

MACHIAVELLI'S MISTAKE:

WHY GOOD LAWS ARE NO SUBSTITUTE FOR GOOD CITIZENS

Samuel Bowles

To my teachers

Alexander Gerschenkron

&

Charles E. Lindblom

Many of the ideas in the pages that follow were presented when the author served as Castle Lecturer in Yale's Program in Ethics, Politics and Economics at Yale University in 2010. The Castle Lectures were endowed by John K. Castle to honor his ancestor, the Reverend James Pierpont, one of Yale's original founders. Castle Lectures are intended to promote reflection on the moral foundations of society and government, and to enhance understanding of ethical issues facing individuals in our complex modern society.

... e' necessaria a chi dispone una repubblica e ordina leggi in quella, presuppone tutti gli uomini rei .. si dice che la fame e la povertà fa gli uomini industriosi, e le leggi gli fanno buoni.

...anyone who would found a republic and order its laws must assume that all men are wicked ... it is said that hunger and poverty make men industrious, and laws make them good.

Nicolò Machiavelli, *Discorsi sopra la prema deca di Tito Livio*. 1513-1518,
I,3 (translation: SB)

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I PROLOGUE

II MACHIAVELLI'S MISTAKE

Aristotle's Legislator and Machiavelli's Prince

A constitution for knaves

Markets as morality free zones

Knaveonomics

Machiavelli's mistake

III MORAL SENTIMENTS AND MATERIAL INTERESTS

Homo economicus has company

Crowding out (and in)

Crowding out: A Taxonomy

A surprise for Aristotle's Legislator

The lab and the street

IV INCENTIVES AS INFORMATION

A neural basis for crowding out?

Incentives as messages

Bad news

Moral engagement

Control: Incentives may compromise autonomy

A puzzle

V A LIBERAL CIVIC CULTURE

The economy produces people

Markets and civic culture

Order in liberal and other societies

Good laws and good citizens

Doux commerce?

A liberal civic culture

VI ECONOMICS REDISCOVERS ARISTOTLE

Why not a constitution for knaves?

A second best world

Smith versus Smith

VII A MANDATE FOR ARISTOTLE'S LEGISLATOR

Getting and becoming: Are incentives to blame?

Moral lessons

A mandate for Aristotle's Legislator

Laws as they might be for citizens as they might be

APPENDICES

WORKS CITED

ENDNOTES

INDEX (to come)

ACKNOWLEDGEMENTS

This work originated a quarter of a century ago in a paper on the unintended cultural consequences of economic policies. The paper had been titled *Mandeville's Mistake*, honoring (albeit dubiously) the author of the *Fable of the Bees*, the early 18th century verses held by some to be the founding work in classical political economy. The paper was received with interest in 1989 by my fellow members of the September Seminar, then meeting at the Philosophy Department of University College London. But empirical support for its primary claims was thin, and I set the paper aside.

The accumulation of experimental and other new data encouraged me to return to the project. The resulting book, now titled *Machiavelli's Mistake*, is based in part on my Castle Lectures at Yale University, where the critical commentary of Bryan Gersten, Phil Gorski, Laurie Santos, Stephen Smith, and Chris Udry resulted in many improvements. This is the second time I have learned from Yale's social science faculty, the first time was as a student in the precursor to the Ethics Politics and Economics major that had sponsored my Castle lectures. My first debt then is to my mentor and inspiration in that program, Charles Lindblom. My subsequent study of economic history as a doctoral student with Alexander Gerschenkron convinced me that the large questions of social organization and its evolution might have answers, though not necessarily the large ones that I was at the time hoping for.

My thinking on these issues has been shaped over the years by the comments of members of the September Seminar (since 1986) and the Santa Fe Institute Working Group on the Coevolution of Behavior and Institutions (since 1998). For their comments on earlier drafts of this work I would particularly like to thank Mahzarin Banaji, Yochai Benkler, Erica Benner, Wendy Carlin, Jung-Kyoo Choi, the late Gerald Cohen, Joshua Cohen, Steven Durlauf, Simon Gächter, Joshua Greene, Jonathan Haidt, Philippe van Parijs, Elizabeth Phelps, John Roemer, Daria Roithmayr, Rajiv Sethi, Seana Shiffrin, Rebecca Saxe, Paul Seabright, Erik Olin Wright, Elisabeth Wood, and Luigi Zingales

The energetic criticism that the book's title aroused at a lecture in Florence (Machiavelli's home town) correctly pointed out that Machiavelli's ethics are far more nuanced and his view of

human nature far less malign than the English language epithet “Machiavellian” usually implies. I am grateful to Chiara Valentini and Susan Karr for pushing me to clarify (in Chapter 2) just what the mistake in the title is.

My collaborators Sung-Ha Hwang and Sandra Polanía Reyes are virtual co-authors of parts of the book; I am grateful to them for permission to use material from our joint work. Chapter V draws upon work published in *Philosophy and Public Affairs* (2011) and I thank them for permission for its inclusion in this work. Chapters III and IV are extensions of ideas published in the *Journal of Economic Literature* (2012).)

Finally, the Santa Fe Institute and the Certosa di Pontignano of the University of Siena have provided unsurpassed environments for research, reflection and writing. Their staffs have made this research both enjoyable and possible, with particular thanks to Margaret Alexander, Joy Lecuyer, and Barbara Kimbell of the Santa Fe Institute Library. I would also like to thank Nicole Villar Hernandez for assistance with the research and the MacArthur Foundation,, the Behavioral Sciences Program of the Santa Fe Institute and the U.S. National Science Foundation for financial support.

Santa Fe, New Mexico
November, 2014

I

PROLOGUE

Two and a half centuries ago Jean-Jacques Rousseau invited his readers to consider “Laws as they might be” for “men as they are.” Aside from the gendered language, the phrase resonates still. We know that governing well requires an understanding of how people will respond to the various laws, economic incentives, information, or moral appeals that make up a system of governance and that a policy maker or constitution writer may consider altering. But what are we to understand by Rousseau’s “men as they are”?

Enter: economic man. An important strand of thinking about the design of public policy and legal systems as well as the organization of firms and other private organizations is based on the idea that people – whether citizens, employees, business partners or potential criminals --are entirely self interested and amoral, aptly going by the name *Homo economicus*. Nobody believes this to be literally true; instead the assumption has been advanced on grounds of prudence.

But I hope to convince you that the policies implemented following this self-interest based paradigm may sometimes make the assumption of universal selfishness more nearly true than it might otherwise be. This is a result of the unintended cultural consequences of policies that economists have favored, including more extensive and better defined private property rights, enhanced market competition and the greater use of monetary incentives to guide individual behavior.

I will show that these and the other policies advocated as necessary for a market economy to function well could at the same time promote self interest and undermine the means by which a society sustains a robust civic culture and could even erode the social norms essential to the workings of markets. Included among the cultural casualties of this crowding out process are such workaday virtues essential to market functioning as truthfully reporting one’s assets and liabilities when seeking a loan, keeping one’s word, and working hard even when nobody is looking. Markets and other economic institutions do not work well where these and other norms are absent. High performance modern economies require cultural underpinnings such as the assurance that a handshake is indeed a handshake; where one doubts this, mutual gains from exchange may be limited by distrust.

The paradoxical idea that “perfecting” markets might make them work less well is not limited to particular policy, the adverse impact of which, should it occur, might be readily reversed. And it applies well beyond markets. A people’s civic mindedness or intrinsic desire to uphold social norms may be squandered, perhaps irreversibly, shrinking the space for more enlightened policies in the future. Thus while some economists imagined that in a distant past *Homo economicus* had invented markets, it could as well have been the other way around: the proliferation of amoral self interest might be one of the consequences of living in the kind of society that economists idealized.

The dilemma facing the policy maker or constitution writer is this: Incentives and constraints that would be optimally designed if “men as they are” were to resemble *Homo economicus* might backfire if they foster the very self interest that they were designed to harness in the service of the public good. The dilemma would not arise if *Homo economicus* were an accurate description of “men as they are.” In this case there would be nothing to crowd out.

But, over the past 2 decades behavioral experiments (reviewed in Chapters 2 and 3) have provided hard evidence that citizens’ ethical and other regarding motivations are common in virtually all human populations. The experiments also showed that these motives are sometimes crowded out by policies and incentives that appeal to the material interests.

In Haifa, at six day care centers, a fine was imposed on parents who were late picking up their children at the end of the day. Parents responded to the fine by doubling the fraction of time they arrived late (Gneezy and Rustichini (2000)). When after 12 weeks the fine was revoked, their enhanced tardiness persisted unabated. Parents’ lateness in the centers in which the fine was imposed and in the control group without the fine are shown in Figure 1.1.

While other interpretations are possible, the counter-productive imposition of the fines illustrates a kind of negative synergy between economic incentives and moral behavior. By placing a price on lateness, as if putting it up for sale, the fine seems to have undermined the parents’ sense of ethical obligation to avoid inconveniencing the teachers. And it led them to think of lateness as just another commodity they could purchase. I do not doubt that had the fine been sufficiently high the parents, would have responded differently. But putting a price on everything, even if this could be done and the right prices could be found, might not be a good idea. In a recent experiment, even the sight of money and the discussion of coins (rather than non monetary

objects) induced children to behave in less pro-social ways and to be less helpful towards others in a natural setting following the experiment (Gasiorowska, Zaleskiewicz and Wygrab (2012).)

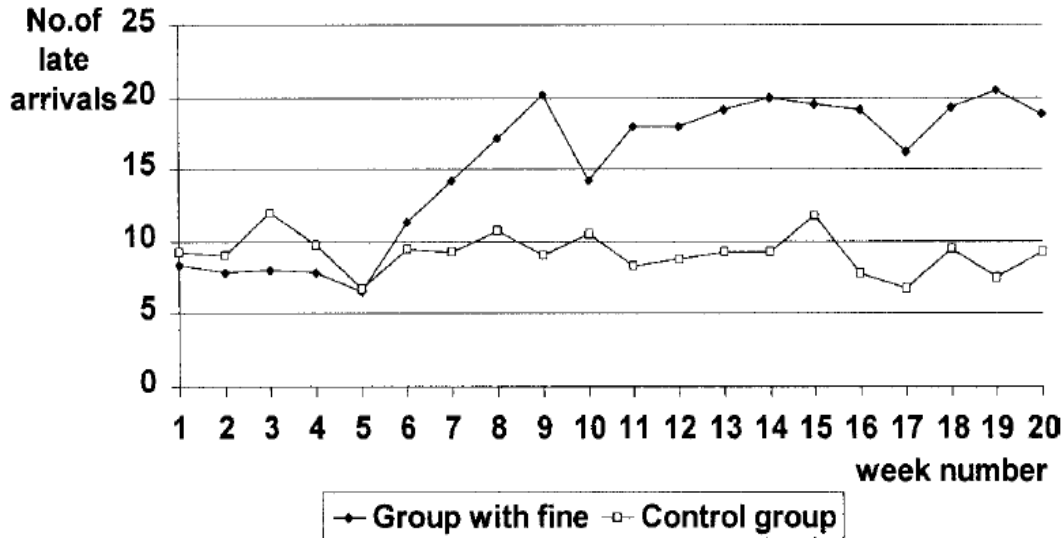


Figure 1.1. The effect of a fine for lateness in Haifa’s day care centers. Source: Gneezy and Rustichini (2000). The fine was imposed in week 5 and retracted in week 17.

In another study, kids less than two years old avidly helped an adult retrieve an out of reach object in the absence of rewards. But after being rewarded with a toy for helping the adult, the helping rate fell by forty percent. Felix Warneken and Michael Tomasello, the authors of the study, conclude: “children have an initial inclination to help, but extrinsic rewards may diminish it. Socialization practices can thus build on these tendencies, working in concert rather than in conflict with children’s natural predisposition to act altruistically.” (Warneken and Tomasello (2008):1787)

Their advice about socialization might be a good lesson for public policy too.

How is the policy maker to respond to the realization that both economic incentives and non economic motives are necessary for effective policy, but the former may diminish the latter? If the two horns of the dilemma are taken for granted, then policy maker may reasonably consider a more limited role for economic incentives in his policy packages. Similar reasoning might lead him to restrict the role of markets in allocating resources in favor of a larger role for governments

or informal organizations. Doing so would be consistent with the main point of Michael Sandel's *What Money Can't Buy: The Moral Limits of Market*, namely, that "putting a price on every human activity erodes certain moral and civic goods worth caring about." (Sandel (2012), Sandel (2013):121) Sandel makes a convincing case for a public debate on "where markets serve the public good and where they don't belong."

But a good case can also be made that incentives and markets *per se* may not be entirely to blame. We will see that crowding out may reflect more fundamental problems stemming from the relationship between the person implementing the incentive and its target such the appearance of a controlling or greedy intent or an untrusting attitude on the part of an employer, or the message inadvertently conveyed by the incentive such as, in Haifa, "it's ok to be late, as long as you pay for it."

If this is the case then the policy maker can do better than limiting the role of incentives and markets so as to manage the horns of the dilemma. He can actually circumvent the dilemma or at least attenuate it by finding policies in which economic incentives and ethical and other non economic motives are synergistic, each enhancing the effect of the other. He may also seek to limit the economic and political disparities that sometimes foster contrary reactions by the targets of incentives.

This is the paradigm for policy making that I wish to advance here.

A complementary part of such a new paradigm would have to take account not only of the importance of the ethical and other regarding motives that feature prominently in the pages that follow, but also of the new evidence on the cognitive processes by which we arrive at our decisions. The work of Thaler and Sunstein (2008), Kahneman (1994), Kahneman and Tversky (2000) and others has made it clear that people are not nearly as far sighted, calculative and consistent in their decision making as has been generally assumed by economists. Instead we are biased towards the status quo, inconsistent in choosing among alternatives occurring at different times, and even after correction we insist on making what economists would consider to be computational mistakes, for example in taking actions in an uncertain situation. People, in short, are not all that good as choosers. This is an essential component of a new policy paradigm.

But I will be less concerned with how we make decisions than with what we value, how incentives and other aspects of public policy may affect what we value, and why this suggests

modification of the dominant policy making paradigm in the social sciences. I will begin by explaining what this dominant paradigm is and telling the remarkable story of how its practitioners came to be either unaware or unconcerned that the policies it favored might crowd out ethical and other social motivations

II

MACHIAVELLI'S MISTAKE

Having noticed a suspicious bunching of sick call-ins on Mondays and Fridays, the Boston Fire Commissioner on December 1, 2001 ended the Department's policy of unlimited paid sick days. In its stead he imposed a 15-day sick day limit; the pay of firemen exceeding that limit would be docked. The firemen responded to the new incentives: those calling in sick on Christmas and New Years Day increased tenfold over the previous year.

The Fire Commissioner retaliated by cancelling their holiday bonus checks (Belkin (2002)). The firemen were unimpressed: the year following they claimed 13,431 sick days; up from 6,432 the previous year (Greenberger (2003)). Many of the firemen, apparently insulted by the new system, abused it, or abandoned their previous ethic of serving the public even when injured or not feeling well.

I admit to some sympathy for the Commissioner's plight, having once offered my teen age kids a price list for household chores as a way of topping up their modest weekly allowance. In response they entirely ceased doing the housework that they had generously done in the absence of the incentives. The Commissioner's difficulties and my failed experiment in home economics are far from exceptional.

As we will see, the imposition of explicit economic incentives and constraints to induce people to act in socially responsible ways is sometimes ineffective or even counterproductive, as the Fire Commissioner discovered. But is this a problem? My hunch is that a larger penalty would have worked. The firemen's massive sick call-in on Christmas and New Years Day does not mean that they had lost interest in money. Had the Fire Commissioner imposed a heavier penalty, the firemen would surely have shaped up, even if the distrust they felt had indeed eclipsed their sense of duty. Economic interest would have substituted for their pride in a job well done or the public well served.

There may be limits to the effectiveness of these and similar constraints and incentives.

Heavy fines or more positive monetary incentives might have deterred phony excessive call-ins, but would they have been sufficient to motivate the more subtle and immeasurable aspects of a fireman's professionalism and bravery? Practical impediments aside, a liberal society may wish to place limits on the kinds and extent of penalties and inducements that the Commissioner might deploy (a topic to which we return in Chapter 6). Instead of letting penalties be a substitute for the firemen's sense of duty, the Commissioner might seek policies that would instead affirm and enhance their civic mindedness.

Whether you think the firemen's response to the Commissioner's incentives is a problem or not lines you up on one or the other side of a venerable and unresolved conflict in the philosophy of governance.

ARISTOTLE'S LEGISLATOR AND MACHIAVELLI'S PRINCE

Political philosophers from Aristotle to Thomas Aquinas, Jean-Jacques Rousseau, and Edmund Burke recognized the cultivation of civic virtue not only as a test of good government, but also as its essential foundation. "Legislators make the citizen good by inculcating habits in them," Aristotle had written in the *Ethics*. "It is in this that a good constitution differs from a bad one." (Aristotle (1962):103) A century earlier Confucius had provided advice about how this might be done, and the pitfalls to be avoided: "Led through laws and disciplined through punishments, people may try to avoid them and never feel a sense of shame. Led through virtue and disciplined through the rules of propriety, they will have a sense of shame and learn self-ordering."

Early in the sixteenth century, Nicolò Machiavelli gave rather different advice: "Anyone who would found a republic and order its laws must assume that all men are wicked [and] ... never act well except through necessity... it is said that hunger and poverty make them industrious, laws make them good." (Machiavelli (2001):30-31). Machiavelli's "laws make them good" might sound a bit like Aristotle's Legislator "inculcating good habits" in the public. But here, as with his "all men are wicked," Machiavelli uses "good" (*buono*) and wicked (*rei*) to describe behavior or actions, not aspects of character or human nature. The task of government for Machiavelli was not to uplift the moral character of the populace, but rather to induce citizens motivated by what he termed the "natural and ordinary humors" to act as if they were good.

Machiavelli makes clear, especially in his *Discourses*, that it is not the morality of its

citizens that makes a well governed city possible, but rather the capacity of a good prince to “order the laws.” (Benner (2009)) By comparison to Italy, he wrote, Spain and France are well governed; but this “ derives not much from the goodness of the people, which is for the most part lacking, ... but from the way that these kingdom are ordered.” “France,” he continued, is “a kingdom moderated more by laws than any that has been known in our times.” (Machiavelli (1984):174,180, chapters 56 & 58). The message was unmistakable: Citizens of ordinary predispositions and desires could nonetheless be well governed if their behavior is “moderated ... by laws.”

Machiavelli's advice that princes and legislators should distinguish between individual motives and society-level consequences was the key insight of Bernard Mandeville's scandalous verses *The Fable of the Bees*, that held that virtue was dispensable, even pernicious for the social order. Mandeville's scandalous hive thrived on licentious greed; and when the bees turned virtuous, collapse and disorder ensued. (Mandeville could not have known that the genus *Apis* are among the most cooperative of all species and are genetically programmed not to compete.) The subtitle of the 1714 edition of his *Fable* announced that the work contained “...several discourses to demonstrate that human frailties...may be turn'd to the advantage of civil society, and made to supply the place of moral virtues,” with the result, he explains in the text (Mandeville (1924):24), that “the worst of all the multitude did something for the common good.”

In case the reader might fail to decipher the verses of the *Fable*, Mandeville provided a prose commentary in which he explained:

Hunger, Thirst and Nakedness are the first Tyrants that force us to stir; afterwards our Pride, Sloth, Sensuality and Fickleness are the great Patrons that promote all Arts and Sciences, Trades, Handicrafts and Callings; while the great Taskmasters Necessity, Avarice, Envy and Ambition ... keep the Members of the Society to their labour, and make them submit, most of them cheerfully, to the Drudgery of their Station; Kings and Princes not excepted. (Mandeville (1988a):366)

To Mandeville, the benign consequences of what Machiavelli would have called the ordinary humors is not a natural fact about human society. Just as Machiavelli saw the foundation of good government in the human capacity to order the laws, Mandeville explained that it was “the dextrous Management of a skilfull Politician” that allowed the “Private Vices” to be “turned into Publick Benefits.” (Mandeville (1988a):369)

In contrast to the Aristotelian view that good laws make good citizens, Mandeville's *Fable* suggested that the right institutions might harness shabby motives even to elevated ends. It was left to Adam Smith to explain how this improbable alchemy might be accomplished.

...he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. Smith (1976 [1776])

Competitive markets and well defined property rights, he explained, would let the invisible hand do its magic: “It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest” (Smith (1976 [1776]), Book 1, Chapter 2). Thus under the right institutions, elevated consequences may follow from ordinary motives.

A CONSTITUTION FOR KNAVES

Novel foundations for law and public policy followed. Thus in his *Essays: Moral, Political and Literary* (1742), David Hume (1964) :117-118 recommended the “maxim” that

in contriving any system of government ... every man ought to be supposed to be a *knave* and to have no other end, in all his actions, than private interest. By this interest we must govern him, and, by means of it, make him, notwithstanding his insatiable avarice and ambition, cooperate to public good.

In similar spirit, Jeremy Bentham offered his “*Duty and Interest* junction principle: Make it each man’s *interest* to observe ... that conduct which it is his *duty* to observe”(Bowring (1962):380). His *Introduction to the Principles of Morals and Legislation* is the first text in what we now call public economics. In it Bentham laid out the public policy implications of Hume’s maxim.

Thinking among jurists paralleled the economists. “If you want to know the law and nothing else,” Oliver Wendell Holmes, Jr, told students of law in 1897 (and every entering law school class since, it appears) “you must look at it as a bad man, who cares only for the material consequences which such knowledge enables him to predict, not as a good one who finds his reasons for conduct, whether inside the law or outside it in the vaguer sanctions of conscience ... The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it — and nothing else.” (Holmes (1897).)

However, the classical economists' response to the constitutional challenge of freedom and

order that still resonates in juridical and economic thinking was not motivated by the belief that economic actors and citizens are indeed amoral. Quite the contrary.

Hume pioneered the study of the evolution of social norms; and in the sentence immediately following the passage about knaves quoted above, he mused that it is “strange that a maxim should be true in politics which is false in fact.” Smith (1976 [1759]):3 in his *Theory of Moral Sentiments*, held that : “How selfish soever man may be supposed, there are evidently some principles in his nature that interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.” In practice, the policies advocated by the classical writers did not overlook appeals to ethical and other regarding motives. Bentham, as we will see in the final chapter, believed that punishments should be “moral lessons.”

Just a few lines before directing law students' attention to the “bad man” Holmes insisted that “The law is the witness and external deposit of our moral life.” Legal practice today, like the classical writers policies, recognizes a broad range of social dispositions rather than simply the self interest of the “bad man.” Market regulation for example combines fines for violations with requirements of disclosure to stockholders or the public of the wrongs done, so as to shame the blameworthy.

Even Machiavelli's famously corrupt citizens were introduced as a prudent assumption for the prince based on a widespread proverbial expression of the time – “*it is said that all men are wicked*” – not as evidence of a malign human nature, an assumption that Machiavelli rejected as a matter of fact : “our reasonings are about those peoples among whom corruption is not very widespread and there is more of the good than of the spoiled” and “very rarely do men know how to be altogether bad or altogether good.” (Machiavelli (2001):101, 139, Benner (2009)). The contrast between the Machiavellian policy paradigm and that of Aristotle does not arise because of a difference in assumptions of innate human goodness: Hume for example endorsed the idea while Bentham rejected it.

Thus, the appeal of the “constitutions for knaves” approach was not that citizens are in fact knaves. Rather it was, first, that the pursuit of self interest was coming to be seen as a benign or at least harmless activity when compared to other more disruptive “passions” such as religious fervor or the pursuit of power, and second that as an empirical matter the virtues alone provided an

insufficient basis for good government.

Avarice had been considered to be among the most mortal of the Seven Deadly Sins during the Middle Ages, and increasingly so with the expansion of commercial activity after the 12th century.(Bloomfield (1952):95). So it is surprising that self interest was to become a respectable motive, and even more surprising that this owed little to the rise of economics, at least initially. The year before Adam Smith penned his lines about the self interest of the butcher, the brewer and the baker putting dinner on the table, Boswell's Dr. Johnson had a different endorsement for the motive: "there are few ways in which a man can be more innocently employed than in getting money." (Boswell (1980):597)

It was the shadow of war and disorder that made self interest an acceptable basis of good government. Deaths in warfare during the 17th century constituted a larger share of all European mortality than in any century for which we have records, including the Century of Total War that happily is now behind us. Writing after a decade of warfare between parliamentarians and royalists, Hobbes sought to determine "The Passions that encline men to Peace," and found them in "Feare of Death; Desire of such things as are necessary to commodious living; and a Hope by their Industry to obtain them."(Hobbes (2005):104 (Chapter 13)). Knaves might be preferable to saints.

The second appeal of the constitutions for knaves approach was more directly related to Machiavelli's work and to the practical turn in political theory that he advocated. "... it seems more appropriate to go the reality of things rather than how we might imagine them," he wrote. . "There are many imagined republics and principalities that have never been seen nor known to exist in reality.(Machiavelli (1900):92, Chapter 15). A century and a half later Spinoza would open his *Tractatus Politicus* with: "no men are ... less fit to govern ... than theorists or philosophers ... [who] sing the praises of a human nature nowhere to be found, [who] ... rail at the sort which actually exist [and] conceive of men not as they are but as they would like them to be." (Spinoza (1958):261). A bit later Mandeville would introduce his *Fable* with virtually identical language.

But it was more than realism about human motives that recommended the retreat from virtue as the *sine qua non* of good government. If the "others" with whom we would seek to govern were our kin, neighbors, or friends, then our concern for their well-being or our desire to avoid

social sanction might induce us to act in ways contributing to the good governance of the city. But the metaphor of the polity as a family had long since ceased to be tenable, and the scope of governance had expanded considerably with the growth of cities and the consolidation of the nation state. With the increasing scope of markets, individuals had come to interact not with a few dozen, but with hundreds and indirectly with millions of strangers.

The Machiavellian paradigm was a response to the concern that when large numbers of strangers interact, ethical and other regarding motives would be an insufficient basis for good government, which therefore would need to adopt a system of constraints and incentives to supplement the civic virtues. It was the unavoidable insufficiency of the civic virtues, not their absence or irrelevance that worried Machiavelli. The classical economists who took up his policy paradigm knew that an economy or social system could not function well in their absence. Even the scandalous Mandeville reassured his readers on this point:

I lay it down as a first Principle, that in all Societies, great or small it is the Duty of every Member of it to be good; that Virtue ought to be encouraged, Vice discountenanced, the Laws obey'd and the Transgressors punished. (Mandeville (1988b):407)

The “natural liberty” that Smith endorsed was constrained by morality. His famous “Every man ... is left perfectly free to pursue his own interest in his own way” was conditioned in that very sentence by the proviso: “as long as he does not violate the laws of justice,” meaning that “every member of the society [is to be protected] from the injustice or oppression of every other member of it.” (Smith (1776):Book IV) Mandeville expressed the same idea equally memorably, analogizing human proclivities to the unruly growth of an untended grape vine: “So Vice is beneficial found, when its by justice lopt and bound.” Mandeville (1924)

The classical economists were thus quite aware that what later came to be called their *Homo economicus* was a simplification at variance with what they knew about human behavior, but one that would clarify among other things how policies that altered economic incentives would affect behavior. Here is John Stuart Mill (1844): 97, among the last of the classical economists laying down the boundaries and key assumptions of the discipline that were to remain with us until very recently::

[Political economy] does not treat of the whole of man's nature ... it is concerned with him solely as a being who desires to possess wealth, ... it predicts only such

...phenomena ...as take place in consequence of the pursuit of wealth. It makes entire abstraction of every other human passion or motive.

He termed this “an arbitrary definition of man.”

The advent of the neoclassical school in the late 19th century did not change the status of self interest as a handy but empirical false abstraction. F. Y. Edgeworth, a founder of the neoclassical paradigm expressed this view in his *Mathematical Psychics*: Edgeworth (1881)104): “The first principle of economics is that every agent is actuated only by self-interest.” In the passage immediately above this statement, Edgeworth refers to the “political struggle for power as well as the commercial struggle for wealth” And he recognizes that as a matter of fact “the happiness of others as compared by the agent with his own, neither counts for nothing, not yet ‘counts for one’”. But political economy could study the effects of incentives that appeal to the wealth maximizing side of individuals without reference to the other motives that Mill and Edgeworth fully recognized but took to be beyond the purview of the discipline.

What the classical economists and most economists since missed is the possibility that moral behavior would be affected -- perhaps adversely -- by incentive-based policies designed to harness self-interest. “Why should it be,” Kenneth Arrow asked, “that the creation of a market for blood would decrease the altruism embodied in giving blood?” (Arrow (1972):351) Until recently, most economists have been sure enough of the answer that they have not bothered with a reply. To Arrow it was “really an empirical question, not a matter of first principles.”

But to most economists, the unspoken first principle was that incentives and morals are additively separable in the mathematical sense of the term, meaning that the effects of variations in the one did not depend on the level of the other. When two things are additively separable they are neither synergistic – each contributing positively to the effect of the other, like a duet being better than the parts taken separately – nor the opposite.

I will call this the separability assumption, and return to it in subsequent chapters. You have already seen where it can go wrong. The firemen’s sense of duty towards the citizens of Boston was not separable from their self interested regard for their own pay: a policy addressed to the latter appears to have diminished the former. The whole in this case was less than the sum of the parts, and it is exactly this possibility that the Machiavellian paradigm overlooked.

Separability and its violations will play a big part in what follows, and it is a mathematical

concept that is hardly an everyday word; so it is worth pausing to consider an example. Amy Wrzesniewski, Barry Schwartz and their co authors studied the motives that led young men and women to attend the U.S. Army's West Point academy (Wrzesniewski, Schwartz, Cong et al. (2014). They used questionnaires administered by the institution to nine annual cohorts of incoming cadets to assess whether the individual had sought admission for instrumental motives ("to get a better job," the "overall reputation of West Point") or intrinsic motives ("desire to be an Army officer," "personal development"). They then followed the 11,320 cadets for a decade after graduation to see if these motives for admission correlated with later success..

Figure 2.1 presents shows how having strong instrumental motives is statistically associated with becoming a commissioned officer (one of the measures of success), for cadets with high intrinsic motives for attending West Point, median intrinsic motives, and low intrinsic motives. What do the data show?

First, notice that for those with a mean level of instrumental motives (the zero on the horizontal axis) having strong intrinsic motives is associated with a substantially increased likelihood of making commission. Second, for those with very low intrinsic motives (the bottom line), stronger instrumental motives were also associated with an enhanced likelihood of making commission . But the big news is the third observation: for cadets with median or high intrinsic motives, having in addition strong instrumental motives was associated with worse performance. Or to put it differently: for those lacking instrumental motives, the strength of intrinsic motives made a substantial difference in performance, while for those with high levels of instrumental motives, having, in addition strong intrinsic motives added little or nothing to the chance of gaining a commission. In this case intrinsic and instrumental motives were not additively separable, they were substitutes.

How might the military academy use this information to train officers? If it believed that intrinsic motives were largely absent from its pool of potential recruits, then it would appeal to the instrumental motives of the cadets, stressing the job opportunities in the armed forces and the value of West Point's reputation in getting jobs outside the military. But if it knew (correctly) that many cadets are highly idealistic young men and women wishing to serve their nation, it would downplay the appeal to instrumental motives.

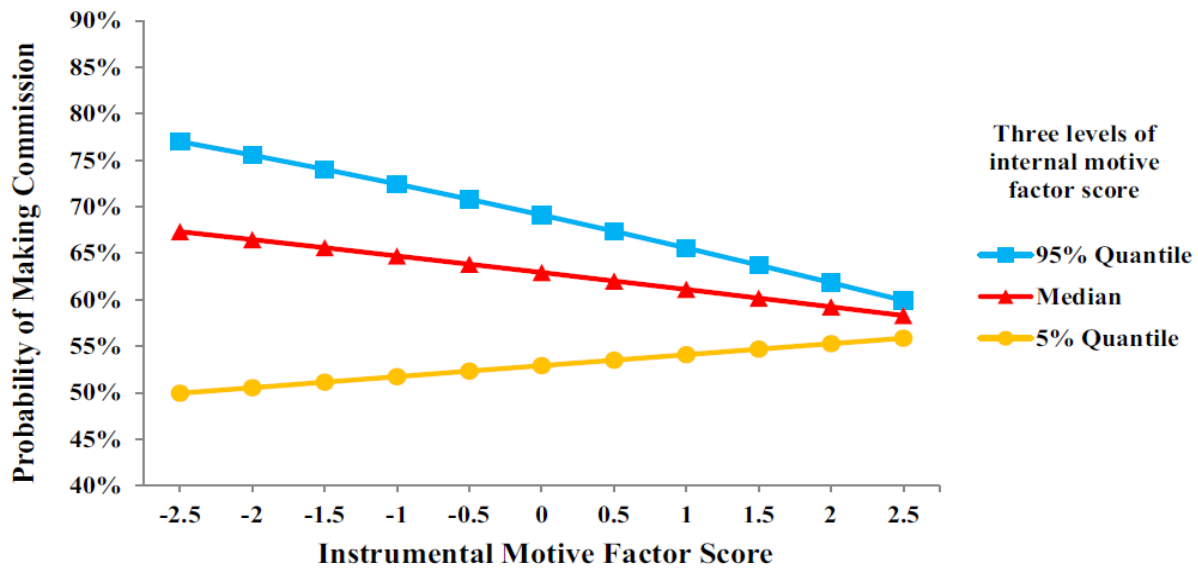


Figure 2.1. Instrumental and intrinsic motives as substitutes in the post graduation performance of West Point cadets. The instrumental motive factor score measures self interested reasons for attending West Point. The internal motive factor score measures the intrinsic motives mentioned in the text. “Making commission” means becoming a commissioned officer.

Economists have routinely either made the first assumption (that intrinsic motives are absent) or that the two sets of motives are separable so that (in the figure) all three lines would have been upward rising and parallel. As a result of this implicit separability assumption economists failed to take account of the conditions under which the ethical and other-regarding concerns would flourish and favorably affect societal outcomes, and how using incentives to harness self-interest to the public good might attenuate civic virtue.

MARKETS AS MORALITY-FREE ZONES.

Since the late 18th century, economists, political theorists and constitutional thinkers who have embraced the Machiavellian paradigm have taken *Homo economicus* as their working assumption about economic behavior. Partly for this reason, competitive markets, well-defined property rights, as well as efficient, (and since the 20th century) democratically-accountable states

are now seen as the critical ingredients of governance. Good institutions thus displaced good citizens as the *sine qua non* of good government. In the economy, prices would do the work of morals.

It was a short step to thinking that while ethical reasoning and concern for others should inform ones actions as a family member or citizen, the same did not go for shopping or in the workaday world of making a living. Lewis Carroll's *Alice* had taken the economists' message to heart. When the Duchess exclaimed "Oh, 'tis love, 'tis love makes the world go round," Alice whispered "Somebody said that it's done by everyone minding their own business." (Carroll (1982):104

How could people minding their own business take the place of love? The classical constitutional challenge posed by Bentham, Hume, Smith and others constitutes the Holy Grail that still motivates policy design. The quest is to find laws and other public policies that would simultaneously facilitate peoples' private pursuit of their own ends, while inducing each to take adequate account of the effects of their actions on others. This can be done if each actor internalizes the entire costs and benefits of an action, including the effects on others, rather than taking account merely of one's private benefits and costs. Morals are one way to pursue this objective (or as the Duchess put it, "love").

The other way – "everyone minding their own business" -- requires two things. First everything that matters to people when they decide what to must have a price. Then, second, taxes, subsidies and other policies must be devised to affect prices so that the price a buyer pays to acquire a good includes all of the costs incurred by anyone as a result of its production, and the price a seller receives likewise includes all of the benefits (to the buyer and to others) of making the good available. In this case, prices measure the entire social costs and benefits of the good's production and distribution (rather than just the private costs to the buyer and seller).

As a result "everyone minding their own business" means that simply by paying attention to prices people will be implementing (albeit unwittingly) the Holy Grail, namely, taking account of the effects of their actions on others. This is exactly what Mandeville had had in mind when he startled his readers with the claim that "the dextrous Management of a skilfull Politician" would allow the "Private Vices" to be "turned into Publick Benefits."

The remarkable idea behind Smith's invisible hand is that under the right institutions "the

dextrous Management of a skilfull Politician” might not even be necessary: prices determined by market competition can do the job, entirely unaided by subsidies, taxes, or other government policies. Smith stressed that competition among buyers and among sellers is critical to this result, and he warned about the ways in which monopolies can disrupt the workings of the invisible hand: “People in the same trade seldom meet together, even for merriment and diversion, but he conversation ends in a conspiracy against the public; or in some contrivance to raise prices.”

Economists since have stressed that the institutional conditions for the invisible hand to work go beyond competition and include the requirement that contracts be complete. This means that every aspect of an exchange – anything valued by either the exchanging parties or anyone else – have a price that is included in a contract governing the exchange. Complete contracts assign claims and liabilities so that each actor “owns” all of the benefits and costs resulting from his or her actions, including those conferred or imposed on others.

If contracts were complete, then competition among self-interested individuals would implement outcomes that are what is termed Pareto-efficient, meaning that there exists no other technically feasible outcome in which at least one individual would be made better off without anyone being made worse off. Kenneth Arrow and Gerard Debreu proved an “invisible hand theorem” (called the Fundamental Theorem of Welfare Economics) to this effect, for which they were awarded a Nobel prize.

The axioms of the invisible hand theorem (including the assumption that contracts are complete) did more than to clarify an idealized world in which one could dispense with the need for government intervention to ensure efficient outcomes in the economy. Less heralded but more important for our purpose is that in this world it seemed that one could also dispense with virtue. Another Nobel Laureate, James Buchanan, described his visit to a farm stand near his home in Blacksburg, Virginia:

I do not know the fruit salesman personally, and I have no particular interest in his well-being. He reciprocates this attitude. I do not know, and have no need to know, whether he is in the direst poverty, extremely wealthy, or somewhere in between... Yet the two of us are able to...transact exchanges efficiently because both parties agree on the property rights relevant to them. (Buchanan (1975):71)

Markets thus came to boast a kind of moral extra-territoriality akin to the hiatus of national sovereignty claimed by foreign embassies. The voluntary nature of transactions and the efficiency

of the results (at least under assumptions of the theorem) made competitive exchange a special domain, one in which one could suspend the substantive normative standards commonly applied to relationships among citizens or family members. Generalizing Buchanan's observation and designating the market as a "moral free zone" the philosopher David Gauthier (1986):84,96 held that: "Morality arises from market failure. ... morality has no application to market interactions under the conditions of perfect competition." And so, avarice, now repackaged as self interest, was tamed, transformed from a moral failing to just another kind of motive, like the taste for ice cream.

KNAVEONOMICS

But what if, unlike Buchanan and his fruit seller, it is not the case that "both parties agree on the property rights" relevant to the exchange? This will be true when contracts are not complete and there are some aspects of an exchange that are not priced: you breathe my second hand smoke; farmer Jones' bees pollinate farmer Brown's apple trees. When Jones exchanges his honey for Brown's apples, he cannot also charge for the free services provided by his bees. Their assistance to farmer Brown is a spillover (termed an external economy), that is, a direct effect between economic actors that is not priced and covered by (is "external" to) the terms of market exchange. The bee services are left out of the contract and are not included in the price at which Jones sells his honey. As a result Jones' private revenues from the sale of apples fall short of the total social benefits that his farm produces (the bee services are not included). The result is a market failure, and as Gauthier noted a role for morality in economic interactions reasserts itself.

In the early 20th century Alfred Marshall and Arthur Pigou spelled out the microeconomic logic for nonetheless letting prices do the work of internalizing the effects of one's actions on others, even when markets fail. Where contracts are incomplete, they advocated taxes on industry for the environmental damages (external diseconomies) it imposed on others, and subsidies for a firm's worker training activities that benefit other firms when workers change jobs. Farmer Jones would get a subsidy equal to the value of his bee's pollination service to Farmer Brown, so that Jones' revenues (including the subsidy) would now measure the full social benefits of his apple production activities. What came to be called optimal taxes and subsidies were those that recompensed an economic actor for the benefits that his actions conferred on others and made him

liable for the costs of his actions borne by others, when these benefits and costs would not otherwise be accounted for in the actor's private revenues and costs.

Green taxes that "make the polluter pay" for environmental spillovers are an example. Where feasible, optimal incentives of this type would exactly implement Bentham's Duty and Interest Junction principle: altering the material incentives under which the individual acts so as to align self interest with public objectives. The optimal taxes and subsidies advocated by Marshall, Pigou and economists since are thus a substitute for complete contracts, extending the logic of the invisible hand to cases in which its assumption are violated.

The result is a coherent guide for Machiavelli's prince, clarifying what is required to induce citizens to act if they were good, namely to provide incentives and constraints such that a self regarding individual would act as if he valued the effect of his actions on others in the same manner that those who are effected would evaluate them. The job description of the wise policy maker in this case would no longer that of Aristotle's Legislator tasked with uplifting the population. The job could be more like that of Machiavelli's prince – ordering the right laws to induce the citizens to act as if they were good.

Bucolic examples like farmer Jones' bees and Brown's apple blossoms were for a century a textbook stable, but incomplete contracts are more the norm than the exception. The reason is that information about the amount and quality of the good or service provided in an exchange is either asymmetric or non-verifiable, that is, it is not known to both parties, or even if known it cannot be used in the courts to enforce a contract. Where this is the case, there will be some aspects of an exchange that are not in the contract, or as Emile Durkheim put it: "Not everything in the contract is contractual..." (Durkheim (1967):189) As a result market failures are not confined to the well known cases of environmental spillovers, but occur in the workaday exchanges essential to the functioning of a capitalist economy: labor markets and credit markets.

