

“Machiavelli’s Mistake”
(pp. 1-12)

Samuel Bowles
Professor, Santa Fe Institute, and Professor of Economics,
University of Siena
Legal Theory Workshop
UCLA School of Law

3/8/2012, 5:00p.m.-6:45p.m., Room 1314

<http://www.law.ucla.edu/home/index.asp?page=817>

*Draft for UCLA Workshop. Please Don't Cite
Or Quote Without Permission.*

MACHIAVELLI'S MISTAKE:

WHY GOOD LAWS ARE NO SUBSTITUTE FOR GOOD CITIZENS

Samuel Bowles

Santa Fe Institute and University of Siena

Forthcoming Yale University Press, 2011

... 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 arrange 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 prima deca di Tito Livio*. 1513-1518 (translation: SB)

[Dedication]

Parts of this book were given as the Castle Lectures in Yale's Program in Ethics, Politics and Economics, delivered by Samuel Bowles at Yale University in 2010. The Castle Lectures were endowed by Mr. John K. Castle. They honor his ancestor, the Reverend James Pierpont, one of Yale's original founders. Given by established public figures, 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.

TABLE OF CONTENTS

PREFACE

I. MACHIAVELLI'S MISTAKE

II. MORAL SENTIMENTS AND MATERIAL INTERESTS

III IS LIBERAL SOCIETY A PARASITE ON TRADITION?

IV THE SOPHISTICATED LEGISLATOR'S DILEMMA

V CODA: GOOD GOVERNMENT AND THE SCIENCE OF HUMAN BEHAVIOR (to come)

APPENDICES (incomplete)

INDEX (to come)

WORKS CITED

PREFACE

Machiavelli's Mistake originated in a paper that never saw the light of day. It showed how better defined property rights, enhanced competition and other policies perfecting the conditions that economists have shown to be necessary for markets to function well could promote self interest and undermine the processes by which a society sustains a robust civic culture. Included among the cultural casualties of idealized markets were such workaday virtues essential to market functioning as keeping one's word and working hard even when nobody is looking. While the textbooks said that *Homo economicus* invented markets, I suggested that it might be the other way around. My claim was an economic analogue to idea that the Hobbesian state might produce Hobbesian man developed by Michael Taylor (1976)).

The paradoxical idea that “perfecting” markets might make them work less well was enthusiastically received in 1989 by my fellow members of the September Seminar, then meeting at the Philosophy Department of University College London. The adverse effects of explicit economic incentives on ethical motivation implied by the argument were quite plausible and widely thought to be true (except among economists). Firsthand experience told in its favor. I had started writing the paper shortly after a failed experiment in home economics: when I offered my teen age kids a price list for household chores as a substitute for a weekly allowance, they entirely ceased doing the housework that they had generously done in the absence of the incentives. Paying for blood, one heard, would discourage donations. But nonetheless I was far from convinced.

The ill-fated paper had been titled *Mandeville's Mistake*, honoring (albeit dubiously) the author of the *Fable of the Bees*, the early 18th century verses that held that virtue was dispensable, even pernicious for the social order. Bernard Mandeville's scandalous hive thrived on licentious greed; and when the bees turned virtuous, collapse and disorder ensued. Mandeville had set the stage for the most revolutionary idea yet proposed by an economist. Elevated consequences may follow from ordinary motives:

...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])

My concern in the paper had not been with the empirical implausibility of the conditions under which Adam Smith's invisible hand would promote economic efficiency, but rather with the cultural consequences that would follow if those conditions were to be more nearly approximated, as is the standard objective of economic policy makers working in the tradition that Smith and the other classical economists initiated.

I returned two decades later to these ideas for two reasons. First, behavioral experiments – a few of my own, but mostly those of others – provided hard evidence that (to use Smith's language) the “moral sentiments” are sometimes crowded out by policies and incentives that appeal to material interests. Paying for blood does indeed sometimes reduce donations

Second, as I reflected on the causes of human suffering and the possibilities for human flourishing I became increasingly convinced that policies that expected little of the citizen beyond his self interest, which could be harnessed to “promote an end that was no part of his intention” were not up to the challenge and might even be part of the problem. There has never been a social order in which this form of economism would have been sufficient; but as we will see in the next chapter, the inadequacy of this approach has grown over time

Dispensing with virtue thus looked like an increasingly bad idea. Yet there had been good prudential and other reasons for the classical economists' advocacy of policies that made few demands on the citizens' moral predisposition. How the moral sentiments and the material interests might work synergistically in the promotion of public ends remained uncharted territory. It was time, I thought, to return to the ideas in my discarded paper.

I decided to rename the project Machiavelli's mistake, shifting the *locus classicus* of the error two centuries earlier to the writer who first made explicit an idea that was only hinted in Mandeville's *Fable*. This was that to secure order the prince must address citizens as they are rather than as they might be, and that as a matter of prudence if not of fact, he should assume that they are self-interested. The mistake in the title is not a flaw in Machiavelli's writing but rather an

erroneous contemporary way of thinking – common among economists – that combines a professed indifference toward the nature of individual preferences with overconfidence in the ability of clever incentives to induce even the entirely amoral and self interested citizen to act in the public interest. Leo Strauss (1988):49 traced this way of thinking to the sixteenth century Florentine: “Economism is Machiavellianism come of age.” The new title was energetically resisted when I presented a synopsis of the book in Florence; my critics correctly pointing out that Machiavelli's ethics are far more nuanced and his view of human nature far less malign than the term Machiavellian usually implies (Benner (2009), Pocock (1975)). But this book is not a critique of Machiavelli, it is a reflection on contemporary ways of thought, the earliest hints of which are found in his work.

The pages that follow are based on my Castle Lectures at Yale University in 2010, where the critical commentary of Bryan Gersten, Phil Gorski, Laurie Santos, Stephen Smith, and Chris Udry resulted in many improvements. 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 Yochai Benkler, Erica Benner, the late Gerald Cohen, Joshua Cohen, Steven Durlauf, Simon Gächter, Josh Greene, Jonathan Haidt, Philippe van Parijs, John Roemer, Daria Roithmayr, Seana Shiffrin, Rebecca Saxe, Erik Olin Wright and Elisabeth Wood. My collaborators Sung-Ha Hwang and Sandra Polanía Reyes are virtual co-authors of parts of the manuscript and I am grateful to them for permission to use material from our joint work. Chapter III draws upon work published in *Philosophy and Public Affairs* (Winter 2011) and I thank them for permission for its inclusion in this work.

I would also like to thank the Behavioral Sciences Program of the Santa Fe Institute, the University of Siena, and the U.S. National Science Foundation for financial support.

Santa Fe, New Mexico
November, 2010

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) Early in the sixteenth century, Nicolò Machiavelli gave rather different advice: “Anyone who would order the laws of a republic 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 (1984):69-70). The task of government for Machiavelli was to 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 goodness of its citizens that makes a well governed city possible but rather our capacity to “order the laws.” (Benner (2009))

Machiavelli's advice that princes and legislators should distinguish between motives and consequences was the key insight of Bernard Mandeville's, lighthearted verses *The Fable of the Bees*, a text considered by some to be the founding work of classical economics. (Mandeville did not know 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 the *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 any 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 chearfully, to the Drudgery of their Station; Kings and Princes not excepted. (Mandeville (1988):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 (1988):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 to elevated ends.

