Bénabou-Tirole *QJE* 2011).
Reluctance to challenge a moral imperative.

- Asking for the price of a transaction that is socially unacceptable (“just to know”)
- Questioning a religious dogma.

Idea: reputational damage from considering a repugnant transaction or thinking about a religious dogma is done even if one ends up behaving morally.

Perceived as a “calculating individual”

[BTW: “*The age of chivalry is gone. That of sophisters, economists, and calculators, has succeeded; and the glory of Europe is extinguished forever.*” Edmund Burke (in 1793), one of the founders of British conservatism.]

return