Contractual incompleteness occurs in these two cases due to the impossibility of writing an enforceable contract that specifies that the employee will work hard and well, and the fact that credit contracts (e.g. the stipulation that the borrower is not to invest the loan in a Ponzi scheme) cannot be enforced if the borrower is broke (Bowles (2004).) Contracts are also incomplete (or non-existent) in team production processes such research and the provision of legal services, and the voluntary provision of public goods such as neighborhood amenities or

adherence to social norms.

The labor and credit market examples share a common structure: a principal (the employer, the lender) wishes to induce the agent (the employee, the borrower) to act in way beneficial to the principal but counter to the agent's interests (work hard, use the money borrowed in a prudent way). But because the information about work effort or the prudent use of the money is either not known to the principal or not admissible in court, the conflict of interest between the two cannot be resolved by specifying the terms of a complete and enforceable contract.

The *de facto* terms of the exchange are determined in important measure by the strategic interaction among the parties, not by the courts. The same problem arises when a farmer pays a share of his crop to the landowner. The problem common to these three cases is that the agent does not own the results of his or her actions: the lender takes the loss if the borrower cannot repay due to the agent's choice of an overly risky project, the employer enjoys most of the benefits of the employee's hard work and likewise cannot recover back wages if the work has not been done. .

Thus in much of a modern capitalist economy the complete contracts assumption of the invisible hand theorem are violated. The great contribution of the mathematical representation of the workings of a market economy, allowing a proof of the "invisible hand theorem" was to clarify just what Adam Smith's idea required. Here is how Arrow later described the contribution of his invisible hand theorem:

There is by now a long and ...imposing line of economists from Adam Smith to the present who have sought to show that a decentralized economy motivated by self interest and guided by price signals would be compatible with a coherent disposition of economic resources that could be regarded in a well-defined sense as superior to a large class of possible alternative dispositions. .. it is important to know not only whether it *is* true but whether it *could be* true. (Original emphasis).(Arrow and Hahn (1971):vi-vii)

The quite restricted conditions under which it "could be true" show how empirically unlikely it is that a policy of laissez faire would implement an efficient outcome. I know from firsthand experience that modern textbook writers struggle to find empirical examples even of a single market that approximates the model on which the theorem is based. (The somewhat lame and probably inappropriate example my textbook co authors and I used is the market for beer.)

MACHIAVELLI'S MISTAKE

But perhaps the kinds of optimal subsidies and taxes proposed by the Marshall-Pigou tradition can provide surrogates for the complete contracts found to be lacking in credit, labor and other markets. If so the prices could nonetheless do the work of morals; the domain in which “people minding their own business” is a good policy would then be vastly expanded

Many ingenious systems of incentives have been proposed to this end by practitioners in a field called mechanism design. But as we will see in the penultimate chapter, the assumptions required to make these clever mechanisms work are about as remote from real economies as are the axioms on which the Fundamental Theorem of Welfare Economics is based. Thus mechanism design has not yet and is not likely to devise incentives that make ethical and other regarding motives redundant.

Excepting on the whiteboards of economics classrooms, in the real world people try to avoid dealings with *Homo economicus*. Employers still prefer hiring workers with a strong work ethic; banks still prefer to lend to people whom they trust to conduct their businesses as proposed, rather than adopting self interested but riskier projects.

“[T]he contract itself is not sufficient,” Durkheim wrote,|. “...regulation of the contract ...is social in origin.” (Durkheim (1967):189) He was repeating the commonplace that handshakes matter; and where they do not, the economy underperforms. In a paper explaining his invisible hand theorem Arrow (1971):22) writes:

In the absence of trust ...opportunities for mutually beneficial cooperation would have to be foregone...norms of social behavior, including ethical and moral codes (may be) ...reactions of society to compensate for market failures

In other words, because contracts are incomplete, morals may sometimes do the work of prices rather than the other way around.

Arrow’s point is that social norms and moral codes can attenuate market failures when they have the effect of internalizing the benefits and costs that one individual’s actions confer or inflict on others. Despite contracts being incomplete, the major markets of a modern economy – the markets for labor, credit, and knowledge --function as well as they do because social norms and other-regarding motives foster a positive work ethic, an obligation to tell the truth about the qualities of a project or a piece of information, and a commitment to keep promises.

The importance of norms and other social motives extends far beyond what we usually

term market failures to encompass any of the arenas of social life in which the effects of one's actions on others is not governed by contract. Included are the long term climatic effects of one's lifestyle choices, the creation of drug resistant "super bugs" through the opportunistic use of antibiotics and the traffic congestion resulting from one's choice of travel plans and ways of getting around.

The importance of social norms in underwriting good governance is likely to increase over time as these and other problems arising from non-contractual social interactions pose ever larger challenges to human well being. The changing nature of work itself – from producing things to processing information and providing care for example – also suggests that contractual incompleteness and hence the essential nature of social norms will increasingly characterize our economies.

The classical economists were right in thinking that ethical and other regarding motives would be insufficient for the good governance of an economy in which many, even most, interactions are among strangers. And nobody now doubts Smith's once remarkable claim that a citizen's self interest could be harnessed to "promote an end that was no part of his intention" But Joseph Schumpeter was also right more than half a century ago when he wrote that "no social system can work ...in which everyone is ...guided by nothing except his own ...utilitarian ends" (Schumpeter (1950):448). Schumpeter was not referring to families or the polity, where the importance of ethical and other regarding motives is widely recognized, but to the workings of a capitalist firm. Thus while the moral sentiments are insufficient, it seems also true that that, consistent with what Arrow wrote, they are nonetheless indispensable to a well governed economy and society. And as will become clear presently, policies to harness self interest may also displace the moral sentiments.

Machiavelli's radical notion (anticipating Rousseau by more than two centuries, and mechanism design by more than four) was that his Prince should devise a structure of governance such that people with "ordinary and natural humors" will nonetheless choose to act in ways that result in a well governed city. The mistake in the title of this book is not a flaw in Machiavelli's writing but rather a radical extension of his notion by economists who combine a professed indifference toward the nature of individual preferences with an overconfidence in the ability of clever incentives to induce even the entirely amoral and self interested citizen to act in the public

interest.

The political philosopher Leo Strauss (1988):49 traced the genesis of this way of thinking to the sixteenth century Florentine: “Economism is Machiavellianism come of age.” But while the origins of the thinking that Strauss criticized can indeed be found in Machiavelli’s writings, Machiavelli, unlike much of modern economics, cautioned against imagining that good governance was possible in a self interested (“corrupt”) citizenry:

... neither laws nor orders can be found that are enough to check a universal corruption. For as good customs have need of laws to maintain themselves, so have laws have need of good customs so as to be observed. (Machiavelli (1996):I.18)

This, however, is not the Machiavelli that was to be the harbinger of a way of thinking about good governance that imagined that one could dispense with virtue entirely.

This is a mistake because ethical and other regarding motives are essential to a well governed society and are likely to be even more so in the future, and because, as we will now show, policies that ignore this fact and are indifferent to the preferences that motivate how people act may compromise these essential moral and other regarding predispositions. This is the reason that, like Aristotle's Legislator, policy makers today should be concerned about the firemen’s response to the Commissioners punitive incentives and about the parents arriving even later at the day care centers after the fine was imposed. In the eyes of the Legislator, the Prince’s tool kit, designed as it was for knaves or the wicked, may be part of the problem. Behavioral experiments provide evidence of why this is indeed the case.

III

MORAL SENTIMENTS AND MATERIAL INTERESTS

Thomas Schelling's email to me about the experiments that I report in this chapter recalled the "exciting and stimulating times" he had spent in the early 1950s White House as a young staffer in the Executive Office of the President. "People worked long hours," he told me, "and felt compensated by the sense of accomplishment, and ... personal importance. Regularly a Friday afternoon meeting would go on until 8 or 9, when the chairman would suggest resuming Saturday morning. Nobody demurred. We all knew it was important, and we were important. ... What happened when the President issued an order that anyone who worked on Saturday was to receive overtime pay...? Saturday meetings virtually disappeared."

Was Schelling's experience atypical?

Incentives work, often affecting the targeted behavior almost exactly as conventional economic theory predicts: textbook examples include the work response to incentives of Tunisian sharecroppers and American windshield installers, as well as experimental subjects (Laffont and Matoussi (1995), Lazear (2000), Falkinger, Fehr, Gaechter et al. (2000).)

But sometimes blackboard economics appears to fail. Substantial rewards for high school matriculation in a randomized experiment in Israel had no impact on boys and little effect on girls, except among those already quite likely to matriculate (Angrist and Lavy (2009).) Large and in most cases immediate cash payment in return for tested scholastic achievement in 250 urban schools in the U.S. were almost entirely ineffective, while incentives for student inputs (reading a book, for example) had the expected, if modest effects (Fryer (2010).) In an unusual natural experiment, the imposition of fines designed to shorten hospital stays in Norway had the opposite effect (Holmas, Kjerstad, Luras et al. (2010)). In England hospital stays were greatly reduced by a policy designed to evoke shame and pride in hospital managers rather than the calculus of profit and loss (Besley, Bevan and Burchardi (2009).)

Jewish West Bank settlers, Palestinian refugees, and Palestinian students were asked how angry and disgusted they would feel or how supportive to violence they might be if their

political leaders were to compromise on contested issues between the groups (Ginges, Atran, Medin et al. (2007)). Those who regarded their group's claims (on Jerusalem, for example) as reflecting "sacred values" (about half in each of the three groups) expressed far greater anger, disgust and support for violence if the compromise were accompanied by a monetary compensation for their own group than if no compensation were offered.

A similar reaction may explain the response of Swiss citizens in a survey of their willingness to accept an environmental hazard: when offered of compensation their resistance to the local construction of a waste facility increased (Frey and Jegen (2001).) Many lawyers believe (and experimental evidence suggests) that inserting explicit contractual provisions in the case of breach increases the likelihood of breach (Wilkinson-Ryan (2010)).

These examples cast doubt on the classical separability assumption – that incentives and the moral sentiments are simply additive in implementing desirable outcomes. And as I will show in this chapter, harder evidence comes from laboratory experiments played for significant sums of money among for the most part anonymous subjects. We will also see later that the experiments appear to predict what people do outside the lab. Among Brazilian fishermen, for example, those who cooperate in prisoners' dilemma experiments on shore, adopt more environment friendly traps and nets when they take to their boats.

The fact that in the lab and the street, appeals to material self interest appear to compromise the moral sentiments would be of no concern to the Legislator if there were little or nothing to crowd out. But this is not the case. Natural observation and recent experimental data indicate that in most populations few individuals are consistently self-interested, and that moral and other regarding motives are common.

HOMO ECONOMICUS HAS COMPANY

In the prisoners' dilemma game, defection maximizes each player's payoffs irrespective of what the other does; and it does not take a game theorist to figure this out. But when the game is played with real people, between 40 and 60 percent of the players cooperate rather than defecting. (Fehr and Fischbacher (2001)). Most subjects prefer the mutual cooperation outcome over the higher material payoff they would get by defecting on a cooperator and they are willing to take a chance that the other feels the same way (and also is also willing to take a chance.). When they

defect, it is most often not because they are tempted by the higher payoffs they would receive but instead because they cannot be sure what the other will do, and hate the idea that other would exploit their cooperation. We know this because when the prisoners' dilemma is played sequentially (one person moves first rather than simultaneously with the other) the second mover more often than not reciprocates the first mover's move, cooperating if they have done so, and defecting if they have not.

The experiments discussed in this and later chapters are listed in table 3.1 along with the pages on which they are most fully described.

Using data from a wide range of experiments, Ernst Fehr and Simon Gächter estimate that between 40 and 66 percent of subjects exhibit reciprocal choices, meaning that they returned favors even when they would gain higher payoffs by not doing so. The same studies suggest that between 20 and 30 percent of the subjects exhibit conventional self-regarding preferences (Fehr and Gächter (2000b), Camerer (2003)). Less than a fifth of experimental subjects made self interested choices in the Trust Game of Armin Falk and Michel Kosfeld described below

George Loewenstein and his coauthors distinguished among the following types in their experiments:

saints consistently prefer equality, and they do not like to receive higher payoffs than the other party even when they are in a negative relationship with the opponent...*loyalists* do not like to receive higher payoffs in neutral or positive relationships, but seek advantageous inequality when ..in negative relationships. ..*ruthless competitors* consistently prefer to come out ahead of the other party regardless of the type of relationships.(Loewenstein, Thompson and Bazerman (1989):433)

Of their subjects, 22 percent were saints, 39 percent were loyalists, and 29 percent were ruthless competitors (the rest could not be classified).

As this experiment makes clear, when people do not act as ruthless competitors, they depart from the standard economic model in a multiplicity of ways. This and other experiments show that some are unconditionally altruistic, simply valuing the benefits received by others, or their well being. Others reciprocate good deeds, expressing a conditional form of altruism. Others dislike inequality, apparently not because of a concern about the well-being of others because of a commitment to justice. Thus experiments show that while *Homo economicus* is among the

dramatis personae on the economic stage, he is not alone, and indeed is often seriously outnumbered.

<i>Game</i>	<i>pp.</i>	<i>Values measured</i>
One-shot Prisoner's dilemma		Player's reciprocity conditional on their beliefs about the actions to be taken by the other; effect of market framing on values
Gift exchange		Reciprocity and expectations of reciprocity
Trust (with and without fines)		Investor: generosity or expectations of reciprocity. Trustee: reciprocity
Dictator		Unconditional generosity
Third-party Punishment		Third party: willingness to pay to punish violations of fairness in the treatment of others. First party: effect of expected punishment
Ultimatum		Proposer: unconditional generosity or belief in the fair-mindedness of the Respondent. Respondent: fairness, reciprocity
Repeated Public Goods		Altruism, reciprocity conditioned on the past actions of others
Public Goods with Punishment		Contributor: unconditional generosity or belief in the willingness of others to punish unfairness, shame when violating a social norm, . Punisher: fairness, reciprocity

Table 3.1 Values indirectly measured in experimental games. The indicated values provide plausible explanations of experimental behavior when this differs from behavior expected of an individual seeking to maximize game payoffs (and believing others to be doing the same). The second column gives the page numbers on which the structure of each game is explained. Table A1 in the appendix gives more detail on the structure of these games, the kinds of other regarding or ethical behavior that may be observed, and what would constitute a violation of the separability assumption.

I use the term “social preferences” to refer to motives such as altruism, reciprocity, intrinsic pleasure in helping others, inequity aversion, ethical commitments and other motives that induce people to help others more than would an individual who like Mill’s citizen sought solely to maximize his own wealth or material-payoff . My use of the term is thus not restricted to cases in

which the actor assigns some value to the payoffs received by another person. I use the broader definition of social preferences because moral, intrinsic, or other reasons unrelated to a concern for another's payoffs often motivate people to help others, adhere to social norms, and act in other pro-social ways even when it is personally costly to do so. A person, for example, may adhere to a social norm not because of the harm that a transgression would do to others, but because of the kind of person she would like to be. Helping the homeless may be motivated by what James Andreoni calls the "warm glow" of giving rather than a concern with the wellbeing of the poor (Andreoni (1990)).¹

Knowing that the use of incentives may undermine one or more of these dimensions of social preferences provides a warning to the Legislator, but not much guidance on how to design incentives and other policies in the presence of crowding out. To decide whether to use incentives, and if so, what they should be the Legislator has to know more about the citizens' behavior in the absence of incentives and her response to the various kinds of incentives that might be implemented. This requires an understanding of how incentives work, and why when crowding out occurs they may work less well or even fail to work.

CROWDING OUT (AND IN)

To provide the Legislator with this information I will consider a paradigmatic problem facing policy makers: how to get citizens to contribute to some public good when it is costly for them to do so. Examples would include paying taxes, obeying traffic regulations, and contributing to neighborhood improvement projects.

This can be represented as a Public Goods Game. An individual may choose to bear a cost in order to take an action – a contribution to the public good such as the disposal of trash in an environment-friendly manner – that confers benefits on others. The individual herself will, like all citizens, benefit from the public good, but to represent the difficulty of securing the provision of public goods we assume that the cost of her contributing is greater than the benefit she will receive. Thus while everyone contributing is the best outcome (it maximizes the total payoff of the public), for each citizen not contributing at all is the individually payoff maximizing choice, and this is true independently of what the other citizens do. (Technically, not contributing is the dominant strategy for a payoff maximizer, just as it is in the Prisoners Dilemma). The Public Goods game is a

prisoners' dilemma with more than two players. In addition to protecting the environment, other problems that often take the form of a public goods problem are the voluntary payment of taxes, limiting one's carbon footprint, upholding social norms, producing new knowledge, maintaining public safety and acting so as to maintain the good reputation of one's group.

The citizen may be encouraged to contribute to the public good by a subsidy or other economic incentive. In what follows I will use the term incentive (without the adjectives explicit, economic, monetary, and so on) to mean an intervention that affects the expected material costs and benefits associated with the action. In the standard economic model, the story ends here: the subsidy reduces the net cost of contributing to the public good and the citizen contributes more as a result. But citizens also have social preferences (I'll nickname this heterogeneous mélange of motives simply as "values" here). And these may motivate taking actions that benefit others even at a cost to oneself.

The challenge facing the Legislator arises because values may be influenced (positively or negatively) by the use of incentives, that is when values and incentives are not separable. To see how, let the extent the action (denoted by a) be represented by a single number, and the same be true of both explicit incentives (s) and values (v). Non-separability occurs when the presence or extent of the incentive affects the individual's values.

This is illustrated in Figure 3.1 where, the upper route from incentives to the action – the pathway via “material cost net of incentive” is the one stressed by the Machiavellian paradigm. On this causal route, the incentive reduces the net costs of the public spirited action, raising the material payoffs that result from contributing, and as a result increasing the actor's motivation to provide the public good. The lower pathway in the figure – passing through “values” -- expresses the possibly negative effect of the incentives on the individual's pro-social values and hence the indirect negative effect of incentives on the citizen's contribution to the public good.

Economists following J. S. Mill in focusing on the citizen “solely as a being who desires to possess wealth” will routinely ignore this indirect effect either because they think it is not there, or because it is not economics. But it is definitely there, and as a result, has to be economics.

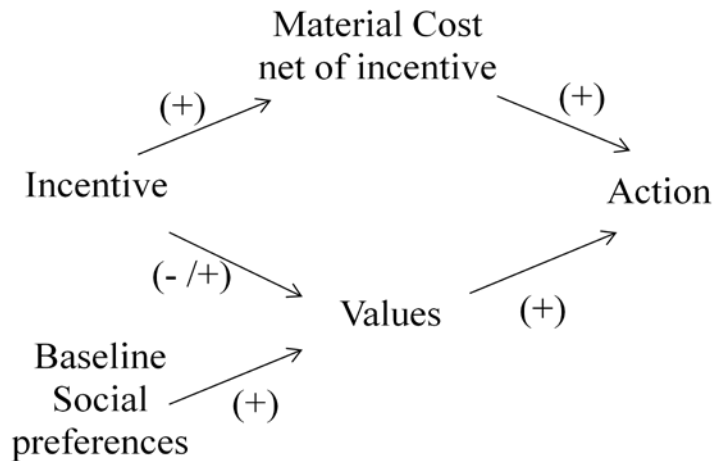


Figure 3.1 Incentives, values, and actions: the problem of non-separability

Due to the effect of incentives on preferences, the total – direct and indirect -- effect of the incentive may fall short of that which works directly on the costs and benefits of the targeted activity. In this case we say that incentives crowd out social preferences and that incentives and social preferences are substitutes: the effect of each on the targeted activity declines the greater is the level of the other. Where the effect on social preferences is positive we have the synergy that the Legislator seeks: crowding in occurs and social preferences and incentives are complements, the level of each enhancing the effect of the other.

The total effect of the introduction of an incentive on the action of the individual Δ^T is the sum of the direct effect of the subsidy (Δ^D , which must be positive) plus the indirect effect of the subsidy operating via its effect on values and their effect on the action, which may be of either sign. We have separability when there is no indirect effect so that $\Delta^T = \Delta^D$. This appears to be true of the effects of incentives on the work activity of American windshield installers, Tunisian farmers and the other” textbook” cases mentioned at the beginning of the chapter.

Where $\Delta^T < \Delta^D$ incentives and social preferences are substitutes (or are “sub-additive” or said to exhibit “negative synergy” or “crowding out”). This may have been the true of the surprisingly modest or even absent effects of financial rewards for school work mentioned at the outset.

Where the indirect effect is negative and large enough to offset the direct effect of the

incentive so that $\Delta^T < 0$, we have the much cited case of incentives that literally “backfire”, that is have the opposite of the intended effect, which I term “strong crowding out.” The response to incentives of the Boston firemen and the Haifa parents are examples. Where $\Delta^T > \Delta^D$ we have synergy between the two effects: incentives and social preferences are complements rather than substitutes, and are sometimes termed super additive.

Table 3.2 summarizes the relevant definitions and gives terms commonly used to refer to violations of separability. Note that crowding out does not require that the total effect of the incentive be negative, but instead only that it be less than would be the case if separability held.

	<i>Indirect effect, Δ^I</i>	<i>Terminology</i>
$\Delta^T = \Delta^D$	None	Additive separability; separability, additivity
$\Delta^T > \Delta^D$	Positive	Complementarity, synergy, super-additivity, crowding in
$\Delta^T < \Delta^D$	Negative	Substitutability, negative synergy, sub-additivity, crowding out
$\Delta^T < 0$	Negative, more than offsets direct effect	Strong crowding out; incentives are counter productive

Table 3.2. Incentives and social preferences: Additive separability and its violations. Note: Δ^T and Δ^D respectively are the total and the direct (partial) effect of the incentive on the action

If the introduction of an incentive were to reduce instead of increase contributions to the public good, we would surely have evidence of crowding out. But because this kind of strong crowding out is just an extreme manifestation of the problem, simply comparing contributions with and without incentives is not sufficient. It would miss all cases of non-strong crowding out, in which the effect of the incentive was in the intended direction, but it was not as large as would have been the case had social preferences and incentives been simply additive. Experimental evidence for crowding out therefore is based on a comparison in the presence or absence of an incentive of how closely the individual’s behavior approximates what an entirely amoral and self regarding person would do.

Here is an example of a subsidy that ‘worked’ but at the same time induced almost entirely selfish behavior in people who, in the absences of the incentive, had acted quite unselfishly. Juan

Camilo Cardenas and his coauthors implemented an experimental “public bads” game called the Common Pool Resource Game that is very similar in structure to the kind of real world commons problem faced by his subjects – rural Colombian eco-system users.(Cardenas, Stranlund and Willis (2000).)

In the experiment, Cardenas let the villagers choose how many “months” they would spend extracting resources from the “forest” (the common pool resource), There was a level of exploitation (one month per year) that if I were practiced by all, would maximize the total payoffs to the group. But each individual would do better by extracting much more than this social optimum. Over-extracting resources is thus a public bad.

The game represents the real problem of resource conservation faced by the villagers, and it has the same structure as a prisoners’ dilemma or a public goods game: each subject in the experiment would earn higher material payoffs by overexploiting the “forest” and this is true irrespective of what the others do. But all would do best if each limited their extraction. The villagers could easily determine the payoffs from the experiment (in real money) that they would get for every combination of what they and the others did. Each villager was randomly assigned to one of 14 groups in which they would play the experiment over a number of periods.

In the first stage of the experiment lasting 7 periods there were no incentives and no communication among the villagers. In this first stage the villagers on average extracted 44 percent less of the experimental “resource” than would have maximized their individual payoffs, given the extraction levels taken by the others.. This discrepancy between what an entirely selfish person would have done and what the villagers in fact did-- foregoing the material gains they would have had by extracting a bit more than three months more than they did, given what the others had done -- is a measure of their social preferences.

In the second stage of game (with 8 periods of play) Cardenas introduced two new treatments. In 9 of the groups the villagers were allowed to communicate with each other briefly before playing anonymously. These groups extracted a bit less under the communication treatment than they had in the no-communication stage, thus deviating even more from what a person who cared only about her own payoffs would do.

Members of the remaining 5 groups, the experimenter explained, were liable to pay a small fine (imposed by the experimenter) if it was found that they had over-extracted the resource when

they were monitored (which would occur with a probability known to the villagers). As expected, villagers in the groups subject to affine initially extracted much less than without the fine, showing that the fine had the intended effect. However, as the second stage of the experiment progressed those in the groups that were subject to the fine considerably raised their extraction levels. By the end of the second stage their levels of extraction were barely (and not statistically significantly) less from what an entirely self interested person would do.

Figure 3.2 shows the extent to which the villagers restricted their extractions from the “forest” to levels below the social optimum. It appears that the incentive worked, but that it almost entirely sidelined whatever motives had led the villagers in the absence of the incentive to forgo substantial individual gain by limiting their extraction levels while enhancing the material payoffs of others in the group.

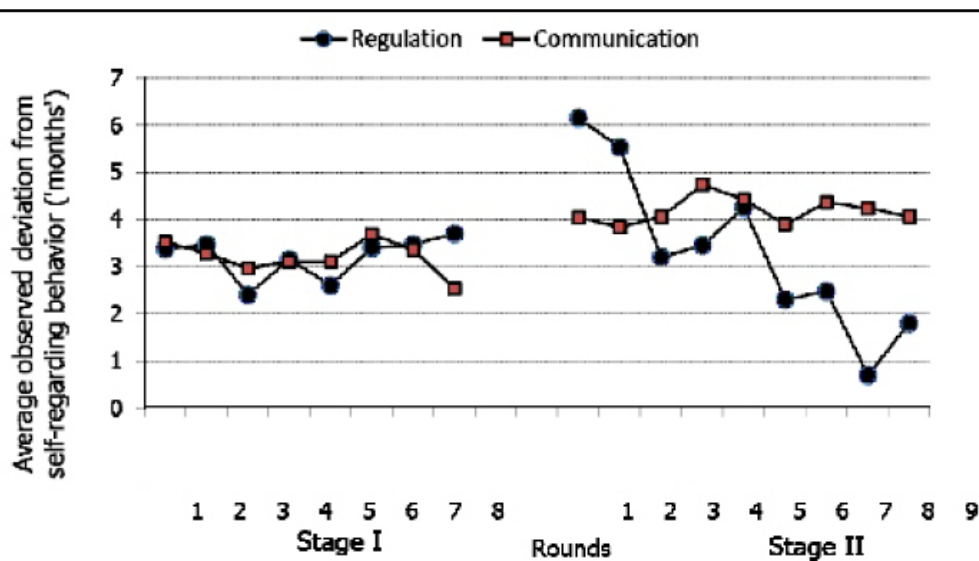


Figure 3.2. The effects of communication and economic incentives on the strength of social preferences Source: (Cardenas, Stranlund, et al. (2000).

CROWDING OUT: WHAT THE LEGISLATOR NEEDS TO KNOW.

What was it that eclipsed the villagers’ green-mindedness once the fines were announced? People often react negatively to the mere presence of incentives rather than their extent giving to charity when tax breaks are involved (whatever their magnitude) may feel different, or send a different signal (to others or oneself) than would be the case in the absence of these incentives. Even if this is the case, a larger incentive will work more than a smaller one. As with the Boston

firemen, a large enough penalty probably would have had the intended effect even though the lesser penalties announced by the Fire Commissioner backfired.

But the extent of the incentive may also matter. Thus the effects of incentives on social preferences may be either marginal (depending on the level of the incentive) or categorical (the presence of incentives affecting social preferences independently of their level) or a combination of the two.

Figure 3.3 shows the effect of variations in a subsidy on the action taken by an individual when either categorical or marginal crowding out holds, and under separability. Each line gives what is termed the individual's 'best response,' meaning the level of the action that will maximize the individual's utility (from both the incentives and from the individual's values) given the rate of subsidy indicated on the horizontal axis. The line is termed the best response function; its slope is the effect of the subsidy on the action. The steeper slope when separability holds indicates that the subsidy has a bigger effect in this case than when marginal crowding out occurs. When strong marginal crowding out holds, the effect is negative (greater subsidy induces a lesser contribution to the public good). Marginal crowding in would be indicated by a line steeper than the "separability" line (not shown.)

The intercept at the vertical axis of the best response function gives the citizen's contribution in the absence of any subsidy (the other regarding citizen contributing more than the self regarding when there is no subsidy.) The intercept labeled "other regarding contribution when $s = \epsilon$ " gives the contribution when a subsidy is offered but it is very small (ϵ means a number as close to zero as you wish, but not zero). The difference in the vertical intercepts under separability and categorical crowding out is measures the extent to which the mere presence of the subsidy diminishes social preferences.

Figure 3.3 would provide the Legislator with just the information he needs, indicating for each subsidy the contribution to the public good that could be expected depending on the nature and extent of crowding out. If estimates of the best response function showed that citizens were other regarding and that strong crowding out would occur were a subsidy to be offered, then the Legislator would abandon the use of incentives. If the Legislator knew that an incentive would categorically crowd out social preferences, then he would either implement a subsidy larger than s' in the figure, or no subsidy at all (subsidies between 0 and s' result in lower contributions to the

public good.

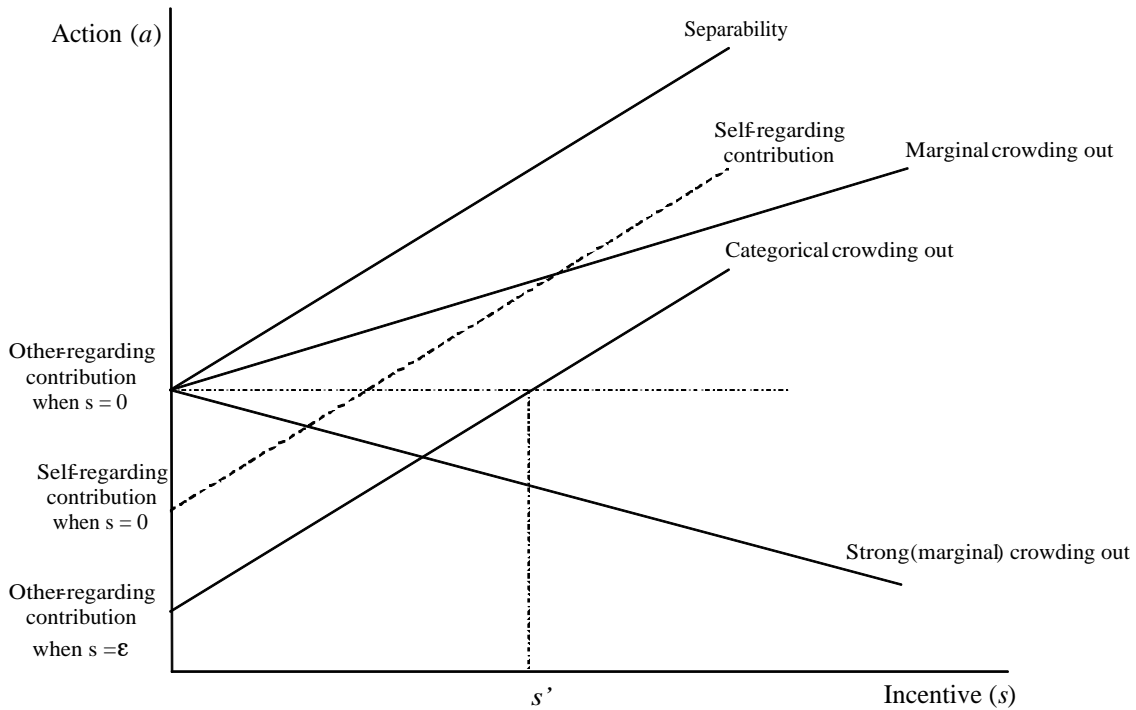


Figure 3.3. Citizen's contribution to the public good (a) under non-separability of incentives and values. Under separability (top line) categorical and marginal incentive effects are additive. Under strong crowding out the use of the incentive is counterproductive; this holds for all levels of the subsidy under the strong (marginal) crowding out function shown. Under categorical crowding out, incentives less than s' are also counterproductive in the sense that contributions are less than they would have been in the absence of incentives.

Under (non-“strong”) marginal crowding out, the incentive “works” in the sense that the direction of the effect is as intended, so the Legislator may decide to implement subsidy. He might even implement a larger subsidy in the presence of crowding out than he would have had separability obtained, a surprising fact that we will explore further when we return to the Legislator’s choice of the optimal subsidy when crowding out occurs in Chapter 7.

Our discussion of Figure 3.3 is not simply a thought experiment. Remarkably, an experiment shows that it can be estimated empirically and that both categorical and marginal

crowding out occur. Bernd Irlenbusch and Gabriele Ruchala implemented a public goods experiment in which the 192 German student subjects faced three conditions: no incentives to contribute and a bonus, given to the highest contributing individual, that was either high or low (Irlenbusch and Ruchala (2008).). Payoffs were such that even with no incentive individuals would maximize their payoffs by contributing 25 units. In the no-incentive case contributions averaged 37 units, or 48 percent above what would have occurred if the participants had been motivated only by their own material rewards from the game. Contributions in the low-bonus case were a bit higher but not significantly different from the no-bonus treatment. In the high-bonus case, significantly higher contributions occurred, but the amount contributed (53 units) barely (and insignificantly) exceeded that predicted for self-interested subjects (50 units).

In Figure 3.4 we use the observed behavior in the high and low bonus case along with the assumption that marginal crowding affects the slope of the citizens' best response function by a given amount (so that the function remains linear as in Figure 3.3) to estimate the marginal effect of the bonus. We find that a unit increase in the bonus is associated with a 0.31 increase in contributions. This contrasts with the marginal effect of 0.42 that would have occurred under separability, that is, had subjects without social preferences simply best responded to the incentive. Crowding out thus affected a 26 percent reduction in the marginal effect of the incentive.

The estimated response to the incentive also gives us the level of categorical crowding out, namely the difference between the observed contributions (37.04) in the absence of any incentive and the predicted contributions had an arbitrarily small incentive been in effect (the vertical intercept of the observed line in Figure 3.4) or 34.55. The incentive thus categorically crowded out 21 percent of the social preferences (measured by 12.04, the excess in contribution levels above the best response for self-interested subjects).

Categorical crowding out is evident in other experiments. In one, reported willingness to help a stranger load a sofa into a van was much lower under a small money incentive than with no incentive at all, yet a moderate incentive increased the willingness to help over the no incentive condition (Heyman and Ariely (2004). Using these data as we did in the Irlenbusch and Ruchala study, we estimate that the mere presence of the incentive reduced the willingness to help by 27 percent (compared to the no incentive condition).

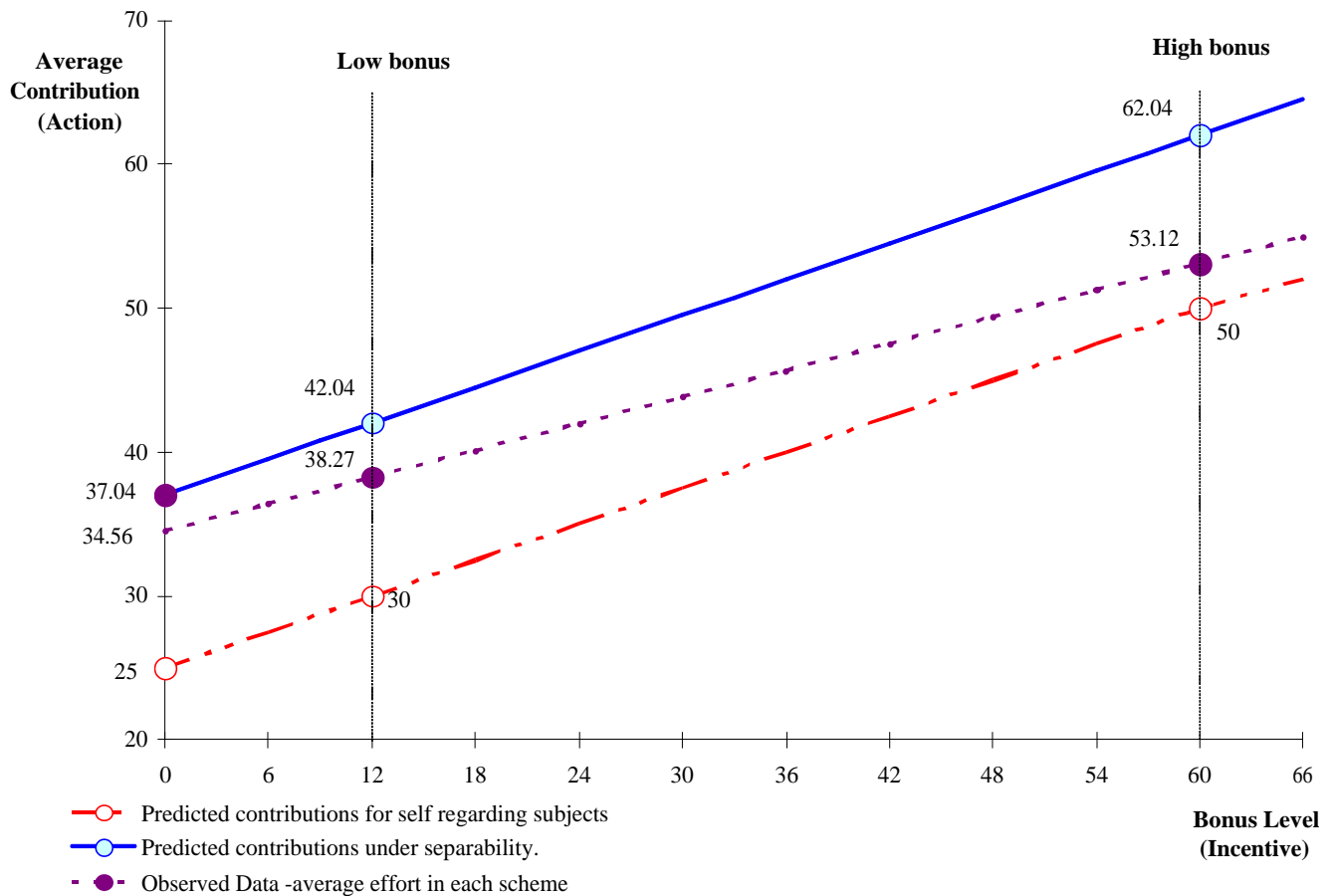


Figure 3.4. Categorical and marginal crowding out From Irlenbusch and Ruchala (2008) . Source: see text. The experimental design is a Public Goods Game comparing no incentive with two team-based compensation schemes with bonus for the highest contributor in the team. The bonus is self-funded (each member pays one-fourth of the bonus). The maximum contribution level is 120. .

Another experiment implemented by Cardenas (2004) allows us to distinguish categorical and marginal crowding, but here (as in some other experiments) we observe categorical crowding in. This is our first evidence that incentives and social preferences can sometimes be complements rather than substitutes. Because bringing this about is the aim of Aristotle’s Legislator, it is worth going explaining the result in some detail.

As in his earlier study Cardenas implemented an experimental Common Pool Resource Game similar to the real life conservation problem faced by his subjects – rural Colombian

eco-system users. In the absence of any explicit incentives, the villagers on average extracted 44 percent less of the experimental “resource” than would have maximized their individual payoffs, providing evidence of a significant willingness to sacrifice individual gain so as to protect the resource and raise group-average payoffs. When they were liable to pay a small fine (imposed by the experimenter) if they over-extracted the resource, as expected, they extracted even less than without the fine, showing that the fine had the intended effect.

The fact that the average extraction under the small fine treatment was 55 percent less than what would have maximized the payoffs of the subjects (when account is taken of the fine) suggests that the fine had increased the salience of the villagers’ social preferences (by 25 percent, if the 44% deviation from the self-interested behavior is taken as the measure of social preferences). Interestingly, raising the fine from a low to a high level had virtually no effect.

Variations in the fine thus did not work as an incentive, but rather (in Cardenas’ view) the very presence of the fine (high or low) was a signal, one that alerted subjects to the public good nature of the interaction and the importance of conserving the resource. We will present other examples of fines as messages; in some, with positive effects, as here, and in others with the more common crowding out effect. These cases hold important lessons for why incentives sometimes are counterproductive and how under well-designed policies incentives can crowd in social preferences. They also provide guidance for Aristotle’s Legislator

A SURPRISE FOR ARISTOTLE’S LEGISLATOR.

The Legislator knows that when private economic interactions fail to provide incentives for the efficient use of a society’s resources, his task is to design optimal taxes, fines, or subsidies. But in light of the evidence just presented there is the added twist that the preferences that will determine citizen’s responses to the policies depend on the incentives deployed. Of course the incentives that are optimal depend on the nature of the citizens’ preferences that result from this process, for these will determine the effects of the incentives.

This is not some impenetrable chicken and egg problem however. It just requires that for any policy under consideration, the effects of incentives on preferences be considered and the expected outcome of the policy take account of these indirect effects. The Legislator is thus not simply selecting, say, a tax rate, but rather a tax rate and the possibly altered distribution of

preferences in the population that is likely to result from its imposition. For policy evaluation, it is the joint effect of the pair [tax rate, preferences resulting from the tax rate] that must be considered.

Armed with the idea that policies may affect preferences, the Legislator can re-think the problem of optimal incentives. His intuition is that because crowding out reduces the effectiveness of incentives, he should use them less than would his Machiavellian counterpart who is unaware of these effects.

If crowding out is so strong that the incentive has an effect the opposite of its intent, the Legislator will of course abandon the use of incentives. So his intuition is correct in this case. But as we have seen, the effect of crowding out need not be literally counterproductive in this sense. And where the effectiveness of incentives is blunted but not reversed by crowding out, the implications for the optimal use of incentives are far from obvious.