It was left to Adam Smith to explain how this improbable alchemy might be accomplished. 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).

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 arguably the first text in what we now call public economics. In it Bentham laid out the public policy implications of Hume's maxim.

In the early 20th century Alfred Marshall and Arthur Pigou spelled out the microeconomic theory underpinning this approach, advocating taxes on industry for the

environmental damages it imposed on others, and subsidizing a firm's worker training activities that benefit other firms when workers change jobs. What came to be called optimal taxes and subsidies were those that recompensed an economic actor for the benefits that his actions conferred on other and made him liable for the costs of his actions borne by others, when these benefits and costs would not otherwise appear in the actor's private revenues and costs. Green taxes that “make the polluter pay” for environmental spillovers are an example. Optimal incentives of this type exactly implement Bentham's Duty and Interest principle: altering the material incentives under which the individual acts so as to align self interest with public objectives.

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).)

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 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*, famously 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.” 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.” Even Machiavelli's famously corrupt citizens were introduced as a prudent assumption for the prince – “it is said that all men are wicked” – not as evidence a malign human nature, an assumption that Machiavelli rejected as a matter of fact : “our

reasonings are about those peoples among whom corruption has not expanded very much and there is more of the good than of the spoiled” and “very rarely do men know how to be altogether wicked or altogether good.” (Machiavelli (1984) see also Benner (2009)).

Like Machiavelli, the classical economists did not assume a malign human nature. Instead they reasoned that when large numbers of strangers interact, ethical behavior 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. As a result, economists, political theorists and constitutional thinkers since the late 18th century have taken amoral and self-regarding *Homo economicus* as their fundamental assumption about behavior, and partly for this reason, have stressed competitive markets, well-defined property rights, as well as efficient, (and since the 20th century) democratically-accountable states 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.

The classical constitutional challenge posed by Bentham, Hume, Smith and others constitutes the Holy Grail that still motivates policy design: 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. In posing the challenge this way they correctly identified the source of the market failures that to this day provide the primary rationale for government interventions in the economy.

In the intervening years economists have considerably sharpened our understanding of what it means to that people “take adequate account of the effects of their actions on others” and why, when they do not, the resulting decentralized allocations will be inefficient. The result is a coherent guide for Machiavelli's prince, clarifying what it might mean 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 that are effected would evaluate them.

If the “others” were our kin, neighbors, or friends, our concern for their well-being or our desire to avoid social sanction might induce us to take account of the effects of our actions on them. Reflecting this fact, an important response to the constitutional challenge – one that long

predates the classical economists but that now seems utopian – is that caring for the well-being of others need not be confined to intimates but ought to be extended to all of those with whom one interacts. However, with the increasing scope of markets over the last half millennium, individuals have come to interact not with a few dozen, but with hundreds and indirectly with millions of strangers. And so, with the maturation of capitalism and growing influence of economic reasoning, the burden of good governance shifted from the task of cultivating civic virtue to the challenge of designing institutions that work tolerably well in its absence. Prices, not ethics would ensure that actors took account of the effects of their actions on others.

What economists call complete contracts ensure this result by assigning 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. Complete contracts thus accomplish exactly the result of the optimal taxes and subsidies advocated by Marshall and Pigou. If contracts were complete, the invisible hand would work: competition among self-interested individuals would implement outcomes that are efficient in the sense that implement an outcome such that there exists no other 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” to this effect, for which they were awarded a Nobel prize.

Less heralded but more important for our purpose is that the assumptions underlying the invisible hand theorem did more than dispense with the need for government intervention in the economy; they also dispensed with virtue. Another Nobel Laureate, James Buchanan, described his visit to a farm stand near 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 optimality of the results (at least under idealized conditions) made competitive exchange a special domain which suspended the substantive normative standards commonly applied to

relationships among citizens or family members. Generalizing Buchanan's observation, and the status of the market economy as a morality-free territory David Gauthier (1986):96 held that: “morality has no application to market interactions under the conditions of perfect competition.”

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 the case when contracts are not complete— you breath 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 termed an environmental spillover (or external dis-economy), that is a direct effect between economic actors that is not covered by the terms of market exchange. The result is a market failure. Bucolic examples like this are 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 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. 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 cannot be enforced if the borrower is broke (Bowles (2004).) Contracts are also incomplete (or non-existent) in team production processes 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 the conflict of interest between the two cannot be resolved by specifying the terms of a complete and enforceable contract. The defacto terms of the exchange are determined 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 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.

In all of these cases the complete contracts assumption of the invisible hand theorem is violated and as a result decentralized allocations implemented by competition among self interested economic actors will be inefficient. The challenge to the policy maker or constitution writer is to find a way to assign to each actor the entire benefits and costs (to themselves and to others) of his actions, thereby providing a surrogate for complete contracts. For example, assigning ownership of the land to the sharecropper (who would then own the entire crop) would accomplish this. Replacing sharecropping by a fixed rent that does not depend on how much is produced would do the same.

This remains the canonical model of policy-making in economics. Hume's maxim about knaves is beautifully illustrated by the theory of mechanism design (Laffont (2000), Maskin (1985), Hurwicz (1975)). This branch of economics seeks to determine the contracts, property rights and other social rules – in short, constitutions – that induce individuals with conventional self-regarding preferences acting in the absence of binding agreements to implement an outcome which is not sought by any of the individual participants, but which is socially valued. Green taxes and training subsidies of the kind advocated by Pigou and Marshall are example, but far more complex and ingenious schemes have been developed. For example, to deter shirking among production team members in cases where their work effort is not observable, pay each member the entire value of the output of the team, minus a constant sum. This seemingly bizarre formula 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.

This initial promise of this approach was that the incentives provided by a combination of market prices, governmental taxes and subsidies, and perhaps more complex mechanisms could implement desirable social outcomes based on individual utility maximization, and that this could be done irrespective of individual preferences. By “giving the invisible hand a helping hand” (as *The Economist* put it, struggling to defang the challenging mathematics of mechanism design), well designed publically provided incentives seemingly would dispense with virtue entirely as a foundation of good government and spared the policy maker or

constitution writer the liberal embarrassment of seeking to foster some values – a concern for the environment or future generations, for example – over others. The job description of the wise policy maker was no longer that of than Aristotle's Legislator tasked with uplifting the population. Thanks to mechanism design, 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.

But just as the “invisible hand theorem” did little to vindicate a laissez faire policy toward the economy and instead demonstrated just how implausible were the axioms required to demonstrate this result, modern mechanism design has clarified the limits of the “constitutions for knaves” approach. The capacity of the mechanism designer to correct market failures depends, we now know, on his access to extraordinary information that is typically private, and his ability to impose unlimited financial or other penalties. Even if one assumed that those designing public policy are paragons of the civic virtues which this approach would like to dispense with in the rest of the population, the scope of mechanism design would be quite limited in 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. I will return to the limits of mechanism design in Chapter IV.

Thus mechanism design is not likely – especially in a liberal society – to provide incentives that make ethical and other regarding preferences redundant. In a paper explaining the 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 the major markets of a modern economy – the markets for labor, credit, and knowledge – complete contracts are the exception. These markets 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. Moreover, 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, affect, and the new forms of wealth characteristic of what is called the “weightless economy.” The same conclusion follows from the fact that the challenges facing the worlds peoples – epidemic spread, climate change, personal security – increasingly arise from global and other large scale interactions that cannot adequately be governed by the incentives and sanctions provided by private contract and governmental fiat.

What I call Machiavelli's mistake, then, is the idea that public policy and economic organization should be designed as if citizens were entirely self interested. It is a mistake not because the as if clause is false – it was never intended as an empirical statement – but because the resulting policies may compromise ethical and other regarding motives that are essential to a well governed society.

II

MORAL SENTIMENTS AND MATERIAL INTERESTS

The good news – given that we cannot dispense with virtue – is that while *Homo economicus* is among the *dramatis personae* on the economic stage, he is not alone, and indeed is often seriously outnumbered.

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. In one-shot prisoners' dilemma experiments, the rate of cooperation is commonly between 40 and 60 percent, even though defecting on one's fellow prisoner maximizes one's payoffs irrespective of what the other does (Fehr and Fischbacher (2001)). Most subjects prefer the mutual cooperation outcome over the higher material payoff they would get by defecting on a cooperator. When they defect, it most often is because they cannot be sure what the other will do and hate allowing the other to exploit their trust. The experiments appear to predict what people do outside the lab: among Brazilian fishermen those who cooperate on shore in prisoners dilemma experiments adopt more environment friendly traps and nets when they take to their boats. I will consider the external validity of these experiments at some length below.

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 not because of a concern about the well-being of others because of a commitment to justice. Where it will not cause confusion I use the omnibus term social preferences and sometimes just “values” or “virtues” to include all of the above, the term encompassing both ethical commitments and other-regarding preferences. (Experimental evidence on the nature of social preferences is surveyed in Bowles and Gintis (2011))

The fact that *Homo economicus* has company would not discomfit the classical economists. We have seen that they did not ignore moral behavior; nor did they imagine than an economy could function well in its absence. Least of all did they assume that as a matter of fact, individuals are the amoral maximizers enshrined in the later concept of *Homo economicus*. Instead the classical economists assumed that moral behavior would be unaffected by incentive-based policies designed to harness self-interest. Because it is so often implicit, it may help to identify what may be its first explicit statement, by John Stuart Mill (1844): 97

[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.

Political economy could study the effects of incentives that appeal to this wealth maximizing side of individuals without reference to the other motives that Mill took to be beyond the purview of the discipline. Incentives and morals were thus separable in the mathematical sense of the term: the effects of the one did not depend on the level of the other. The classical economists'

Homo economicus was not so much amoral and self interested as schizophrenic.

As a result of this implicit separability assumption they failed to take account of the conditions under which social preferences would flourish and favorably affect societal outcomes and how harnessing self-interest to the public good might attenuate civic virtue. The contrast between the economic approach and that of Aristotle and those who followed him does not arise because of a difference in assumptions of innate human goodness: Hume for example endorsed the idea while Bentham rejected it. Instead, the contrast concerns whether, like Aristotle's Legislator, policy makers today should be concerned about the effects of public policy and law on moral and other regarding motivations.

The bad news is that the classical separability assumption is not generally true. Material incentives often crowd out common moral sentiments, dampening and even reversing their intended effects. Sometimes crowding in also occurs, incentives seemingly activating rather than compromising social preferences. The classical separability assumption – that incentives and social preferences are additive in implementing desirable outcomes – is frequently violated both in the lab and in the street. Thus policies designed for a citizen seen “solely as a being who desires to possess wealth” are less effective than expected counter-productive when incentives crowd out morals, or even counter-productive.

The hardest evidence for this failure of the separability hypothesis is from experiments, but here are some real world examples.

On December 1, 2001 the Boston Fire Department terminated its policy of unlimited paid sick days, replacing it with a 15-day sick day limit; pay would be docked for firemen exceeding the limit. 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 all of their holiday bonus checks (Belkin (2002)). The firemen were unimpressed: the year following they claimed 13431 sick days; up from 6432 the previous year (Greenberger (2003)). Many of the firemen, apparently angered by the new system, abused it, or abandoned their previous ethic of serving the public even when injured or not feeling well

Representative samples of Jewish West Bank settlers in 2005, Palestinian refugees in 2005, and Palestinian students in 2006 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. Some Iranians now regard their country's nuclear program as "sacred" with a strong taboo against material compensation as an inducement to compromise with those urging its restriction.

A similar reaction may explain the reaction of Swiss citizens in a survey of their willingness to accept an environmental hazard: when offered of compensation their resistance to the local siting 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)).

Incentives backfire for reasons other than their adverse effect on other social preferences (Seabright (2009).) For a person with an income target, increased monetary incentives may allow target attainment with less effort. Camerer, Babcock, Loewenstein, *et al.* (1997) suggest that this may explain why New York City taxi drivers work fewer hours when they are making more per hour. But in dozens of experiments incentives have had unexpected, often perverse, effects, apparently because appeals to the material interests affect the moral sentiments. These interactions are far from simple: Aristotle's Legislator will need a little economic model to understand the problem of non-separability.

COMPLEMENTS AND SUBSTITUTES

Consider an individual who may bear a cost 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. Taking the action may be encouraged by a subsidy or other economic incentive. In what follows I will use the term incentive (without the adjectives explicit, economic 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 call them "values" here) that may motivate taking actions that benefit others even at a cost to oneself.

Where separability does not hold there are interaction effects between values and explicit incentives such that the behavioral effects of values may be influenced (positively or negatively) by the use of explicit incentives. To see how, assume that for a given individual the extent the action (denoted by a) and both explicit incentives (s) and the intensity of values (v) can be represented by single numbers. Non-separability occurs when the presence or extent of the incentive affects the intensity of the values so that $v = v(s)$. Then we describe the interrelationships of incentives and values by a function indicating the individual's choice of an action: $a = \mu(s, v(s))$, where all other influences on the action of the individual are taken to be exogenous, and ignored. Figure 2.1 illustrates the relationship among these variables.

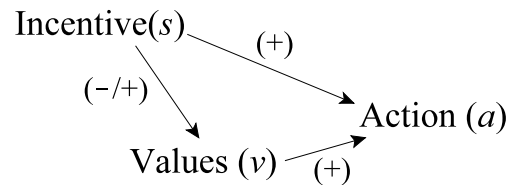


Figure 2.1: Incentives, Values and Actions: How separability fails.

An increase in v contributes to the public spirited action (holding incentives constant), as do incentives (holding values constant). The problem for the Legislator is that incentives may affect values positively or negatively. 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.

Where $\Delta^T > \Delta^D$ then incentives and social preferences are synergistic and are termed complements (meaning that the effect of the one is greater the higher is the level of the other). Where the reverse is true the two influences on the action are substitutes (or are said to exhibit

“negative synergy” or “crowding out”). Where the indirect effect is negative and sufficient in magnitude to offset the direct effect of the incentive so that $\Delta^T < 0$, we have the much cited case of incentives that “backfire”, here termed “strong crowding out.” Table 2.1 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, only that it be less than would be the case if additivity held.