There are two approaches the Legislator could take. He could seek to define incentives and other policies, which, rather than crowding out ethical and other regarding motivations, would instead have synergistic effects, crowding them in. We will consider how this might be done in the final chapter, but here we will describe a more modest approach that the Legislator might adopt. This is to simply take crowding out as a given and then determine the optimal level or mix of incentives, taking account of their effects on preferences.

To see how the Legislator would proceed we return to the Public Goods game studied in the previous two sections.² To make the problem interesting we assume that the citizens are not so ethical or other regarding that they fully internalize the benefits that their contributions confer on others. Were they to do this public goods provision would no longer be termed a problem. But assuming this not to be the case, under-contribution will occur: everyone would be better off if everyone contributed more. This provides the reason for considering introducing an incentive, which as before is some amount s proportional to the citizen's contribution. Citizens will typically differ one from another in their social preferences and hence in their susceptibility to crowding out, but we will postpone considering this aspect of the Legislator's policy making challenge until the last chapter.

The Legislator knows that, as in Figures 3.3 and 3.4, the effect of the subsidy depends on the nature and the extent of the crowding out problem. Consider the case (as in the empirically estimated citizen's best response function in Figure 3.4) where both categorical and marginal crowding out will occur if he implements an incentive. As a result the citizen's best response function is as given by the solid upward rising line in Figure 3.5. The Legislator then knows that he can choose any point on that line, including no subsidy at all ($s = 0$), and expect the citizens to contribute the amount given by the height of the line. His naïve counterpart, who I will call the Machiavellian planner, falsely believes that incentives and the citizens values are simply additive as motivations for contributions to the public good and as a result he thinks that his policy options lie along the dashed line labeled "separability".

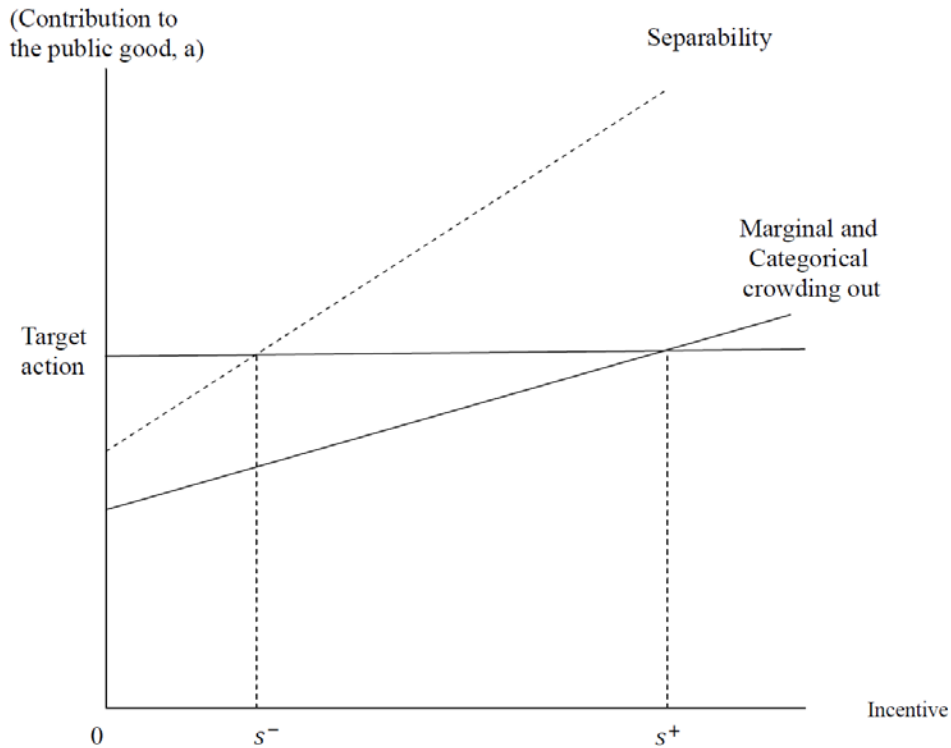


Figure 6.2. Underuse of the incentive by the Machiavellian mechanism designer. The Aristotelian Legislator, aware of the crowding out problem would select the subsidy, s^+ , greater than that chosen by Machiavellian mechanism designer who is unaware that the incentive and social preferences are not separable, s^- .

Because marginal crowding out occurs, the true effectiveness of the subsidy (known to the Legislator) is less than what his naïve counterpart believes (the best response function is flatter

than it would be under separability) and this could lead the Legislator to adopt a lesser subsidy, just as he imagined.. This is correct; but there is a second often overlooked effect.

Because the incentive is less effective (either categorically or marginally) that it would be in the absence of crowding out it follows that for any given level of the subsidy adopted, the extent of under-provision of the public good will be worse. The true best response function of the citizen is uniformly below the best response function imagined by the Machiavellian planner. As a result, the degree of public goods underprovision anticipated by the Legislator is greater than that the shortfall anticipated by his Machiavellian counterpart.

This is important because in a great many cases there are decreasing returns to the level of provision of the public good, meaning that increases in its provision are especially beneficial when it is more underprovided and correspondingly less valuable when underprovision is less. Where this is the case, the benefit of further increasing the the citizens' contribution is therefore correspondingly greater in the eyes of the Legislator than to the Machiavellian planner. This is true for any given level of the subsidy. The Legislator thus has reason to adopt a greater subsidy than would be adopted by the Machiavellian planner. This is the second effect of taking account of non-separability.

Whether the Legislator will indeed implement a larger subsidy than the Machiavellian planner, or, as the Legislator at first thought, the opposite, depends on which of these two effects is larger. The Legislator will choose a larger subsidy if the greater benefit of changing the citizens' behavior more than offsets the decreased marginal effectiveness of the subsidy.³ While the fact that the Legislator's recognition of the crowding out problem would lead to greater rather than lesser use of the subsidy seems odd, it is not. Think about the doctor who discovers that a treatment is less effective than he thought. Will he prescribe a lesser dose? Not necessarily; he may opt for a stronger rather than weaker dose, or for abandoning the treatment in favor of an alternative.

The logic that explains why the Legislator would implement a greater subsidy than his naïve counterpart is quite general; but it is especially transparent in cases where the Legislator would like to meet some specific target level of action. An example would be: every citizen should take at least four hours of first aid training (four hours would then be the target action). The Legislator believes that further training was of little additional benefit, and that those with less than four hours training were really not much different in their ability to help others during emergencies

from those who had not attended the course at all. This is just an extreme version of decreasing returns; there is no benefit of additional time in training greater beyond four hours. Then we have the situation depicted by the horizontal (“target action”) line in Figure 3.5, from which it is clear that to meet this target the Legislator will adopt the subsidy s^+ greater than s^- the subsidy adopted by the planner who is unaware of the crowding problem.

The same logic applies to the doctor and the Legislator: one may make more use of an instrument precisely *because* it is less effective. But the awareness of its ineffectiveness would also lead the doctor and the Legislator to seek other ways of accomplishing the desired end. Attendance at first aid courses might be promoted, for example, by clarifying how important it is during natural disasters that most people know the elements of first aid. The Legislator would be more likely to explore alternatives to the use of incentives than would be the Machiavellian planner. The reason is that he would be aware of the greater cost of hitting the target while relying exclusively on incentives.

So in the end, the Legislator’s intuition was half right: he is more likely to abandon the use of incentives entirely than is the naive Machiavellian planner who believes that social preferences and incentives are separable; but if he uses the incentive at all, whether he uses it more or less will depend on the nature and extent of the crowding out problem and the extent of decreasing returns to public goods provision.

THE LAB AND THE STREET

The experimental evidence for crowding out and the guidance it might give the Legislator would be of little interest if lab results did not reflect behaviors outside the lab. Testing for separability in natural settings is difficult, but generalizing directly from experiments even for phenomena much simpler than separability is a concern in any empirical study (Falk and Heckman (2009)) and is often unwarranted (Levitt and List (2007)). Consider, for example, the Dictator Game: typically more than 60 percent of the dictators allocate a positive sum to the recipient, and the average given is about a fifth of the endowment initially granted by the experimenter to the dictator. But we would be sadly mistaken if we inferred from this that 60 percent of individuals would spontaneously transfer funds to an anonymous passerby, or that the same subjects would offer a fifth of the bills in their wallet to a person who is homeless asking for help. Another

example: while pro-social behavior in an experiment by Benz and Meier (2008) was correlated with non-experimental behavior, subjects who reported that they had never given to a charity allocated 65 percent of their endowment to a named charity in a lab experiment.

A possible explanation of these discrepancies between experimental and real world behavior is that most individuals are strongly influenced by the cues of appropriate behavior offered by the situation in which an action is taken, and there is no reason to think that experiments are an exception to this context-dependent aspect of individual behavior. An experiment about giving may prompt giving.

External validity concerns arise from four aspects of human behavioral experiments that do not arise in most well-designed natural science experiments. First, experimental subjects typically know they are under an unknown researcher's microscope, possibly inducing different behaviors than would occur under total anonymity or under the scrutiny of neighbors, family or workmates. Second, experimental interactions with other subjects are typically anonymous and without opportunities for ongoing face to face communication, unlike many social interactions of interest to economists and policy makers. Third, subject pools— to date, overwhelmingly students -- may be quite different from the real-world populations of interest, due to age effects and to the process of recruitment and self-selection.

Finally, the experimental games involved are social dilemmas – prisoners dilemmas or public goods games – or tasks involving sharing with others – the Ultimatum Game and Dictator Game. These are settings in which social preferences are likely to be important and therefore there is something to be crowded out. But while we would be right in concluding from the experimental evidence that that incentives may crowd out blood donations or participation in community service projects, we might also wonder if the experimental evidence has as much to say about the effect of incentives on behavior while shopping or cleaning hotel rooms.

It is impossible to know whether these four aspects of behavioral experiments bias experimental results in ways relevant to the question of separability. For example, the fact that in most cases subjects are paid a “show up fee” to participate in an experiment might attract the more materially oriented who may be less motivated by social preferences subject to crowding out. Conversely, if potential subjects knew that the topic of the experiment was cooperation those who signed up might be atypically civic minded.

We can do more than speculate about these problems. Nicole Baran and her coauthors studied University of Chicago Graduate School of Business students who had acted with greater reciprocity in the Trust Game (those who as trustees most generously reciprocated large transfers by the investor). Baran asked if these reciprocators were also those most likely to contribute to the University upon graduation. They were (Baran, Sapienza and Zingales (2010)).

Similarly, among the Japanese shrimp fishermen that Jeffrey Carpenter and Erika Seki studied, those who contributed more in a public goods experiment were more likely to be members of fishing cooperatives that shared costs and catch among many boats than to fish under the usual private boat arrangements (Carpenter and Seki (2010)). A similar pattern was found among fishermen in the Brazilian north east, where some fish offshore in large crews whose success depends on cooperation and coordination, while those exploiting inland waters fish singly. The ocean fishers were significantly more generous (in Public Goods, Ultimatum and Dictator Games) than the inland fishers (Leibbrandt, Gneezy and List (2010)).

A better test of the external validity of experiments would include a behavior-based measure of how cooperative the individuals were, not simply whether they took part in a cooperation-sensitive production process like offshore or cooperative fishing. The Brazilian fishers provide just such a test. Shrimp are caught in large plastic bucket-like contraptions; holes are cut in the bottom of the traps to allow the immature shrimp to escape, thereby preserving the stock for future catches. The fishermen thus face a real world social dilemma: the present value of expected income of each would be greatest if they cut only small holes in their own traps (increasing their own catch) while others cut large holes in theirs (preserving future stocks). Small trap holes are a form of defection, and just as in the Prisoners' Dilemma or Public Goods Game defection maximizes the individual's material payoffs irrespective of what the others do (it is the dominant strategy). But a shrimper might resist the temptation to defect if he were both public spirited towards the other fishers and sufficiently patient to value the future opportunities that they all would lose were he to use traps with smaller holes.

Fehr and Andreas Leibbrandt implemented both a Public Goods Game and an experimental measure of impatience with the shrimpers. They found that the shrimpers with both greater patience and greater cooperativeness in the experimental game had punched larger holes in their traps, thereby protecting future stocks for the entire community (Fehr and Leibbrandt

(2011)). The effects, controlling for a large number of other possible influences on hole size, were substantial. A shrimper whose experimentally measured patience and cooperativeness is a standard deviation greater than the mean is predicted to cut holes in his traps that are half a standard deviation larger than the mean.

Additional evidence of external validity comes from a set of experiments and field studies with 49 groups of herders of the Bale Oromo people in Ethiopia who were engaged in forest commons management. Devesh Rustagi and his coauthors implemented public goods experiments with a total of 679 herders. They also studied the success of the herders' cooperative forest projects. The most common behavioral type in the experiments, constituting a bit more than a third of the subjects, were "conditional cooperators" who responded positively to higher contributions by others. Controlling for a large number of other influences on the success of the forest projects, the authors found that groups with more conditional cooperators were more successful, in terms of number of new trees planted, than groups with fewer conditional cooperators. This was in part because members of groups with more conditional cooperators spent significantly more time monitoring the use of the forest by others. As in the case of the Brazilian shrimpers, the effects of group composition were large. A 10% increase in the fraction of experimentally identified conditional cooperators in a group was associated with an increase in trees planted or time spent monitoring by members of the group of about 3% (Rustagi, Engel and Kosfeld (2010)).

The available evidence suggests that students volunteering for experiments are not more pro-social in their orientations than other students; nor are student subjects more pro-social than non-students, indeed the reverse seems to be the case. An example comes from the trust game in which subjects in the role of "investor" are provisionally given a sum from which they transfer some amount to another subject, called the "trustee". This amount was then tripled by the experimenter. The trustee, knowing the investor's choice, could in turn "back-transfer" some (or all, or none) of this tripled amount, returning a benefit to the investor. When Fehr and John List played a trust game with both students and chief executive officers in Costa Rican businesses, they found that the businessmen both trusted more and also reciprocated trust by the investor to a far greater degree than did the students, as can be seen from Figure 3.5.⁴

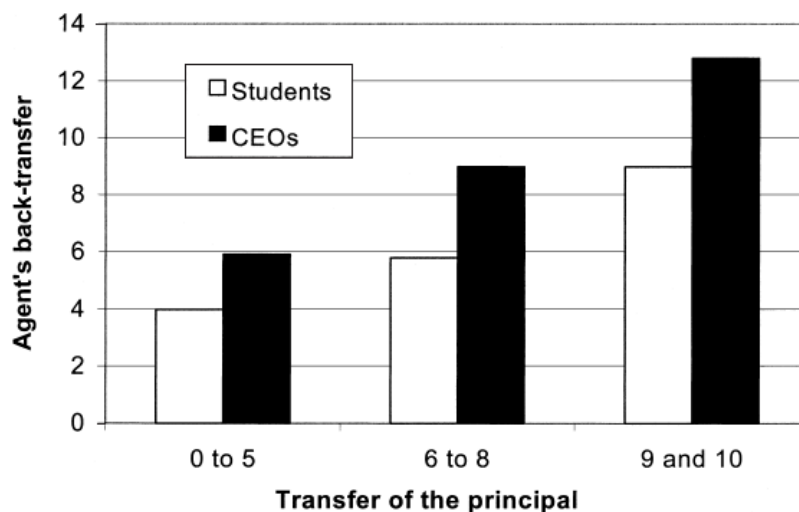


Figure 3.5. Reciprocation of trusting offers in the trust game among Costa Rican students and CEOs (Fehr and List (2004))

CONCLUSION

It appears, then, that J.S. Mill took the field a step backward when he narrowed the subject of political economy to the study of the individual “solely as a being who desires to possess wealth.” Mill’s exclusion of ethical and other regarding motives would have been a harmless simplification facilitating the analysis of the subject matter of economics, were it the case that the excluded ethical and other regarding motives were absent (not something that Mill would have supposed) or that that the effects of incentives on the one hand and the excluded motives were simply additive (which is what Mill must have thought). But as we have seen, neither of these two possible justifications for Mill’s exclusion can be sustained.

Care should be taken in generalizing experimental evidence for this conclusion to the real world. But none of the external validity concerns considered above is sufficient to dismiss the experimental evidence here. We have seen that far from being absent, motives such as reciprocity, generosity and trust are common and that these preferences may be crowded out by the use explicit incentives. We have also shown how information about the nature (categorical, marginal, strong) and extent of crowding out could guide the Legislator in his choice of the level of incentives, possibly limiting their use. The best response functions in Figure 3.3 and 3.4 are definitely part of the Legislator’s new toolbox for the use of incentives when crowding out occurs.

But the Legislator would surely want to go beyond simply designing appropriate policies given the crowding out problem. He would want to know if he could devise policies that would attenuate the crowding problem or perhaps even reverse it, so that crowding in occurred, as in Cardenas' experiment in rural Colombia. Categorical and marginal crowding in would transform Figure 3.5, vastly improving his policy options.

Is there some way that he could transform the non separability of incentives and social preferences from a curse to a blessing?

IV

INCENTIVES AS INFORMATION

Machiavelli sought to design policies so that the self interested would act as if they were “good.” This was and remains an excellent idea. But the result is sometimes that the good come to act as if they were “wicked”. This is not news to people who have succeeded in business, like David Packard:

In the late 1930s, when I was working for General Electric ..., the company was making a big thing of plant security. ... guarding its tool and parts bins to make sure employees didn't steal. ...many employees set out to prove this obvious display of distrust justified, walking off with tools and parts whenever they could. ...

He left behind GE’s self fulfilling low expectations of employees when he founded his own company:

When Hewlett Packard got under way.. I determined that our parts bins and storerooms should always be open....which was advantageous to HP in two important ways. ... the easy access to parts and tools helped product designers and others who wanted to work out new ideas at home or on weekends. ... and open bins and storerooms were a symbol of trust, a trust that is central to the way HP does business (Packard 1995, pp 135)

Perhaps the greatest challenge facing Aristotle’s Legislator is to develop policies under which social preferences will be synergistic with incentives that appeal to economic self interest, each enhancing rather than diminishing the effect of the other. Even if the Legislator were to adopt to a more modest “do no harm” creed, and seek to design incentives that are simply additive to social preferences he will need to learn more about the conditions under which crowding out occurs and how they might be reversed. To design adequate policies the Legislator will need to know a lot more about why this crowding out occurs, that is about the arrow in Figure 2.1 from Incentive to Values.

THE MEANING OF INCENTIVES

My strategy for deciphering the causes of crowding out is to take a page from F. A. Hayek

who taught economists to consider prices as messages (Hayek (1945).) The usual application of this insight is that when, say, the price of bread goes up, the message is: “bread is now more scarce and should be economized, so you should put potatoes or rice on the table tonight.” Incentives are a form of a price, or they affect prices. So the message sent by a fine for lateness in picking up your child at the day care center might be: “your tardiness is inflicting costs on our staff, so you should try a little harder to come on time.” But while the bread price generally conveys the right information, the lateness price for the Haifa parents must have sent a mixed message.

To understand the difference, Sandra Polanía Reyes and I set out to collect all of the evidence from experimental economics bearing on the separability assumption. We eventually found 51 studies using over 100 subject pools with a total of 26 thousand subjects in 36 countries. (Bowles and Polania-Reyes (2012)) The resulting data set includes more than a hundred different subject pools, over twenty-six thousand subjects from 36 countries, playing Dictator, Trust, Ultimatum, Public Goods, Third Party Punishment, Common Pool Resource, Gift Exchange and other principal-agent games. These are all settings in which one’s actions affect the payoffs of others so that social preferences may affect a subject’s experimental behavior.

We found evidence of non separability in all of these games, and of crowding out in most of them. More important, the experiments provide clues about why policies designed on the premise that citizens or employees are entirely self interested often induce people to act as if they were.

With few exceptions the experiments were not designed to test the effects of incentives on preferences or the causal pathways accounting for these effects, but instead to assess the nature and extent of non-selfish preferences. In the experiments that we reported, however, we found a way to test hypotheses about the effects of incentives on social preferences.

Our empirical strategy (based on experimental results) was to observe the total effect of incentives on behavior and to note whether this differs from the predicted direct effect (the top arrows in Figure 2.1) in order to infer the effects of incentives on (unobserved) social preferences and thereby on actions (the bottom two arrows). Our data set includes all the economic experiments we were able to locate that allow this or some other test of the separability assumption.

The entire data set presented in Bowles and Polania-Reyes (2012) shows that non

separability is common in these experimental settings, and that crowding out is a pervasive phenomenon whose effects are large enough to focus the attention of the Legislator.

There are two broad classes of causal explanations of crowding out effects. Incentives may affect preferences because they provide cues to the nature of the situation in which a person finds herself so that a different set her preferences apply. (“If you are shopping it’s ok to be entirely self interested; if you are with your family it is not”). In this case we say that preferences are situation dependent (an economist would say “state dependent” the state with an incentive being distinct from the state in which an incentive is present.)

The second broad class of incentive effects on preferences is that incentives may deliberately or unwittingly alter the process of cultural transmission by which people come to acquire preferences. In this case we say that preferences are endogenous and incentives affect which preferences people adopt. Evidence from experiments may have a bearing on the effects of incentives on cultural transmission; but for the most part the duration of a typical experiment (a few hours at most) is far too brief to capture the social learning or socialization that would make preferences endogenous. I will consider the case of endogenous preferences in the next chapter. In this chapter I will use experimental evidence to consider cases in which preferences are situation dependent, and incentives are part of how a situation is defined or cued.

We know that our behavior is acutely sensitive to the nature of the decision situation (Ross and Nisbett (1991), Tversky and Kahneman (1981)) and, as we will see, the presence or extent of incentives provides information about the situation. Situation-dependence arises because actions are motivated by a heterogeneous repertoire of preferences –from spiteful to payoff-maximizing to generous, for example -- the salience of which depends on the nature of the decision situation – interacting with a domineering supervisor, shopping, or relating to one’s neighbors, for example. The boss may bring out spite while meeting with neighbors, generosity.

To see how this works, think about gifts (Healy (2006).) Economists know that money is the perfect gift – it replaces the giver’s less well-informed choice of a present by the recipient’s own choice. But when the holidays come around few economists give money to their friends, family and colleagues. This is because we also know that money cannot convey thoughtfulness, romantic interest, concern, whimsy, or any of the other messages that non-monetary gifts sometimes express. A gift, we know, is more than a transfer of resources; it is a signal about the giver and her

relationship to the recipient, and money changes the signal.

Can the same be said of incentives? A long tradition in psychology has concluded that it can:

The multiple meanings of ... tangible rewards are reflected in our everyday distinction among bribes and bonuses, incentives and salaries. ... they carry different connotations concerning, for example, (i) the likely conditions under which the reward was offered, (ii) the presumed motives of the person administering the reward, and (iii) the relationship between the agent and the recipient of the reward (Lepper, Sagotsky, Dafoe, et al. (1982) lower case roman numerals added).

All three of the pieces of information conveyed by incentives –“the likely conditions,” “the presumed motives” and the “relationship” -- may adversely affect the social preferences of the target of the incentives. Sometimes the news is not good.

BAD NEWS

Incentives are implemented for a purpose, and because the purpose is often evident to the target of the incentives, the target may also infer information about the person who designed the incentive, about his or her beliefs concerning the target, and the nature of the task to be done (Benabou and Tirole (2003), Fehr and Rockenbach (2003)). As Mark Lepper and his coauthors say, incentives may affect preferences for reasons familiar to economists, that is because they indicate “the presumed motives of the person administering the reward.” By implementing an incentive, one reveals information about one’s intentions (own payoff maximizing vs. fair-minded, for example) as well as beliefs about the target of the incentives (hardworking or not, for example) and the targeted behavior (how onerous it is, for example.) This information, in turn, may then affect the target’s motivation to undertake the task at hand.

The Fire Commissioner’s threat to dock the pay of those with more than 15 sick days conveyed the information that he did not trust that the firemen were indeed doing their very best to come to work (especially on Mondays and Fridays). The new situation – working for a boss who does not trust you – seems to have altered the motivation of the firemen. Of course we cannot know just what accounted for the spike in sick call-ins. That is why we use experimental information as well as natural observation to attempt to understand why crowding out occurs.

This ‘bad news effect’ commonly occurs in relationships between principals, who design incentives to induce agents to behave in the principal’s interest to a greater extent than they would

do in the absence of the incentive. Here is an example.

As in the trust game played with Costa Rican CEO's and students, in this experiment German students in the role of "investor", the principal, were given the opportunity to transfer some amount to the agent, called the "trustee". This amount was then tripled by the experimenter. The trustee, knowing the investor's choice, could in turn "back-transfer" some (or all, or none) of this tripled amount, returning a benefit to the investor (Fehr and Rockenbach (2003)).

But there was a new treatment. When the investor transferred money to the trustee, he or she also specified a desired level of the back-transfer. The experimenters implemented an incentive condition in which the investor also had the option of declaring that he would impose a fine if the trustee's back-transfer were less than the desired amount. The investor could also decline the use of the fine, the choice of using or declining the fine option being known to the trustee and taken prior to the trustee's decision. There was also a "trust" condition in which no such incentives were available to the investor.

In the experiment, Trustees reciprocated generous initial transfers by investors with greater back-transfers. But the use of the fine reduced return transfers conditional on the investor's transfer, while renouncing the use of the fine when it was available to the investor increased back-transfers. Only one-third of the investors renounced the fine when it was available; their payoffs were 50 percent greater than the investors who threatened use of the fines. Figure 4.1 summarizes their results. The bad news interpretation suggested by Fehr and Rockenbach is that in the trust condition, or when the fine was renounced by the investor, a large initial transfer signaled that the investor trusted the trustee. The threat of the fine, however, conveyed a different message and diminished the trustee's reciprocity.⁵

Crowding out as the result of the "bad news" mechanism may be prevalent in principal agent settings and can be averted where the principle has a means of signaling trust or fairness. Not surprisingly crowding out affects individuals who are intrinsically motivated or fair-minded; for own payoff maximizers, it appears there is nothing to crowd out.

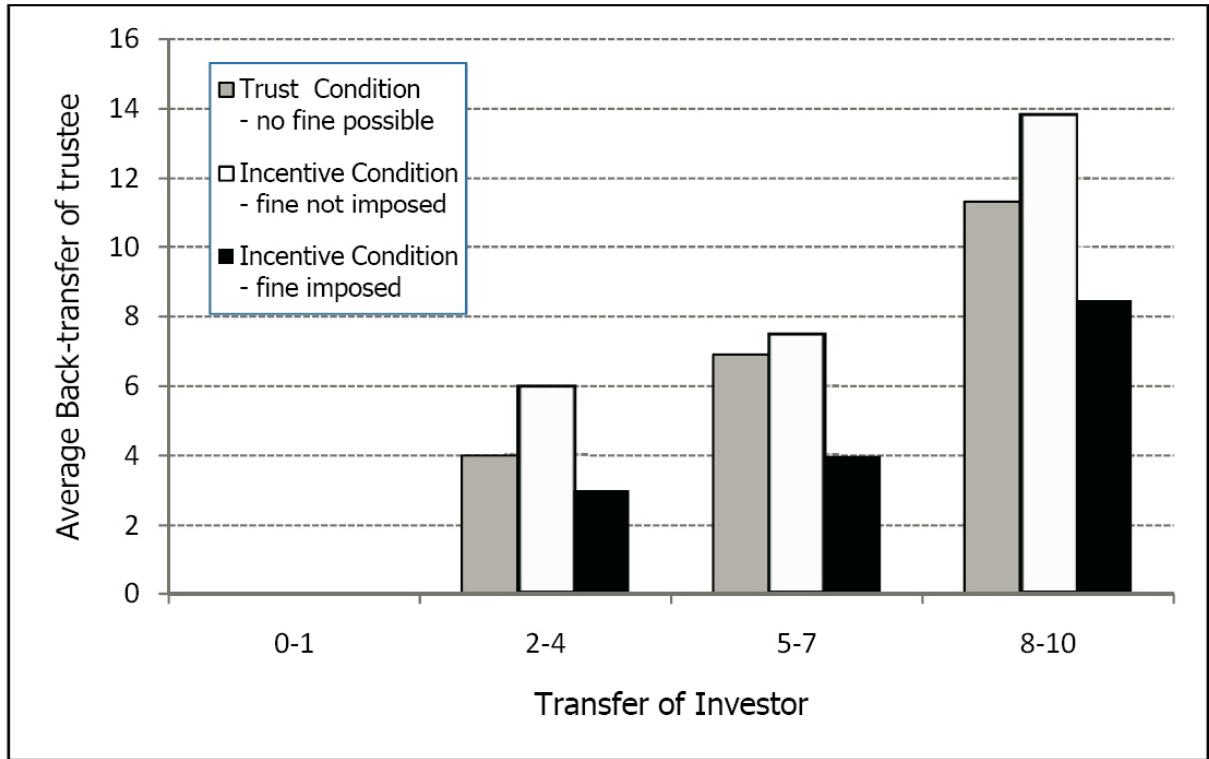


Figure 4.1. Incentives crowd out reciprocity in the trust game Source: (Fehr and Rockenbach (2003))

MORAL DISENGAGEMENT.

There is another reason, less familiar to economists, for crowding out due to the messages that incentives send. In most situations people look for clues of appropriate behavior, and incentives often provide them. We will see that a plausible explanation of some of the framing effects of incentives observed in experiments is that it occurs because market-like incentives trigger what psychologists term “moral disengagement” (Bandura (1991), a process that occurs because “people can switch their ethicality on and off” (Shu, Gino and Bazerman (2009):31).

In this case the incentive effect does not work via information about the principal, and it may be at work even in non-strategic settings. Incentives provide information about (as Lepper and his coauthors put it) “the likely conditions under which the reward was offered:” and thus may offer cues for appropriate behavior. In the experimental evidence, this second mechanism is distinguished from the first by the fact that in the former the incentives are implemented by a

principal who is a player in the game; while in the latter the targets of the incentive are not playing against the incentive designer; rather the incentives are introduced by the experimenter.

Situational cues may be very subtle, and our responses to them unwitting. When experimental subjects had the opportunity to cheat on a test and as a result to gain higher monetary rewards, less than quarter did so when the room was brightly lit, but more than half cheated when the room was slightly less well lit (the variations in lighting had no effect on the observability of cheating.) In another experiment subjects who wore (nonprescription) dark glasses were much less generous to their partner in a Dictator Game than were those outfitted with clear glasses (Zhong, Gino and Bohns (2010)). The dark glasses and darkened room gave the subjects a sense of anonymity, the researchers found. But it was entirely illusory: it is difficult to imagine that a subject could really think that his own wearing dark glasses would make him less observable, especially given that the experiment was conducted at computer terminals in closed cubicles. Even dogs steal more when its dark; but the reason is probably not moral disengagement (Kaminski, Pitsch and Tomasello (2013).)

The dark glasses were just a cue to anonymity. The degree of anonymity changes dramatically as we move between family, workplace, marketplace and other domains of social interaction. Fiske (1992) provides taxonomy of four psychological models corresponding to distinct kinds of social relationships: authoritarian, communal, egalitarian and market, each with culturally prescribed patterns of appropriate behavior. Depending on the information they convey, incentives may signal that the situation corresponds to any one of these four types, and because the degree of anonymity differs, therefore evoke distinctive responses. Because incentives are common in markets and markets are associated with anonymity, incentives may be a cue to real or imagined anonymity.

Here is an example. Armin Falk and Nora Szech investigated the willingness of University of Bonn students to accept a monetary payment in return for allowing a healthy young mouse whose life had been entrusted to their care to instead be gassed to death. Before making their decision the subjects were shown a very disturbing video of a mouse being gassed and a picture of one of the very cute mice. They hypothesized that students who were reluctant to surrender their mouse when this was simply an individual choice, might be more willing to let the mouse die if the mouse in their trust were instead purchased by another subject (resulting in the

certain death of the mouse).

Forty-six percent of the subjects in the individual (non-market) treatment were willing to surrender their mouse for 10 Euros or less (Falk and Szech (2013).) When the mouse trustee could “sell” the mouse to a ‘buyer’ with whom the trustee bargained, however, seventy-two percent to let their mouse die at this price or lower.

Falk and Szech asked the subjects the minimum payment at which they would be willing to give up their mouse . Using this information they were able to calculate the least payment such that the fraction willing to let their mouse die in the individual treatment would be as great as was the case in the market treatment for 10 Euros or less.

What they found out is astounding: to induce the same fraction of subjects in the individual treatment to part with their mouse as had been the case in the market treatment for a price of 10 Euros, the subjects would have to have been offered 47.50 Euros. This almost five fold difference between the least price at which seventy-two percent of the subjects would let their mouse be gassed is a measure of the moral disengagement affected by the market setting.

Notice that the market did not reduce the price at which the students would let the mouse die because of incentives, for in both the individual and the market setting there was a monetary incentive to let the mouse die. What differed was the situation. . A similar finding comes from Hoffman and her co authors (1994) who illustrated the framing power of names. Generosity (proposer offers) and fair-minded behavior (responder rejections of low offers) were both diminished by simply re-labeling an Ultimatum Game the “Exchange Game” and re-labeling proposers and responders “sellers” and “buyers” (Hoffman, McCabe, Shachat, et al. (1994)). The renaming did not affect the incentives at work in the game, these were identical in the two treatments. The renaming instead affected the subjects sense of appropriate behavior independently of the incentives. The power of names has been confirmed in many (but not all) experiments since then (Zhong, Loewenstein and Murnighan (2007)) but in some cases (Ellingsen, Johannesson, Möllerström, et al. (2011)) the framing effect appears to have altered subjects beliefs about the actions, of others rather than their preferences.

But literally naming the game is not necessary for framing effects to occur. Incentives alone may provide powerful frames for the decision maker. A year before the first reality TV Survivor show, Andrew Schotter and his coauthors found that market-like competition for “survival”

among subjects reduced their concern for fairness in an Ultimatum Game experiment (Schotter, Weiss and Zapater (1996)). In this game Player 1 is given an endowment and asked to propose a division of it with Player 2. Player 2, knowing the size of the endowment, decides whether to accept or reject the division. If Player 2 accepts, then the proposed division is implemented. If Player 2 rejects, both players receive zero. As is commonly observed in the Ultimatum Game, Player 1 made quite generous offers and low offers were frequently rejected.

But when the experimenters told the subjects that those with lower earnings would be excluded from a second round of the game, Player 1 subjects offered less generous amounts to Player 2, and Player 2 accepted lower offers. The authors' interpretation was that: "...the competition inherent in markets...offers justifications for actions that, in isolation, would be unjustifiable." (p.38).

While plausible, direct evidence for this "moral disengagement" explanation is lacking because the social preferences that apparently accounted for fair behavior in the non-survival condition of the experiment were not measured. There are cases, however, in which the reduction in the salience of ethical reasoning induced by the presence of incentives can be explicitly measured.

A large team of anthropologists and economists implemented both Dictator and Third Party Punishment Games in 15 societies ranging from Amazonian, Arctic and African hunter gatherers to manufacturing workers in Accra, Ghana and US undergraduates (Barr, Wallace, Ensminger, et al. (2009), Henrich, Ensminger, McElreath, et al. (2010)). In the Dictator Game an experimental subject is assigned a sum of money and asked to allocate some all or none of it to a passive recipient. The Third Party Punishment Game is a Dictator Game with an active onlooker (the third party) who observes the dictator's allocation. If the third party deems the dictator's allocation worthy of punishment, he or she may then pay to impose a monetary fine on the dictator. One would expect that in the presence of a third party and the incentives provided by the prospect of being fined, the dictators would adjust their allocations upwards (compared to the two party standard Dictator Game) and thus avoid being fined. But this is not what happened.

Surprisingly, in only two of the 15 populations were the offers significantly higher in the Third Party Punishment Game than in the Dictator Game, and in four of the populations the allocations were significantly (and in some cases substantially) lower. In Accra, for example, where 41 percent of the dictator's allocations resulted in fines by the third party, the allocations

were 30 per cent lower in the Third Party Punishment Game than in the Dictator Game. The incentives provided by the fine did not induce higher allocations, but rather had the opposite effect.

Crowding out of ethical motives is suggested by the fact that the dictator's adherence to one of the world's religion (Islam or Christianity, including Russian Orthodoxy) raised allocations in the Dictator Game by 23 percent (compared to those unaffiliated with a world religion.). But in the Third Party Punishment Game, the estimated "religion effect" virtually disappeared, and was not significantly different from zero. The presence of the incentive based on the fine appears to have defined the setting as one in which the moral teachings of these religions were not relevant. Consistent with a crowding out interpretation of these results, the negative effect on the dictator's allocations of his or her economic need (number of children, conditional on a given level of income and wealth) was substantial (and statistically significant) in the Third Party Punishment Game, but in the Dictator Game this "economic need effect" was an order of magnitude smaller and not significantly different from zero.

In the Accra sample (Barr (2004)) in the standard two party Dictator Game, the dictator's allocation co-varied significantly with the frequency of attendance at church or mosque. But this large "religion effect" vanished in the Third Party Punishment Game. The incentives implicit in the Third Party Punishment Game appear to have substituted economic motivations for moral concerns. These experiments are also consistent with the expectation that crowding out operates via an effect of incentives on the behavior of those with pre-existing social preferences. It does not work on those with little or nothing to crowd out.

CONTROL: INCENTIVES COMPROMISE AUTONOMY.

The third reason why incentives may crowd out social preferences is that incentives (or constraints) may compromise a subject's sense of autonomy and reduce an intrinsic motivation to undertake the targeted task. (Lepper and Greene (1978), Deci and Ryan (1985), Deci, Koestner and Ryan (1999)). The underlying psychological mechanism appears to be a fundamental desire for "feelings of competence and self-determination" that are associated with intrinsically motivated behavior (Deci (1975)). These effects occur, as we will see, in strategic situations where the bad news that incentives convey concerns the desire of a principal to control the agent. But

most of the experimental evidence for this third crowding out mechanism comes from non-strategic settings (the experimenters, not a principal implements the incentive.) This self-determination mechanism thus differs from the previous two mechanisms -- bad news about a principal and moral disengagement-- because it arises from the target's desire for autonomy and does not depend on the target inferring negative information about a principal or clues about appropriate behavior.

Lepper and his co authors explained why “a person induced to undertake an inherently desirable activity as a means to some ulterior end should cease to see the activity as an end in itself.”

When an individual observes another person engaging in some activity, he infers that the other is intrinsically motivated... to the extent that he does not perceive salient, unambiguous, and sufficient extrinsic contingencies to which to attribute the other's behavior....

An important inference is that the presence of incentive may lead observers to conclude that some seemingly generous action was not done for the intrinsic pleasure of helping others, but as an instrumental response to the incentive. But the more direct effect is the effect of the incentive on the individual's perceptions about his own motivation:

a person engages in similar processes of inference about his own behavior and its meaning. To the extent that the external reinforcement contingencies controlling his behavior are salient, unambiguous, and sufficient to explain it, the person attributes his behavior to these controlling circumstances.. But if external contingencies are not perceived...the person attributes his behavior to his own dispositions, interests, and desires. (Lepper, Greene and Nisbett (1973):7)

According to this interpretation, where people derive pleasure from an action per se in the absence of other rewards, the introduction of incentives may 'over-justify' the activity and reduce the individual's sense of autonomy. Consistent with this view, recent experiment, we have seen in the Prologue, shows the negative effects of incentives specifically on altruistic behavior in infant children. The authors of that study conclude: “Perhaps when rewards are offered children simply come to perceive a formerly self-sufficient activity as merely a means to some more valuable end.” (Warneken and Tomasello (2008):1788)

In an iconic “over-justification” experiment, a reward was offered by the experimenter if the subject were to engage in a particular activity in which they had previously enthusiastically engaged in the absence of anticipated reward. Subjects had been selected for their prior interest the activity, namely painting (Lepper, Greene, et al. (1973)). Those who chose painting received a “Good Player Award” with the child’s name on it. Those who chose some other play activity received no reward. In a control condition no rewards were promised or given.

A week or two after the experiment the subjects’ play activities were observed in their ordinary school setting. Those who had been offered a reward conditional on their engaging in painting took up painting only half as frequently as those who had not anticipated a reward. Moreover, those who had been promised a reward also painted less than they had painted prior to the promise of reward treatment. Finally, in the promised reward treatment (during the experiment, not in their ordinary classes later) the subjects’ paintings were judged to be of substantially lower quality than those painted by the students in the no reward condition.

The experimenters designed the experiment to isolate whether the negative impact occurred because an activity had been rewarded or because the subject had chosen the activity knowing that it would be rewarded. The latter hypothesis was strongly supported. In a third treatment children who had already chosen to paint were later rewarded in the same way as those who had been promised a reward. Consistent with the over justification hypothesis, children in this unanticipated reward condition did not subsequently eschew painting; they later painted a bit more than the unrewarded subjects (but not significantly so). Thus the affect appears to be associated with compromised autonomy, not the receipt of rewards per se.

The fact that these cases the incentive was a reward rather than a penalty suggests that it did not convey negative information about the incentive designer, but instead altered the meaning of the activity itself from one that expressed autonomy to one that expressed compliance. The interpretation that self-determination is involved in the negative response to incentives is consistent with the fact that close supervision or arbitrary temporal deadlines for completion of an otherwise enjoyable activity have effects very similar to financial or other rewards. Lepper and his co authors point out that the “detrimental effects of unnecessarily close adult supervision or the imposition of unneeded temporal deadlines suggest strongly that the effects ... are the result of

superfluous constraints on children's actions, not a specific function of the use of tangible rewards.” (Lepper, Sagotsky, Defoe et al. (1982)).

Unlike the “over-justification” experiments by psychologists where incentives are typically implemented by the experimenter, economists have studied strategic interactions in which the incentives are implemented by a player in the game. In these strategic settings the same apparently control averse reaction occurs, but because the incentives are designed by a principle to affect the behavior of an agent, the adverse reaction may also result from “bad news” about the principal. It is likely that in these and other cases more than one mechanism is at work when crowding out or in occurs. Here is an example.