$\Delta^T = \Delta^D$	Separability, additivity
$\Delta^T > \Delta^D$	Complementarity, synergy, super-modularity, crowding in
$\Delta^T < \Delta^D$	Substitutability, negative synergy, sub-modularity, crowding out
$\Delta^T < 0$	Strong crowding out

Table 2.1. 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

To design adequate policies the Legislator will need to know a lot more about the arrow from Incentive to Values, about its causes and nature. There are two quite different reasons why incentives may compromise values.

In the first case, as we will see, incentives provide cues to appropriate behavior or information about the intentions of the individual deploying the incentives or his beliefs about the target of the incentive. When this occurs, we say that the cause of non separability is the state-dependence of preferences (psychologists would term them situation-dependent (Ross and Nisbett (1991)). The incentives are part of the state (or situation) so that the values brought to bear on the action vary with the kinds of incentive that are present if any.

But incentives may affect values in a more durable way, altering the process by which we take on new values through learning from parents, other elders, or peers (Bisin and Verdier (2010), Boyd and Richerson (1985), Cavalli-Sforza and Feldman (1981), Bowles (1998)). In this second case we say that preferences are endogenous. The key difference from state-dependent preferences is the long term persistence of the effect of the incentive on preferences, which endures even if the incentive is altered or withdrawn. The durability of endogenous preference

change arises because the updating process on which cultural transmission is based typically occurs during youth and its effect persist over decades if not entire lifetimes.

There are also two quite different kinds of effect of incentives on values. Sometimes the very presence of an incentive alters an individual's evaluation of an action; sometimes the effect depends on the extent of the incentives. Non-separability may thus be either categorical (the presence of incentives affecting values independently of their level) or marginal (the effect of incentives on values depending continuously on the extent of the former) or a combination of the two. Separability and its violations are illustrated in Figure 2. 2.

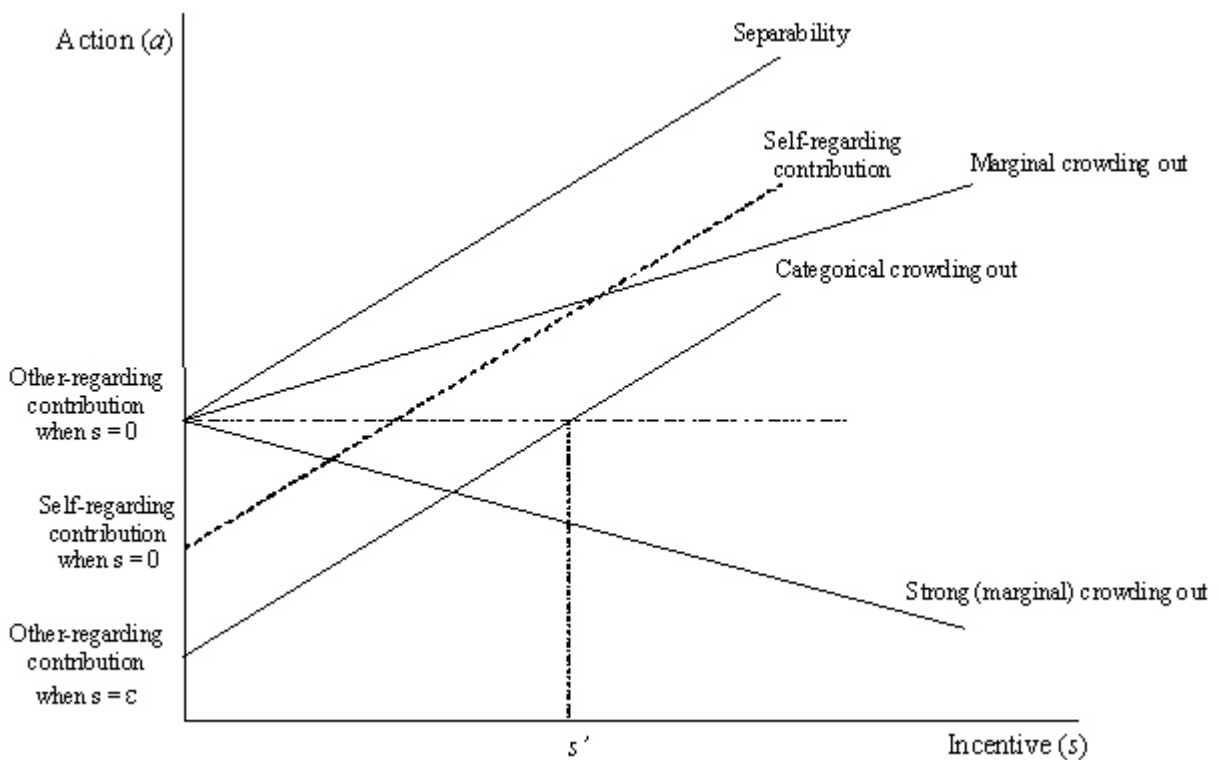


Figure 2.2. Citizen's contribution to the public good (a^*) under non-separability of incentives and values. Under separability (top line) categorical and marginal incentive effects are additive. Under strong crowding out the use of the incentive is counterproductive; this holds for all levels of s under the 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.

An experiment allows an estimate of both categorical and marginal crowding out.

Irlenbusch and Ruchala (2008) implemented a Public Goods experiment with German student subjects. The Public Goods game is an n-player prisoners' dilemma thought to capture the structure of many so called social dilemmas – payment of taxes, participating in political activities, reducing one's environmental impact – in which individual and group interests conflict. The n players are each provided by the experimenter with a sum of money called their “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 receive $2/5$ which clearly does not justify foregoing the 1; yet if everyone contributed 1, then each would receive 2).

In the Irlenbusch and Ruchala experiment, the game was modified in two ways. First, the costs and benefits of contributing were such that a payoff-maximizing subject would make a positive contribution, but much less than the amount that would maximize the total payoffs were all group members to do the same. Second, subjects faced three conditions: no incentives to contribute (as above) and a bonus given to the highest contributing individual that was either high or low (results are shown in Figure 2.3). In the no-incentive case contributions averaged 48 percent above the Nash equilibrium (25) that would have occurred if the participants had been motivated only by the material rewards of the public project. Contributions in the low-bonus case were not significantly different from the no-bonus treatment. In the high-bonus case significantly higher contributions occurred, but the amount contributed barely (and insignificantly) differed from that predicted for self-regarding subjects.

In Figure 2.3 I use the observed behavior in the high and low bonus case to estimate the marginal effect of the bonus, finding 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. 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 observed contributions (37.04) minus the predicted contributions had an arbitrarily small incentive been in effect (the vertical intercept of the observed line in figure 2.3), or 34.56. The incentive thus categorically crowded out 21 percent of the effect of social preferences (measured by the excess in contribution levels above Nash equilibrium for self interested subjects, 12.04.)