Armin Falk and Michael Kosfeld implemented a principal-agent game with adult experimental subjects to explore the idea that ‘control aversion’ based on the self-determination motive may be a reason why incentives sometimes degrade performance (Falk and Kosfeld (2006)). Experimental agents in a role similar to an employee chose a level of ‘production’ that was costly to them to provide and beneficial to the principal (the employer). The agent's choice effectively determined the distribution of gains between the two, with the agent’s maximum payoff occurring if he produced nothing. Before the agent's decision, the principal could elect to leave the choice of the level of production completely to the agent's discretion, or impose a lower bound on the agent's production (three bounds were varied by the experimenter across treatments, the principal’s choice was simply whether or not to impose it.) The principal could infer that a self-interested agent would perform at the lower bound or, in the absence of the bound, at zero, and thus imposition of the bound would maximize the principal’s payoffs.

But in the experiment agents provided a lower level of production when the principal imposed the bound. Apparently anticipating this negative response, fewer than a third of the principals opted for the bound in the moderate or low-bound treatments. This minority of “untrusting” principals earned on average half of the profits of those who did not seek to control the agents' choice in the low-bound treatment, and a third less in the intermediate bound condition.

Control aversion and the desire for self-determination are not the only effects of the principal’s seeking to bind the agent. As anticipated by our discussion of the information content of incentives above, the imposition of the minimum in this experiment gave the agents remarkably accurate information about the principals' beliefs about them. In post-play interviews, most agents

agreed with the statement that the imposition of the lower bound was a signal of distrust; and the principals who imposed the bound in fact had substantially lower expectations of the agents. The untrusting principals' attempts to control the agents' choices induced over half of the agents (in all three treatments) to contribute minimally, thereby affirming the principals' pessimism.

Depending on the distribution of principal's prior beliefs about the agents, a population with preferences similar to these experimental subjects could support either a trusting or an untrusting outcome in each of which the prior trusting and untrusting beliefs are perpetuated and hence each of which could persist indefinitely .

A NEURAL BASIS FOR CROWDING OUT?

It is worth exploring whether the Aristotelian Legislator could go beyond the insights of the “incentives as messages” approach to explore the proximate neural basis for crowding out. Could he map how an incentive acts as a stimulus of our deliberative and affective neural activity, and then use this information to design incentives and other policies that would activate rather than deactivate the neural pathways associated with cooperative and generous behavior? For example, on the basis of neuro-imaging and other evidence one might be tempted to conclude that incentives activate more deliberative and less affective cognitive processes, and that deliberation tends to result in self interested action. Could incentives be made synergistic with social preferences, then, by framing subsidies and fines to stimulate emotional and visceral rather than calculative responses?

The idea is not as far fetched as it sounds. To identify the proximate causes of the crowding out behavior in the trust game of Fehr and Rockenbach (2003) described above Li, et al. (2009) compared evidence of activation of distinct brain regions of trustees when faced with an investor who had threatened to sanction the trustee for insufficient back transfers and an investor who had not threatened a sanction. As in the Fehr and Rockenbach experiment, the effect of the threat was to reduce rather than increase back transfers, another example of strong crowding out. The fMRI showed that the threatened sanctions de-activated the Ventromedial Prefrontal Cortex (VMPFC, a brain area correlated with higher repayment in this experiment) as well as other areas relating to the processing of social rewards. The threat activated the parietal cortex, an area thought to be associated with cost-benefit analysis and other self-interested optimizing. The interpretation by Li

and his co authors is that the sanctions induced a “perception shift” favoring a more calculative and hence self-interested response.

If it were confirmed that incentives such as the threat of the sanction in the trust game activate brain regions associated with calculative self interest there might be a way to design incentives that did not have this effect. The perception shift in the subjects facing the threat involved two quite different ways of responding to a stimulus. These are termed affective (or emotional) and deliberative (or cognitive). Or, as the neuroscientist Joshua Green puts it “The human brain is like a dual-mode camera with both *automatic settings* and a *manual mode*.” (Greene (2014)) Dual process theory in psychology seeks to understand how their interactions influence behavior. (Greene, Sommerville, Nystrom et al. (2001) Loewenstein and O'Donoghue (2007), Sanfey, Loewenstein, McClure et al. (2006))

Why would incentives stimulate deliberation rather than emotion and why would this (in many circumstances) override pro-social emotions such as sympathy? The first part of the question is easy: an incentive is an invitation to do benefit cost calculations. One needs to determine if the incentive is sufficient to motivate the targeted activity. In this incentives differ from the strongly valenced aspects that are associated with emotions like sympathy, pain avoidance, or fear. If one has a painful burning sensation in one's hand one does not have to deliberate much about whether getting away from the fire is the right course of action.

But why might deliberation when it is induced by an incentive result in less pro-social behavior? There is no good answer to the second part of the question above, as yet.

There some evidence that this dual process approach can identify the proximate causes of behavior not only in the Trust Game but also in the other experiments . Sanfey, Loewenstein, et al. (2006) interpreting an experiment reported in Sanfey, Rilling, Aronson et al. (2003) write:

A neuro-imaging study examining the Ultimatum Game found two brain regions that were particularly active when the participant was confronted with an unfair offer, the anterior insula and the dorsolateral prefrontal cortex (dlPFC) . Activation in these areas has been shown to correlate with emotional and deliberative processing, respectively, and it was found that if the insular activation was greater than the dlPFC activation, participants tended to reject the offer, whereas if the dlPFC activation was greater, they tended to accept the offer. This offers neural evidence for a two-system account of decision-making in this task

Small, Loewenstein and Slovic (2007) found that a picture of a needy girl induces more

charitable giving than do statistics on need and that providing both statistics and the picture reduces giving (compared to showing the picture needy girl alone). They conclude: “when thinking deliberatively, people discount sympathy towards identifiable victims but fail to generate sympathy toward statistical victims” Here the affective system appears to promote generosity but it is overridden by the deliberative system when the presentation of statistics stimulate that process.

But it would be a mistake to think that deliberation is the enemy of generosity. It can be quite the opposite. Among liberals (in the study of Skitka, Mullen, Griffin et al. (2002) as described by Loewenstein and O'Donoghue (2007)) the deliberative process is more generous towards AIDS affected people, and less influenced by considerations of “responsibility” than is the affective. “The study found that subjects were less likely to advocate subsidized treatment under conditions of high [cognitive] load, which we would interpret as evidence that deliberative reactions are more concerned than affective reactions to AIDS victims. More interestingly, under conditions of high load, both liberals and conservatives were less likely to provide subsidized treatment to those deemed responsible (relative to those deemed not responsible), whereas under conditions of low load, liberals treated both groups equally whereas conservatives continued to favor groups who were seen as less responsible for contracting the disease. These findings are consistent with our framework if affective and deliberative reactions were consistent for conservatives — so cognitive load has no effect — but conflicting for liberals.”

As in all animals, the genetic basis for human visceral and other non deliberative reactions to stimuli has evolved under the influence of natural selection; and it would be surprising if these did not induce self interested behavior. Pain avoidance and fear of danger reactions are examples. This being the case one could expect that self interested behavior could be aligned with the neural pathways associated with emotion no less than with deliberation. There thus does not appear to be any simple mapping from the “self regarding”-“other regarding” distinction in behavior to the deliberation- emotion distinction in cognitive processing or the prefrontal cortex - limbic system distinction neuroscience. While (as Jonathan Cohen (2005) says) the prefrontal cortex “may be a critical substrate for *Homo economicus*” the deliberative decision maker, it probably is no more implicated in the self-interest assumed in most economic models than is the limbic system.

Thus it may be that incentives – like statistics in the needy girl example, and the opposite of

cognitive load in the HIV AIDS case – make the deliberative processes more salient. Whether this deliberation results in more generous behavior (crowding in: liberals in the AIDS case) or the opposite (crowding out in the giving to the girl case) depends on whether the imperative to generosity resulting from the deliberative process (utilitarian or other concepts of duty to others, fairness, for example) is stronger than the generosity inducing emotions (sympathy, for example) of the affective process. A taxonomy of these cases suggesting that there is no simple mapping between the deliberative-emotional distinction and social preferences appears in Table 4.1

	<i>Emotional</i>	<i>Deliberative</i>
<i>Ethical, other regarding</i>	Sympathy to those harmed; anger and those who harm; “moral disgust” or fear of “sinful” acts	Account taking of one's actions effect on others
<i>Self regarding</i>	Hunger, other appetites; fear of personal danger	Maximizing own-expected gains or well being

Table 4.1 No simple mapping between deliberative and emotional cognitive processing and ethical or other regarding behavior.

A related idea is advanced by Joshua Greene (2010) in two parts. First the deliberative processes are outcome based (consequentialist) and utilitarian, while the affective processes support judgments (non-consequentialist, deontological) such as duty or the conformity of an action to a set of rules, rather than its consequences. Second, these ways of behaving are associated with activation in different brain regions (respectively prefrontal cortex vs limbic system.) The above neuro-scientific evidence would then imply that economic incentives induce consequentialist reasoning (and implicitly reduce the salience of deontological judgments) and the behavioral experimental evidence of crowding out would then imply that often (but not always) consequentialist reasoning is less “pro social” than deontological judgment, resulting in the observed apparent crowding out of generosity and fairness concerns by the use of incentives. But we are left wondering why consequentialist reasoning is more likely to induce self interest and deontological logic to favor ethical and other regarding behavior. If I take pleasure both from helping people in need and eating ice cream and I cannot at the moment do both, it is not at all clear why

activating my deliberative processes would lead me to go for the ice cream, while were I to respond viscerally I would instead help the person in need.

A PUZZLE

The three reasons why incentives may crowd out social preferences – bad news, moral disengagement and control aversion – provide some of the information that Aristotle’s Legislator will need to design policies such that incentives can provide complements rather than substitutes for a desire to uphold social norms and act generously towards ones fellow citizens. In each case policies can be devised to minimize the crowding out problem and even to induce crowding in. In considering the Legislator’s policy options in the final two chapters we will provide examples of highly effective incentives deployed by peers who have nothing to gain personally and whose mutually beneficial purposes are clearly conveyed. But we cannot go directly to the Legislator’s policy challenges and options.

Evidence that social preferences are common and that they underwrite mutually beneficial exchanges, but are often crowded out by use of explicit economic incentives presents us with the following puzzle. The adverse effect of incentives on generosity, reciprocity, the work ethic and other motives essential to well functioning institutions including markets, would seem to portend instability and dysfunction for any society in which explicit economic incentives are widely used.

Have societies somehow avoided the seemingly unavoidable vicious cycle in which markets and other incentive driven institutions erode the cultural foundations on which they depend, leading to the intensification of the use of incentives to compensate for the increasing deficiency of ethical and other regarding preferences needed to underwrite economic exchange and other mutually beneficial social intercourse?

Why did the Boston Fire Commissioner and the firemen not wind up in a kind of economic and cultural arms race to the bottom, with the Commissioner just upping the ante, imposing increasingly draconian pay deductions and the firemen eventually abandoning their civic obligations entirely in favor of precisely the opportunistic self interest that the Commissioner had attributed to them?

Why only the Fire Department? Why would precisely this dynamic not play out right across any market based economy? And wouldn't this dynamic end with exactly the constitution for knaves that Hume advocated, but in contrast to Hume's account, with a citizenry of knaves as well? Or has something like this actually happened?

V

A LIBERAL CIVIC CULTURE

If incentives sometimes crowd out both ethical reasoning and intrinsic motivations, and if the status of markets as a morality free zone is celebrated by leading thinkers, it seems just a short step to Marx's broadside condemnation of capitalist culture:

Finally there came a time when everything that men had considered as inalienable became an object of exchange, of traffic and could be alienated. This is the time when the very things which till then had been communicated, but never exchanged, given but never sold, acquired but never bought: virtue, love, conviction, knowledge, conscience— when everything passed into commerce. It is the time of general corruption of universal venality. Marx, (1956):32

But a century and a half later “universal venality” misses the mark for the cultures of northern Europe where capitalism was born and also for the distant North American and other offshoots of these populations.

New York City's diplomatic immunity from prosecution for traffic violations provides a natural experiment to test Marx's prediction (Fisman and Miguel (2006).) There the diplomats of 146 nations and distinctive cultures face the same external incentives and constraints, and we can measure the frequency with which they willingly break the law and inconvenience others by parking illegally and not paying the resulting ticket.

In the five years prior to November 2002 the average number of violations per diplomat was 19, with the diplomats of Egypt (140 violations per diplomat), Bulgaria (117), Albania (84) and Pakistan (69) among the most flagrant. Over the same period the 31 diplomats from the United Kingdom, where capitalism was born, committed exactly zero violations; and the same was true of Sweden, Norway, Canada, and the Netherlands, the second birthplace of capitalism. Other early capitalist nations posted modest numbers of transgressions per member of their diplomatic contingent of this period: 1 for Germany 2.7 and for Belgium. Late comer capitalist nations were likewise epitomes of parking probity:

Japan's 47 diplomats posted not a single violation, and Korean diplomats committed just 0.4 violations on average.

Do not read too much into the parking violations evidence. Adam Smith, you will see presently, warned that in contrast to merchants diplomats were not to be trusted in any case, so perhaps the people with **D** license plates are not representative of their cultures. We will shortly turn to more convincing data based on cross cultural comparisons of behavior in experimental games of cooperation and social norms enforcement.

This and other evidence is consistent with the view that the oldest capitalist societies have indeed sustained vibrant civic cultures characterized by widespread conformity to cooperative and generous social norms, at least by comparison to many other societies in which the reach of market institutions has been more recent and restricted. In the early capitalist European and European offshoot economies, incentives fit for J.S. Mill's acquisitive *Homo economicus* have for two centuries or more been widely used and even celebrated. That this has not brought on the corruption of which Marx warned seems puzzling in light of the experimental evidence that explicit economic rewards and penalties sometimes drive out social preferences.

The puzzle would be easily resolved if the behavioral effects of incentives observed in experiments were either ephemeral or strictly limited in domain, so that, say, the moral disengagement when offered incentives for a particular task at work did not spillover to other domains of life, for example family life or citizenship.

But we will see that the economy is a great teacher, and its lessons are neither fleeting nor confined. And I will also provide reasons why living in a highly incentivized world might have long term effects on the process of cultural evolution above and beyond those stemming from the adverse situational cues sometimes associated with incentives, as documented in the previous chapter.

Further thought about the puzzle thus only deepens the mystery. But the puzzle can be resolved by considering the capitalist societies just mentioned not simply as economies, but instead as liberal societies.

By liberal society I mean one characterized by extensive reliance on markets to allocate economic goods and services, formal equality of political rights, the rule of law,

public tolerance, and few barriers to mobility based on race, religion, or other accidents of birth (in contrast to societies loosely termed “traditional” or more broadly “non-liberal. In the empirical studies below, examples of liberal societies are Switzerland, Denmark, Australia, the U.S. and the U.K., while examples of non-liberal societies (lacking at least one of the above attributes of liberal societies) are Saudi Arabia, Russia, Ukraine, and Oman as well as the small scale societies of hunter-gatherers, herders and low technology farmers to be considered presently. The economy of a liberal society need not be capitalist – it could be an economy of independent producers in which the employment of labor with the intention of making a profit (the definition of a capitalist economy) plays a limited role. But the economies of all of the empirical examples of liberal societies mentioned above are capitalist.

THE ECONOMY PRODUCES PEOPLE

The idea that the social interactions occurring in markets and other institutional environments durably shape social norms and preferences, which are then generalized to non economic domains of life has long been recognized. Marx was not alone in holding this view. The royalist Edmund Burke reasoned in the same way when he lamented that the French Revolution had ushered in “the age of sophisters and economists.” The long term effects on preferences to which Marx, Burke and others refer are quite different from the incentive effects studied in the previous chapter, where incentives affect preferences because they provide cues, altering the situation in which a person finds herself ,increasing the salience of some preferences and diminishing others. But in contrast to these situation-dependent effects, incentives may also alter the process by which people come to acquire new tastes, habits, ethical commitments and other motivations, making preferences endogenous.

Acquiring new preferences (like a new accent) often takes place early in life and the learning process is strongly attenuated thereafter. The key difference between endogenous and situation-dependent (or framing sensitive) preferences is that in the former case incentives may affect a long term learning process, the results of which persist over decades if not entire lifetimes. When preferences are situation dependent, a new the situation – the

withdrawal of an incentive, for example – changes which of the person's repertoire of preferences will motivate behavior. I say that incentives *affect* preferences in both the situation-dependent and endogenous preference case; but the two effects differ: in the former case the incentive is a reversible signal about the principal or the situation; in the latter the incentive alters the preference-updating process.

The developmental processes involved in learning new preferences typically include population-level effects such as conformism, schooling, religious instruction and other forms of socialization that are not readily captured in experiments. We cannot therefore hope for the kinds experimental evidence on the evolution of preferences that were possible in the previous chapter in studying the effects of incentives as messages about situations. . But historical and ethnographic data⁶, while not directly related to the use of incentives, are quite consistent with the view that economies structured by differing incentives are likely to produce people with differing preferences. Here is some of the evidence.

Over a period of three decades Melvin Kohn and his collaborators have studied the relationship between one's position in the authority structure of the workplace-- giving as opposed to taking orders, designing incentives or being their target -- and the individual's valuation of self-direction and independence in their children, as well as one's own intellectual flexibility, and personal self-directedness (Kohn (1969), Kohn and al. (1983), Kohn (1990)). They concluded that "...the experience of occupational self-direction has a profound effect on people's values, orientation, and cognitive functioning." (Kohn, Naoi, Schoenbach et al. (1990):967) The studies take account of the possibility that personality is affecting choice rather than vice versa and provide convincing evidence that there is a causal arrow from job to preferences.

His collaborative study of Japan, the U.S. and Poland (Kohn, Naoi, et al. (1990)) yielded cross culturally consistent findings: people who exercise self-direction on the job also value self-direction more in other realms of their life (including child-rearing and leisure activities) and are less likely to exhibit fatalism, distrust, and self-deprecation. Kohn and his co-authors reason that "...social structure affects individual psychological functioning mainly by affecting the conditions of people's own lives." Kohn concludes that:

The simple explanation that accounts for virtually all that is known about the effects of job on personality ... is that the processes are direct: learning from

the job and extending those lessons to off-the-job realities. (1990a):59

Additional evidence comes from a study by the anthropologists Herbert Barry, Margaret Child and Irvin Bacon. They categorized 79 mostly non-literate societies according to the prevalent form of livelihood (animal husbandry, agricultural, hunting and fishing) and the related ease of food storage or other forms of wealth accumulation, the latter being a major correlate of dimensions of social structure such as stratification (Barry, Child and Bacon (1959)). Food storage is common in agricultural societies but not among foragers.

Bacon and his co authors also collected evidence on forms of child-rearing, including obedience training, self-reliance, independence and responsibility. They found large differences in the recorded child-rearing practices. These co-varied significantly with economic structure, controlling for other measures of social structure such as unilinearity of descent, extent of polygyny, levels of participation of women in the predominant subsistence activity, and size of population units. They concluded, "knowledge of the economy alone would enable one to predict with considerable accuracy whether a society's socialization pressures were primarily toward compliance or assertion." The causal relationship is unlikely to run from child-rearing to economic structure, as the latter is dictated primarily by geography in the sample of simple societies under study.

These society-level studies cannot isolate the effects of incentives per se, as this would involve finding what would be highly unlikely ever to exist: a sample of otherwise similar societies with measurably different incentive structures. The most the cross cultural studies show is that preferences vary with the manner in which societies organize their economic life. Surprisingly, in light of their inability to capture long term learning effects, experiments can isolate at least some short term learning effects of incentives per se.

THE PERSISTENT EFFECTS OF INCENTIVES

In the experiments presented thus far there is no reason to think that the unexpected effects of incentives – crowding out – would persist even after the incentive was removed. But as the persisting tardiness of the Haifa day care parents shows, this is what sometimes occurs, suggesting that incentives have durable effects on preferences

Donning the hat of the Legislator, Joseph Falkinger and his co authors designed an incentive system to induce subjects in an experiment to contribute to a public good. The system s caused subjects to contribute almost exactly the amount predicted for a own-material-payoff-maximizing individual. (This is what we saw in Chapter 3: when a high incentive for public goods contribution was offered in the experiment by Irlenbusch and Ruchala's, subjects contributed almost exactly what the economics textbook model predicts. A similar result occurred in the Cardenas et al experiment when fines for over extraction of the "forest" were imposed.)

It would be tempting to conclude from this evidence that the subjects were indeed material payoff maximizers. But this would be mistaken. In the Falkinger experiment (as in those reviewed in Chapter 3) in the absence of the incentive subjects contributed significantly more than would have been optimal for a payoff maximizing individual. This shows, unsurprisingly in light of the evidence of the previous chapters, that even when incentives "work" they sometimes eclipse social preferences. In this case the eclipse was complete: whatever motives induced the subjects to contribute generously in the absence of the incentives were entirely absent once the incentives were imposed.

Even more interesting from the standpoint of the durable influence of incentives on preferences is the effect of having previously experienced the incentive system on how the subjects behaved after the incentives were withdrawn. If the effect of the incentive was simply to provide a situational cue then play in the absence of an incentive should not depend on the prior experience with an incentive system. If the prior experience matters then either the framing effect of the incentive is not immediately erased (which is certainly possible), or the incentive had a learning effect, or both.

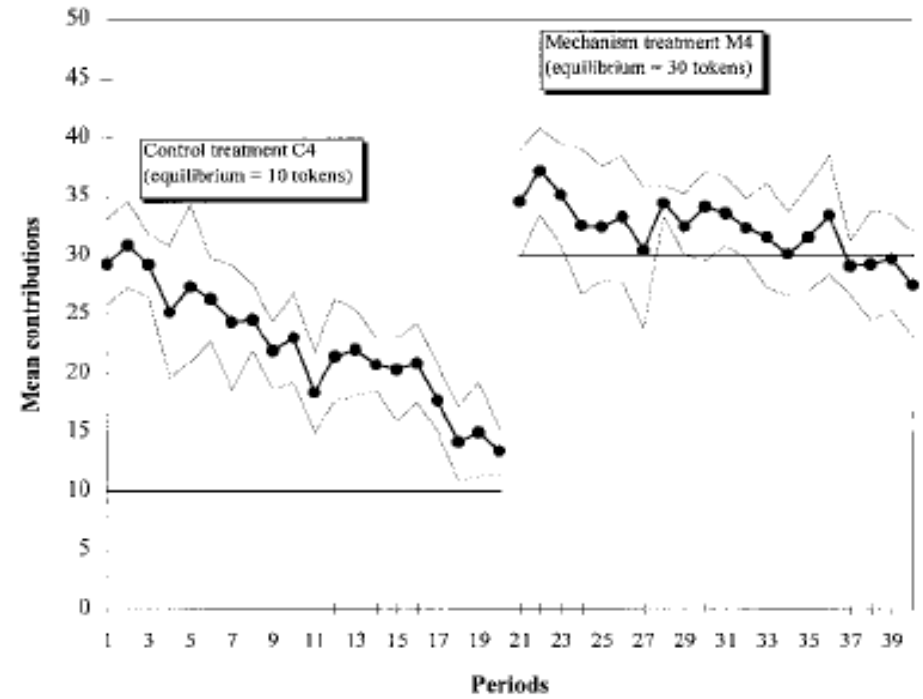
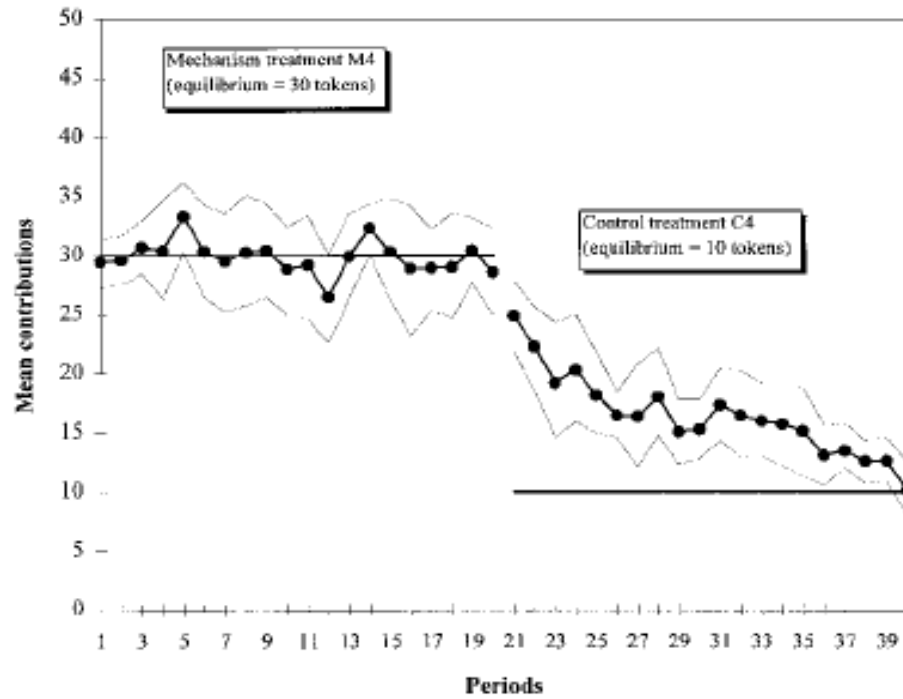
What happened in the experiment (shown in Figure 5.1) is that in the absence of the incentive, the subjects who had previously experienced the incentive system contributed 26 per cent less than those who had never experienced it. (Falkinger, Fehr, et al. (2000).) Because of the short duration of experiments we are unable to distinguish here between the possibility that the framing effect of the incentives persisted long enough to affect subsequent play and the alternative explanation that the incentive actually altered the individual's preferences in a durable way.

A durable negative effect of the experience of incentives occurs, too, in an experimental Gift Exchange game implemented by Simon Gaechter and his co authors (Gaechter, Kessler and Konigstein (2010)) The game is a sequential prisoners' dilemma in which the "employer" chooses a wage to offer the "employee" who then can either accept the wage or not. "Workers" who accept the wage then select a level of "production" that is costly for the worker to provide (it reduces the 'worker's' payoffs) and beneficial for the principal. Selfish agents would obviously accept any positive wage and then produce nothing.

In the two 'incentive' treatments of this standard game, the principal can offer a contract specifying not just the wage, but also the desired level of output, making the wage payment conditional on the employer's output target being fulfilled by the worker. In the "fine" treatment under this setup, failure to meet the target is penalized by a wage reduction, while in the "bonus" treatment, meeting the target is rewarded by a wage increase. The standard setup without targets, bonuses or fines is called the 'trust' treatment because a principal would offer a positive wage only if he trusted the agent to reciprocate by providing sufficient production to more than offset the wage.

The authors suspected – based on earlier experiments – that in the standard game principals would indeed trust and agents would reciprocate,. And they did as the solid line in Figure 5.2 shows, although the workers production declined over time. By contrast, incentives (either fines or bonuses) sustained high levels of production over the entire first phase.

Figure 5.1. Incentives work; but they also crowd out social preferences in subsequent periods. Source: Falkinger, Fehr, et al. (2000).



But the authors' main interest was whether this trusting and reciprocal behavior in the trust treatment would be affected by having first played the game under the fine or bonus treatments. The results are in the right panel of the figure. When subjects who had participated in ten periods of the trust treatment played another ten periods of the same game, as before with partners chosen randomly after each round (the solid line), generous wages were common and were reciprocated, so that production levels remained high (with as before a modest decline over time). But when subjects who had experienced either the bonus or fine treatment for the first ten periods played the trust treatment (that is, in the absence of incentives) in the second set of ten periods, production fell far below the levels of those who had never experienced the incentive treatments. The difference appears to be explained by the destruction of reciprocal motivations: conditional on a given wage being offered by the principal, the production level offered by the worker was substantially (and statistically significantly) lower among those who had experienced the incentives in the first period.

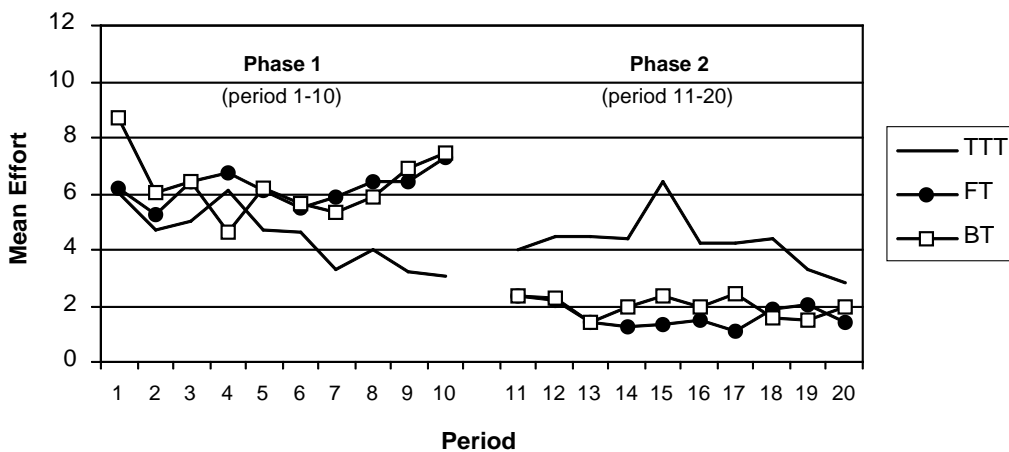


Figure 5.2 Effects of the prior experience of incentives on trust and reciprocity. Each treatment is a series of ten one shot sequential prisoners' dilemma (see text), each with a randomly selected partner. TT refers to the subjects who played the game (without fines or bonuses) in both phases. FT and BT are respectively those who played during the first phase with incentives (respectively fines and bonuses) and in the second phase without incentive. The difference in the second set of periods between the TT subjects and the FT or BT subjects is statistically significant. Source: Gaechter, Kessler, et al. (2010)

I say “appears to be” because we do not have any observations on the subjects’ motivations. Remarkably, the best clues about the motivational learning that takes place under the influence of incentives comes from what may be the first experiment on the way that incentives may crowd out ethical reasoning designed (not surprisingly) by two political scientists.

Norman Frohlich and Joe Oppenheimer implemented a 5- person public goods game under two conditions. In addition to the usual game, they introduced an ingenious Rawlsian “veil of ignorance” treatment which aligns self interest and the interests of all members of the group, thus abolishing the social dilemma. In the Rawlsian treatment each of the five players chooses how much to contribute to the public good, but then receives the payoffs of a randomly chosen one of the five players, that is, whatever he has contributed, with equal probability he receives with equal probability the payoff associated with how he played the game and the payoff associated with how each of the other four played. In the usual game as in the Prisoners Dilemma, an individual maximizes her payoffs by contributing nothing. In the Rawlsian game however, the best one can do is to contribute the amount that will maximize the average payoffs of all five members of the group as this will maximize each player’s expected payoffs. Players had no difficulty figuring out that this requires contributing the maximum.

Ten groups played each of these treatments (conventional Prisoners Dilemma and the Rawlsian veil of ignorance game), half of them with a brief period of discussion among the players prior to their (anonymous and simultaneous) play. No communication was allowed in the other groups. All of the groups then entered a second phase in which no communication was allowed and only the standard Prisoners’ Dilemma Game was played.

Not surprisingly, in the first phase and in the absence of communication, subjects in the veil of ignorance game contributed much more than those in the standard game, just as those playing under the Falkinger mechanism described above contributed substantially to the public good. Also unsurprisingly, when communication was allowed, contributions in both the veil of ignorance and the usual game were significantly higher. But the effect of communication was much greater in the standard game, and those in this treatment (standard game x communication) contributed slightly (but statistically significantly) more than those playing the veil of ignorance game with communication.

The authors, however, were more interested in the second phase of the experiment, in which everyone played the standard game. They wanted to know whether playing the veil of ignorance game would induce the subjects to act more generously towards others. They hypothesized that those who had earlier played behind the veil of ignorance would recognize what the fair and socially optimal outcome was and would be more motivated by fairness considerations when they came to play the normal game.

But this is not at all what they found.⁷ In the groups that had not communicated in phase 1 there was no difference in the levels of contributions between those who had played the veil of ignorance game and the standard game when they both played the standard game in phase two. But among those who had communicated during phase one, those who had played the standard game contributed twice as much in phase two than those who had earlier played the veil of ignorance game. The experience of having played the game in which self interest was a good guide to what was the best for all group members apparently made people less able to articulate or respond to arguments of fairness when they were facing a real social dilemma and were allowed to communicate with one another.

When surveyed, subjects playing behind the veil did indeed recognize that they were playing fairly. But in the second phase when they played the standard game their concerns about fairness were uncorrelated with (and apparently had no effect on) how much they contributed. By contrast those who had played the standard game in the first phase self reported that they had played less fairly, but in the second phase fairness concerns were a strong predictor of contributing larger amounts.

To interpret the results, note that the Rawlsian treatment is what economists call an incentive compatible mechanism in which prices do the work of morals so that the pursuit of self interest implements the socially optimal outcome. The authors' explanation of the results is then:

When subjects ... do not have to think about ethical dilemmas (because the problem has been solved by the imposition of an incentive compatible device) they do not pay as much attention to the ethical dimension of their choice. As a result, when they subsequently have to make ethically problematic decisions they are more likely to downplay the ethical component than are those who have had to confront the dilemma. (Frohlich and Oppenheimer (1995))

This is a "use it or lose it" interpretation of the eclipse of moral reasoning as a result of

exposure to prices doing the job of morals. It is entirely at variance with the view, common among economists that incentive-compatible mechanisms like ideal competitive markets are to be admired precisely because they economise on “the scarce resources of altruistic motivation” that might otherwise be used up. (Arrow (1972):3)

Competition among self regarding individuals on competitive markets for goods with complete contracts is widely recognized as an incentive compatible mechanism with attributes similar to the Rawlsian veil of ignorance treatment in the experiment. This is the reasoning underlying Buchanan’s indifference towards the condition of his fruit seller, and Gauthier’s claim that morality has no role in evaluating market outcomes. If Frohlich and Oppenheimer are right that such settings citizens do not have to “flex their ethical muscles” (Frohlich and Oppenheimer (1995)) then one wonders why Marx’s dire prediction on the evolution of culture in capitalist economies turned out to be mistaken.

The experimental evidence on the effects of incentives and other aspects of the economy on preferences thus does not help much to resolve the puzzle with which we began. It seems that ethical crowding out effects are substantial and that what the economy teaches is long lasting and tends to be generalized from one’s economic experiences to other domains of life. The puzzle deepens considerably when we use experiments to compare behavior across countries, and to ask if societies in which markets play a larger role also suffer a deficiency of social preferences.

MARKETS AND CIVIC VIRTUE

As the result of three large cross-cultural behavioral experiments across a broad range of economic and political systems we now have behavioral measures concerning individuals’ cooperativeness, fair-mindedness and other social preferences. In addition to the Third Party Punishment Game, Dictator Game, and Trust Game mentioned above, the Ultimatum Game, and the Public Goods with Punishment Game (described below) also provide behavioral measures of generosity, willingness to sacrifice personal benefits to uphold fairness and other social norms and to contribute to a public good. These three studies provide evidence that these behaviors flourish in market based societies, though to varying degrees.

The most surprising evidence comes from the experimental Ultimatum Game played by subject pools in 15 isolated small-scale societies (Henrich, Boyd, Bowles et al. (2005), not the

same 15 as in the study just described, incidentally). In this game my collaborators and I anonymously paired subjects for a single interaction. One is the “responder,” the other the “proposer.” The proposer is provisionally awarded an endowment (‘the pie’), known to the responder, to be divided between proposer and responder. The proposer then offers a certain portion of the pie (including none) to the responder. If the responder accepts, the responder gets the proposed portion, the proposer keeps the rest, and the game is over. If the responder rejects the offer, both get nothing and the game is over.

Entirely self-regarding proposers who believe that respondents are also self-regarding will anticipate that no positive offer will be rejected and so will offer the least possible amount. This prediction from the assumption of self interest rarely has been observed in literally hundreds of experiments in dozens of countries. Our study was no exception. Most proposers offered substantial amounts to the responder; and low (but nonetheless positive) offers were rejected. We interpreted the rejection of positive offers as evidence of a social preference for fair outcomes (termed inequality aversion by economists).

The substantial offers made by most proposers could also be explained by the proposers inequality aversion, or altruism towards the responder, or simply an ethical commitment that it is ‘the right thing to do.’ But substantial offers could also be made by a propose attempting to maximize her expected payoffs taking account of the fact that low offers are more likely to be rejected. Thus calling these substantial offers generous refers to their amount not their motivation. But they are nonetheless evidence of the strength of social preferences, either those of the proposer herself, or her beliefs about the likelihood of a rejection due to the inequality aversion of the responder.

In our study of hunter-gatherers, herders, and low technology farmers (horticulturalists), the groups with greater exposure to markets on average both made more generous offers as proposers in the Ultimatum Game and as respondents were more willing to reject low offers and as a result receive nothing rather than accept a highly unequal division of the pie. The two least market-exposed groups – the Tanzanian Hadza hunter gatherers and Amazonian Quichua horticulturalists – offered a quarter and a third of the pie (respectively) in contrast to the highly market-integrated Indonesian Lamalera whale hunters, who offered on average more than half of the pie to the respondent. Considering all of the groups, a standard deviation difference in our

measure of market exposure was associated with about half a standard deviation increase in the mean Ultimatum Game offer.

These were eyebrow raising findings among anthropologists and other social scientists among whom it is not uncommon to think that markets make people selfish. *The Wall Street Journal*, unsurprisingly, saw things differently. They headlined their January 24, 2002 front page story describing our results as “The Civilizing Effect of the Market.”

A second phase of this project studied primarily rural peoples in Africa, Oceania, and South America (Henrich, McElreath, Barr et al. (2006), Henrich, Ensminger, McElreath et al. (2010)). (This is the project that produced the evidence about the crowding out of religion in the Third Party Punishment Game in Accra). The correlation of Ultimatum Game offers and the extent of market exposure found in the first phase was reproduced in the second phase (of approximately the same magnitude), and a similar positive market correlation was found for offers in the Dictator Game and the Third Party Punishment Game.

These results are not inconsistent with the experimental evidence that I presented in the previous chapter. The same Accra workers for whom monetary incentives apparently reduced the salience of religion and resulted in less generous behavior were among the most market-exposed in this study (they acquired all of their food by purchase, for example) and also among the most generous, offering well above the average of the 15 subject pools in the Dictator and Ultimatum game.

Unlike the first phase of this project, the second included one market-based liberal society: a rural population in Missouri (USA). We can gauge the Missourians’ fair-mindedness in the Ultimatum Game by the minimum offer (fraction of the pie) that would accept. This was reported to the experimenter (at the outset of the game). This so called minimum acceptable offer (MAO) is also the greatest amount the subject is willing to forgo in order not to accept an unfair offer. The MAO thus captures at once the subject’s “willingness to pay” for fairness and the least advantageous division of the pie that the subject considers to be fair enough to not reject. The Missourians’ MAO was the third highest among the 15 subject pools. Controlling for subjects’ age, sex, schooling, and the average income, the Missourians minimum acceptable offer was 2.6 times the average of the other groups, and 2.4 times the MAO of the famously egalitarian Hadza hunter-gatherers (Woodburn (1982).) In the Dictator Game, virtually all of the Missourians

offered half the pie, making them the most generous of the populations (recall that the Hadza subjects offered a quarter, on average).

More comprehensive evidence comes from experiments with an usually diverse set of (also coincidentally 15) subject pools, including some from quintessentially liberal societies (U.S., U.K., Switzerland, Germany, Denmark, Australia) and others (Turkey, Russia, Saudi Arabia, China, Oman, South Korea). This study also provides (as we will see in the next section) an idea that may help resolve the puzzle with which we began. Cultural differences among the subject pools in this study may be somewhat attenuated, however, because (unlike the previously-mentioned field experiment studies with herders, hunter gatherers, farmers and such) the subjects are university students (Herrmann, Thoni and Gaechter (2008a)). The common experiment implemented (by the same experimenter) in these sites is a Public Goods with Punishment Game.

This is a modification of the Public Goods Game, the n-player prisoners' dilemma described in the previous chapter. The n players are each awarded an endowment and given the opportunity anonymously to contribute some, all or none of this to a common pot (the public good), the amount in which (after all the contributions are made) is doubled or tripled and then distributed in equal parts to the players, irrespective of the amounts they contributed. This describes a public goods game if the group size and the multiplication factor is such that the individual would maximize payoffs by contributing nothing irrespective of what the others do, and yet that total payoffs (summing over the group) are maximized if everyone contributes the entire endowment. (For example if there are 5 members of the group and the multiplication factor is two, then by contributing 1 to the public pot one would increase their payoff from the distribution of the common pot by $2/5$ which clearly does not justify foregoing the 1; yet if everyone contributed 1, then each would receive 2).

The punishment modification of this game is that after all players have made the allocation to the common pot, each is provided with information about the contributions of each other player (the identities are not given, just an ID number known only to the experimenter) and given the opportunity to pay (reduce one's own payoff) in order to reduce the payoff of any other member in the group. This procedure is followed on each of the periods of the game (often ten). The presence of an option to punish a player who violates a norm of generosity towards others makes the Public

Goods Game with Punishment similar to the Third Party Punishment Game played in Accra in the last chapter, but in that game the punisher was a third party, not a one whose payoffs are directly affected by the behavior of the possible target of punishment.

A summary of the results is in Figure 5.3 where the groupings of the subject pools (though somewhat arbitrary) dramatizes the considerable cultural differences evident in the experimental play.

This game provides information on three behaviors motivated by social preferences: willingness to contribute to a public good (public generosity) and to penalize those who do not (upholding social norms) both at a cost to oneself, and the degree of positive response to punishment by others (shame at one's violation of a social norm). Where all three of these dispositions are present, contributions to the public good will be substantial.

As expected, cultural differences in game play among the subject pools were significant, but in all of them (as is common with other experiments (Fehr and Gaechter (2000a)) subjects contributed substantial amounts in the first period. But in the absence of the punishment option, in subsequent periods cooperation unraveled. However (also as expected from other experiments) when the punishment option was available it was widely used, especially in the early periods, and as a result the unraveling of contributions did not occur in any of the 15 subject pools. In the experiment with punishment, the subject pools with the highest average contributions were (in order) Boston, Copenhagen, St. Gallen (Switzerland), Zurich, and Nottingham; the lowest average contributions were in Athens, Riyadh, Muscat (Oman), Dnipropetrovs'k (Ukraine) , and Samara (Russia).

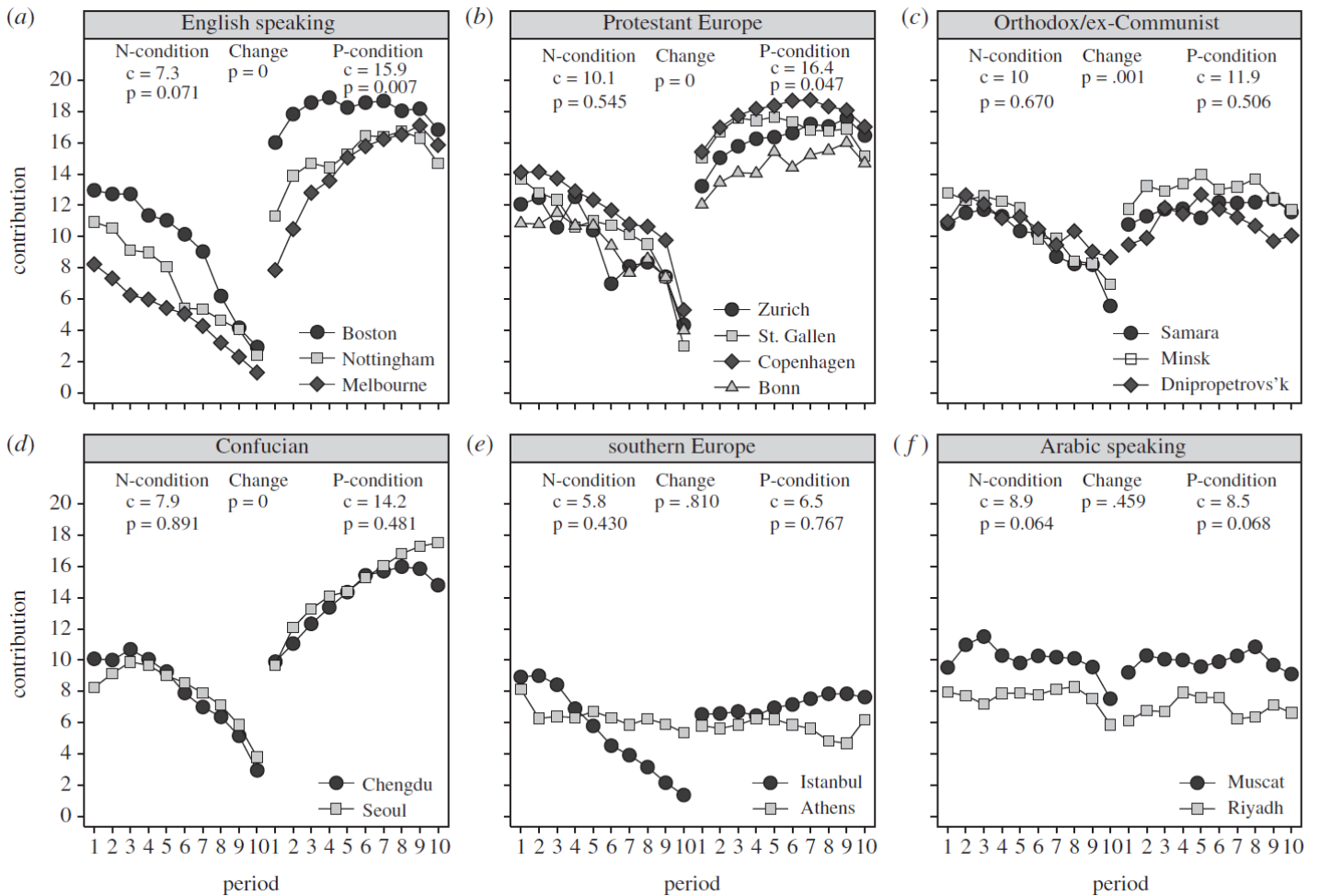


Figure 5.3. Cultural differences in the response to punishment in a public goods game.
Source: Gaechter, Herrmann and Thoni (2010). See text.

Average contribution levels in the subject pools correlated positively with measures (for the populations from which the subjects were drawn) of the rule of law (the correlation coefficient among these two measures, $r = 0.53$), democracy ($r = 0.54$), individualism ($r = 0.58$), and social equality ($r = 0.65$). Positive correlations were also found, as expected, with survey measures of trust ($r = 0.38$).⁸

Individually costly voluntary contribution to a public good to be shared with strangers is surely a measure of the civic virtues that Marx thought would not thrive in a market economy. The same is true of rejections of positive offers in an ultimatum game, which show that people are willing to sacrifice their own material gain in order to punish those who violate social norms of

fairness. That these behaviors are greater in nations characterized by a greater extent of market interactions is puzzling. Understanding why these cross countries correlations occur will cast further doubt on idea that market based economies will promote “universal venality.”

SUSTAINING SOCIAL ORDER IN LIBERAL AND OTHER SOCIETIES

The difference between the cooperating and free-riding subject pools in the cross cultural study just described is due to the use of punishment and the response to being punished. In the experiment without the punishment option, subjects in Samara, Dnipropetrovs’k and Muscat contributed more than those in Boston, Nottingham and Zurich. The reason why these subject pools did less well in the punishment version of the game is that a significant amount of punishment was directed not only at shirkers but also at high contributors. The latter may have occurred as a vendetta-like retaliation against punishment received in earlier rounds by subjects who believed that it was the high contributor who were doing most of the punishment (Figure 5.4.) The authors termed punishment of those contributing the same or more than the subject “anti-social punishment.” Other experiments have found the same patterns.

The extent of anti-social punishment was significantly and inversely correlated with the previously mentioned societal measures of the rule of law ($r = - 0.53$), democracy ($r = - 0.59$) individualism ($r = -0.63$), social equality ($r = - 0.72$). In the five high-contributing subject pools(Boston, Copenhagen, St. Gallen, Zurich, and Nottingham), shirkers who were punished responded by significantly increasing their contributions in subsequent periods. In only one of the 5 low contributing subject pools (Athens, Riyadh, Muscat, Dnipropetrovs’k, and Samara) did shirkers respond positively to punishment (in four the response was not significantly different from zero.)

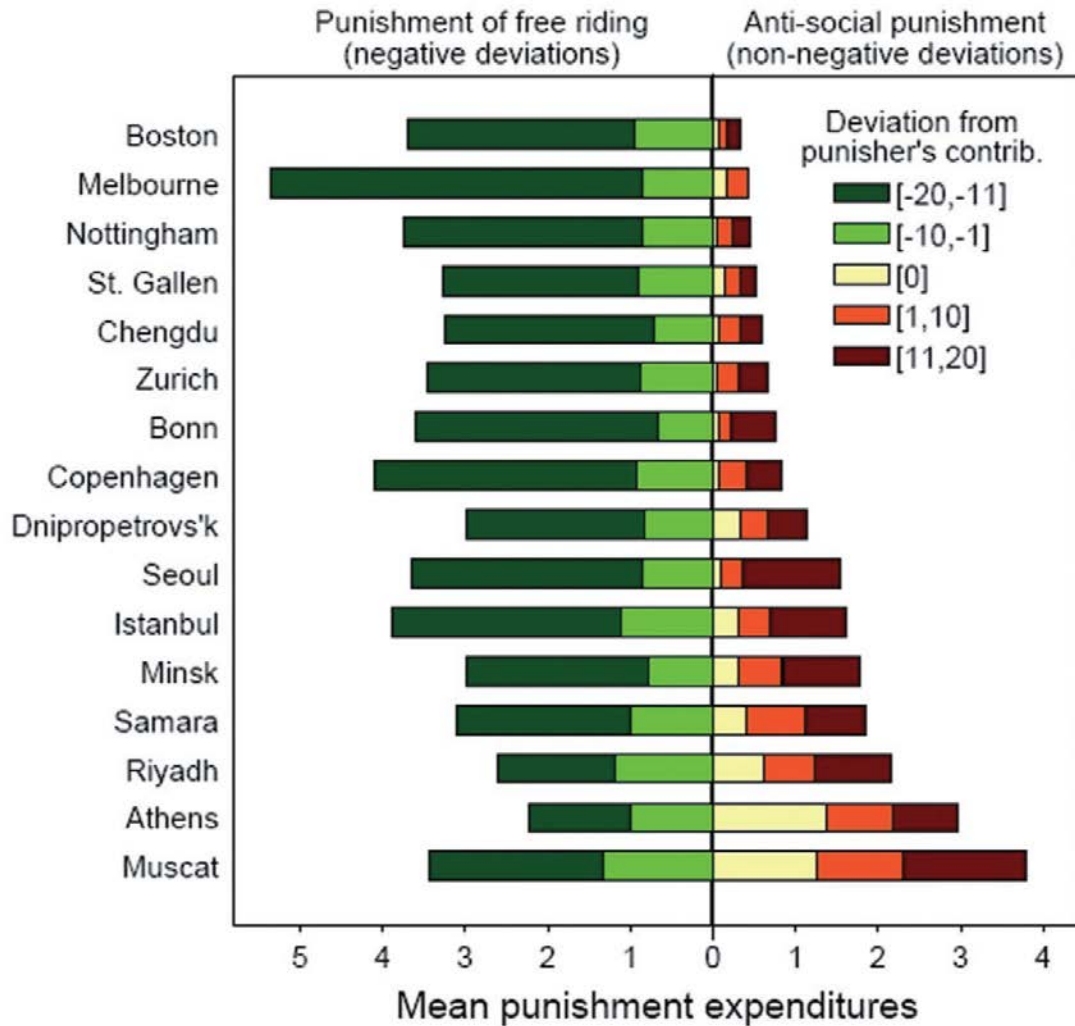


Figure 5.4 Anti-social punishment in a Public Goods Game. Bars to the left of 0 indicate the extent of punishment of those who contributed less than the punisher. To the right are punishments targeted on those who contributed the same or more (darker shades) than the punisher (anti social punishment). Source: Herrmann, Thoni, et al. (2008a)

A plausible explanation of the differing uses of punishment and reactions of its targets is that punishment works only if it is regarded as legitimate, conveying the signal that the target has violated widely held norms. It appears that punishment of free riders – even by complete strangers – is legitimate and evokes shame, not anger, in Boston and Copenhagen but in Muscat and Samara it is seen differently.

An experimental exploration of the effect of legitimacy on the efficacy of punishment by

Ertan, Page, and Putterman is consistent with this interpretation. Prior to playing the Public Goods Game, each group of experimental subjects in Providence (USA) was invited to deliberate and to vote on whether punishment should be allowed, and if it should be restricted in any manner. Unlike the highly individualistic set up of the more common Public Goods Game with Punishment, the Ertan et al experimental protocol thus matches real world practices of cooperation and norm enforcement as revealed in ethnographic and other studies (Mahdi (1986), Wiessner (2005)). Here is what they found: “no group ever allowed punishment of high contributors, most groups eventually voted to allow punishment of low contributors, and the result was both high contributions and high efficiency levels” (Ertan, Page and Putterman (2009).) Apparently the determination of the punishment system by majority rule made the punishment of shirkers not only an incentive but also a signal of group norms.

This result suggests the following hypothesis to explain the contrasting levels of cooperation sustained by peer punishment in experiments with subject pools from liberal and other societies. Consider the structure of what anthropologists call a lineage-segmented society. Lineages are the fundamental social unit, composed of families sharing a (perhaps quite distant) common ancestor and performing essential functions of risk pooling and redistribution. These segments are also responsible for the moral instruction and behavior of their members, and for the appropriate rectification of any transgressions towards members and non-members alike, including punishment and compensation where appropriate (Mahdi (1986), Boehm (1984).) Punishment by a non-member for a member’s misbehavior may itself be considered a transgression, requiring rectification or inviting retaliation. Ernst Gellner’s description of pastoralists as “a system of mutually trusting kinsmen” is an example. These are “strong, self-policing, self-defending, politically participating groups...They defend themselves by means of indiscriminate retaliation against the group of any aggressor. Hence they also police themselves and their own members, for they do not wish to provoke retaliation.” (Gellner (1988):144-145)

By contrast, in liberal societies the tasks of moral instruction and the maintenance of order are routinely entrusted to individuals who are unrelated and at least initially unknown to those who they teach, police, and judge. Inverting the moral code of lineage segmented societies, the legitimacy of these teachers and police and court officers is based on their anonymity and lack of

relationship to those with whom they interact, enhanced by their uniforms, degrees, and official titles acquired (at least ideally) through a process of fair competition, not family favor. Perhaps this explains why when Boston subjects who contributed less than the average in the public goods game were punished, they substantially increased their contributions, while under the same conditions subjects in Dnipropetrovs'k actually reduced theirs (though not by a significant amount). While the incentive to contribute more was no doubt salient in both cases, the signal may have differed. Boston subjects may have read the fine as disapproval by fellow citizens, while for those in Dnipropetrovs'k it was perhaps an insult.

This hypothesis has yet to be tested empirically; but if it were found to have merit, it would direct attention not to the cultural consequences of markets, but rather to liberal political, judicial and other non-market institutions as the key to liberal civic culture. This not the usual explanation of the civic culture of liberal societies, namely the *doux commerce* hypothesis that credits the exchange process itself.

DOUX COMMERCE?

Voltaire was astounded that in the mid 18th century, at the Stock Exchange of London “the Jew, the Mohammedan, the Christian deal with one another as if they were of the same religion, and give the name infidel only to those who go bankrupt ... the Presbyterian trusts the Anabaptist and the Anglican takes the Quaker at his word” and “upon leaving this peaceful and free assembly, some withdraw to the synagogue, ... others retire to their churches... others to have a drink ... and everyone is happy.” (Voltaire (1961):18). Perhaps he was indeed observing “the civilizing effect of markets” that the *Wall Street Journal* celebrated with the help of our experiments.

Understanding why people of the more market exposed societies in our study made more generous offers in the game, and were more likely to reject low offers requires two pieces of information. First in many of the populations that we studied, interactions with strangers are often fraught with danger, but this is less true where regular market exchanges occur for the simple reason that some of the strangers one meets provide opportunities for mutual gains through exchange. And second, our experimental subjects played anonymously, which may have cued

them to play in a way that would have been appropriate behavior towards strangers. A plausible explanation of the more generous and fair-minded experimental behavior in the more market oriented societies then, is the following: people learn from their market experiences that fair dealing with strangers is often quite profitable. Maybe the members of the Stock Exchange who so impressed Voltaire had learned something similar.

This suggests that there may be something to the key hypothesis of the *doux commerce* thinkers, namely, that “where there is commerce the ways of men are gentle” as Montesquieu put it (Montesquieu (1961):81). Smith comes the closest to providing a causal explanation of why this would be the case, using as an example, the probity of the merchant (in contrast with ambassadors):

When a person makes perhaps twenty contracts in a day, he cannot gain so much by endeavoring to impose on his neighbors, as the very appearance of a cheat would make him lose. When people seldom deal with one another, we find that they are somewhat disposed to cheat, because they can gain more by a smart trick than they can lose by the injury which it does their character (Smith (1896)):254-5.

This is a reputation-based variant of a large class of game theoretic models in which, where contracts are incomplete or promises unenforceable, frequent or durable exchanges with known individuals provide incentives for otherwise self-regarding individuals to adopt norms such as honesty and diligence towards their partners, thereby underwriting exchanges, the mutual benefits of which would otherwise be compromised by malfeasance (Bowles (2004)):232-249)

If Smith is right, then, markets in which exchanges take place with a restricted number of people and are repeated over a long period might promote honest dealing. And just as Kohn showed that the lessons learned at work about authority and independence are generalized to child rearing values and other realms, the associated social norms might become more generally diffused. Perhaps this explains why when the anthropologist arrived with an odd game to play and with real money on the table, our more market exposed experimental subjects were more concerned about fairness and were more generous to their partners.

Smith’s argument that repeated interactions with many individuals all known to each other may promote honest-dealing makes sense. But it does not provide an explanation of why markets

would provide a more favorable setting for this dynamic, as opposed to other institutions such as families, states, or teams of people who regularly work together. In these other settings the number of people with whom one interacts is fewer than in most markets and the expected repetition of interactions much greater. I will return to Smith's honest merchant in the next chapter. For now it is sufficient to say that I do not think that Smith's version of the *doux commerce* hypothesis (or any of its other variants) adequately accounts for the civic minded citizenry of many of the highly market oriented societies, which may. If I am right, the explanation has everything to do with non market aspects of liberal social orders.

A LIBERAL CIVIC CULTURE

Here is my proposed resolution of the puzzle of the robust civic cultures in societies in which markets play a major role. I will first explain the idea, and then present some evidence.

Liberal states have neither the information nor the coercive reach to eliminate opportunism and malfeasance, but they can protect citizens from worst-case outcomes, whether these are personal injury, loss of property or other calamities. The result, writes Norbert Elias (2000) is a "civilizing process" based on the fact that "the threat which one person represents for another is subject to stricter control...everyday life is freer of sudden reversals of fortune [and] physical violence is confined to the barracks..." This attenuation of calamity is accomplished through the rule of law, occupational and other forms of mobility, and (in the past half century or so) by social insurance.

By minimizing risk, the state thus becomes a substitute for familial and parochial ties on which lineage-segments and other traditional identities. Because these ties are therefore less valuable, they are less likely to be sought, and maintained. The result is a cultural environment favorable to the evolution of more universal norms that apply to strangers as well as family and clan. De facto insurance against worst-case outcomes may also free people to act on their social preferences by assuring them that those who conform to moral norms will not be exploited by their self-interested fellow citizens.

While this risk-reduction aspect of the liberal state affects the entire panoply of social interactions, I will illustrate it by showing how it would promote trust in a market exchange. (A

game theory model of this argument is in appendix 2.) Consider a population composed of a large number of people who interact in pairs to engage in an exchange in which they may either behave opportunistically (e.g. attempt to steal the other's goods) or exchange goods to their mutual benefit. Call these strategies "defect" and "cooperate." Suppose that a defector takes the goods of the cooperator, but runs the risk of taking a beating in process, so that cooperating is a best response to being paired with a known cooperator. Defecting is always the best response to a defector. Though mutual cooperation (and exchange) maximizes total payoffs (and also the individual payoffs for both individuals), a trader paired with a unknown stranger would defect in the absence of a reasonable assurance that the stranger is a cooperator.

How confident would you need to be that your exchange partner is trustworthy in order motivate you to "cooperate"? This will depend on the consequences of inadvertently cooperating with a defector. If being exploited by a defector inflicts serious costs on the hapless cooperator, you would cooperate only if you were virtually certain that he was trustworthy. What the liberal state did was to make the consequences of trusting a defector much less dire. The rule of law thus lowers the bar for how much one would have to know about ones partner before trusting him. Thus the rule of law could promote the spread of trusting expectations and hence of trusting behavior in a population. John Rawls (1971) provides a complementary argument: "when it is dangerous to stick to the rules when others are not" (p. 336) "public institutions" may penalize defectors, thereby reducing their numbers, lowering the probability that a cooperator will be exploited by a defector, and so minimizing the would-be cooperator's motivation to preemptively defect as a risk minimizing strategy

Markets, of course, are part of this story, one that is complementary to the risk reduction provided by the rule of law. In the example just given, the occasion for a trusting relationship between the buyer and seller would not have arisen were it not for the possibility of mutual gain through exchange. The synergistic effects of markets and the rule of law in favoring the evolution of trust among strangers may account for Voltaire's observations about cooperation among different religions on the Stock Exchange and the surprising results of our cross cultural experiments.

Markets may have been part of “the civilizing process” in further ways. In many cases the spread of markets contributed to the emergence of national states bound by the rule of law, and if the above argument is correct, this favored the evolution of generalized trust. The expansion of markets also favored the proliferation of more universal social norms by promoting the national systems of schooling-by-strangers that Gellner termed *exo-socialization* (Gellner (1983)). Indeed Gellner writes that markets could regulate a division of labor at the national level only if the multiplicity of parochial traditional cultures were replaced by more universal values consistent with the extensive interaction with strangers in market environments. The resulting national standardization of language and culture facilitated occupational and geographical mobility, rendering individuals’ income-earning assets less specific to place and craft. The result was to complement the other literal and *de facto* forms of insurance provided by liberal institutions (D’Antoni and Pagano (2002), Bowles and Pagano (2006).)

Consistent with my proposed resolution of the puzzle, the emergence of the rule of law appears to have been associated with a parallel shift from trust in kin and other particular individuals to generalized trust, consistent with social psychologist Toshio Yamagishi’s “emancipation theory of trust” (Yamagishi, Cook and Watabe (1998), Yamagishi and Yamagishi (1994), Gambetta (2008)). Tabellini (2008), for example, shows that generalized (rather than familial) trust appears to thrive in countries with a long history of liberal political institutions. The strong inverse association between measures of political participation such as signing petitions or participating in demonstrations or boycotts on the one hand, and on the other, measures of the extent of one’s obligation to respect and care for one’s children and parents in a large sample of immigrants to Europe is also consistent with this view (Alesina and Giuliano (2009).)

This process of states substituting for familial or parochial norms appears to have been at work in the 11th century Mediterranean trading system, which, with the expanding geographical reach of markets, witnessed the eclipse of familial, communal and other parochial systems of so-called “collectivist” contract enforcement by more universalistic state-based “individualist” systems (Greif (1994).) These are some of the reasons why market-based societies may exhibit high levels universalism in the definition and application of social norms.

The other way I suggested that the rule of law may support social preferences is by

protecting those who conform to social norms from exploitation by defectors. This could occur both because defectors would be fewer under the rule of law (Rawls' argument) or because the assurance that defectors would be punished for their transgressions might reduce the appeal of a pre-emptive defection by an individual who would otherwise prefer to conform to the norm.

This second crowding in effect is evident among the Hokkaido University subjects who cooperated more in a public goods experiment when assured that others (but not themselves) would be punished if they did not contribute sufficiently, despite the fact that this had no effect on the subject's own material incentives (Shinada and Yamagishi (2007)). They apparently wanted to be cooperative but wished even more to avoid being the sucker who is exploited by defectors, exhibiting what Iris Bohnet and her co authors call "betrayal aversion" (Bohnet, Greig, Herrmann, et al. (2008)). The fact that defectors would be punished by a third party (the experimenter, a surrogate for the state) attenuated the subject's fear that a defector would gain at her expense. Similar synergies occur in natural settings: social norms support observance of traffic regulations, but these may unravel in the absence of state-imposed sanctions on flagrant violations. These, or similar causal mechanism may well account for the few experiments in which material incentives and moral motives were complements rather than substitutes, the former enhancing the salience of the latter.

I have suggested that the puzzle of vibrant civic cultures in many market based societies may be resolved by attention to the ways in which geographical and occupational, mobility, rule of law and other aspects of the liberal state represent a distinctive way of preserving social order, one that does this in part by sustaining civic virtues. If I am right, then, the kinds of incentives and constraints that people face in a liberal democratic and market based society sometimes affect a kind of crowding in of social preferences rather than the opposite more commonly observed in experiments.

This is just what the Aristotelian Legislator was hoping to accomplish: to design public policies such that the structure of incentives and constraints facing people worked synergistically rather than at cross purposes with peoples' ethical and other regarding dispositions. It is time, he thinks, to visit a few economics faculties to see what help he can on this score. He will be surprised by the warm welcome he receives.

VI

ECONOMICS DISCOVERS ARISTOTLE

Since Richard Titmuss' *The Gift Relationship: From Blood Donations to Social Policy* (1971), economists have been intrigued but for the most part un-persuaded by the claim that policies based on explicit economic incentives may be counter-productive because they induce people to adopt a 'market mentality' and thus compromise pre-existing values to act in socially beneficial ways (Arrow (1972), Solow (1971), Bliss (1972).)

At the time of its publication there were two reasons to doubt Titmuss' claim. First, there was little hard evidence that the social preferences such as altruism, fairness, and civic duty that are said to be compromised by economic incentives are important influences on individual behavior or are in any way essential to the functioning of a market-based economy. Second, even if these social preferences were thought to be important influences on behavior, there was even less evidence in the Titmuss (1971) book that explicit economic incentives undermine them.⁹

Another reason why economists for the most part did not share Titmuss' concern that incentives might drive out ethical motivations was the then burgeoning field of mechanism design. This body of research was an extension of the earlier Marshall-Pigou tradition in what is called welfare economics that addressed market failures by means of green taxes, training subsidies and other incentives. Mechanism design held out the promise that well designed incentives could, along with market prices, implement efficient outcomes, even among entirely self-regarding citizens. At the time when Titmuss' book appeared, virtue, it seemed, was something that economists could safely ignore as J.S. Mill had advised.

But research during intervening years changed upset this complacency. As we have seen, experimental and other evidence has made it clear that ethical and other regarding motivations are common and moreover that, as Titmuss had claimed, they are sometimes crowded out by the use of incentives. At the same time economic theory had turned to the study of the exchange process when contracts are incomplete, that is, when it was not the case that everything that mattered could be stipulated in an enforceable agreement. New models of the labor market, for example,

recognized that work itself was not something that could be contracted for so that getting a job done well depends at least in part on the employee's work ethic. The new microeconomics of labor, credit and other markets, enumerated the ways that, as Arrow had earlier said, social norms and moral codes could induce economic actors to internalize the costs and benefits of their actions on others when contracts failed to accomplish this result.

Another advance in economic theory – in the seemingly unrelated field of macroeconomics – also motivated second thoughts about the Machiavellian policy paradigm. A generation ago Robert Lucas (1976) rocked economics with a simple observation: taxes and other government interventions in the private economy affect not only (as intended) the costs and benefits of actions that citizens may take, but also their beliefs about the actions of others (including the government) now and in the future. Thus just as in the case of citizens values illustrated in Figure 3.1 incentives may have an indirect effect (in this case via citizens' beliefs) as well as the intended direct effect. Here is an example: announcing stiffer penalties for non-payment of taxes provides an incentive to pay up, but it also may convey the information that non-compliance is a common practice, leading erstwhile honest citizens to cheat.

Lucas reasoned that the effect of a policy intervention can be predicted only if one takes into account these indirect effects on beliefs, and then studies the joint outcome in which both citizens' beliefs and the targeted economic actions are mutual dependent. His point was that a new economic policy is not an intervention in a given model of how the economy works, but rather a change in the model itself (Lucas (1976):41-42)

given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure ... relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models.

He concluded: "... policy makers, if they wish to forecast the response of citizens, must take the latter into their confidence" The idea was taken to be so important that economists now express it in upper case letters, bestowing an honor on the Lucas Critique withheld even from the invisible hand. Here I apply Lucas' logic to cases in which incentives affect not only beliefs (as Lucas stressed) but also preferences, and may thus have unintended effects.

Not surprisingly in light of the new experimental evidence and developments in the theory

of contracts and public policy, the discipline of economics, which had spurned Titmuss a generation earlier, rediscovered him: A paper appearing in a top economic journal (Mellstrom and Johannesson (2008)) asked "Was Titmuss right?" (The paper gives an affirmative answer for women but not for men). Much earlier, Albert Hirschman had chided his fellow economists, who, he said, propose

... to deal with unethical or antisocial behavior by raising the cost of that behavior rather than proclaiming standards and imposing prohibitions and sanctions. The reason is probably that they think of citizens as consumers with unchanging or arbitrarily changing tastes in matters civic as well as commodity-related behavior. ... A principal purpose of publically proclaimed laws and regulations is to stigmatize antisocial behavior and thereby to influence citizens' values and behavioral codes. (Hirschman (1985):10)

Economists, some of them at least, had not only reread Titmuss; they had rediscovered Aristotle, asking what his Legislator ought to do given this new knowledge about the possible anti-synergy between incentives and the ethical motives or other social motives that were now recognized as essential to a well functioning economy.¹⁰

Not all economists, however, were convinced that the importance of social preferences and the crowding out problem warranted a new approach to policy. One could, they thought, take on board the new experimental evidence and advances in economic theory without abandoning the Machiavellian policy paradigm. To see this, suppose that the policy maker recognizes the need to attenuate market failures, and also understands that explicit incentives may crowd out the ethical and other regarding motivations that might otherwise induce actors to internalize the costs and benefits that their actions confer on others, thereby mitigating the market failure arising when contracts are incomplete. The policy maker might nonetheless seek to devise systems of subsidies, penalties and constraints to implement socially desired outcomes, even in a population of entirely amoral and self regarding citizens. To think this possible, one would have to bar strong crowding out (so as to preclude cases in which incentives are literally counter-productive) and concede the necessity that the policy maker not be entirely self interested. But accepting these two caveats, a constitution for knaves just might do the job, as Hume had recommended, even (as Hume never intended) in a world of not at all hypothetical flesh and blood knaves.

WHY NOT A CONSTITUTION FOR KNAVES IN A (LIBERAL) SOCIETY OF KNAVES?

This remains the canonical model of policy-making in economics, and it is worth exploring whether it could possibly work. To do this we need the theory of mechanism design (Laffont (2000), Maskin (1985), Hurwicz, Schmeidler and Sonnenschein (1985)). The founding works in this branch of economics appeared at about the same time as Titmuss' book. (Gibbard (1973, Hurwicz (1974, 1972)). The growing concern with environmental, public health, and other market failures provided new urgency for the mechanism designer's attempt at "giving the invisible hand a helping hand," as *The Economist* put it, struggling explain the challenging mathematics of the field when three of its leading scholars – Leo Hurwicz, Eric Maskin, and Roger Myerson --were awarded the Nobel Prize.

Mechanism design seeks to address market failures by providing a set of contracts, property rights and other social rules – in short, a constitution – that can be implemented even when important bits of information about individuals are known only privately and cannot be used by the mechanism designer to implement the incentives, constraints and other aspects of the proposed policy. This privacy of information restriction excludes utopian solutions in which the designer could simply wish away the underlying reason for the market failure. If information about how hard a person worked, for example, or one's true valuation of some good or service could be used by the mechanism designer, then it also could have provided the basis for a private contract among the parties that would have eliminated the market failure that the mechanism designer had been called in to address. So, to address the problem of free riding in a work team, for example, the mechanism designer cannot simply propose an enforceable contract that all work hard and well, or even that each member must truthfully report how hard she worked. If the mechanism designer could do this, then the team members could have done the same, without the help of the mechanism designer and there would have been no need for the mechanism designer in the first place.

The nature of the problem having been defined in this way, three conditions for a solution then define the challenge of mechanism design. The first is that the resulting allocation be Pareto efficient. (Recall that a Pareto efficient outcome one for which there exists no alternative outcome that is feasible given existing resources and technologies in which at least one person is better off

and nobody is worse off.)

Second, policies must rely on the voluntary participation of individuals in their economic activities. They may choose their own course of action guided by their own preferences, including opting out of any possible exchange or other interaction.

The third condition is that there are no restrictions on the preferences that people may have, including those attributed to *Homo economicus*.

Call these three conditions: efficiency, voluntary participation, and preference neutrality. The first imposes a minimal collective rationality condition (minimal because it ignores considerations of justice, for example). The second and third express the standard liberal commitments to individual liberty and to the neutrality of the state on matters concerning individuals' ends (sometimes termed liberal neutrality, Goodin and Reeve (1989)).

Mechanism design allows us to consider how a constitution for knaves would work because included in the unrestricted motivations of the actors must be knavish preferences, including greed and deceit. If mechanism design had succeeded in its quest for privacy-respecting rules that met the three conditions above, then we would have to conclude that a constitution for knaves would work, at least in this limited sense.

But more than four decades of research in mechanism design showed something quite the opposite.

A mechanism is a set of rules for determining how an economy's resources will be used, one that the designer might impose on a population to induce the desired behaviors. The fines and subsidies in the experiments reviewed in the previous chapters are part of the mechanism designer's toolbox. So too are such familiar ways of determining resource use as majority rule, buying and selling on competitive markets, and as we will see, more exotic rules as well. The set-up is strikingly like the problem faced by Machiavelli's benign prince charged with governing well through the imposition of enlightened laws upon a citizenry of "natural and ordinary humors."

The mechanism designer as a later day Machiavellian prince is clear in the "rotten kid" theorem devised by Gary Becker (1974). Becker identifies conditions under which an altruistic household head (the mechanism designer) can impose a rule governing his transfers of income within the family that will induce all family members to act as if they cared only about the total well being of the family as a whole.

Sufficient “love” by one member guarantees that all members act as if they loved other members as much as themselves. As it were, the amount of “love” required in a family is economized: sufficient “love” by one member leads all other members by “an invisible hand” to act as if they too loved everyone. Becker (1974):1080

This is exactly what *The Economist* had in mind: providing an analogue for the invisible hand where Adam Smith’s variant does not work. Becker concludes “Armed with this theorem, I do not need to dwell on the preferences of the nonheads” not even, apparently, the one after whom the theorem is named. The theorem works, Becker explains, because the altruistic family head can arrange things so that every family member will “internalize fully all within-family ‘externalities’ regardless of how selfish ...these members are.”

Exporting Becker’s rotten kid theorem to the world of public policy, well designed publically provided incentives seemingly could dispense with virtue entirely as a foundation of good government (at least among the “nonheads.”) As important, it spared the policy maker or constitution writer the liberal embarrassment of seeking to foster some values – a concern for the environment or for future generations, or even for one’s own family, for example – over others.

But just as proof of the “invisible hand theorem” demonstrated just how implausible were the axioms required to advocate a laissez faire policy on efficiency grounds, modern mechanism design has clarified the limits of even the cleverest ways of giving the invisible hand the helping hand of public policy. It turned out that the rotten kid theorem, for example, relied on highly unrealistic assumptions, limiting its application to a set of rather special cases, not to the general problem faced by family heads, princes and mechanism designers alike (Bergstrom (1989).)

To see why a constitution for knaves could not work, let me set aside the objection that the constitution writer had better not be a knave, and that life would be unpleasant surrounded by knaves (although perhaps less so if one were one.) The basic idea in mechanism design is to align individual incentives with the objective of efficiency, defined over a large group of people. This may be done, as we saw in Chapter 2, if a way can be found to induce each person to act as if he or she internalizes all of the benefits and costs resulting from his or her actions that are conferred or imposed on others. This will be the case of course for a solitary individual like Robison Crusoe on

his island; he alone “owns” the results of the work he has done, the risks he has taken, and the knowledge he has learned. Somehow getting people to act as if they were Robinson Crusoe is the name of the game.

Just how difficult this is to do for a large group of individuals with unrestricted preferences will be clear from the problem of team production that arises in any process where individuals together contribute to a single output. Think about a group of software engineers working together to write code for new applications and to eliminate bugs. To pose this as a problem of Machiavellian mechanism design, we can invoke the preference neutrality condition and suppose that each team member is interested only in their own material payoff, however determined. The civic-minded prince (as we will call the mechanism designer) can observe the output of the team, but not how hard each member has worked (respecting the privacy of information restriction).

The problem the prince faces is that if the compensation of the team members is simply an equal share of the output of the team, then a worker’s payoff to providing an amount of work effort that results in one unit of additional output will not be one unit of compensation, as it would be for Robinson Crusoe, but instead $1/n$ units, where n is the number of team members. The worker thus does not own the full benefit that his efforts have created, and so if he can choose on his own (the voluntary participation condition) he will provide less effort for the team than would be efficient. This is because he would not take into account the benefits that his effort confers on other team members. In this situation every worker could be better off if each of them worked a little harder. Of course the prince knows that they would indeed voluntarily work harder if they valued the contributions that their efforts made to the material payoffs of the other team members, but the preference neutrality and voluntary participation conditions prohibit his imposing these altruistic preferences on the team or compelling team members to act counter to their preferences.

If he must design a mechanism that works for self interested individuals, then he would like team members to act as if they were Robinson Crusoe, owning the results of their work. Each would then work up to the point that the discomfort of working a little harder is just compensated by the additional benefit of her work to the members of the entire team (or equivalently, that the marginal disutility of work is equal to the marginal product of work). But if the prince cannot require the team members to exert any particular level of work (that is private information) meeting this objective seemingly would be impossible.

But it is not. The prince need only announce that he will pay each member the entire value of the output of the team, minus a given constant sum determined by the prince. This bizarre mechanism ensures that any contribution by a member to the output of the team will be exactly compensated, giving each member the Robinson Crusoe incentives of an isolated individual who owns the fruits of his labor. The subtraction of a constant sum is necessary to allow the prince to balance his budget; otherwise he would be paying out in compensation n times the total output of the team.

Problem solved?

Not yet. To see why, introduce some real world risk and say that the team's realized production depends not only on the sum of the team members' efforts but also on some other (positive or negative) shocks to production. Suppose, realistically, that as in the case of members' effort, these other influences are not observable by the prince, so he cannot determine if the team's unexpectedly low output in some year is the result of bad luck or shirking workers. The contract offered by the prince would of course have to assure that team members received an *expected* income (averaging over "good" shocks and "bad") that at least as great as their next best alternative (the wage in some other job, unemployment insurance or such). However, given the nature of output, for teams of any significant size each member's *realized* income in any period could be a large multiple of that figure *of either sign*. This is because each member is residual claimant on the entire team's *realized* output, and shocks to total output in which would realistically dwarf any individual's average compensation in the long run. A contract under which in some periods a team member would not be paid and instead would be required to pay the team a substantial multiple of her expected salary is not likely to be attractive for any but extraordinarily wealthy members with virtually unlimited access to credit. As a result for all but very wealthy, no contract of this type would be voluntarily accepted by the team members, thus violating the voluntary participation.

This is not the end of the story, of course. Far more complicated mechanisms might address the problems that confront the prince's simple "pay each the whole output" plan. But the verdict of the now vast literature in mechanism design is that the prince's problems are endemic, not specific to the team production example that I have used as an illustration. When in 2007 the Prize Committee of the Royal Swedish Academy of Sciences honored Maskin, Hurwicz and Myerson

for their contributions to mechanism design they explained what the field had discovered. Hurwicz's paper published the year after Titmuss' book had proved the following "negative result" concerning voluntary participation (the "participation constraint" and the requirement that the outcome be self motivated namely, "incentive compatible"): In the presence of private information "no incentive compatible mechanism which satisfies the participation constraint can produce Pareto-optimal outcomes." Referring to Maskin's joint work with Laffont, and Myerson's joint work with Satterthwaite, they wrote: "In a large class of models Pareto efficiency is incompatible with voluntary participation, even if there are no public goods."

The result is surprising because it is generally thought that Pareto efficiency is not difficult to sustain even without the help of the mechanism designer, as long as there are none of the familiar impediments to efficient bargaining such as ill defined property rights or non-excludability of some aspect of the goods involved, as in the case of public goods). In this case a large number of buyers or sellers in one shot interactions knowing only their own preferences and the prices offered or asked by their trading counterparts should be able to bargain their way to an efficient allocation, eventually exhausting all potential gains from trade and hence by definition implementing a Pareto efficient outcome. .

But this is not the case, as was been shown by Chatterjee and Samuelson (1983) and Myerson and Satterthwaite (1983), and others. The problem is that when traders meet, they have no incentive to reveal their true valuations of the goods that may be exchanged. The reason is that, in any plausible model of bargaining, their stated valuations will influence the realized prices. Knowing this, the potential seller of a good will overstate the lowest price at which he will sell. The potential buyer will adopt the analogous bargaining ploy and understate the highest price at which she is willing to buy. The result is that unless the true valuations of buyers and sellers are always widely separated or one of the two traders gets to set the price unilaterally, some mutually beneficial exchanges will not be consummated. Buyers and sellers will walk away with money left on the table so Pareto efficiency will elude the bargainers.

The key result of these investigations is that it is never in the interest of either party to truthfully reveal their evaluations, even if the other party is reporting truthfully. In similar manner, the fact that individuals may benefit by misrepresenting their preferences prevents a benign social planner from eliciting the information he needs to provide incentives for the efficient

provision of a public good (Gibbard (1973), Laffont and Maskin (1979)).

Chatterjee (1982) provided a mechanism for the so called double auction case of the two exchange partners described above for which a trader's best response (the action that maximizes her payoffs) is to truthfully report her valuations and as a result all mutually beneficial trades occur. The mechanism implementing this Pareto efficient outcome requires upfront payments between the traders that depend only on the announced values, irrespective of whether a trade ensues. The size of the payments (quite intuitively) depends on the losses that each trader's misrepresentation of the true valuation would have imposed on the other had the other responded truthfully. The author describes this mechanism as “the payment to each player of the expected externality generated by his action ...” The payment is effectively a tax on false revelation that is just sufficient to make telling the truth a best response.

The problem is that if traders know their own valuations, some would do better by just refusing to make the upfront payment and withdrawing from the mechanism. It therefore works only if the parties improbably do not know in advance how much they value their own goods, or barring this odd situation, participation is involuntary (violating the voluntary participation constraint).¹¹

Figure 6.1 summarizes what I call the Liberal Trilemma. Setting aside questions of distributive justice, the three conditions of liberal constitutional design and public policy – neutrality with respect to preferences, voluntary participation, and Pareto efficiency – are not generally compatible. Let’s consider each dyad and see why the excluded condition is not feasible.