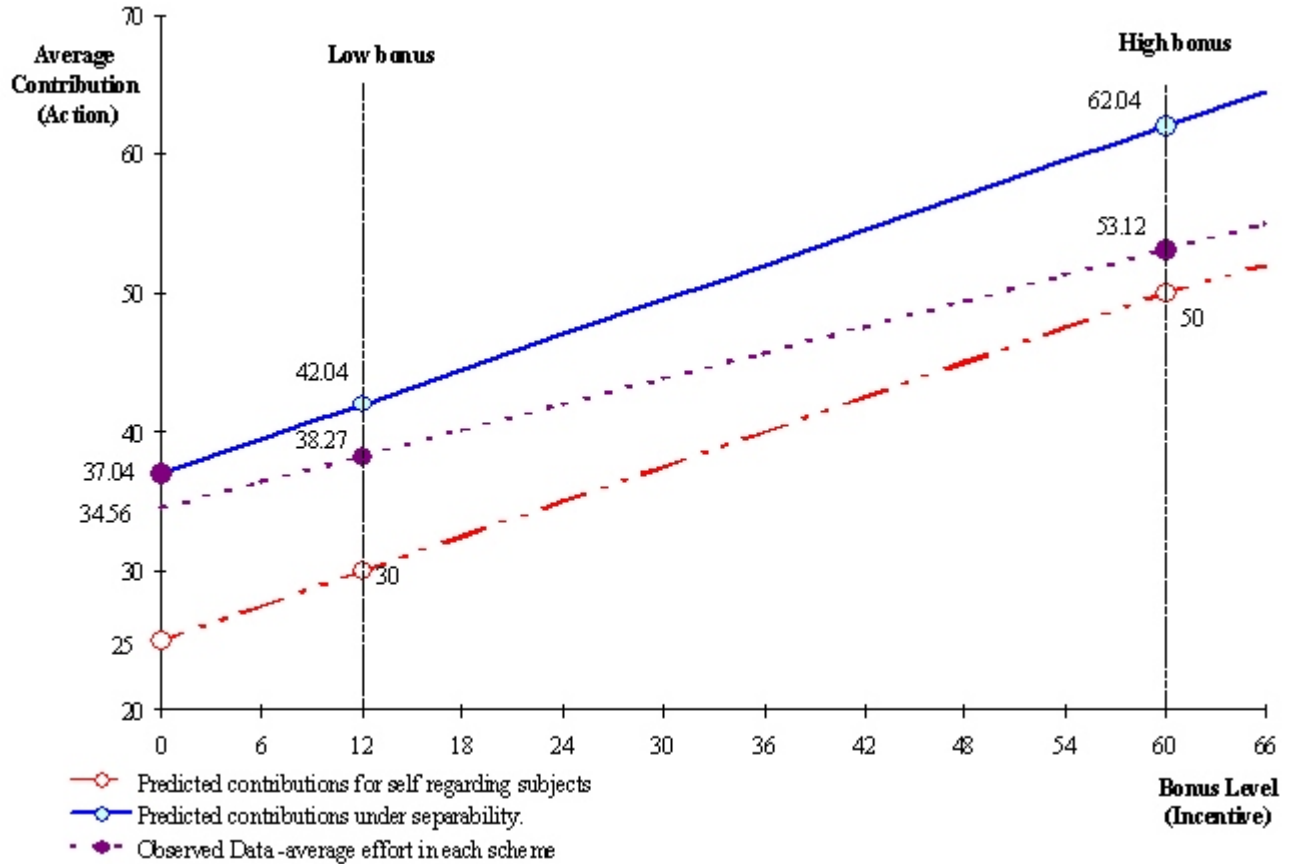


Figure 2.3. Categorical and marginal crowding out From Irlenbusch and Ruchala (2008) . Source: see text. The experimental design is an adapted Voluntary Contribution Mechanism game comparing two team-based compensation schemes without and with a relative reward (or bonus) for the highest contributor in the team. The bonus is self-funded (each member pays one-fourth of the bonus). Each subject simultaneously decides an effort level from the interval [0, 120].

Categorical crowding out is also evident in three experiments by Heyman and Ariely (2004). 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). Using these data as I did in the Irlenbusch and Ruchala study, I estimate that the mere presence of the incentive reduced the willingness to help by 27 percent (compared to the no incentive condition).

Another experiment that allows us to distinguish categorical and marginal crowding Cardenas, Stranlund, and Willis (2000), but here (as in some other experiments) we observe categorical crowding in. Cardenas implemented an experimental Common Pool Resource game a close relative of the Public Goods game that very similar in structure to the kind of commons problem faced by his subjects – rural Colombian rural eco-system users. In the absence of any explicit incentives, the subjects extracted 44 per cent less of the experimental “resource” than would have maximized their payoffs, 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 fine treatment was 55 percent less than the Nash equilibrium for self-interested 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 deviation from the self-interested Nash behavior is taken as the measure of social preferences). Interestingly, raising the fine from a low to a high level had virtually no effect. In this case the small fine did not work as an incentive, but rather in Cardenas’ view rather as a signal, one that alerted subjects to the public good nature of the interaction. These and similar cases of crowding in hold important lessons about how well-designed policies can make incentives and social preferences complements rather than substitutes, a problem to which we will return in chapter IV.

Many experiments provide evidence of strong crowding out (literally counter-productive incentives) but cannot distinguish the blunted marginal incentives of marginal crowding out from additivity or even crowding in. The reason is that they do not establish the response to incentives that would be observed under separability and thus are able to detect only strong crowding out (based on the sign of the effect) and not weak (based on the size of the effect). A common

misinterpretation of these experiments is that $\Delta^T > 0$, as was found in the Irlenbusch and Ruchala (2008) experiment or similar findings that an incentive had an effect in the intended direction, is evidence against crowding out

SEPARABILITY FAILS.

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 Polanía Reyes (2010)). In the three-quarters of the cases allowing the nature of non separability to be identified, incentives and social preferences are substitute rather than complements.

<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
Ultimatum		Proposer: unconditional generosity or belief in the fairmindedness 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

Tale 2.2 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.

In addition to the possibly adverse effect incentives on the learning of social preferences (the endogenous preference case) there are three plausible (state-dependent) interpretations of the failure of the separability assumption in these experiments: incentives may frame a decision problem and thereby suggest self interest as the appropriate behavior, or compromise the individual's sense of autonomy, or convey information affecting behavior. These processes – termed endogenous preferences, framing, over-determination, and the information content of incentives – often work jointly and sometimes with opposite effect.

The experimental games studied included the public goods, ultimatum, public goods with punishment, gift exchange and trust games, a brief account of the games appears in Table 2.2 (Table A1 in the appendix adapted from Camerer and Fehr (2004) 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.)

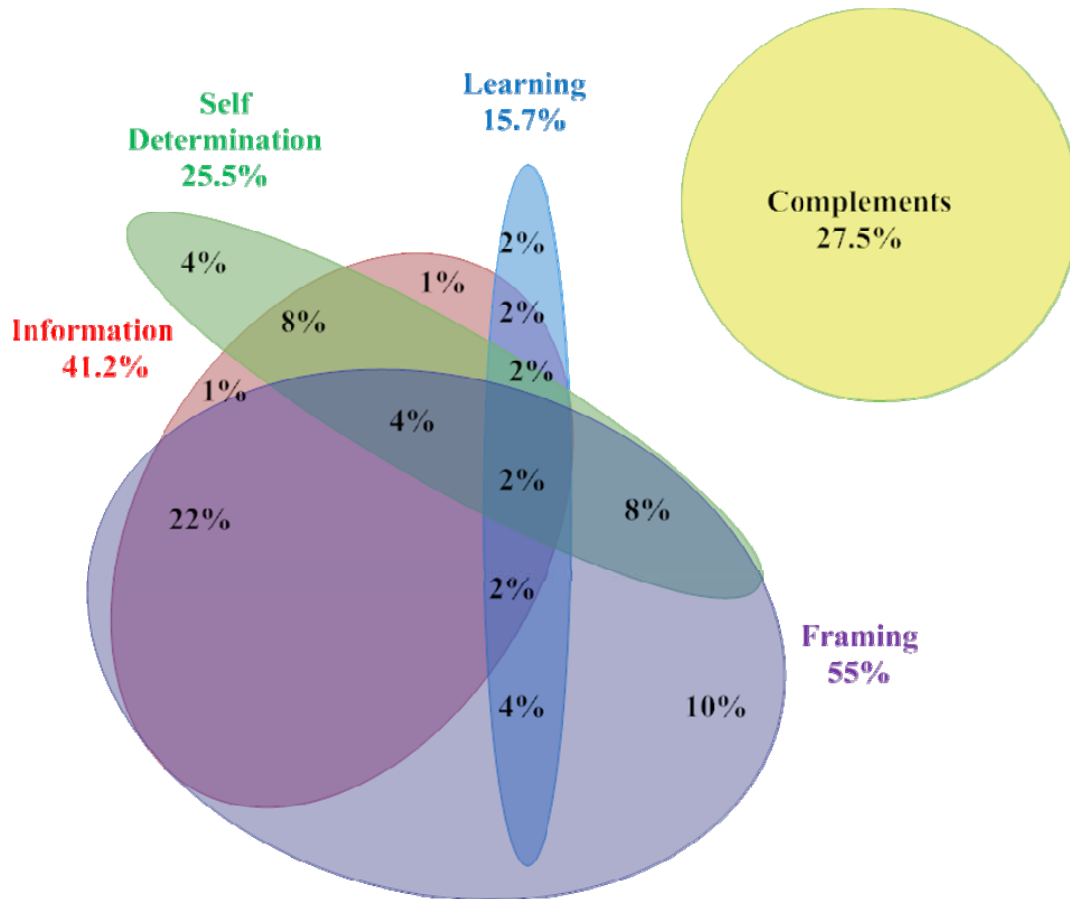
Thus, some results presented in the next 4 sections may be explained by more than one of these four mechanisms involved in crowding out, either because multiple mechanisms are at work, or because the experiment does not provide sufficient information to say which one accounts for the evidence of non-separability. As a preview, Figure 2.4 presents a summary of our findings, the size of the ellipses indicating the total number of studies that exhibit each of the four crowding out mechanisms in question, and the intersections giving the cases where multiple mechanisms may be involved. The crowding in cases are shown as complements.

Figure 2.4. Summary of experimental evidence on the four crowding out mechanisms and crowding in. The numbers indicate the percentage of the total of 51 studies that exhibit the mechanisms indicated.

Framing. Incentives are part of how a decision situation is represented and may signal appropriate behavior (Tversky and Kahneman (1981)), as seems to have been the case with the Colombian subjects in Cardenas' experiment. Framing is also at work when simply using market terminology (“exchange”) to describe an experiment reduces fair-minded behavior (Hoffman, McCabe, Shachat, *et al.* (1994)) or in which market-like competition “offers justifications for actions that in isolation would be unjustifiable” (Schotter, Weiss, and Zapater (1996)).

But the frame-shifting effects of incentives may occur in cases of third-party imposed fines or subsidies, too.

Figure 2.4. Summary of experimental evidence on the four crowding out mechanisms and crowding in. The numbers indicate the percentage of the total of 51 studies that exhibit the mechanisms indicated.



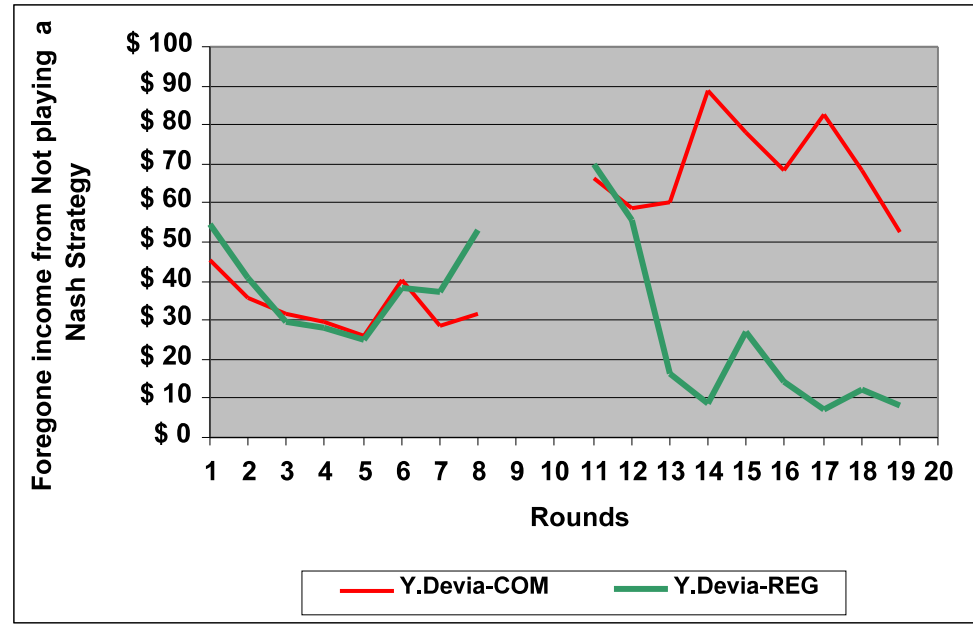
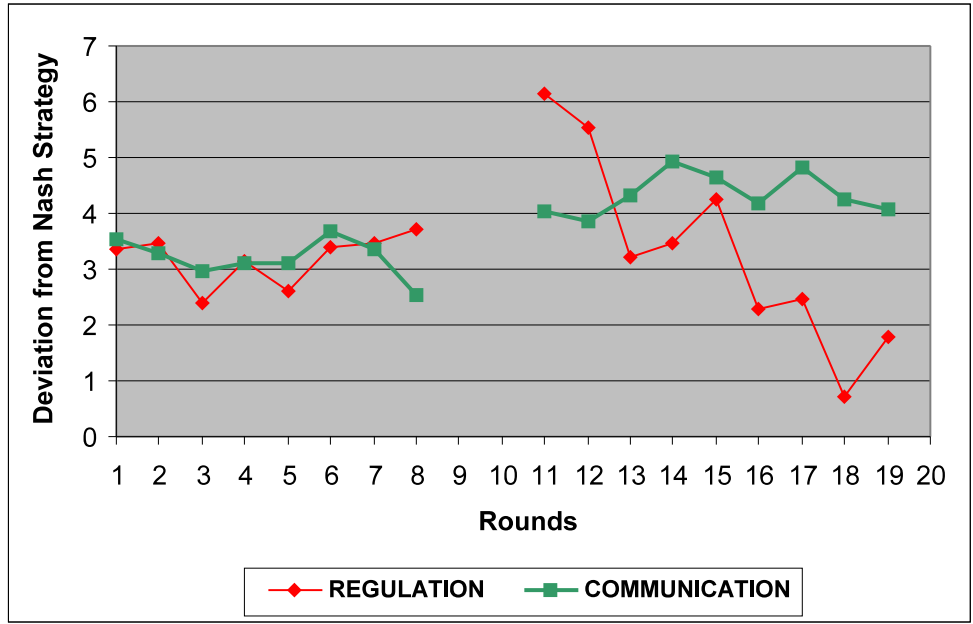


Figure 2.5. Effects of social preferences with communication or fines. Shown are two measures of the salience of social preferences. The top panel shows the average deviation (in 'months of exploitation of the forest') from the level that would have maximized the individual's material payoff, given what others in the group did. For example, in round 1 of stage I, both groups exploited the 'forest' a little more than three months less than would have maximized their individual payoffs. Implementing the social optimum would have required deviating about six and a half months from the self-interested response (not shown). The second measure is the income foregone by the individual by withdrawing less from the forest than would have maximized his income. Source: Cardenas, Stranlund, and Willis (2000) Used with permission.