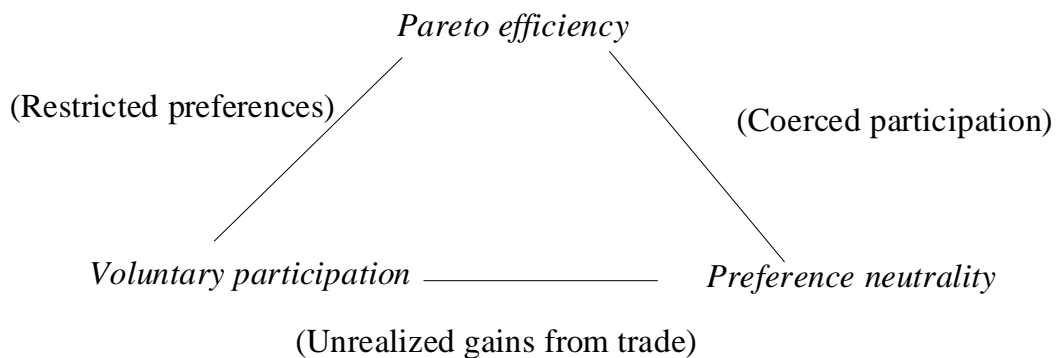


Figure 6.1. The liberal trilemma: impossibility of Pareto efficiency, preference neutrality, and voluntary participation.

If voluntary participation is combined with preference neutrality – the bottom of the triangle – potential buyers will underreport their valuation of the good in question, as we just saw, and potentially sellers will under report the least price at which they would be willing to part with their good. As a result there will be some unrealized gains from trade, violating Pareto efficiency, as Chatterjee and Samuelson and Myerson and Satterthwaite showed.

Pareto efficiency and preference neutrality are jointly feasible – the right hand edge of the triangle -- as Chatterjee and d'Aspremont and Gerard-Varet showed, if people can be compelled to participate in a mechanism even if (once they know their endowments) they would prefer not to, violating the voluntary participation condition

And voluntary participation and Pareto efficiency are jointly feasible if the traders were predisposed to report honestly their true valuations. Then, in the Chatterjee Samuelson set up individuals would never walk away from a potential mutually beneficial trade. But this of course requires a violation of preference neutrality, requiring traders to place a high value on honesty.

Thus, even if one assumed that those designing public policy are paragons of the civic virtues which the modern day Machiavellian approach would like to dispense with in the rest of the population, mechanism design will fail to produce a constitution for knaves fit for a liberal society. Among the reasons are the legal and moral constraints – against whipping the lazy, or imprisoning the debtor – combined with the fact that an individual's ability to pay fines is limited by his wealth, as well as the limited extent of intrinsically private information that may be placed in the hands of those designing incentives. Machiavelli anticipated this point in his *Discourses* when he concluded that to induce the truly bad to act as if they were good “would be either a very cruel enterprise or altogether impossible” (Machiavelli (1996):51).

The literature on mechanism design reads as one piece of bad news after the other. The negative results are important contributions, and as they accumulated many in the field beat a strategic retreat, weakening the standard of efficiency. Again the Swedish Nobel committee: “Thus in settings where participants have private information, Pareto optimality in the classical sense is in general not attainable, and we need a new standard of efficiency...” Thus mechanism design has (for the most part) abandoned the standard Pareto criterion in favor of “incentive-efficiency” which means simply the best that can be done given the way people respond

to incentives, or in other words, given existing preferences. Getting over this bar no longer requires that there be no technically feasible gains from exchange or from a reallocation by fiat that would benefit at least one party without harming anyone.

This seemingly technical watering down of the time honored objective of the field concedes that the nature of preferences indeed does matter. As we have seen, they matter in the sense that, for example, in the problem studied by Myerson and Satterthwaite, if people were committed to truthfully reporting how much they value goods that may be exchanged (and if this were common knowledge) then mutually beneficial trades that would otherwise not occur could be executed. The same would be true if instead of a predisposition to honesty, the traders placed some value on the gains that the exchange would confer on person with whom they are seeking to trade. And in the case of the hypothetical team of software developers, as we have seen, a bit of altruism among the members would have attenuated their failure to implement an efficient outcome,

If efficient use of economic resources among entirely self interested citizens could be accomplished through well designed incentives, then the fact that social preferences may be crowded out as a result might be lamentable for other reasons, but would not constitute an impediment to economic efficiency. But this as we have seen is not the case. Many of his fellow economists may now agree, if ruefully, with Tim Besley: “Perhaps then the solution can only lie in creating better people.” (Besley (2013):492)

The Legislator will take away an uncomfortable set of five facts about incentives.: Incentives are essential to a well governed society; they alone will not implement a fully efficient use of economic resources; as a result ethical and other social preferences are essential; unless designed to at least “do no harm,” incentives may stand in the way of “creating better people”; and as a result, public policy must be concerned about the nature of individual preferences and the possibility that incentives may affect them adversely.

A SECOND BEST WORLD

As a result, the Legislator has some difficult tradeoffs to consider: policies that would more effectively harness self interest to public ends may compromise the ethical and other regarding preferences on which the success of the Legislator’s constitution must also depend. And,

as we will see, the reverse is also true: policies to support the proliferation and expression of ethical and other regarding motivations will sometimes reduce the effectiveness of explicit incentives in implementing efficient outcomes.

These are just cultural variants of a venerable if paradoxical idea in economics. This is that policies to address market failures that move the economy in the “right” direction may be misguided unless they can go all the way and implement the ideal type market and property rights on which the invisible hand theorem is based. This “all or nothing” advice to the policy maker is called the general theorem of the second best (advanced by Lipsey and Lancaster (1956-1957)).

Here is the idea. In a competitive economy of the type represented by the fundamental welfare (“invisible hand”) theorems, suppose there are two violations of the assumptions under which market equilibrium results in a (Pareto-) efficient allocation of resources. Imagine, for example, the existence of a textbook monopoly that restricts sales and sets prices higher than the marginal cost of production. The same firm also contributes to environmental degradation, so that the private marginal cost of production to the owner of the firm is less than the social marginal cost (taking the costs of the environmental external dis-economy into account) Call this firm British Petroleum.

Then the second best theorem shows that the correction of one of these market failures (making the oil industry competitive, for example, if this could be done) may take the economy farther away from an efficient outcome. The competitive firms that would as a result make up the industry would produce more than the erstwhile monopoly (they would expand production until their marginal cost equaled market price, rather than restricting sales to sustain higher prices.) This would correct the standard problem of monopoly: firms sell too little so as to profit from high prices. But increased production would worsen the environmental problems. Letting the company remain a monopoly might have been a better policy. Analogously, imposing on the firm, the social costs of their environmental spillovers would lead them to restrict sales even more than in the unregulated monopoly case, thus exacerbating the monopoly problem. If adopting both anti-trust and environmental policies simultaneously (“going all the way”) is for some reason precluded then one cannot be sure that adopting either policy singly will improve things.

The intuition behind this result is that the allocational distortions caused by the violation of one of the efficiency conditions can generally be attenuated by distortions induced by other

violations. The remarkable result is that bringing the economy closer to the fulfillment of the standard efficiency conditions may result in an efficiency loss.

A similar result follows from the general non-separability of incentives and social preferences: where market failures arise because contracts are incomplete (and hence norms may be important in attenuating market failures), public policies and legal practices that more closely approximate idealized incentives associated with complete contracting may exacerbate the underlying market failure by undermining socially valuable norms such as trust or reciprocity and may result in a less efficient allocation.

We know that these norms will remain socially valuable under any conceivable set of mechanisms because there is no constitution for knaves that works. “Going all the way” is not an option. Aristotle’s Legislator thus lives in a second best world in which interventions that would be called for in an ideal world, may not only underperform, they may even make things worse.

One of these tradeoffs stems from the possible cultural downside of policies to perfect the workings of markets and to extend their role in determining the way a society uses its human and material resources. To understand these concerns we return to the idea in Chapter V that social institutions are teachers, that is, environments in which people learn new motivations and discard previously held ones. What people learn, and the process of emulation of the motivations of others that determine which motivations will proliferate depends on how people interact, that is on who meets whom to engage in what kind of interaction, with what kinds of payoffs.

Look at markets through this lens. To see how markets – at least those in economics textbooks – differ from other institutions, I will define two dimensions of institutions as learning environments: first, interactions may be either durable or ephemeral, and second, interactions may be personal or anonymous. A key characteristic of markets, first advanced by Max Weber and since stressed by advocates of an expanded role for markets, is that markets do not require either personal affection or long term relationships among people engaged in exchange. (Weber (1978):636.)

Contrast markets with what the mid 20th century sociologist Talcott Parsons called the "two principal competitors" of the market: "requisitioning through the direct application of political power" and "non-political solidarities and communities" (Parsons (1967):507). Centralized bureaucratic allocations (“requisitioning through.. political power”) are in some respects as

impersonal as markets -- at least if we compare idealized renditions of these institutions. But membership in the group defining the allocation is generally given, entry and exit costs are high (often involving a change in citizenship or at least residence, or political identification), and contacts are far from ephemeral. By contrast to both states and markets, kin-like directly-interacting communities with stable membership – often described by the terms organic solidarity, clan, generalized reciprocity, or *gemeinschaft* -- are neither ephemeral nor impersonal (Ouchi (1980);Sahlins (1974);Durkheim (1967);Tonnies (1963)). Parsons “non-political solidarities and communities” are based on long term face to face interactions among known partners.

Table 6.1 presents these three ideal types along with a fourth of perhaps limited scope -- ephemeral and personal social interaction –in which one’s racial or some other identity at birth is important. Racially segmented day-labor markets are an example, as they are personal (the racial identities of the participants matter) but the contact among participants is not on-going.

	<i>Anonymous</i>	<i>Personal</i>
<i>Ephemeral</i>	Markets	Racially segmented markets
<i>Durable</i>	Bureaucracies	Communities, clans, families

Table 6.1. Markets and other institutions as learning environments

Seen in this light, markets would not seem to be much of a school for generosity or concern for others. In their study of generosity toward the needy, Loewenstein and Small (2007) found that emotional concern is greater for “victims who share our own affective state, who are geographically or socially proximate, who are similar to us or are presented in a vivid fashion...”). It is rightly said that it is a feature of markets, not a bug, that they allow mutually beneficial interactions among individuals who share none of these aspects of closeness or vividness. But this arms-length nature of market interactions (at least the impersonal and ephemeral ideal type) attenuates the vividness or even possibility of concerns for the good or harm to others that one's actions may cause.

Of course, many markets are quite unlike the impersonal and ephemeral ideal. In the case of labor markets, for example, we have a continuum from day labor spot markets for farm workers where interactions are neither personal nor durable, to the lifetime employment practices famously associated with some Japanese firms, or the intimate relationships among those making up a small family owned firm.

To see why the difference matters, return to Adam Smith's observation that merchants are more trustworthy than ambassadors. His reasoning was that merchants are likely to interact with many people, each of whom knows how the merchant treated each of the others. Cheating one of them, Smith observed, would incur a costly loss in reputation with all of the merchant's other customers. Ambassadors, on the other hand, with more ephemeral interactions, "can gain more by a smart trick than they can lose by the injury which it does their character."

SMITH VERSUS SMITH

Smith's reasoning about the reasons for the probity of the merchant and his low opinion of ambassadors seems to be correct. I am not sure if he is right about the results, but he could be: don't forget that Costa Rican CEO's were more reciprocal than students in the trust game, and the diplomatic corps at the United Nations is no exemplar of parking probity.

Consistent with Smith's reasoning, we know from a series of experiments and natural observation that the more markets deviate from the spot market, with its idealized "flexibility" and arms length impersonality, the greater is the tendency to form loyalties to exchange partners and to reciprocate fair treatment and trust by those with whom one exchanges

The sociologist Peter Kollock (1992):341 investigated "the structural origins of trust in a system of exchange, rather than treating trust as an individual personality variable". Using an experimental design based on the exchange of goods of variable quality, he found that trust in and commitment to trading partners as well as a concern for one's own and others' reputations emerges when product quality is variable and non-contractible but not when quality is contractible. So incomplete contracts encourage trust.

Like Kollock, Brown, Falk, and Fehr (2001) used a market experiment to explore the effects of contractual incompleteness on the pattern of trading. The good exchanged varied in quality, with higher quality more costly to provide. In the complete contracting condition, the level

of quality promised by the supplier was enforced by the experimenter, while in the incomplete contracting condition by paying the associated cost the supplier could provide any level of quality (irrespective of any promise or agreement with the buyer). Buyers and sellers knew the identification numbers of those they were interacting with, so they could use information they had acquired in previous rounds as a guide to whom they would like to interact with, the prices and quality to offer, whether to switch partners, and the like. Buyers had the opportunity to make a private offer (rather than broadcasting a public offer) to the same seller in the next period, thus attempting to initiate an on-going relationship with the particular seller.

Very different patterns of trading emerged under the complete and incomplete contracting conditions. In the first, 90 percent of the trading relationships lasted less than three periods (and most of them were single-shot). By contrast, only 40 percent of the relationships were this brief under the incomplete contracting condition and most traders formed trusting relationships with their partners. For example, buyers in the incomplete contracting condition offered prices considerably in excess the supplier's cost of providing quality thus deliberately sharing the gains from trade. When Buyers were disappointed by the quality supplied, they terminated the relationship, thereby withdrawing the implied rent from the supplier. The differences were particularly pronounced in later rounds of the game, suggesting that the traders learned from their experiences, and updated their behaviors accordingly. As in the Kollock experiment people learned to trust trading partners and remain loyal to them (not switching when better deals were available elsewhere) when contracts were incomplete but not otherwise.

The experiments document a synergistic two-way relationship between incomplete contracts and social preferences such as trust. As we have seen, social preferences contribute to well functioning markets in cases when contracts are incomplete. And the experiments show that it is also that case that exchange with incomplete contracts tend to produce conditions – such as the durable and personal interactions that Smith said where the daily life of merchants but not ambassadors – under which people tend to adopt social preferences.

This represents a kind of virtuous circle: the trust that is essential to mutually beneficial exchange when contracts are incomplete appears to be learned precisely in the kinds of trading relationships that evolve when contracts are incomplete. The virtuous circle's vicious cousin also exists, of course: lack of trust among trading partners will induce traders to design as complete as

possible contracts, thereby making the evolution of trusting behavior unlikely. If the contracts could really be made complete, then the disappearance of trust would not be an impediment to trade. But from our exploration of the constitution for knaves we know that this is not generally possible.

The complementarity of trust and contractual incompleteness in this case illustrate the main message of the theory of the second best. Exchanges are durable and personal because the market participants do not abandon their current exchange partner the moment that a better deal appears, an example of what economists term “market inflexibility” and consider to be an impediment to the efficient functioning of markets. But what are the consequences of making markets more flexible, in the experiment, for example, not allowing traders to know the identity of their partners, so that long term relationships could not occur except by accident? The population could well tip from the virtuous to the vicious circle.

To see how policies to more implement impersonal and ephemeral ideal type market may degrade the economy as an environment in which ethical or other regarding preferences are leaned, return to Smith’s merchants-more-trustworthy-than-ambassadors logic. His reasoning can be understood as a repeated Prisoners’ Dilemma.¹² We know that if social interactions are sufficiently long-lasting and people sufficiently patient, then individuals with cooperative preferences (strictly: initially predisposed to cooperate and then to retaliate against those who do not cooperate in the previous round) will do better than non cooperators, as long as these so called “tit for tat” cooperators are sufficiently common. If those who do well materially in a society are more likely to be copied, either because they occupy prominent positions as cultural models, because people adapt their preferences and beliefs to reduce cognitive dissonance or for other reasons, then long lasting economic interactions with given individuals will support a society of cooperators.

Now consider a group of villagers whose livelihood depends on common pool resource like a forest or a fishery that is subject to a tragedy of the commons type of overexploitation. These could well be the Colombian forest users in the experiments of Juan Camilo Cardenas and his coauthors. Recall that they cooperated in the sustainability of their experimental “resource” especially when they could communicate with one another. Suppose that in real life the resource has been sustained over centuries because most villagers are tit for tat cooperators, resisting the

temptation to free ride on the restraint of their fellow villagers because a defection by one would provoke a defection by all.

This works because everyone in the village knows that they will be interacting with each other in the future, as will their grand children. The common ownership of the natural asset increases the expected duration of interactions among community members because those who leave the village surrender their claim on the asset. This provides exactly the conditions for effective disciplining of defectors in the management of the common pool resource.

Privatizing the asset – giving each member a marketable share in the forest – provides each with an incentive to sustain the resource and to monitor those who over-exploit it. But this will also make it easier to leave and thus enhance exit options and reduce the expected duration of interactions and the maximum severity of punishments, possibly sufficiently to make non-cooperation the more rewarding strategy. The result would be to favor the evolution of self regarding preferences.

The example is hypothetical, but what appear to be similar processes are not difficult to come by in field studies by historians, economists and anthropologists. A curious twist in the history of land rights in highland Peru provides a natural quasi experiment illustrating these effects (Braaten (2014).) In an area long dominated and managed by owners of the large haciendas, the leftist military coup in 1968 initiated a series of land reforms making farmers the *defacto* owners of the land they tilled. Because the government favored cooperative forms of organization it also initiated a time consuming titling process by which communities were given formal joint ownership of the land, intending that eventually all rural land holders' land rights would be recognized in this manner. Though pushed in varying degrees by local land reform directors, the recognition of joint land titles in the *comunidades campesinas reconocidas* (as they were termed) generated little interest on the part of the *campesinos*, for the formal legal status of the land had no practical import for the farmers at the time. There was no market in land in any case, and land rights appear to have been understood as jointly held in all communities.

When civilian rule was reestablished ten years after the coup, the joint titling process was halted in midstream. In both the 'recognized' villages and those that the military government had failed to reach before being thrown out, farmers continued their independent cultivation as *defacto* owners. In both sets of villages, too, a traditional community assembly of almost entirely male

household heads deliberated questions of governance. Among the assembly's tasks was the organization of *faenas*, communal work parties that maintained the farmers' complex collective irrigation systems, roads, public buildings and other common resources, including specifying the number of days – typically of males -- each household is required to contribute. The assemblies also disciplined those who reneged on their community labor responsibilities, their admonishments backed up by the real threat that the land operated by the free rider's could be confiscated. Men also volunteered to help neighbors in a traditional custom of reciprocal sharing of farm work called *ayni*. The joint titling appeared to have changed little; shortly after the private titling began, a field researcher noted that most farmers did not know if their community was “recognized” or not.

Things started changing dramatically in the late 1990s with the introduction of formal legal title to individual land holdings, including the right to sell titles and hence a market in land ownership and the use of land as collateral for loans. By 2011 the Special Land Titling and Cadaster Project had issued a million and a half individual titles. But because the recognized communities already held formal joint ownership, the new land laws did not apply to those holdings. By the time Ragnhild Haugli Braaten arrived in the highlands to implement her Public Goods Game experiments, the ownership status of each community was well known, and differences between the individually titled areas, which were now called “private communities,” and the joint title areas had become apparent.

Braaten wanted to know if the form of land ownership was associated with differences in the degree of cooperation among the *campesinos*. She interviewed and performed a public goods experiment with 570 people, half of them from seven jointly owned communities and half from 8 “private” communities. Other than their very different land rights histories, the two sets of communities were barely distinguishable in literacy, area per household, degree of poverty, mean income, elevation, and even degree of trust (extent of agreement with the standard “One can trust the majority of people” statement). But people in the private communities contributed fewer than half the number of *faena* days to the communal projects and substantially and significantly fewer days to the reciprocal mutual farming help *ayni* days as well.

The experimental public goods game that she implemented was immediately recognized by the *campesino* subjects as something like the *faenas*. Among men, those from the joint

ownership villages contributed over a third more to the experimental public good than those from the private communities, controlling for individual and community characteristics. (Women from the joint ownership communities were in this respect indistinguishable from those from the private communities, a finding that Braaten attributes to the fact that the communal governance institutions as well as both *faena* and *anyi* work are almost entirely male activities.) She concludes that the “recent formalization of individual land rights has ... weakened the traditional forms of cooperation.”

As with individual land titling, the development of modern labor markets in the Peruvian highlands also appears to have made the traditional contributions of communal labor the calling of chumps. Those exploiting the exit option increased their payoffs by simply ignoring what in the past had been regarded as a community norm (Mallon (1983)) Other ethnographic and historical studies –from India, medieval traders in the Mediterranean, Mexican and Brazilian shoe makers -- suggest that Braaten’s conclusion may have broad relevance.¹³

These cases, like Braaten’s study provide examples of crowding out at the societal rather than individual level. The Legislator’s dilemma is that implementing the conditions under which markets would work well in the case that contracts were complete – such as flexibility and mobility—may compromise the cultural conditions necessary for mutually beneficial exchange to take place when contracts are not complete. Even such a simple intervention as eliminating some aspects of contractual completeness – such as establishing ownership rights in a lake or a plot of land may make it less likely that people will come to have the exchange-supporting norms and other values essential to good governance. More closely approximating the ideal type market with complete contracts can therefore make things worse, not better.

Getting Smith’s invisible hand to work a little better may make it less likely that the social norms of his trustworthy merchant can fill in the gaps left where contracts remain incomplete. To figure out how to address the challenge of this second best world, Aristotle’s Legislator returns to his roots.

VII

A MANDATE FOR ARISTOTLE'S LEGISLATOR

When in 325 BCE the Athenian citizens' Assembly decided to set up a colony and naval station in the Adriatic far to the west of Greece, they took on a project of enormous proportions: thousands would undertake the mission in 289 ships (Ober (2008):124-134). And they had little time to spare as the window for safe navigation around the Peloponnesus en route would close in a matter of weeks. Neither the personnel nor the ships were at the moment under public orders; the settlers, oarsmen, navigators, or soldiers would have to be recruited from their private lives, and the ships outfitted for the mission (some would carry horses, as cavalry were involved). We know how they accomplished this because the Assembly decree was preserved.

Trierachs (ship commanders and equippers) were to be appointed from among Athens' wealthy and required to bring a ship fully outfitted and crewed to the docks at Piraeus by a given date. Those who felt unjustly burdened could appeal their assignment (called a liturgy) by challenging some (also presumably wealthy) individual to take on their liturgy, or else to exchange all of their real and personal property holdings. If the target of the challenge refused to do either, then a popular jury would determine which man's estate was the larger and should therefore bear the costs of the liturgy(Christ (1990)). This ingenious provision allowed citizens to use their own private information about their own and their neighbors' wealth to mitigate any injustices in the initial assignment of the liturgy, and thus to avoid the consequent opposition to the provisioning of the mission.

The decree continues: "The demos is to crown the first (*trierach*) to bring his ship with a crown of 500 drachmas and the second with a crown of 300 drachmas and the third with a crown of 200 drachmas," adding that "the herald of the Council [of 500] is to announce the crowns at the contest of the Thargelia [a festival], in order that the competitive zeal of the *trierachs* towards the demos may be evident." The daily wage for a skilled worker at the time was about one drachma, so these were substantial rewards, even while representing a miniscule fraction of the total costs of

executing the liturgy. Others responsible for the timely dispatch of the mission would also be honored.

Lest there be any doubt about the elevated purpose served by these incentives, the decree went to considerable lengths spelling out the benefits expected from the Adriatic naval base: “the demos may for all future time have its own commerce and transport of grain and [a] guard against the Tyrrhenians [Etruscan pirates].”

And for those unmoved by the honors and rewards that timely performance of one's duty as a citizen would garner, there was a warning: “but if anyone to whom each of these things has been commanded does not do them in accordance with this decree, whether he be a magistrate or a private individual, the man that does not do so is to be fined 10,000 drachmas” with the proceeds going to honor Athena. (The prizes for timely arrival of the ships in Piraeus would also have most likely been given by the prizewinner as an offering to Athena.)

Collectively, the Athenian *polis* was an accomplished mechanism designer; and had a time machine allowed the citizens a peak at “Machiavellianism come of age” in contemporary economics (Strauss (1988)), they would have considered risible the idea that material incentives and moral sentiments are simply additive. They are the first Aristotelian Legislators, though there is no evidence that Aristotle (who died 3 years after the Adriatic mission was mounted) was involved.

Recall the experiment in which the imposition of fines on parents arriving late to pick up their children at day care centers in Haifa resulted in a doubling of the number of tardy pickups (described in Chapter 1). Now imagine that the Athenians' time machine had allowed them to visit Haifa in the late 1990s and to design the day care centers' policy for dealing with late parents.

The sign that they posted on the door of the centers would not have read “Since some parents have been coming late we (with the approval of the Authority for Private Day Care Centers in Israel) have decided to impose a fine on parents who come late to pick up their children. As of next Sunday a fine of NIS 10 [about \$3 at the time, in New Israeli Shekels] will be charged every time a child is collected after 16.10.”

Instead the sign would have announced: “The Council of Parents wish to thank everyone for arriving on time to pick up your children, as this reduces the anxiety that the children sometimes feel and allows our staff to leave in a timely manner to be with their own families. We

will recognize all parents who have a perfect record unblemished by lateness for the next 3 months with an award of 500 NIS, to be given at our annual parents and staff Holiday party, with an option to contribute your award to the school's Teacher of the Year celebration.”

But that would not have been all: “Those who arrive more than 10 minutes late, however, will pay a fine of 1000 NIS, with the payment of the fine publically transmitted also at the Holiday party. In the unlikely event that the occasion for such a fine arises, the payment will also support the Teacher of the Year celebration.” And the message would have closed with: “Of course sometimes it is impossible for reasons beyond your control to arrive on time; and should this occur, you may explain the circumstances before a committee of parents and staff, and in case of unavoidable lateness or if the fine would cause extreme hardship, the lateness will be publically reported but no fine will be imposed.”

In fairness to the Haifa children's centers, their cryptic sign informing the parents of the fines that I just cited was not their idea of good public policy; it was part of an experimental design to avoid framing the lateness problem as a moral issue. The experimenters wanted to know the effect of the incentive per se, and wished to avoid confounding the effect of the fine by adding the effect of the moral message. The experimental design, in other words, was based on the idea that the parent’s response to the fines did not depend on how the new policy was explained.

I wonder what would have happened if on the doors of all of the day care centers had been posted messages like the one that the time-traveling Athenians proposed, explaining the ethical problem of lateness, except that, as in the experiment, in the control centers, no fine had been announced. Would the presence of the fine for lateness in conjunction with the moral message at the treatment day care centers have enhanced the salience of the moral message, crowding in social preferences and resulting in a greater reduction in tardiness than occurred at the control centers whose parents read only the moral message? Would this Athenian version of the experiment have reversed the crowding out that apparently occurred in the absence of moral framing?

It might have.

GETTING AND BECOMING

When people engage in trade, produce goods and services, save and invest, they are not only attempting to *get* things, they are also trying to *be* someone, both in their own eyes and in the

eyes of others (Cooley (1902), Yeung and Martin (2011)). This commonplace idea among psychologists and sociologists had been missed by most economists until George Akerlof and Rachel Kranton alerted the discipline to the possibility in their *Identity Economics* (2010). (Akerlof and Kranton (2010)). Sometimes the constitutive and the acquisitive motives are closely aligned, as with Adam Smith's merchant who (one imagines) would like to act in ways that constitute him (in his own eyes) as an honest man, the reputation for which (in the eyes of others) over time will also underwrite profitable exchanges. But constitutive motives cannot be reduced to self interest with a long time horizon.

The acquisitive and the constitutive reasons for actions may sometimes clash. Crowding out is sometimes a consequence of the fact that we act as a way of being and not simply as a means for getting; monetary inducements addressed to our acquisitive desires may dampen or impede the pursuit of our constitutive aspirations. Among the reasons, we have seen, are that in addition to affecting the costs and benefits of an action, incentives also provide information about the person imposing the incentive, suggest appropriate behavior by framing decision situations, and may compromise the target's sense of autonomy. Responding to an incentive in the manner intended (that is, as a payoff maximizer) may compromise an individual's constitutive objectives and contribute to all three of these reasons for crowding out. The same reasoning, we will see, suggests ways of making incentives and social preferences synergistic.

Constitutive objectives may override acquisitive motives prompting an adverse reaction to an incentive if a payoff maximizing response would make the individual a chump or a victim, with crowding out the result. An adverse response to an incentive may also arise because of the political nature of the incentives which are often transparently an attempt to control the target (Grant (2012)). It is no less true, of course, that responding in a self interested way to an incentive may constitute the actor a good citizen or an intelligent shopper, so that the constitutive and acquisitive motives are closely aligned, as with Smith's merchant. But often acquisitive ends compete with the constitutive motives that J.S. Mill proposed that economists should ignore, sometimes overriding them. Which of these it is may explain why incentives sometimes work exactly as economists predict on the basis of unmitigated self interest (or even if social preferences are present, assuming that they are simply additive with incentives.)

This may explain some of the experimental evidence for crowding out. Recall that in the

Trust Game implemented by Fehr and Rockenbach (2003) the investor could announce that he would fine the trustee if the trustee's back transfer was not sufficient, and that threatening the use of fines reduced the level of reciprocity of the trustee. Conditional on the amount of the investor's transfer to the trustee, back-transfers were lower under the threat of the fine (Figure 4.1).

On closer scrutiny, however, the threatened fine *per se* seems not to have been the main problem. Looking at the data to determine who it was among the trustees that responded negatively to the incentive, it is clear that this example of strong crowding out occurred almost exclusively as a reaction to the seemingly unfair intent of the investor. Crowding out occurred among trustees who were subjected to a demanded back transfer, which if acceded to would have transferred most of the joint surplus (total payoffs of the two) to the investor. There was no backlash against investors who announced modest levels of desired returns, demands such that the investor and the trustee would both substantially share in the joint surplus. In these cases, the use of the fines reduced back-transfers by an insignificant amount.

The use of the fine in these conditions signaled the unfair intent of the investor, not simply his distrust of the trustee. The key to the difference in explaining the effects of threat of the fine was its message. Where the stipulated back-transfer would have captured most of the surplus, the threatened fine conveyed greed. Where it would have split the surplus more equally, the fine conveyed a commitment to fairness, and perhaps a desire by the investor not to be exploited by the trustee. The use of the fine in conjunction with a seemingly unfair back transfer demand provided an acquisitive motive to reciprocate but it also may have transformed the meaning to the trustee of conceding a substantial back transfer. A large back transfer no longer constituted her as a cooperative and ethical person; it may have appeared instead to be the behavior of a victim or a sucker.¹⁴

The suspicion that it was the relationship between the investor and the trustee, not the threatened fine *per se* that was the source of strong crowding out in the Fehr and Rockenbach trust game is reinforced by a diametrically opposite reaction to fines in a Public Goods with Punishment experiment, similar to the cross cultural study surveyed in Chapter V. Fines imposed by peers at a cost to themselves who had nothing to gain personally appear to have crowded in social preferences.

Recall that in this game fellow group members have the opportunity to pay (reduce their

own payoff) in inflict a punishment (reduce the payoffs of) others in their group once each member's contributions to the public good are revealed. In some variants of this experiment group membership is shuffled after each period so that in subsequent periods a punisher it is very unlikely to be in the same group with the target of his punishment. In this so called “stranger” treatment, the punisher cannot benefit should the target respond by subsequently contributing more. Both the punisher and target know this, and importantly, the target knows that the punisher knows this. Punishment in this case is an altruistic act as it benefits others at the expense of the punisher, and hence it cannot be interpreted as a signal of the punisher’s intent to garner a larger slice of the pie.

Figure 7.2 shows the period by period contributions in a stranger treatment of the Public Goods game implemented by Fehr and Gaechter. The first ten periods were the standard game with no punishment option, and in the subsequent periods peer punishment was allowed. When punishment was not an option, average contributions decline to virtually nothing at the end of ten periods. But as is in the “English speaking,” “Protestant Europe,” and “Confucian” populations in the cross cultural study, Figure 7.2 also shows that when the punishment option is introduced, contributions start at around the level of the first period of the no punishment treatment, but then steadily rise. (The results not different if the game with the punishment option is played in the first 10 periods, and the standard Public Goods Game is played in the second half of the experiment.)

Why is the fine by the principal (investor) in the trust game counterproductive while the same incentive imposed by peers in the Public Goods Game so effective. A plausible explanation is that, when punished by a peer who had nothing to gain by doing so, those who have contributed less than others interpreted the fine as a signal of public-spirited social disapproval by fellow group members.. As a result targeted free riders feel shame, which they redress by subsequently contributing more. The positive effect may be even greater for free riders who escaped punishment because they got the moral message without being angered by being punished. If this is the case the incentive (prospect of peer imposed fines) has crowded in social preferences.

But is this “crowding in” explanation true?

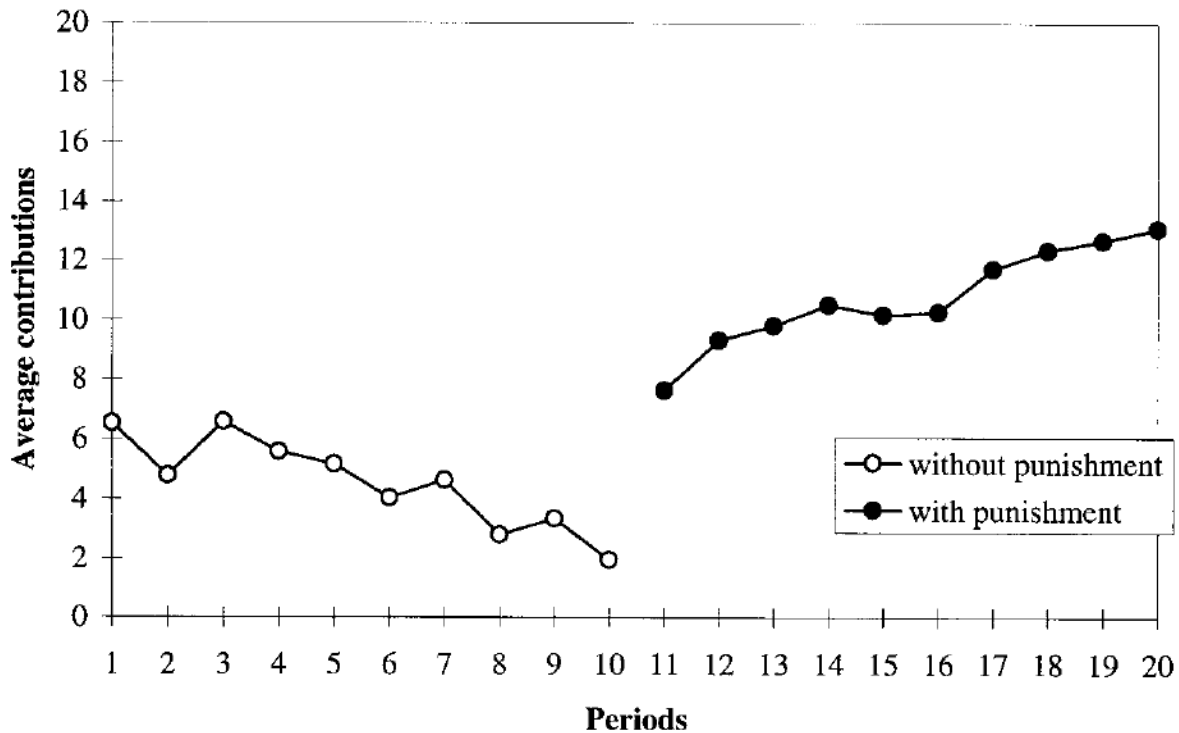


Figure 7.1 Contributions without and with peer punishment in the public goods game.
 Source Fehr and Gaechter (2000a)

The fact the contributions were higher on average in presence of fines could have occurred because when fines may be imposed, free riding no longer maximized a subject's expected material gain. Taking account of the likelihood that low contributions provoke punishment, self interest may have dictated contributing more. If so, the evidence would not demonstrate that a fine crowded in social preference. The increase in contributions would have been evidence that fines sometimes substitute for social preferences.

My co authors and I were able to explore this possibility. In our Public Goods with Punishment experiment, we found that the responses of the targeted free riders cannot be explained by self interest. The reason is that taking account of the likelihood of being punished that could have been expected by free riders in the experiment given the actual punishment behavior by their peers; contributing more did not increase one's expected payoffs (Carpenter, Bowles, Gintis et al. (2009)). The extent of punishment was not sufficient to induce an expected payoff maximizer to contribute more. The experience of being punished, nonetheless, strongly affected free riders'

subsequent behavior. Notice that for this to have worked, the erstwhile free rider had to have had some moral qualms about his self-interested behavior, which when publically recognized resulted in feelings of shame.

More direct evidence that incentives imposed by peers activate shame or other social preferences is provided by the fact that in other experiments purely verbal messages of disapproval have a substantial positive effect on free riders' subsequent contributions (Barr (2001), Masclet, Noussair, Tucker, et al. (2003)). Abigail Barr found that among the Tanzanian rural people who played her Public Goods Game, free riders often received verbal rebukes for their stinginess from other subjects in the experiment ('Now I know why I never get offered food when I drop by your house!'). The targets contributed more in subsequent rounds of the game. Others who had escaped rebuke even though they had contributed as little as the targets of the rebukes responded even more positively.

Like the fines imposed by seemingly civic minded peers in the more standard Public Goods with Punishment game, rebukes from neighbors may have raised the salience of the constitutive motives of the Tanzanian subjects making it be more important to affirm one's generous nature (in the eyes of ones neighbors and oneself alike) rather than to take home a few thousand more Tanzanian shillings at the end of the game.

A variant of the "control aversion" experiment of Falk and Kosfeld described in section chapter IV is also consistent with the interpretation that crowding out does not follow from the use of incentives *per se*, but rather from the relationship between the target of the incentive and the designer of the incentive. When agents themselves (rather than a principal) implemented controls on their fellow members, the negative control averse response did not occur.¹⁵ (Schnedler and Vadovic (2011) Apparently the imposition of controls by peers, especially if it is the result of a deliberative process legitimated by voting, pose no threat to people's desire to constitute themselves as autonomous individuals.

Can the getting and becoming framework provide the Legislator with a guide to legislation and policy?

MORAL LESSONS: ARE INCENTIVES TO BLAME?

As an early example of Machiavellian mechanism design, at the outset I cited Jeremy

Bentham's "Duty and Interest junction principle." It is a guide to how proper incentives should harness self-interested objectives for public ends. But, like Hirschman, Bentham also understood the constitutive side of action and the need to design incentives that, unlike the Haifa day care centers fines, are complements of the moral sentiments rather than their substitutes:

A punishment may be said to be ...a moral lesson, when by reason of the ignominy it stamps upon the offence, it is calculated to inspire the public with sentiments of aversion towards those pernicious habits and dispositions with which the offence appears to be connected; and thereby to inculcate the opposite beneficial habits and dispositions (Bentham (1970 [1789]) : p.26.)

The idea of making punishments "a moral lesson" has a long and checkered history. Public executions and whippings, pillories, ducking stools (a predecessor to water-boarding), and other forms of public humiliation, branding and even burning all have been advocated on these grounds. In the U.S. punishments designed to shame the convicted are again practiced, particularly for commercial and sex offenses (embezzlement and prostitution) and some minor other crimes (shop lifting). Included are requirements to publically display signs advertising the transgression: a former stockbroker advertising his embezzlement conviction "I am a convicted thief" or, "I got caught possessing cocaine" or "Convicted: DWI"(Garvey (1998). The National Health Service in the U.K announced that it would list the names (labeled with a red flag) of general practitioners who repeatedly missed early signs of cancer in their patients; green ratings would be given to doctors with a record of timely recommendations that their patients see a specialist (Times (2014).)) In Kansas City a community access TV channel ran a popular show called "John TV" that displayed the names and photographs of men convicted of soliciting a prostitute.

Aside from concerns that these punishments may compromise the dignity of the punished individual, they seem poorly designed for the educative role that Bentham advocated. The targets of the public aversion that Bentham sought to inspire were, he made quite clear, the habits and dispositions accounting for the transgressions, not the transgressors themselves.

Other unconventional forms of punishment are more attuned to this end: at a compulsory "School for Johns" in San Francisco, women explain the hardships of their former livelihoods as prostitutes; a burglar was required by a Memphis judge to permit his victim to enter the burglar's home unannounced (accompanied by a police officer) and remove any item of a value equivalent

to the stolen object. In both cases the judges sought to make the offense more personal and more vivid in the eyes of the convicted, which, following the reasoning of Loewenstein and Small (2007), would correct the moral disengagement apparently associated with their transgression.

But these eye-catching punishments are a distraction from the more workaday world of the Legislator, where as we have seen, more ordinary incentives and constraints often provide a normative frame that sometimes adversely affects how people respond. But this is far from always the case.

From an economic point of view, the small tax on plastic grocery bags enacted in Ireland in 2002 might appear similar to the fine for lateness at the Haifa day care centers: it slightly raised the cost of the action that the incentive sought to deter. But effect of the tax could not have been more different: in just two weeks following its introduction the use of the plastic bags dropped by 94 percent (Rosenthal (2008)). The tax may have crowded in social preferences: carrying a plastic grocery bag home appeared to have joined wearing a (genuine) fur coat in the closet of anti-social practices.

Like the contrast between the effects of the use of fines by principals in the trust game and by peers in the in the public goods game, the difference between the lateness fine and the plastic bag tax is puzzling but instructive. In the Haifa case, in the announcement of the fine provided no justification for the punishment. There was no “moral lesson.” In the absence of an explicit normative frame, the price on lateness provided a default frame: lateness was something for sale. Moreover in the eyes of other parents, an occasional lapse could well have occurred for reasons beyond the tardy parent’s control, rather than as the result of a deliberate disregard for the inconvenience it caused the teachers. The announcement of the fine itself alerted parents to the fact that ‘some parents’ were arriving late, perhaps making tardiness a less serious transgression of what may have been seen as a general social norm of punctuality. Finally, lateness was not so common as to be widely noticed by other parents, and at least among parents, the only observers of one’s tardy arrival at the day care were themselves also late.