Here is another example from rural Colombia by Cardenas and his co authors. Experimental subjects whose livelihoods depend on easily-depleted forest were asked to individually and anonymously choose how much to withdraw from a mutually beneficial common pool analogous to 'the forest' (Cardenas, Stranlund, and Willis (2000)). Payoffs were such that the level of withdrawal that maximized the gains of the group as a whole was substantially less than the level that maximized the gains of the individual acting singly. The experiment thus captured a common market failure in which self-interested actions by each would over-exploit a common pool resource and reduce the well-being of all.

Groups of subjects played 8 rounds of this game without communication, withdrawing on average amounts that were about midway between the individually self-interested and the group-beneficial levels (Figure 2.5). Their substantial deviation from the individually selfish level is a measure of the subjects' other-regarding or ethical values. The experimenters then changed the rules. In subsequent play for some groups face-to-face communication was allowed (but there was no way to make promises binding). Groups in this "communication" treatment modestly improved their performance, extracting less from the 'forest', thereby deviating more from self interest, and gaining higher benefits.

The other treatment precluded communication but simulated a "government regulation" Withdrawals were not to exceed the announced group-optimum level, and subjects would be monitored and fined for over-exploitation. The regulation reduced the level of withdrawal that would be chosen by an entirely selfish individual, but the expected fines were such that some overexploitation of the common pool remained the payoff maximizer's optimal choice. In this "regulation" treatment, subjects initially responded by restricting their withdrawals to close to the group optimum. But after two periods their behavior increasingly conformed to self-interest, and for the last three rounds their choices were almost entirely self-interested sacrificing only one-fifth as much individual payoff to protect the 'forest' as subjects in the communication treatment. The fine, while insufficient to enforce the social optimum, apparently all but extinguished the subjects' ethical predispositions that in the earlier rounds had induced them to withdraw much less than would maximize their own payoffs.

Self-determination. Where people derive pleasure from an action *per se* in the absence of

other rewards, the introduction of explicit incentives may 'over-justify' the activity and reduce the individual's sense of autonomy. 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). There is a substantial empirical literature on the psychology of intrinsic motivations (Deci, Koestner, and Ryan (1999), as well as non-experimental studies in economics (surveyed in Frey and Jegen (2001)). Recent experiments by economists are consistent with this view.

An experiment by Falk and Kosfeld (2006) explored the idea that ‘control aversion’ may be a reason why incentives degrade performance. Experimental agents in a role similar to an employee chose a level of ‘production’ that was costly to them 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 whether or not to impose it.) The principal could infer that a self-regarding agent would perform at the lower bound and thus imposition of the bound would maximize the principal’s payoffs.

But in the experiment, agents chose a lower level of production when the principal imposed the bound. Apparently anticipating this response, fewer than a third of the principals opted for its imposition in the moderate or low bound treatments. The 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. In post-play interviews, most agents agreed with the statement that the imposition of the lower bound was a signal of distrust. Fifty-seven percent of the agents were 'control averse' – they contributed on the average 73 percent more when the principal did not seek to control their behavior than when the bound was imposed. “Selfish” individuals – those providing the minimum possible under all conditions made up only 18 percent of the total; the rest gave more than dictated by self interest but were not influenced by whether the principal imposed the bound or not.

Control aversion and the desire for self-determination are not the only effects of the principal's seeking to bound the agent. The imposition of the minimum in this experiment gave the agents remarkably accurate information about the principals' beliefs concerning the agents: those who imposed the bound had substantially lower expectations of the agents. Their consequent attempt to control the agents' choices induced over half of the agents (in all three treatments) to contribute minimally, thereby affirming the principals' pessimism. This illustrates our fourth reason why the separability assumption may fail.

Incentives convey information. Principals select incentives based on their own objectives and their beliefs about how well the agent will perform his task under each possible incentive. Thus the incentives selected necessarily reveal information about the principal's preferences, the nature of the task, and his beliefs concerning the agent (Sliwka (2007), Benabou and Tirole (2003).) The incentives selected may indicate that the principal is seeking to profit at the expense of the agent, or that the principal believes the agent to be otherwise not committed to performing well, or that the job is onerous, or, as we have seen, that he does not trust the agent.

This predicament for the principal is nicely illustrated in a Trust Game experiment by Fehr and Rockenbach (2003). German students in the role of "investor" chose to transfer some of their endowment to the other player (which was then doubled by the experimenter). The "trustee" now with this doubled transfer, and knowing the investor's choice, could in turn provide a personally costly "back-transfer," returning a benefit to the investor. When the investor transferred money to the trustee, he also specified a desired level of the back-transfer. In one treatment, the experimenters implemented an incentive condition in which the investor 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 taken prior to the trustee's decision. There was also a "trust" condition in which no such incentives were available to the investor.

The use of the fine reduced return transfers, while renouncing the fine when it was available increased return transfers. Only one-third of the investors renounced the fine; their payoffs were 50 per cent greater than the investors who threatened use of the fines. The authors'

interpretation is that trusting elicited a positive reciprocal response that was extinguished by the threat of the fine. (See figure 2.6).

Figure 2.6. Average trustee's back-transfer by level of investor's transfer Larger investors' transfers are reciprocated by larger trustees' back-transfers, but the average back-transfer is least when the fine is imposed, and greatest when the fine was available to the investor, but was renounced. Source: Fehr and Rockenbach (2003). Used with permission.

The proximate causes of the negative impact of incentives in this case are suggested by evidence on the neural responses of the trustees in a Trust Game (Li, Xiao, Houser, *et al.* (2008)) As in the experiment of Fehr and Rockenbach the investor's threat of sanctions negatively affected back-transfers by trustees. To identify the proximate causes of this result, Li and his co-authors used functional magnetic resonance imaging (fMRI) to compare the 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. Threatened sanctions de-activated the Ventromedial Prefrontal Cortex (VMPFC), a brain area correlated with higher back-transfers 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 processes. The interpretation by Li and his coauthors is that the sanctions induced a "perception shift" favoring a more self-interested response.

ENDOGENOUS PREFERENCES.

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 or in a similar situation differing only in the absence of an incentive. But incentives may also durably change motivations. Preferences are endogenous in this sense if one's experiences– working under a particular kind of contract, for example, or getting paid for good grades – result in durable changes in motivations and hence a change in behavior that persists even after the experiences accounting for the change have ended.