By contrast, the introduction of the Irish plastic bag tax was preceded by an extended public deliberation followed by a substantial publicity campaign dramatizing the bags’ role in blighting the environment. Moreover, unlike lateness at the day care center, the result perhaps of the parent’s omission in having failed to take the proper precautions, the use of the bags required a

deliberate act of commission by the shopper, made in a highly public condition. In the Irish case the monetary incentive was introduced jointly with a message of explicit social obligation, and it apparently served as a reminder of the larger social costs of the use and disposition of the bags. In Haifa the fine seems to have said: “Lateness is OK as long as you pay for it” while in Ireland the message was explicitly framed as something like “Don’t trash the Emerald Isle!”

To sharpen the difference between the two cases, recall that I wondered what would have happened if the time-traveling Athenians instead of our modern day behavioral economists had designed the day care policy towards tardy parents. Figure 7.1 contrasts actual and the hypothetical experiments, with my guess about how the Athenian experiment might have worked out. I call the design of the experiment that was actually implemented Machiavellian because it is based on the idea that the effect of an incentive can be tested independently of the message with which it is packaged. This is equivalent to assuming separability of the moral and material motives.

My hunch is that the Athenian experiment, like the Irish bag tax, would have yielded quite different results from what actually took place in Haifa. Comparing across the “Athenian’s message” row, while the moral message alone would have had some positive effect (“better” than “status quo”), its impact would have been enhanced by the fine (“best” much better than “worse”). And comparing down the “Fine” column, the message would have altered the effect of the fine from negative to positive (from “worse” than “status quo” without the message to “best” superior to “better” with the message.) In other words, in the absence of the moral message, the fine crowded out social preferences, and with the moral message, instead, the parents’ ethical concerns were crowded in. In this case the fine and the moral message are complements, rather than substitutes.

Message	No fine	Fine	<i>Effect tested by the row</i>
No moral message	Status quo	Worse	<i>Fines without message</i>
Athenian’s message	Better	Best	<i>Fines with moral message</i>
<i>Effect tested by the column</i>	<i>Message without fine</i>	<i>Message with fine</i>	

Table 7.1 Lateness at Haifa: the (actual) Machiavellian and (hypothetical) Athenian experiments compared.

The hypothetical entries in the “Athenian’s message” row of the table are not entirely fanciful. A similar crowding in process was apparently at work in a public goods experiment by Roberto Galbiati and Pietro Vertova.¹⁶ Consistent with the crowding resulting from the presence (but not extent) of a fine in the Cardenas experiment described in Chapter III, Galbiati and Vertova found that the effect of a stated (non-binding) obligation to contribute a certain amount to the public good was greater when it was combined with a weak monetary incentive than when no incentives were offered. But a stronger monetary incentive did not result in an increase in contributions. The strong monetary incentive also had no effect on behavior in the absence of the stated obligation. The authors’ interpretation is that the explicit incentives enhanced the salience of the stated obligation.

Bentham’s advice that punishments must send moral messages does not recommend abandoning Hume’s objective of harnessing self-regarding preferences to public ends: self interest is a powerful and ubiquitous motive. What Bentham’s view of punishments as both incentives and moral signals instead suggests is that legislators and other policy makers may sometimes find ways to turn the separability problem on its head, making incentives and morals complements rather than substitutes.

A MANDATE FOR ARISTOTLE’S LEGISLATOR

Can the Legislator take away any concrete policy advice from what he has learned? I think he can.

Table 7.2 brings together the main policy implications of the fact that the material inducements and constraints stressed by the Machiavellian paradigm are messages as well as incentives, that the moral sentiments are an essential foundation of good government, and that both crowding out and crowding in may take place, depending in part on the policies adopted.

The Legislator should not expect everyone to applaud his taking up this advice. Though hard evidence on their extent is difficult to come by, incentives that crowd out social preferences are probably fairly common. It would be surprising if those implementing these incentives are entirely unaware of the non-separability of social preferences and incentives and that as a result the use of incentives sometimes actually reduces the joint surplus available in an economic interaction. It seems that incentives with this dubious feature are adopted by well informed private

and public sector managers operating in highly competitive environments.

Why do we ever observe pie-shrinking incentives in practice? The pie metaphor gives away the answer. Even if incentives reduce the total expected surplus associated with an economic interaction between a principal and agent, such as taking out a loan or employing a worker, the use of incentives may give the principal a sufficiently larger slice of the smaller pie to justify their use.

This is what occurred in an experiment by Fehr and Gaechter (reported in Fehr and Falk (2002)) with Swiss students. The experiment, similar to the Fehr and Rockenbach trust game, was constructed so that had subjects responded as if they had entirely self-regarding preferences, the joint surplus (sum of payoffs of employer and employee) would have been more than twice as great under the incentive treatment as under the trust treatment (where no fines were allowed)..

But the negative synergy between the incentive and social preferences was so strong that the total surplus was much higher in the trust treatment than when incentives were introduced. This was true even in those cases where principals offered exactly the kind of contract that a mechanism designer would recommend. But under these "optimal" contracts, profits were more than double what they were in the trust treatment, while the payoffs to employees were less than half. The incentive treatment allowed employers to save enough in wage costs to offset the reductions in work effort and hence the shrunken surplus.

Thus one of the reasons why constitutive aspirations sometimes lead agents to respond negatively to incentives – namely, that responding to the incentive in the way intended will benefit the principal disproportionately -- also explains why just these incentives may nonetheless be used by profit-maximizing principals, even when their use results in a smaller pie. Simply informing the principals that the result is a smaller pie will not be effective. If a mutually acceptable division of the pie could be decided in advance (and enforced ex-post) this problem would not arise, because incentives could then be devoted exclusively to increasing the size of the pie rather than being hijacked to enlarge one of the slices at a cost to the others. Prior agreement on the division of the pie is often possible in partnerships and cooperative firms, and it was (and still is) common among hunter gatherer groups that target large game.

<i>Advice to the Legislator</i>	<i>Examples, Evidence</i>
When crowding out occurs (unless it is “strong”) greater use of incentives may be warranted than in the absence of crowding out, or abandoning incentives in favor of some other policy (e.g. appeals to social preferences) may be called for.	Chapter III
Seeking to perfect the conditions for markets to work well in the absence of social preferences (e.g. by making contracts more complete and replacing “missing markets” with well defined property rights) cannot succeed entirely and may crowd out social preferences and degrade economic performance.	The constitution for knaves, evolution of trust and cooperation where contracts are incomplete, Peruvian highlands, Public Goods Games and Trust Games pre and post incentives
Protect civic minded citizens from exploitation by the self interested; minimize the worst case outcomes that may occur for trusting or generous citizens (rule of law, mobility, insurance).	Third party punishment of shirkers in a public goods game may liberate people to act on their generous motivations
Avoid moral disengagement: If the policy has a public purpose that would be endorsed by the citizen make sure that the moral message is clear.	Athenian Assembly’s mobilization of resources for the new colony; Irish plastic bag tax, lateness fines in Haifa
Avoid “bad news:” Incentives designed to control or take unfair advantage of the target may not work; ensure fairness in the implementation of incentives	Contrast of trust game with fines with public goods games with peer punishment; Athenian Assembly
Avoid control aversion: Incentives and constraints implemented by peers after deliberation may avoid negative reactions.	Irish plastic bag tax, experimental games with deliberation on punishments or controls
In the presence of categorical crowding out, avoid small fines	Chapter III
Provide opportunity for people to express their pro-sociality both in their own actions and in peer punishment or comment on norm transgressions, especially when these not only mobilize shame but also educate.	Public goods games with punishment (including verbal)
Where appeals to social preferences and incentives are substitutes specialize in one or the other; where they are complements, use both.	Public goods game with and without “obligation” and fines. Athenian assembly.
Where there are no social preferences to crowd out adopt the standard economic approach to incentives; but the fact that incentives work as predicted for citizens without social preferences does not mean that they are absent or that crowding out has not occurred.	Without incentives experimental subjects act generously in a public goods game, with incentives they behave as the standard model of incentives predicts.

Table 7.2. Advice for the Aristotelian Legislator

Another challenge for the Legislator that I have not yet addressed arises because populations are made up of individuals with differing mixtures of self-interest and the many forms of social preferences. Consider the simple if unrealistic case in which there are just two types, the entirely self interested and those with a degree of altruism toward others, The Legislator would like to encourage contributions to a public good in this population. A monetary incentive (subsidy for contributing or fine for free riding, for example) will encourage contributions from the self-interested, but if incentives and social preferences are substitutes, it may work less well or even backfire among the altruists, as we have seen. The altruists, however, might respond well to hearing about the substantial benefits that the public good will confer on others. But this moral message would be wasted on the self interested.

An obvious strategy for the Legislator is to separate the two populations, addressing each with the appropriate policy. But this will be difficult because the individuals' types (altruistic or not) are not readily known to the Legislator, nor are people likely to voluntarily sort themselves into the two obvious boxes that he would like to impose. Where self selection into the relevant categories can be induced, however, separation strategies can sometimes succeed at least approximately. This occurs, for example, when the modest salaries at environmental and other public service non-governmental organizations deter those unmoved by the mission of the organization to apply, while the prospect of association with the mission itself provides enough additional compensation to attract the committed (Besley and Ghatak (2005)).

The mixed nature of populations poses further challenges once one considers the diversity of motives that are commonly grouped under the heading "social preferences." For the Legislator, the complication of designing policies for a population with multiple seemingly pro-social motivations is worse than an embarrassment of riches. He will quickly find that not all good things go together. A perverse alchemy can convert good motives into unwanted social outcomes.

In an experimental Public Goods Game my co authors and I observed (as expected) that high levels of cooperation were sustained by peer punishment. But when we identified the types of the players – "altruistic" or "reciprocal" -- we found that while the altruists contributed more to the public good, they were less likely to uphold the social norm of contributing by inflicting peer punishment on free riders. Instead, the altruists, were themselves free riding on others who

willingly paid the cost to punish low contributors. (Carpenter, Bowles, et al. (2009)). Reciprocal types were more likely to punish low contributors.

Stimulated by this result Sung-Ha Hwang and I titled a paper “Is altruism bad for cooperation?” We found that under quite plausible conditions, it certainly could be (Hwang and Bowles (2014b).) The Legislator could increase public goods contributions by if he could only find a way (without incurring the wrath of the population) to segregate the self interested types along with the reciprocators into one sub population, letting the altruists have their own group, each group with its own dedicated policies.

But do not think that the idea of separating populations according to their motivations and designing incentives and punishments accordingly is entirely whimsical. The legal scholar Lynn Stout has proposed that we think along similar lines about the appropriate compensation for injuries due to faulty products, for example, in U.S. law:

... the traditional tort pattern of under compensation of victims does not necessarily pose a problem when we are dealing with humans, the vast majority of whom ... have an ‘internal’ incentive, in the form of conscience, to take modest care to avoid harming others ... For natural persons, partial liability may be enough, when added to internal sanction of conscience to motivate most to take care of harming others. But the same under compensation pattern may produce too little deterrence when applied to corporations. ...we may want corporate defendants to pay victims more in damages than human defendants must pay.”Stout (2011):171-2

Stout’s argument, of course, is not that the humans who make corporate decisions are less moral than others (remember the Costa Rican CEOs). It is, instead that when deciding on an appropriate level of care re towards others (in product design for example), the responsibility of managers to maximize profits on behalf of shareholders should induce them to take into account the expectation of under compensation of damages, should these occur. Stout is merely repeating the argument of Milton Friedman in his “The Social Responsibility of Business is to Increase its Profits” when he wrote that “Only people can have responsibilities.”(Friedman (1970))

She might have added, on the basis of hard experimental evidence, that the diffused responsibility of the decision making process and competitive pressure for firm survival faced by managers would work in the same direction (Schotter, Weiss and Zapater (1996), Falk and Szech (2013) .) Stout points out that antitrust law is applied almost entirely to business organizations, and

by contrast to under compensation for damages typical of tort law, it provides that “treble damages” be paid meaning three times the harm caused by the firm’s anti-competitive actions.

The final challenge for the Legislator follows from the just mentioned heterogeneity of populations, along with the observation that the outcome of the Legislator’s policy interventions will almost never be a simple average of a behavior characteristic of each type. Instead the outcome will depend on the composition of the population and the rules governing social interactions that determine how the individual’s actions add up to aggregate outcomes.

This whole-is-not-the-sum-of-the-parts idea is not new. It has been around economics since Adam Smith explained how by the alchemy of the invisible hand the self interest the brewer, the baker and the butcher would put everyone’s dinner on the table.

The problem confronting the Legislator is that under some rules of the game, the actions of the self interested induce even the civic minded to act as if all they care about is their own gains. His opportunity is that under different rules, the reverse occurs. We have already seen this in the Public Goods with Punishment experiment (Figure 7.1) In the absence of the peer punishment opportunity, even those predisposed to contribute substantial amounts eventually came to act as if they were self interested. But once peer punishment of free riders was allowed, the very same population converged to a substantial contribution level, one that in the later periods of the game was sustained with very little actual punishment because egregious free riding had all but disappeared. By some combination of the incentive to avoid the punishment and the shame of having experienced it in previous periods, the self interested were now acting “as if they were good.”

The Legislator’s aim in situations like this is to design rules—like the Public Goods with Punishment game -- so that the civic minded, not the self interested determine the outcome. To see the rudiments of what this requires, imagine that there are just two citizens who will interact once in a symmetrical Prisoners’ Dilemma game (symmetry means that the payoff matrix for the game is identical for the two). One of them is known (by the Legislator and by the other player) to wish simply to maximize her payoffs in the game, while the other has reciprocal preferences (also known to all). He would prefer to cooperate, but only if the other were to do so as well. Is there a way to play this game so that the efficient and equal sharing outcome – both cooperate – will occur?

If the conventional game is played, the two make their choices simultaneously, and both will defect (the reciprocator knows that the other will defect, and as a result, reciprocating, so does he.) But the Legislator could change the rules so that one player goes first. Which should it be? Let it be the selfish. She knows that because the reciprocator will move second, and out of reciprocity will mimic whatever she does, there are now only two possible outcomes of the game, in which either both cooperate or both defect. By cooperating, she can bring about the former outcome in which both do better, which she then does, and the other reciprocates.

Have we come full circle? Is this, in the end, what Machiavelli had in mind? It is not, because essential to the happy outcome in the sequential Prisoner's Dilemma, and in the public goods experiment too, is the presence of ethical or other regarding citizens, who under the proper rules induce the "wicked" to act as if they were not. That is why, in my subtitle, *Good laws are no substitute for good citizens*.

The Legislator will note, too, that a constitution that induces the self interested to act in cooperative and socially beneficial ways will affect the process of cultural evolution itself, inducing larger numbers of the next generation to adopt preferences that provide a direct motive for acting as most citizens do. The possibility that the right incentives, laws and other aspects of the rules of the game might crowd in social preferences rather than the opposite suggests that the Legislator might hope to do better than limiting himself to Rousseau's seemingly prudent injunction to take "people as they are and laws as they might be" (Rousseau (1762)). If the Legislator needs a bumper sticker, he might borrow from the subtitle that Stout chose for her book on law and morality: *Good laws make good people* (Stout (2011))

LAWS AS THEY MIGHT BE FOR CITIZENS AS THEY MIGHT BE

Making good people is a jarring expression, but it is hardly a novel idea. Parents try to do this, as do teachers, religious communities and others. It is hard to imagine a viable society in which activities aimed at making good people are not widely practiced. Certainly there is no ethnographic or historical record of such a society indifferent to virtue. What is novel is that making good people should be an objective of public policy. Compulsory schooling was initially and continues to be advocated as a way to teach the prevailing social norms, apparently with some

success. But the historical experience of state projects of cultural transformation is hardly encouraging.

An example is provided by the German Democratic Republic, which like many Communist Party ruled societies invested considerable energy and resources in the creation of a more solidaristic and less selfish citizen. But in a recent experiment, adults born in East Germany cheated for monetary gain twice as frequently as did those born in West Germany, and this was particularly true of those who had reached maturity prior to the fall of the Berlin Wall. (Ariely, Garcia-Rada, Hornuf et al. (2014))

Nonetheless the Legislator's amendment of Rousseau's injunction is now being taken up in surprising quarters, and in pursuit of objectives quite different from the "new socialist man." Contrast two prominent bits of advice on the appropriate role of morality in shaping a response to financial instability. In the aftermath of the stock market crash of 1987 the *New York Times* editorial board took a page (and even some phrases) from Machiavelli, and Mandeville:

Perhaps the most important idea here is the need to distinguish between motive and consequence. Derivative securities attract the greedy the way raw meat attracts piranhas. But so what? Private greed can lead to public good. The sensible goal for securities regulation is to channel selfish behavior, not thwart it. (Times (1988))

To make sure that nobody missed the point they chose a headline worthy of David Hume's maxim about knaves: "Ban Greed? No: Harness It."

As the housing bubble burst in 2008 and the financial crisis unfolded, many U.S. home owners found that their property was worth less than their mortgage obligation to the bank. Some of these so called "under water owners" did the math and strategically defaulted on their loans, giving the bank the home and "walking away." In contrast to the *New York Times* two decades earlier, the executive vice president of Freddy Mac, the Federal Home Loan Mortgage Corporation, sounded a distinctly Aristotelian plea for moral behavior in the economy

While a personal financial strategy might argue for a strategic default, entire communities and future home buyers can be harmed as a result. And that is why our broader social and policy interests will be best served by discouraging strategic defaults. (Bisenius (2010))

And rather than trusting that by getting the prices right the market would be sufficient to induce people to internalize the external effects of their actions, Freddy Mac urged that "borrowers

considering a strategic default should recognize the damaging impact their actions can have on others.”

He was hoping, in short, that morals would do the work of prices, inverting the mantra of the Machiavellian paradigm in economics.

There was no shortage of moral reasoning on the question. Large majorities of those surveyed held that strategic default is immoral (Guiso, Sapienza and Zingales (2013)) Most of the defaults were not strategic and instead were impelled by job loss or other circumstances making payment impossible. But Freddy Mac’s plea for morality among under water debtors could not have been very effective among those who accused the financial institutions of invoking a double standard, having single mindedly pursued their own pecuniary interests for decades and then imploring the home owners to act otherwise when their house of cards tumbled (White (2010).) The main economic determinants of strategic defaulting were economic: namely, how “far” underwater the property was. But moral concerns – about fairness and predatory banking for example – were prominent among people’s stated reasons for condoning strategic default (Guiso, Sapienza, et al. (2013).)

A renewed attention to virtue among policy makers is being driven in part by the nature of the challenges now facing the world’s populations. With economic inequality on the increase in the world’s major economies, one may doubt Dr. Johnson’s reassurance that “there are few ways in which a man can be more innocently employed than in getting money.” The idea of an economy of knaves now appears to be anything but harmless for other reasons, too.

The force of Arrow’s argument about the essential role of social norms in the economy is likely to increase as the wealth of nations shifts from steel, grain and other goods readily subject to contract, to intangible knowledge, caring for the young and the elderly, and the other new forms of wealth characteristic of what is called the “weightless economy.” The same conclusion follows from the fact that the many of the greatest challenges facing the world’s peoples today and in the future – epidemic spread, climate change, personal security, governing the knowledge based economy – increasingly arise from global and other large scale human interactions that cannot adequately be governed by channeling entirely self interested citizens to do the right thing by means of incentives and sanctions provided by either private contract and governmental fiat. The reason is that for none of these problems is it even remotely feasible that the essential interactions

could be governed by the complete contracts that underpin the invisible hand theorem.

But Freddy Mac style moral exhortation alone is also likely to be less effective than policies inspired by Bentham's punishments that convey a more resonant "moral lesson." Perhaps Bentham had in mind the charivaris of early modern Europe: neighbors, typically women, would surround the home of a philandering husband, a price-gouging baker, or a local dignitary exploiting his status for commercial gain, beating pots and pans to express their moral indignation (Tilly (1981)). The tradition lives on: the municipal commissioner of the Indian city of Rajahmundry (in Andhra Pradesh) hired ten drummers and directed them to beat non-stop outside the homes of tax evaders (Farooq (2005)) The drummers said nothing, but the fact that they were accompanied by tax collectors and other officials conveyed a clear enough message. The policy was highly effective, apparently by invoking the shame of the tax evaders at their transgression of a social norm.

In Bogota, Mayor Antanas Mockus recruited hundreds of mimes in white face and clown outfits and sent them into the chaotic city traffic to poke fun at jay-walkers and to lampoon drivers violating the black and white pedestrian cross-walks ("zebras"). Taxi drivers were particularly notorious for their offences.¹⁷ Citizens were invited to nominate a taxi driver for pedestrian-friendly driving, and the first 140 nominated became the founding members of the Knights of the Zebra. At the induction ceremony Mayor Mockus -- a mathematician and philosopher by training -- gave each of the Knights a plastic zebra to hang from his or her rear view mirror (yes, there was at least one Lady of the Zebra.)

The thumbs down cards he issued in the hundreds of thousands were enthusiastically flashed to violators, like referees' red cards penalizing players for fouls in soccer matches. There were also thumbs up cards to recognize acts of traffic kindness. At the same time the city adopted stricter police action to enforce traffic regulations. The year the mayor took office, the incidence of traffic related deaths exceeded the national average by a considerable amount; the year he left it had fallen to well below the national average (and both relatively and in absolutely continued falling after his departure, in a few years reaching just one third of the number of fatalities of the earlier level.)

Like the tax on plastic grocery bags in Ireland with its surrounding publicity, like drumming back taxes out of evaders in India, the mayor's highly effective campaign to tame

Bogota's traffic imbued formal enforcement and material incentives with an unmistakable moral message: driving aggressively was not cool. This too is what the Athenian Assembly had done in mounting its Adriatic mission. The objective in all four of these cases is worthy of the Aristotelian Legislator: to encourage civic action using appeals to both the material interests and the moral sentiments, framed in ways that the two work synergistically rather than at cross purposes.

There are limits, of course: not all subsidies can be awarded as prizes and not all fines can be flashed like red cards. There are more fundamental limits too. A world hoping for saints might reflect on how the constitution for knaves idea came about in the first place, and why it gained such wide acceptance. An important part of this process was the realization that saints are not the only alternative to knaves; there are also zealots.

When the economist Charles Schultze wrote that "market-like arrangements reduce the need for compassion, patriotism, brotherly love, and cultural solidarity" (Schultze (1977)) he considered this to be a feature, not a bug. One could agree with him, not only because in an economy of strangers there might not be enough love to make "the world go around" as Alice feared in her whispered response to the Duchess. One might also worry that a social order requiring, say, patriotism and cultural solidarity for its proper functioning might also provide justification for divisive and intolerant sentiments, just as the arguments in favor the invisible hand had assisted in the social acceptance of motivations that had previously been termed avarice.

The idea of a constitution for knaves long predated Smith, but his reasoning about the invisible hand and advances in economic thinking since added something important: a mechanism albeit imperfect for directing self-interest to public ends. Recent evidence that other regarding and ethical motives are common greatly enhances the space of feasible policy interventions to include, for example, the wise combination of positive incentives and punishments with a moral lesson adopted by the Athenian Assembly.

But the fact that people have social as well as selfish preferences also presents a challenge. Thanks to economists from Smith to Arrow and Debreu, we do know what a constitution for knaves looks like. And now that we understand what it takes to make it work, we find it wanting not only for its inability to implement an efficient use of resources but also, I have here suggested, for its likely cultural effects. While we need an alternative, we do not yet have any similar conception of a constitution for a population made up of knaves, saints and zealots, in which the

constitution itself will influence the proportion of these and other sentiments in a population.

Considered dynamically in this way and taking account of the dark side of social preferences, the Legislator's challenge is far more difficult than heading Hume's maxim. Some of the social preferences may be more difficult than self interest to channel into socially valued or at least harmless ends. And the positive social preferences of generosity, fair-mindedness and the other civic virtues are a fragile resource for the policy maker, one that may be either empowered by legislation and public policy, or irreversibly diminished. This suggests the following extension of Hume's maxim about knaves: Good policies and constitutions are those that support socially valued ends not only by harnessing self interest, but also by evoking, cultivating and empowering public-spirited motives.

Table A: Experimental games useful for measuring social preferences and the effects of incentives

Game	Definition of the Game	Real life Example	Predictions with self-regarding players	Experimental regularities, References	Interpretation									
Prisoners' dilemma Game	Two players, each of whom can either cooperate or defect. Payoffs are as follows: <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td></td> <td>Cooperate</td> <td>Defect</td> </tr> <tr> <td>Cooperate</td> <td>H,H</td> <td>S,T</td> </tr> <tr> <td>Defect</td> <td>T,S</td> <td>L,L</td> </tr> </table> <p style="text-align: center;">H>L, T>H, L>S</p>		Cooperate	Defect	Cooperate	H,H	S,T	Defect	T,S	L,L	Production of negative externalities (pollution, loud noise), exchange without binding contracts, status competition.	Defect	50% choose Cooperate. Communication increases frequency of cooperation Dawes (1980)**	Reciprocate expected cooperation
	Cooperate	Defect												
Cooperate	H,H	S,T												
Defect	T,S	L,L												
Public Goods Game	n players simultaneously decide about their contribution g_i . ($0 \leq g_i \leq y$) where y is players' endowment; each player i earns $\pi_i = y - g_i + mG$ where G is the sum of all contributions and $m < 1 < mn$.	Team compensation, cooperative production in simple societies, overuse of common resources (e.g., water, fishing grounds)	Each player contributes nothing, i.e. $g_i = 0$.	Players contribute 50% of y in the one-shot game. Contributions unravel over time. Majority chooses $g_i = 0$ in final period. Communication strongly increases cooperation. Individual punishment opportunities greatly increase contributions. Ledyard (1995)**.	Reciprocate expected cooperation									
Ultimatum Game	Division of a fixed sum of money S between a Proposer and a Responder. Proposer offers x. If Responder rejects x both earn zero, if x is accepted the Proposer earns $S - x$ and the Responder earns x.	Monopoly pricing of a perishable good; "11 th -hour" settlement offers before a time deadline	Offer $x = \varepsilon$; where ε is the smallest money unit. Any $x > 0$ is accepted.	Most offers are between .3 and .5S. $x < .2S$ rejected half the time. Competition among Proposers has a strong x-increasing effect; competition among Responders strongly decreases x. Güth et al (1982)*, Camerer (in press)**	Responders punish unfair offers; negative reciprocity									

Dictator Game	Like the ultimatum game but the Responder cannot reject, i.e., the “Proposer” dictates ($S-x, x$).	Charitable sharing of a windfall gain (lottery winners giving anonymously to strangers)	No sharing, i.e., $x = 0$	On average “Proposers” allocate $x=.2S$. Strong variations across experiments and across individuals Kahneman et al (1986)*, Camerer (in press)**	Pure altruism
Trust Game	Investor has endowment S and makes a transfer y between 0 and S to the Trustee. Trustee receives $3y$ and can send back any x between 0 and $3y$. Investor earns $S - y + x$. Trustee earns $3y - x$.	Sequential exchange without binding contracts (buying from sellers on Ebay)	Trustee repays nothing: $x = 0$. Investor invests nothing: $y = 0$.	On average $y = .5S$ and trustees repay slightly less than $.5S$. x is increasing in y . Berg et al (1995)*, Camerer (in press)**	Trustees show positive reciprocity.
Gift Exchange Game	“Employer” offers a wage w to the “worker” and announces a desired effort level e . If worker rejects (w, e) both earn nothing. If worker accepts, he can choose any e between 1 and 10 . Then employer earns $10e - w$ and worker earn $w - c(e)$. $c(e)$ is the effort cost which is strictly increasing in e .	Noncontractibility or nonenforceability of the performance (effort, quality of goods) of workers or sellers.	Worker chooses $e = 1$. Employer pays the minimum wage.	Effort increases with the wage w . Employers pay wages that are far above the minimum. Workers accept offers with low wages but respond with $e = 1$. In contrast to the ultimatum game competition among workers (i.e., Responders) has no impact on wage offers. Fehr et al (1993)*	Workers reciprocate generous wage offers. Employers appeal to workers’ reciprocity by offering generous wages.
Third Party Punishment Game	A and B play a dictator game. C observes how much of amount S is allocated to B. C can punish A but the punishment is also costly for C.	Social disapproval of unacceptable treatment of others (scolding neighbors).	A allocates nothing to B. C never punishes A.	Punishment of A is the higher the less A allocates to B. Fehr and Fischbacher (2001a)*	C sanctions violation of a sharing norm.

Source: Adapted from Camerer and Fehr (2004) Note: ** denotes survey papers, * denotes papers that introduced the respective games.

APPENDIX 2. TRUST AND THE LIBERAL RULE OF LAW

Here I present a model of the relationship between the liberal state and trust illustrating part of the argument at the end of Chapter 5. Consider a population composed of a large number of people who interact in pairs to engage in an exchange in which they may either behave opportunistically (e.g. steal the other’s goods) or exchange goods to their mutual benefit. Call these strategies “defect” and “cooperate,” with payoffs describing an assurance game, as in the top payoff matrix in Figure A2.. Expected payoffs for cooperators and defectors are π_C and π_D and they are both increasing in the probability (p) that one’s partner is a cooperator as shown in the right panel of the figure.

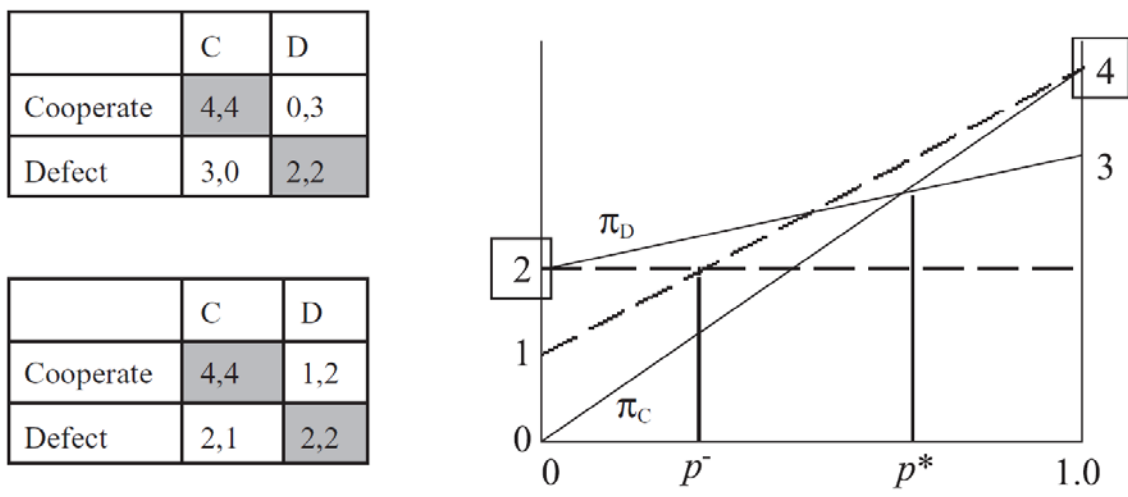


Figure A.2. The rule of law and cooperative norms. Left panel: payoffs in the exchange game (upper without, lower with rule of law); right panel expected payoffs based on the type of one’s partner (solid lines without, dashed lines with rule of law). The two Nash equilibria are mutual defect and mutual cooperate (shaded cells in the payoff table, boxed payoffs in the right panel). Because in the absence of the rule of law, the critical value, p^* , exceeds one half, defection maximizes the expected payoffs of an individual who believes that his or her partner is equally likely to cooperate or defect (this called the risk dominant strategy.) In a large randomly paired population, p^* is termed the risk factor of the cooperative equilibrium, the robustness to instability of which is measured by $1-p^*$. The rule of law (dashed lines) makes cooperating the risk dominant equilibrium, meaning the outcome in which each individual plays the risk dominant strategy

What is the smallest value of p (the probability that one's partner is a cooperator) such that the expected payoff to cooperating exceeds that to defecting? We can see from Figure 4 that one would have to believe that this is the case with a probability not less than p^* (which given the payoffs in the top matrix and in the figure the solid lines is two thirds) in order for cooperating to be the expected payoff-maximizing strategy. Where p^* is substantial and information about one's trading partner minimal, mutual defection would result, replicating the common condition in most of human history, namely that strangers represent dangers, not opportunities for mutual benefit. But if the liberal institutions that attenuate the worst case outcomes are in force (the lower payoff matrix) the cooperator whose partner defects now has a payoff of 1 rather than zero, and the defectors payoff in this case is reduced from 3 to 2. The rule of law reduces the critical value of p to \bar{p} (equal to one-third) so that a trader thinking that the partner is equally likely to be a cooperator or a defector would cooperate.

WORKS CITED

- AGHION, PHILIPPE; YANN ALGAN AND PIERRE CAHUC.** 2008. "CAN POLICY INTERACT WITH CULTURE: MINIMUM WAGE AND THE QUALITY OF LABOR RELATIONS."
- AKERLOF, GEORGE A. AND RACHEL KRANTON.** 2010. *IDENTITY ECONOMICS: HOW OUR IDENTITIES SHAPE OUR WORK, WAGES, AND WELL-BEING*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- ALESINA, A. AND PAOLA GIULIANO.** 2009. "FAMILY TIES AND POLITICAL PARTICIPATION," *IZA*. BONN:
- ANGRIST, JOSHUA AND VICTOR LAVY.** 2009. "THE EFFECTS OF HIGH STAKES HIGH SCHOOL ACHIEVEMENT REWARDS: EVIDENCE FROM A RANDOMIZED TRIAL." *AMERICAN ECONOMIC REVIEW*, 99(4), 1384-414.
- ARIELY, DAN; XIMENA GARCIA-RADA; LARS HORNUF AND HEATHER MANN.** 2014. "THE (TRUE) LEGACY OF TWO REALLY EXISTING ECONOMIC SYSTEMS," *UNIVERSITY OF MUNICH, DEPARTMENT OF ECONOMICS*.
- ARISTOTLE.** 1962. *NICOMACHEAN ETHICS*. INDIANAPOLIS: BOBBS-MERRILL.
- ARROW, KENNETH J.** 1972. "GIFTS AND EXCHANGES." *PHILOSOPHY AND PUBLIC AFFAIRS*, 1(4), 343-62.
- ARROW, KENNETH J.** 1971. "POLITICAL AND ECONOMIC EVALUATION OF SOCIAL EFFECTS AND EXTERNALITIES," M. D. INTRILIGATOR, *FRONTIERS OF QUANTITATIVE ECONOMICS*. AMSTERDAM: NORTH HOLLAND, 3-23.
- ARROW, KENNETH J. AND FRANK H. HAHN.** 1971. *GENERAL COMPETITIVE ANALYSIS*. SAN FRANCISCO: HOLDEN-DAY.
- AXELROD, ROBERT AND WILLIAM D. HAMILTON.** 1981. "THE EVOLUTION OF COOPERATION." *SCIENCE*, 211, 1390-96.
- BAR-GILL, OREN AND CHAIM FERSHTMAN.** 2004. "LAW AND PREFERENCES." *JOURNAL OF LAW, ECONOMICS AND ORGANIZATION*, 20(2), 331-53.
- BAR-GILL, OREN AND CHAIM FERSHTMAN.** 2005. "PUBLIC POLICY WITH ENDOGENOUS PREFERENCES." *JOURNAL OF PUBLIC ECONOMIC THEORY*, 7(5), 841-57.
- BARRY, HERBERT III; IRVIN L. CHILD AND MARGARET K. BACON.** 1959. "RELATION OF CHILD TRAINING TO SUBSISTENCE ECONOMY." *AMERICAN ANTHROPOLOGIST*, 61, 51-63.
- BECKER, GARY.** 1974. "A THEORY OF SOCIAL INTERACTIONS." *JOURNAL OF POLITICAL ECONOMY*, 82, 1063-93.
- BELKIN, DOUGLAS.** 2002. "BOSTON FIREFIGHTERS SICK - OR TIRED OF WORKING," *BOSTON GLOBE*. BOSTON: B1.
- BENABOU, ROLAND AND JEAN TIROLE.** 2006. "INCENTIVES AND PROSOCIAL BEHAVIOR." *AMERICAN ECONOMIC REVIEW*, 96(5), 1652-78.
- BENABOU, ROLAND AND JEAN TIROLE.** 2003. "INTRINSIC AND EXTRINSIC MOTIVATION." *REVIEW OF ECONOMIC STUDIES*, 70, 489-520.
- BENNER, ERICA.** 2009. *MACHIAVELLI'S ETHICS*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- BENTHAM, JEREMY.** 1970 [1789]. *AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION*. ATHLONE PRESS.
- BERGSTROM, THEODORE C.** 1989. "A FRESH LOOK AT THE ROTTEN KID THEOREM-- AND OTHER HOUSEHOLD MYSTERIES." *JOURNAL OF POLITICAL ECONOMY*, 97, 1138-59.

- BESLEY, T.** 2013. "WHAT'S THE GOOD OF THE MARKET: AN ESSAY ON MICHAEL SANDEL'S WHAT MONEY CAN'T BUY." *JOURNAL OF ECONOMIC LITERATURE*, 1, 478-93.
- BESLEY, TIMOTHY; GWYN BEVAN AND KONRAD BURCHARDI.** 2009. "ACCOUNTABILITY AND INCENTIVES: THE IMPACTS OF DIFFERENT REGIMES ON HOSPITAL WAITING TIMES IN ENGLAND AND WALES," *LSE*.
- BESLEY, TIMOTHY AND MAITREESH GHATAK.** 2005. "COMPETITION AND INCENTIVES WITH MOTIVATED AGENTS." *AMERICAN ECONOMIC REVIEW*, 95, 616-36.
- BISENIUS, DON.** 2010. "A PERSPECTIVE ON STRATEGIC DEFAULTS,"
[HTTP://WWW.FREDDIEMAC.COM/NEWS/FEATURED_PERSPECTIVES/20100503_BISENIUS.HTML:](http://www.freddiemac.com/news/featured_perspectives/20100503_bisenius.html)
- BLISS, CHRISTOPHER J.** 1972. "REVIEW OF R.M. TITMUSS, THE GIFT RELATIONSHIP: FROM HUMAN BLOOD TO SOCIAL POLICY." *JOURNAL OF PUBLIC ECONOMICS*, 1, 162-65.
- BLOOMFIELD, MORTON.** 1952. *THE SEVEN DEADLY SINS*. EAST LANSING: MICHIGAN STATE UNIVERSITY PRESS.
- BOEHM, CHRISTOPHER.** 1984. *BLOOD REVENGE: THE ENACTMENT AND MANAGEMENT OF CONFLICT IN MONTENEGRO AND OTHER TRIBAL SOCIETIES*. LAWRENCE: UNIVERSITY PRESS OF KANSAS.
- BOSWELL, JAMES.** 1980. *LIFE OF JOHNSON*. OXFORD: OXFORD UNIVERSITY PRESS (FIRST PUBLISHED 1791).
- BOWLES, SAMUEL.** 1998. "ENDOGENOUS PREFERENCES: THE CULTURAL CONSEQUENCES OF MARKETS AND OTHER ECONOMIC INSTITUTIONS." *JOURNAL OF ECONOMIC LITERATURE*, 36(1), 75-111.
- BOWLES, SAMUEL.** 2004. *MICROECONOMICS: BEHAVIOR, INSTITUTIONS, AND EVOLUTION*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- BOWLES, SAMUEL AND HERBERT GINTIS.** 2011. *A COOPERATIVE SPECIES: HUMAN RECIPROCITY AND ITS EVOLUTION*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- BOWLES, SAMUEL AND SUNG-HA HWANG.** 2008. "SOCIAL PREFERENCES AND PUBLIC ECONOMICS: MECHANISM DESIGN WHEN PREFERENCES DEPEND ON INCENTIVES." *JOURNAL OF PUBLIC ECONOMICS*, VOL 92(8-9), 1811-20.
- BOWLES, SAMUEL AND UGO PAGANO.** 2006. "ECONOMIC INTEGRATION, CULTURAL STANDARDIZATION, AND THE POLITICS OF SOCIAL INSURANCE," S. BOWLES, P. BARDHAN AND M. WALLERSTEIN, *GLOBALIZATION AND EGALITARIAN REDISTRIBUTION*. PRINCETON: PRINCETON UNIVERSITY PRESS,
- BOWLES, SAMUEL AND SANDRA POLANIA-REYES.** 2012. "ECONOMIC INCENTIVES AND SOCIAL PREFERENCES: SUBSTITUTES OR COMPLEMENTS?" *JOURNAL OF ECONOMIC LITERATURE*, 50(2), 368-425.
- BOWRING, JOHN ED.** 1962. *THE WORKS OF JEREMY BENTHAM: VOLUME 8*. NEW YORK: RUSSELL AND RUSSELL.
- BRAATEN, RAGNHILD HAUGLI.** 2014. "LAND RIGHTS AND COMMUNITY COOPERATION: PUBLIC GOODS EXPERIMENTS FROM PERU." *WORLD DEVELOPMENT*, 61, 127-41.
- BUCHANAN, JAMES.** 1975. *THE LIMITS OF LIBERTY*. CHICAGO: UNIVERSITY OF CHICAGO PRESS.
- CAMERER, COLIN.** 2003. *BEHAVIORAL GAME THEORY: EXPERIMENTAL STUDIES OF STRATEGIC INTERACTION*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- CAMERER, COLIN AND ERNST FEHR.** 2004. "MEASURING SOCIAL NORMS AND PREFERENCES USING EXPERIMENTAL GAMES: A GUIDE FOR SOCIAL SCIENTISTS," J. HENRICH, S. BOWLES, R. BOYD, C. CAMERER, E. FEHR AND H. GINTIS, *FOUNDATIONS OF HUMAN SOCIALITY: ECONOMIC EXPERIMENTS AND ETHNOGRAPHIC EVIDENCE FROM FIFTEEN SMALL-SCALE SOCIETIES*. OXFORD: OXFORD UNIVERSITY PRESS,

- CARDENAS, JUAN CAMILO; JOHN K. STRANLUND AND CLEVE E. WILLIS.** 2000. "LOCAL ENVIRONMENTAL CONTROL AND INSTITUTIONAL CROWDING-OUT." *WORLD DEVELOPMENT*, 28(10), 1719-33.
- CARPENTER, JEFFREY; SAMUEL BOWLES; HERBERT GINTIS AND SUNG-HA HWANG.** 2009. "STRONG RECIPROCITY AND TEAM PRODUCTION: THEORY AND EVIDENCE." *JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION*, 71(2), 221-32.
- CARROLL, LEWIS.** 1982. *ALICE'S ADVENTURES IN WONDERLAND*. NEW YORK: HARCOURT BRACE JANOVICH.
- CHATTERJEE, K.** 1982. "INCENTIVE COMPATIBILITY IN BARGAINING UNDER UNCERTAINTY." *QUARTERLY JOURNAL OF ECONOMICS*, 97(1), 717-26.
- CHATTERJEE, K AND W SAMUELSON.** 1983. "BARGAINING UNDER INCOMPLETE INFORMATION." *OPERATIONS RESEARCH*, 31(5), 23-39.
- CHRIST, MATTHEW.** 1990. "LITURGY AVOIDANCE AND ANTIDOSIS IN CLASSICAL ATHENS." *TRANSACTIONS OF THE AMERICAN PHILOSOPHICAL ASSOCIATION*, 10, 147-69.
- COHEN, JONATHAN.** 2005. "THE VULCANIZATION OF THE HUMAN BRAIN: A NEURAL PERSPECTIVE ON INTERACTIONS BETWEEN COGNITION AND EMOTION." *JOURNAL OF ECONOMIC PERSPECTIVES*, 19(4), 3-24.
- COOLEY, CHARLES HORTON.** 1902. *HUMAN NATURE AND THE SOCIAL ORDER*. NEW YORK: CHARLES SCRIBNER'S SONS.
- D'ANTONI, M AND UGO PAGANO.** 2002. "NATIONAL CULTURES AND SOCIAL PROTECTION AS ALTERNATIVE INSURANCE DEVICES." *STRUCTURAL CHANGE AND ECONOMIC DYNAMICS*, 13, 367-86.
- D'ASPROMONT, CLAUDE AND LOUIS-ANDRE GERARD-VARET.** 1979. "ON BAYESIAN INCENTIVE COMPATIBLE MECHANISMS," J. J. LAFFONT, *AGGREGATION AND REVELATION OF PREFERENCES*. AMSTERDAM: NORTH HOLLAND, 269-88.
- DURKHEIM, EMILE.** 1967. *DE LA DIVISION DU TRAVAIL SOCIAL*. PARIS: PRESSES UNIVERSITAIRES DE FRANCE (FIRST PUBLISHED IN 1902).
- EDGEWORTH, F.Y.** 1881. *MATHEMATICAL PSYCHICS: AN ESSAY ON THE APPLICATION OF MATHEMATICS TO THE MORAL SCIENCES*. LONDON: C. KEGAN PAUL AND COMPANY.
- ELIAS, NORBERT.** 2000. *THE CIVILIZING PROCESS*. OXFORD: BLACKWELL [BASEL, 1939].
- ERTAN, ARHAN; TALBOT PAGE AND LOUIS PUTTERMAN.** 2009. "WHO TO PUNISH? INDIVIDUAL DECISIONS AND MAJORITY RULE IN MITIGATING THE FREE-RIDER PROBLEM." *EUROPEAN ECONOMIC REVIEW*, 3, 495-511.
- FALK, A. AND NORA SZECH.** 2013. "ORGANIZATIONS, DIFFUSED PIVOTALITY, AND IMMORAL OUTCOME," *UNIVERSITY OF BONN*.
- FALK, ARMIN AND NORA SZECH.** 2013. "MORALS AND MARKETS." *SCIENCE AND SOCIETY*, 340, 707-11.
- FALKINGER, JOSEF; ERNST FEHR; SIMON GAECHTER AND RUDOLF WINTER-EBMER.** 2000. "A SIMPLE MECHANISM FOR THE EFFICIENT PROVISION OF PUBLIC GOODS." *AMERICAN ECONOMIC REVIEW*, 90(1), 247-64.
- FAROOQ, OMER.** 2005. "DRUMMING TAX SENSE INTO EVADERS," BBC NEWS,
- FEHR, E. AND BETTINA ROCKENBACH.** 2003. "DETRIMENTAL EFFECTS OF SANCTIONS ON HUMAN ALTRUISM." *NATURE*, 422(13 MARCH), 137-40.

- FEHR, ERNST AND ARMIN FALK.** 2002. "PSYCHOLOGICAL FOUNDATIONS OF INCENTIVES." *EUROPEAN ECONOMIC REVIEW*, 46(4 - 5), 687-724.
- FEHR, ERNST AND URS FISCHBACHER.** 2001. "WHY SOCIAL PREFERENCES MATTER," *NOBEL SYMPOSIUM ON BEHAVIORAL AND EXPERIMENTAL ECONOMICS*. STOCKHOLM:
- FEHR, ERNST AND SIMON GAECHTER.** 2000A. "COOPERATION AND PUNISHMENT IN PUBLIC GOODS EXPERIMENTS." *AMERICAN ECONOMIC REVIEW*, 90(4), 980-94.
- FEHR, ERNST AND SIMON GAECHTER.** 2000B. "FAIRNESS AND RETALIATION: THE ECONOMICS OF RECIPROCITY." *JOURNAL OF ECONOMIC PERSPECTIVES*, 14(3), 159-81.
- FEHR, ERNST AND JOHN LIST.** 2004. "THE HIDDEN COSTS AND RETURNS OF INCENTIVES: TRUST AND TRUSTWORTHINESS AMONG CEOS." *JOURNAL OF THE EUROPEAN ECONOMIC ASSOCIATION*, 2(5), 743-71.
- FISMAN, RAY AND EDWARD MIGUEL.** 2006. "CULTURES OF CORRUPTION: EVIDENCE FROM DIPLOMATIC PARKING TICKETS," *NBER*.
- FREY, B. AND R JEGEN.** 2001. "MOTIVATION CROWDING THEORY: A SURVEY OF EMPIRICAL EVIDENCE." *JOURNAL OF ECONOMIC SURVEYS*, 15(5), 589 - 611.
- FRIEDMAN, MILTON.** 1970. "THE SOCIAL RESPONSIBILITY OF BUSINESS IS TO INCREASE ITS PROFITS." *NEWYORK TIMES*, SEPTEMBER 13.
- FROHLICH, NORMAN AND JOE A. OPPENHEIMER.** 1995. "THE INCOMPATIBILITY OF INCENTIVE COMPATIBLE DEVICES AND ETHICAL BEHAVIOR: SOME EXPERIMENTAL RESULTS AND INSIGHTS." *PUBLIC CHOICE STUDIES*, 25, 24-51.
- FROHLICH, NORMAN AND JOE OPPENHEIMER.** 2003. "OPTIMAL POLICIES AND SOCIALLY ORIENTED BEHAVIOR: SOME PROBLEMATIC EFFECTS OF AN INCENTIVE COMPATIBLE DEVICE." *PUBLIC CHOICE*, 117, 273-93.
- FRYER, ROLAND.** 2010. "FINANCIAL INCENTIVES AND STUDENT ACHIEVEMENT: EVIDENCE FROM RANDOMIZED TRIALS," *NBER*.
- GAECHTER, SIMON; BENEDIKT HERRMANN AND CHRISTIAN THONI.** 2010. "CULTURE AND COOPERATION." *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B*, 365, 2651-61.
- GAECHTER, SIMON; ESTHER KESSLER AND MANFRED KONIGSTEIN.** 2010. "DO INCENTIVES DESTROY VOLUNTARY COOPERATION?," *UNIVERSITY OF NOTTINGHAM, SCHOOL OF ECONOMICS*.
- GALBIATI, ROBERTO AND PIETRO VERTOVA.** 2014. "HOW LAWS AFFECT BEHAVIOUR: OBLIGATIONS, INCENTIVES AND COOPERATIVE BEHAVIOR." *INTERNATIONAL REVIEW OF LAW & ECONOMICS*, 38, 48 - 57.
- GALBIATI, ROBERTO AND PIETRO VERTOVA.** 2008. "OBLIGATIONS AND COOPERATIVE BEHAVIOR IN PUBLIC GOOD GAMES." *GAMES AND ECONOMIC BEHAVIOR*, 164(1), 146-70.
- GAMBETTA, DIEGO.** 2008. *DO STRONG FAMILY TIES INHIBIT TRUST?* UNIVERSITY OF OXFORD.
- GARVEY, STEPHEN P.** 1998. "CAN SHAMING PUNISHMENTS EDUCATE? ." *UNIVERSITY OF CHICAGO LAW REVIEW*, 65, 733-94.
- GASIOROWSKA, AGATA; TOMASZ ZALESKIEWICZ AND SANDRA WYGRAB.** 2012. "WOULD YOU DO SOMETHING FOR ME? THE EFFECTS OF MONEY ACTIVATION ON SOCIAL PREFERENCES AND SOCIAL BEHAVIOR IN YOUNG CHILDREN." *JOURNAL OF ECONOMIC PSYCHOLOGY*, 33(2012), 603-08.

- GAUTHIER, DAVID.** 1986. *MORALS BY AGREEMENT*. OXFORD: CLARENDON PRESS.
- GELLNER, ERNEST.** 1983. *NATIONS AND NATIONALISM*. ITHACA: CORNELL UNIVERSITY PRESS.
- GELLNER, ERNEST.** 1988. "TRUST, COHESION, AND THE SOCIAL ORDER," D. GAMBETTA, *TRUST: MAKING AND BREAKING COOPERATIVE RELATIONS*. OXFORD: BASIL BLACKWELL, 142-57.
- GIBBARD, ALLAN.** 1973. "MANIPULATION OF VOTING SCHEMES: A GENERAL RESULT." *JOURNAL OF ECONOMIC THEORY*, 41(4), 587-601.
- GINGES, J; SCOTT ATRAN; DOUGLAS MEDIN AND KHALIL SHIKAKI.** 2007. "SACRED BOUNDS ON RATIONAL RESOLUTION OF VIOLENT POLITICAL CONFLICT." *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE*, 104(18), 7357-60.
- GNEEZY, URI AND ALDO RUSTICHINI.** 2000. "PAY ENOUGH OR DON'T PAY AT ALL." *QUARTERLY JOURNAL OF ECONOMICS*, 115(2), 791-810.
- GOODIN, ROBERT E. AND ANDREW REEVE** EDS. 1989. *LIBERAL NEUTRALITY*. LONDON: ROUTLEDGE.
- GRANT, RUTH.** 2012. *STRINGS ATTACHED: UNTANGLING THE ETHICS OF INCENTIVES*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- GREENBERGER, SCOTT.** 2003. "SICK DAY ABUSES FOCUS OF FIRE TALKS," *BOSTON GLOBE*. B7.
- GREENE, JOSHUA.** 2014. *MORAL TRIBES: EMOTION, REASON, AND THE GAP BETWEEN US AND THEM*. UNPUBLISHED MANUSCRIPT.
- GREENE, JOSHUA.** 2010. "THE SECRET JOKE OF KANT'S SOUL," *CH 2 OF FORTHCOMING BOOK*.
- GREENE, JOSHUA D.; R. BRIAN SOMMERVILLE; LEIGH E. NYSTROM; JOHN M. DARLEY AND JONATHON D. COHEN.** 2001. "AN FMRI INVESTIGATION OF EMOTIONAL ENGAGEMENT IN MORAL JUDGEMENT." *SCIENCE*, 293, 2105-08.
- GREIF, AVNER.** 1994. "CULTURAL BELIEFS AND THE ORGANIZATION OF SOCIETY: AN HISTORICAL AND THEORETICAL REFLECTION ON COLLECTIVIST AND INDIVIDUALIST SOCIETIES." *JOURNAL OF POLITICAL ECONOMY*, 102(5), 912-50.
- GREIF, AVNER.** 2002. "INSTITUTIONS & IMPERSONAL EXCHANGE: FROM COMMUNAL TO INDIVIDUAL RESPONSIBILITY." *JOURNAL OF INSTITUTIONAL AND THEORETICAL ECONOMICS*, 158(1), 168-204.
- GUISSO, LUIGI; PAOLA SAPIENZA AND LUIGI ZINGALES.** 2013. "THE DETERMINANTS OF ATTITUDES TOWARD STRATEGIC DEFAULT ON MORTGAGES." *JOURNAL OF FINANCE*, LXVII, 1473-515.
- HAYEK, FRIEDRICH A.** 1945. "THE USE OF KNOWLEDGE IN SOCIETY." *AMERICAN ECONOMIC REVIEW*, 35(4), 519-30.
- HEIFETZ, A.; E. SEGEV AND E. TALLEY.** 2007. "MARKET DESIGN WITH ENDOGENOUS PREFERENCES." *GAMES AND ECONOMIC BEHAVIOR*, 58, 121-53.
- HENRICH, JOE; ROBERT BOYD; SAMUEL BOWLES; COLIN CAMERER; ERNST FEHR; HERBERT GINTIS; RICHARD MCELREATH; MICHAEL ALVARD; ABIGAIL BARR; JEAN ENSMINGER; NATALIE SMITH HENRICH; KIM HILL; FRANCISCO GIL-WHITE; MICHAEL GURVEN; FRANK MARLOWE; JOHN PATTON AND DAVID TRACER.** 2005. "'ECONOMIC MAN' IN CROSS-CULTURAL PERSPECTIVE: BEHAVIORAL EXPERIMENTS IN 15 SMALL-SCALE SOCIETIES." *BEHAVIORAL AND BRAIN SCIENCES*, 28, 795-855.
- HENRICH, JOSEPH; JEAN ENSMINGER; RICHARD MCELREATH; ABIGAIL BARR; CHRIS WALLACE; CLARK**

- BENNETT; ALEXANDER BOLYANATZ; JUAN CAMILO CARDENAS; MICHAEL GURVEN; EDWINS GWAKO; CAROLYN LESOROGOL; FRANK MARKLOWE; DAVID TRACER AND JOHN ZIKER.** 2010. "MARKETS, RELIGION, COMMUNITY SIZE AND THE EVOLUTION OF FAIRNESS AND PUNISHMENT." *SCIENCE*, 327, 1480-84.
- HENRICH, JOSEPH; RICHARD MCELREATH; ABIGAIL BARR; JEAN ENSMINGER; CLARK BARRETT; ALEXANDER BOLYMATZ; JUAN CAMILO CARDENAS; MICHAEL GURVEN; EDWINS GWAKO; NATALIE HENRICH; CAROLYN LESOROGOL; FRANK MARLOWE; DAVID TRACER AND JOHN ZIKER.** 2006. "COSTLY PUNISHMENT ACROSS HUMAN SOCIETIES." *SCIENCE*, 312, 1767-70.
- HERRMANN, BENEDIKT; CHRISTIAN THONI AND SIMON GAECHTER.** 2008A. "ANTISOCIAL PUNISHMENT ACROSS SOCIETIES." *SCIENCE*, 319(7 MARCH 2008), 1362-67.
- HERRMANN, BENEDIKT; CHRISTIAN THONI AND SIMON GAECHTER.** 2008B. "SUPPORTING ONLINE MATERIAL FOR 'ANTISOCIAL PUNISHMENT ACROSS SOCIETIES'." *SCIENCE*, 319(7 MARCH 2008), 1362-67.
- HIRSCHMAN, ALBERT O.** 1985. "AGAINST PARSIMONY:THREE WAYS OF COMPLICATING SOME CATEGORIES OF ECONOMIC DISCOURSE." *ECONOMICS AND PHILOSOPHY*, 1(1), 7-21.
- HOBBS, THOMAS.** 2005. *LEVIATHAN*. LONDON: CONTINUUM [1651].
- HOLMAS, TOR HELGE; EGIL KJERSTAD; HILDE LURAS AND ODD RUNE STRAUME.** 2010. "DOES MONETARY PUNISHMENT CROWD OUT PRO-SOCIAL MOTIVATION? A NATURAL EXPERIMENT ON HOSPITAL LENGTH OF STAY." *JOURNAL OF ECONOMIC BEHAVIOR & ORGANIZATION*, IN PRESS.
- HOLMES, OLIVER WENDEL, JR.** 1897. "THE PATH OF THE LAW." *HARVARD LAW REVIEW*, 10(457), 457-78.
- HUME, DAVID.** 1964. *DAVID HUME, TH PHILOSOPHICAL WORKS*. DARMSTADT: SCIENTIA VERLAG AALEN.
- HURWICZ, LEONID.** 1974. "THE DESIGN OF MECHANISMS FOR RESOURCE ALLOCATION," M. D. INTRILLIGATOR AND D. A. KENDRICK, *FRONTIERS OF QUANTITATIVE ECONOMICS II*. AMSTERDAM: NORTH HOLLAND PRESS, 3-42.
- HURWICZ, LEONID.** 1972. "ON INFORMATIONALLY DECENTRALIZED SYSTEMS," R. RADNER AND B. MCGUIRE, *DECISION AND ORGANIZATION*. AMSTERDAM: NORTH-HOLLAND PRESS,
- HURWICZ, LEONID; DAVID SCHMEIDLER AND HUGO SONNENSCHNEIN.** 1985. "SOCIAL GOALS AND SOCIAL ORGANIZATION: ESSAYS IN MEMORY OF ELISHA PAZNER," L. H. A. D. S. A. H. SONNENSCHNEIN, *SOCIAL GOALS AND SOCIAL ORGANIZATION: ESSAYS IN MEMORY OF ELISHA PAZNER*. CAMBRIDGE, UK: CAMBRIDGE UNIVERSITY PRESS,
- HWANG, SUNG-HA AND SAMUEL BOWLES.** 2014A. "INCENTIVES, SOCIALIZATION, AND CIVIC PREFERENCES," SANTA FE INSTITUTE:
- HWANG, SUNG HA AND SAMUEL BOWLES.** 2014B. "OPTIMAL INCENTIVES WITH STATE-DEPENDENT PREFERENCES." *JOURNAL OF PUBLIC ECONOMIC THEORY*, IN PRESS.
- IRLENBUSCH, BERND AND G.K. RUCHALA.** 2008. "RELATIVE REWARDS WITHIN TEAM-BASED COMPENSATION." *LABOUR ECONOMICS*, 15, 141-67.
- KAHNEMAN, DANIEL.** 1994. "NEW CHALLENGES TO THE RATIONALITY ASSUMPTION." *JOURNAL OF INSTITUTIONAL AND THEORETICAL ECONOMICS*, 150(1), 18-36.
- KAHNEMAN, DANIEL AND AMOS TVERSKY.** 2000. *CHOICES, VALUES AND FRAMES*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- KAMINSKI, J; A. PITTSCH AND MICHAEL TOMASELLO.** 2013. "DOGS STEAL IN THE DARK." *ANIMAL COGNITION*, 16, 385-94.

- KOHN, MELVIN.** 1969. *CLASS AND CONFORMITY*. HOMewood, IL: DORSEY PRESS.
- KOHN, MELVIN L.** 1990. "UNRESOLVED ISSUES IN THE RELATIONSHIP BETWEEN WORK AND PERSONALITY," K. ERIKSON AND S. P. VALLAS, *THE NATURE OF WORK: SOCIOLOGICAL PERSPECTIVES*. NEW HAVEN: YALE UNIVERSITY PRESS, 36-68.
- KOHN, MELVIN L. AND CARMİ SCHOOLER ET AL.** 1983. *WORK AND PERSONALITY: AN INQUIRY INTO THE IMPACT OF SOCIAL STRATIFICATION*. NEW JERSEY: ABLEX PUBLISHING CORPORATION.
- KOHN, MELVIN; ATSUSHI NAOI; CARRIE SCHOENBACH; CARMİ SCHOOLER AND KAZIMIERZ SŁOMCZYNSKI.** 1990. "POSITION IN THE CLASS STRUCTURE AND PSYCHOLOGICAL FUNCTIONING IN THE U.S., JAPAN, AND POLAND." *AMERICAN JOURNAL OF SOCIOLOGY*, 95(4), 964-1008.
- LAFFONT, JEAN JACQUES.** 2000. *INCENTIVES AND POLITICAL ECONOMY*. OXFORD: OXFORD UNIVERSITY PRESS.
- LAFFONT, JEAN JACQUES AND ERIC MASKIN.** 1979. "A DIFFERENTIABLE APPROACH TO EXPECTED UTILITY-MAXIMIZING MECHANISMS," J. J. LAFFONT, *AGGREGATION AND REVELATION OF PREFERENCES*. AMSTERDAM: NORTH HOLLAND, 289-308.
- LANJOUW, PETER AND NICHOLAS STERN** EDS. 1998. *ECONOMIC DEVELOPMENT IN PALANPUR OVER FIVE DECADES*. DELHI: OXFORD UNIVERSITY PRESS.
- LEPPER, MARK R. AND DAVID GREENE.** 1978. *THE HIDDEN COSTS OF REWARD : NEW PERSPECTIVES ON THE PSYCHOLOGY OF HUMAN MOTIVATION*. HILLSDALE: L. ERLBAUM ASSOCIATES.
- LEPPER, MARK R.; DAVID GREENE AND RICHARD E. NISBETT.** 1973. "UNDERMINING CHILDREN'S INTRINSIC INTEREST WITH EXTRINSIC REWARD: A TEST OF THE 'OVERJUSTIFICATION' HYPOTHESIS." *JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY*, 28(1), 129-37.
- LEPPER, MARK R.; GERALD SAGOTSKY; JANET DEFOE AND DAVID GREENE.** 1982. "CONSEQUENCES OF SUPERFLUOUS SOCIAL CONSTRAINTS: EFFECTS ON YOUNG CHILDREN'S SOCIAL INFERENCES AND SUBSEQUENT INTRINSIC INTEREST." *JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY*, 42(1), 51-65.
- LIPSEY, R. AND K. LANCASTER.** 1956-1957. "THE GENERAL THEORY OF THE SECOND BEST." *REVIEW OF ECONOMIC STUDIES*, 24(1), 11-32.
- LOEWENSTEIN, GEORGE F.; LEIGH THOMPSON AND MAX H. BAZERMAN.** 1989. "SOCIAL UTILITY AND DECISION MAKING IN INTERPERSONAL CONTEXTS." *JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY*, 57(3), 426-41.
- LOEWENSTEIN, GEORGE AND TED O'DONOGHUE.** 2007. "THE HEAT OF THE MOMENT: MODELING INTERACTIONS BETWEEN AFFECT AND DELIBERATION," *SOCIAL AND DECISION SCIENCES, CMU*.
- LOEWENSTEIN, GEORGE AND DEBORAH SMALL.** 2007. "THE SCARECROW AND THE TIN MAN: THE VICISSITUDES OF HUMAN SUFFERING AND CARING." *REVIEW OF GENERAL PSYCHOLOGY*, 11(2), 112-26.
- MACHIAVELLI, NICOLO.** 1996. *DISCOURSES ON LIVY*. CHICAGO: UNIVERSITY OF CHICAGO PRESS [
- MACHIAVELLI, NICOLO.** 1900. *IL PRINCIPE*. FIRENZE: G.C. SANSONI (FIRST CIRCULATED IN 1513, TRANSLATION BY THE CURRENT AUTHOR).
- MACHIAVELLI, NICOLÒ.** 2001. *DISCORSI SOPRA LA PREMA DECA DI TITO LIVIO*. ROMA: SALERNO EDITRICE (FIRST PUBLISHED IN 1513-1517, TRANSLATION BY THE PRESENT AUTHOR).
- MACHIAVELLI, NICOLÒ.** 1984. *DISCORSI SOPRA LA PREMA DECA DI TITO LIVIO*. MILANO: RIZZOLI (FIRST PUBLISHED IN 1513-1517, TRANSLATION BY THE PRESENT AUTHOR).

- MAHDI, NILOUFER QASIM.** 1986. "PUKHUTUNWALI: OSTRACISM AND HONOR AMONG PATHAN HILL TRIBES." *ETHOLOGY AND SOCIOBIOLOGY*, 7(3-4), 295 - 304.
- MALLON, FLORENCIA E.** 1983. *THE DEFENSE OF COMMUNITY IN PERU'S CENTRAL HIGHLANDS: PEASANT STRUGGLE AND CAPITALIST TRANSITION 1860 - 1940*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- MANDEVILLE, BERNARD.** 1924. *THE FABLE OF THE BEES, OR PRIVATE VICIES, PUBLIC BENEFITS*. OXFORD: CLARENDON PRESS.
- MANDEVILLE, BERNARD.** 1988A. "A SEARCH INTO THE NATURE OF SOCIETY," F. B. KAYE, *THE FABLE OF THE BEES*. INDIANAPOLIS: LIBERTY FUND, 323-70.
- MANDEVILLE, BERNARD.** 1988B. "A VINDICATION OF THE BOOK, FROM THE ASPERSIONS CONTAIN'D IN A PRESENTMENT OF THE GRAND JURY OF MIDDLESEX," F. B. KAYE, *THE FABLE OF THE BEES*. INDIANAPOLIS: LIBERTY FUND, 381-412.
- MARTIN, GERARD AND MIGUEL CEBALLOS.** 2004. *BOGOTA: ANATOMIA DE UNA TRANSFORMACION: POLITICAS DE SEGURIDAD CIUDADANA 1995-2003*. BOGOTA: EDITORIAL PONTIFICIA UNIVERSIDAD JAVERIANA.
- MASKIN, ERIC.** 1985. "THE THEORY OF IMPLEMENTATION IN NASH EQUILIBRIUM: A SURVEY," L. HURWICZ, D. SCHMEIDLER AND H. SONNENSCHNEIN, *SOCIAL GOALS AND SOCIAL ORGANIZATION; ESSAYS IN MEMORY OF ELISHA PAZNER*. CAMBRIDGE: CAMBRIDGE UNIVERSITY PRESS, 173-341.
- MELLSTROM, CARL AND MAGNUS JOHANNESON.** 2008. "CROWDING OUT IN BLOOD DONATION: WAS TITMUS RIGHT?" *JOURNAL OF THE EUROPEAN ECONOMIC ASSOCIATION*, 6(4), 845-63.
- MILL, JOHN STUART.** 1844. *ESSAYS ON SOME UNSETTLED QUESTIONS OF POLITICAL ECONOMY*. LONDON: JOHN W. PARKER.
- MYERSON, ROGER B. AND MARK A. SATTERTHWAITE.** 1983. "EFFICIENT MECHANISM FOR BILATERAL TRADING." *JOURNAL OF ECONOMIC THEORY*, 29, 265-81.
- OBER, J.** 2008. *DEMOCRACY AND KNOWLEDGE: INNOVATION AND LEARNING IN CLASSICAL ATHENS*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- OUCHI, WILLIAM.** 1980. "MARKETS BUREAUCRACIES AND CLANS." *ADMINISTRATIVE SCIENCE QUARTERLY*, 25, 129-41.
- PARSONS, TALCOTT.** 1967. *SOCIOLOGICAL THEORY AND MODERN SOCIETY*. NEW YORK: FREE PRESS.
- RAWLS, JOHN.** 1971. *A THEORY OF JUSTICE*. CAMBRIDGE: HARVARD UNIVERSITY PRESS.
- RIANO, YVONNE.** xxxx. "ADDRESSING URBAN FEAR AND VIOLENCE IN BOGOTA THROUGH THE 'CULTURE OF CITIZENSHIP': SCOPE AND CHALLENGES OF A UNIQUE APPROACH.," 209-25.
- ROSENTHAL, ELISABETH.** 2008. "MOTIVATED BY A TAX, IRISH SPURN PLASTIC BAGS," *NEW YORK TIMES*. NEW YORK:
- ROUSSEAU, JEAN-JACQUES.** 1762. *THE SOCIAL CONTRACT: OR PRINCIPLES OF POLITICAL RIGHT*.
- SAHLINS, MARSHALL.** 1974. *STONE AGE ECONOMICS*. CHICAGO: ALDINE PUBLISHING COMPANY.
- SANDEL, MICHAEL.** 2013. "MARKET REASONING AS MORAL REASONING: WHY ECONOMISTS SHOULD RE-ENGAGE WITH POLITICAL PHILOSOPHY." *JOURNAL OF ECONOMIC PERSPECTIVES*, 27, 121-40.
- SANDEL, MICHAEL.** 2012. *WHAT MONEY CAN'T BUY: THE MORAL LIMITS OF MARKETS*. NEW YORK: FARRAR, STRAUS AND IROUX.

- SANFEY, ALAN; GEORGE LOEWENSTEIN; SAMUEL MCCLURE AND JONATHAN COHEN.** 2006. "NEUROECONOMICS: CROSS-CURRENTS IN RESEARCH ON DECISION-MAKING." *TRENDS IN COGNITIVE SCIENCES*, 10(3), 108-16.
- SANFEY, ALAN; JAMES RILLING; JESSICA ARONSON; LEIGH NYSTROM AND JONATHAN COHEN.** 2003. "THE NEURAL BASIS OF ECONOMIC DECISION-MAKING IN THE ULTIMATUM GAME." *SCIENCE*, 300, 1755-58.
- SCHMITZ, HUBERT.** 1999. "FROM ASCRIBED TO EARNED TRUST IN EXPORTING CLUSTERS." *JOURNAL OF INTERNATIONAL ECONOMICS*, 48, 138-50.
- SCHOTTER, ANDREW; AVI WEISS AND INDIGO ZAPATER.** 1996. "FAIRNESS AND SURVIVAL IN ULTIMATUM AND DICTATORSHIP GAMES." *JOURNAL OF ECONOMIC BEHAVIOR AND ORGANIZATION*, 31(1), 37-56.
- SCHULTZE, CHARLES L.** 1977. *THE PUBLIC USE OF PRIVATE INTEREST*. WASHINGTON, D.C: BROOKINGS INSTITUTION.
- SCHUMPETER, JOSEPH.** 1950. "THE MARCH INTO SOCIALISM." *AMERICAN ECONOMIC REVIEW*, 40(2), 446-56.
- SEABRIGHT, PAUL.** 2009. "CONTINUOUS PREFERENCES AND DISCONTINUOUS CHOICES: HOW ALTRUISTS RESPOND TO INCENTIVES." *THE B.E. JOURNAL OF THEORETICAL ECONOMICS*, 9, ARTICLE 14.
- SHINADA, MIZUHU AND TOSHIO YAMAGISHI.** 2007. "PUNISHING FREE RIDERS: DIRECT AND INDIRECT PROMOTION OF COOPERATION." *EVOLUTION AND HUMAN BEHAVIOR*, 28, 330-39.
- SHUBIK, MARTIN.** 1959. *STRATEGY AND MARKET STRUCTURE: COMPETITION, OLIGOPOLY, AND THE THEORY OF GAMES*. NEW YORK: WILEY.
- SKITKA, L.; E MULLEN; T. GRIFFIN; S HUTCHINSON AND B CHAMBERLIN.** 2002. "DISPOSITIONS, IDEOLOGICAL SCRIPTS OF MOTIVATIONAL CORRECTIONS? UNDERSTANDING IDEOLOGICAL DIFFERENCES IN ATTRIBUTIONS FOR SOCIAL PROBLEMS." *JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY*, 83, 470-87.
- SMALL, DEBORAH; GEORGE LOEWENSTEIN AND P. SLOVIC.** 2007. "SYMPATHY AND CALLOUSNESS: THE IMPACT OF DELIBERATIVE THOUGHT ON DONATIONS TO IDENTIFIABLE AND STATISTICAL VICTIMS." *ORGANIZATIONAL BEHAVIOR & HUMAN DECISION PROCESSES*, 102, 143-53.
- SMITH, ADAM.** 1776. *AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS*. LONDON,: W. STRAHAN AND T. CADELL.
- SMITH, ADAM** ED. 1976 [1776]. *AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS*. OXFORD: CLARENDON PRESS.
- SMITH, ADAM.** 1896. *LECTURES ON JUSTICE, POLICE, REVENUE, AND ARMS*. OXFORD: CLARENDON PRESS.
- SMITH, ADAM.** 1976 [1759]. *THEORY OF MORAL SENTIMENTS*. OXFORD: CLARENDON PRESS.
- SOLOW, ROBERT.** 1971. "BLOOD AND THUNDER." *YALE LAW JOURNAL*, 80(8), 1696-711.
- SPINOZA, BENEDICT DE.** 1958. *THE POLITICAL WORKS*. OXFORD: THE CLARENDON PRESS.
- STOUT, LYNN.** 2011. *CULTIVATING CONSCIENCE: HOW GOOD LAWS MAKE GOOD PEOPLE*. PRINCETON: PRINCETON UNIVERSITY PRESS.
- STRAUSS, LEO.** 1988. *WHAT IS POLITICAL PHILOSOPHY*. CHICAGO: UNIVERSITY OF CHICAGO PRESS.
- TAYLOR, MICHAEL.** 1976. *ANARCHY AND COOPERATION*. LONDON: JOHN WILEY AND SONS.
- THALER, RICHARD AND CASS SUNSTEIN.** 2008. *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH AND*

- HAPPINESS*. NEW HAVEN: YALE UNIVERSITY PRESS.
- TILLY, CHARLES**. 1981. "CHARIVARIS, REPERTOIRES AND URBAN POLITICS," J. M. MERRIMAN, *FRENCH CITIES IN THE NINETEENTH CENTURY*. NEW YORK: HOLMES AND MEIER, 73-91.
- TIMES, NEW YORK**. 1988. "BAN GREED? NO: HARNESS IT." *NEW YORK TIMES*, 20 JANUARY, EDITORIAL PAGE.
- TIMES, THE**. 2014. "DOCTOS WHO MISS CANCER TO BE NAMED." *THE TIMES*, 30 JUNE, 3.
- TITMUSS, RICHARD M.** 1971. *THE GIFT RELATIONSHIP: FROM HUMAN BLOOD TO SOCIAL POLICY*. NEW YORK: PANTHEON BOOKS.
- TONNIES, FERDINAND**. 1963. *COMMUNITY AND SOCIETY*. NEW YORK: HARPER AND ROW.
- TRIVERS, R. L.** 1971. "THE EVOLUTION OF RECIPROCAL ALTRUISM." *QUARTERLY REVIEW OF BIOLOGY*, 46, 35-57.
- UPTON, WILLIAM EDWARD III**. 1974. "ALTRUISM, ATTRIBUTION, AND INTRINSIC MOTIVATION IN THE RECRUITMENT OF BLOOD DONORS." *DISSERTATION ABSTRACTS INTERNATIONAL*, 34(12), 6260-B.
- WEBER, MAX**. 1978. *ECONOMY AND SOCIETY: AN OUTLINE OF INTERPRETIVE SOCIOLOGY*. BERKELEY: UNIVERSITY OF CALIFORNIA PRESS.
- WHITE, BRENT**. 2010. "TAKE THIS HOUSE AND SHOVE IT: THE EMOTIONAL DRIVERS OF STRATEGIC DEFAULT," UNIVERSITY OF ARIZONA SCHOOL OF LAW,
- WIESSNER, POLLY**. 2005. "NORM ENFORCEMENT AMONG THE JU/'HOANSI BUSHMEN: A CASE OF STRONG RECIPROCITY?" *HUMAN NATURE*, 16(2), 115-45.
- WILKINSON-RYAN, TESS**. 2010. "DO LIQUIDATED DAMAGES ENCOURAGE EFFICIENT BREACH: A PSYCHOLOGICAL EXPERIMENT." *MICHIGAN LAW REVIEW*, 108.
- WOODBURN, JAMES**. 1982. "EGALITARIAN SOCIETIES." *MAN*, 17, 431-51.
- WOODRUFF, CHRISTOPHER**. 1998. "CONTRACT ENFORCEMENT AND TRADE LIBERALIZATION IN MEXICO'S FOOTWEAR INDUSTRY." *WORLD DEVELOPMENT*, 26(6), 979-91.
- WRZESNIEWSKI, AMY; BARRY SCHWARTZ; XIANGYU CONG; MICHAEL KANE; AUDREY OMAR AND KOLDITZ**. 2014. "MULTIPLE TYPES OF MOTIVS DON'T MULTIPLY THE MOTIVATON OF WEST POINT CADETS." *PROC. NATL. ACAD. SCI. USA*.
- YAMAGISHI, T.; K. S. COOK AND M WATABE**. 1998. "UNCERTAINTY, TRUST, AND COMMITMENT FORMATION IN THE U.S. AND JAPAN." *AMERICAN JOURNAL OF SCIOLOGY*, 104, 165-94.
- YAMAGISHI, TOSHIO AND M YAMAGISHI**. 1994. "TRUST AND COMMITMENT IN THE UNITED STATES AND JAPAN." *MOTIVATION AND EMOTION*, 18, 9-66.
- YEUNG, KING-TO AND JOHN LEVI MARTIN**. 2011. "THE LOOKING GLASS SELF: AN EMPIRICAL TEST AND ELABORATION." *SOCIAL FORCES*, 93(3), 843-79.

ENDNOTES

¹ Experimental evidence on the nature and extent of social preferences is surveyed in Bowles and Gintis (2011).

² Hwang and Bowles (2014b), Bowles and Hwang (2008)) Related studies addressing similar problems of incentives when people are motivated by social preferences are those of Benabou and Tirole (2006), Seabright (2009), Bar-Gill and Fershtman (2005), Bar-Gill and Fershtman (2004) and Heifetz, Segev and Talley (2007).

³ Sung Ha Hwang and I explore conditions under which under use of the incentive by the Machiavellian planner will occur (depending on the nature of the crowding out problem) in three papers (Hwang and Bowles (2014b), Bowles and Hwang (2008)) Hwang and Bowles (2014a).

⁴ Further evidence on the external validity of the behavioral experiments presented here is in Bowles and Gintis (2011).

⁵ Similar cases of crowding out due to the “bad news” conveyed by the incentive are at work in experiments among student subject pools in Switzerland, U.S., Italy, France and Costa Rica (as well as Germany) and in a diverse set of games including Gift Exchange, Public Goods, and a charity giving setting similar to a Dictator Game. Costa Rican businessmen also responded negatively to the bad news that incentives conveyed.

⁶ Some of it surveyed in Bowles (1998)

⁷ The results reported here are from Frohlich and Oppenheimer (2003)

⁸ These and the statistics reported below are calculated from data in Herrmann, Thoni and Gaechter (2008b).

⁹ A Cornell University dissertation two years later did suggest that monetary incentives substantially reduced highly motivated potential donors' likelihood of giving blood; but the work was never published and little read (Upton (1974)).

¹⁰ A few economists turned to the design of public policy in cases where social preferences and incentives were not separable, and in most cases were less than additive in their effects (Bowles (1989), Aaron (1994), Frey (1997), Bar-Gill and Fershtman (2005), (2004), Cervellati, Esteban and Kranich (2008), Sobel (2005), and Aghion, Algan and Cahuc (2008).)

¹¹ Chatterjee's results are an application to the double auction of a similar result demonstrated by d'Aspremont and Gerard-Varet (1979) in the case of the revelation of preferences for a public

good. Under their mechanism truthful revelation is incentive compatible, but the mechanism requires that participation be mandatory.

¹² Studied by Axelrod and Hamilton (1981), refining and extending the earlier insights of Shubik (1959), Trivers (1971), and Taylor (1976).

¹³ In the village Palampur (in Uttar Pradesh, India) the extension of the labor market (and resulting increased geographical mobility) appears to have reduced the costs of exit and hence the value of one's reputation, with the effect that the informal enforcement of lending contracts was undermined (Lanjouw and Stern (1998):570).. Similar cases in which greater mobility and hence anonymity of traders induced and facilitated by market incentives undermined the ethical and other regarding social norms that underpinned the preexisting norm of contractual enforcement come from long distance traders in early modern Europe (Greif (1994), Greif (2002)) and shoe manufacturers in Brazil and Mexico (Woodruff (1998), Schmitz (1999)).

¹⁴ Interpretations other than this "unfair intent" explanation are possible, however, because the larger was the desired back transfer, the more costly was compliance. Thus for larger demands, simply returning nothing and paying the fine (as many of the subjects did) might have been attractive to self-interested subjects, who had they been faced with a lower demand would have maximized payoffs by complying. It seems from this and similar experiments that fines may have negative effects even if imposed to implement a fair outcome and even when the decision to use the fine is not made by the investor, but rather by chance (Fehr and List (2004), Houser, Xiao, McCabe, et al. (2008b)). In these cases the fine appears to have compromised the trustee's sense of autonomy. But the experiments also are consistent with the idea that threats deployed in self-interested ways can backfire. When they do it is probably because accepting unfair treatment is inconsistent with an individual's project to constitute herself as person who is not easily taken advantage of.

¹⁵ A large number of additional experiments have found positive effects of incentives imposed by the decision of the targets of the incentives rather than by the experimenter or by a principal (Cardenas (2005), Tyran and Feld (2006), Kroll, Cherry and Shogren (2007), Ertan, Page and Putterman (2009), Kosfeld, Okada and Riedl (2009), Mellizo, Carpenter and Matthews (2011), Sutter, Haigner and Kocher (2011)).

¹⁶ Galbiati and Vertova (2014) In my taxonomy their result is a case of categorical crowding in (See also Galbiati and Vertova (2008)).

¹⁷ This account is based on Martin and Ceballos (2004) and Riano (xxxx).