Because learning is involved, an experiments of just a few hours duration is unlikely to

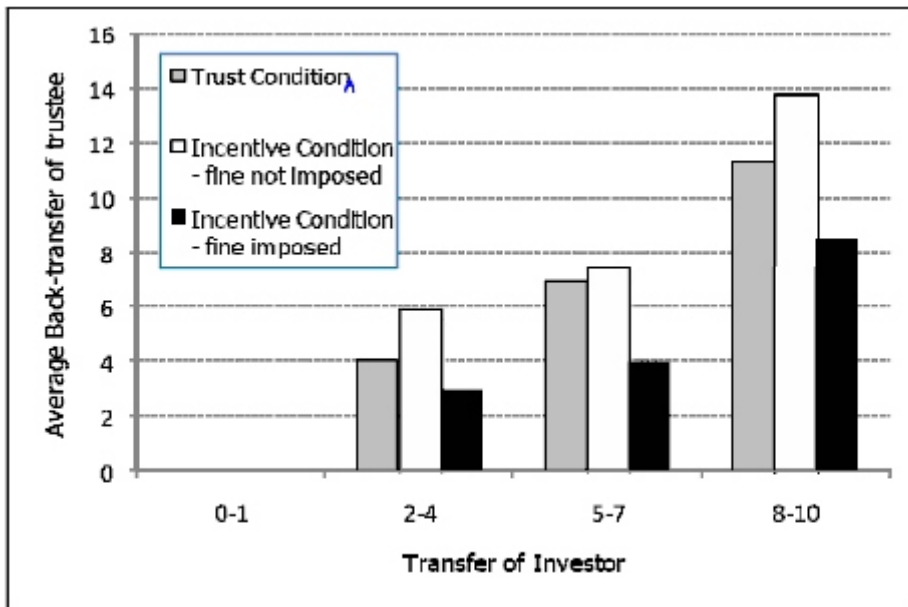


Figure 2. 6. Average trustee's back-transfer by level of investor's transfer Larger investors' transfers are reciprocated by larger trustees' back-transfers, but the average back-transfer is least when the fine is imposed, and greatest when the fine was available to the investor, but was renounced. Source: Fehr and Rockenbach, (2003). Used with permission.

uncover the causal mechanisms at work. This is because adopting new preferences is often a slow process more akin to acquiring an accent than to choosing an action in a game. The developmental processes involved typically include population-level effects such as conformism, schooling, religious instruction and other forms of socialization that are not readily captured in experiments. I will return the question of endogenous preferences in Chapter IV when I ask how a sophisticated Legislator might design optimal incentives in cases where incentives affect the evolution of preferences. Here I provide some modest empirical evidence about the crowding out by endogenous preferences process.

The idea that the social interactions occurring in markets and other institutional environments shape social norms and preferences has long been recognized (as did Edmund Burke when he lamented that the French Revolution had ushered in “the age of sophisters and economists.”) But this commonplace flies in the face of an economic canon that holds that on either empirical or prudential grounds, or for reasons for theoretical convenience or discipline one should follow Hobbes when he asked us to “consider men as if... sprung out of the earth, and suddenly, like mushrooms, come to full maturity without any kind of engagement to each other.” Prominent in this tradition is a paper by George Stigler and Gary Becker “*De Gustibus Non Est Disputandum*” probably the most influential single essay by economists on the topic of individual preferences: “one does not argue about tastes for the same reason that one does not argue about the Rocky Mountains – both are there, and will be there next year, too, and are the same to all men.”(Becker and Stigler (1977):76) They were repeating, in less poetic terms, Hobbes’ point about mushrooms. Becker would later write a book – *Accounting for Tastes* – in which he described how preferences might change under the influence of changing economic environments, taking up the topic pioneered by Herbert Gintis in his doctoral dissertation a generation earlier (Becker (1996), Gintis (1972))

Incentives change preferences because they affect both the range of alternative preferences to which one is exposed and the economic rewards and social status of those with preferences different from one's own (Bowles (2004)). For example, if the relevant incentives allow the selfish to exploit the civic-minded, then the latter are less likely to be copied. Other effects are less obvious: a competitive markets with complete contracts leaves little scope for

acting on ethical, reciprocal or generous preferences, even among those so inclined (Sobel (2007)). I show how this might work in more detail in chapter IV.

Consistent with the view that economies structured by differing incentives are likely to produce people with differing preference, historical, anthropological, social psychological and other data (some of it surveyed in Bowles (1998)) show that differing economic structures are associated with differing parental child-rearing values, personality traits rewarded by higher grades in school, and other developmental influences. Here is some of the evidence.

The effects of workplace organization on personality, including child rearing values, illustrate this process. Over a period of three decades Melvin Kohn and his collaborators have studied the relationship between one's position in the authority structure of one's 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 job structure rather than vice versa. His collaborative study of Japan, the U.S. and Poland (Kohn, Naoi, Schoenbach, *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

As the personality dimensions mentioned by Kohn are part of individuals' preferences explaining how they raise their children, what kind of leisure activities they engage in and the like, this is strong evidence that preferences are endogenous with respect to workplace organization.

Additional evidence comes from a study by 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. They 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, 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 not 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. These experiments show that incentives sometimes have durable effects that persist even in the absence of the incentives.

In the public goods experiment an incentive system designed by Joseph Falkinger and his co authors induced subjects to contribute almost exactly the amount predicted for a own-material-payoff-maximizing individual. It would be tempting to conclude from this evidence that the subjects were indeed payoff maximizers; but this would be mistaken: in the absence of the incentive subjects contributed significantly more than would have been optimal for a payoff maximizing individual. Even more interesting from the standpoint of the durable effect of incentives on preferences, 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, Gaechter, *et al.* (2000).)

This negative effect of the experience of incentives occurs, too, in an experimental Gift Exchange game implemented by Simon Gaechter and his co authors. The game is a sequential prisoners dilemma in which the “employer” chooses a wage to offer the “employee” who can either accept the wage or not. Agents that accept then select a level of “production” that is costly for the agent and beneficial for the principal. Selfish agents would obviously accept any positive wage and then produce nothing. In the two 'incentive' variants 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 failure to meet the target was penalized by a wage reduction, while in the “bonus” treatment meeting the target was rewarded by a wage increase. The standard setup without targets, bonuses or fines is called the 'trust' treatment for a principal would only offer a positive wage 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 Figure 2.7 shows, although the workers production declined over time. By contrast, incentives (of either type) sustained high levels of production over the entire first phase. 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 treatment. When subjects who had participated in ten rounds of the trust treatment played another ten rounds of the same game (with partners chosen randomly after each round), generous wages were common and were reciprocated, so that production levels remained high (with as before a modest decline over time). When subjects who had experienced either the bonus or fine treatment for the first ten rounds played the trust treatment (that is, in the absence of incentives) in the second set of ten runs, production fell far below the levels of those who had never experienced the incentive treatments. The difference is explained by the destruction of reciprocal motivations: conditional on a given wage being offered by the principal, the production level offered by the agent was substantially (and significantly, $p < 0.01$) lower among those who had experienced the incentives in the first period.

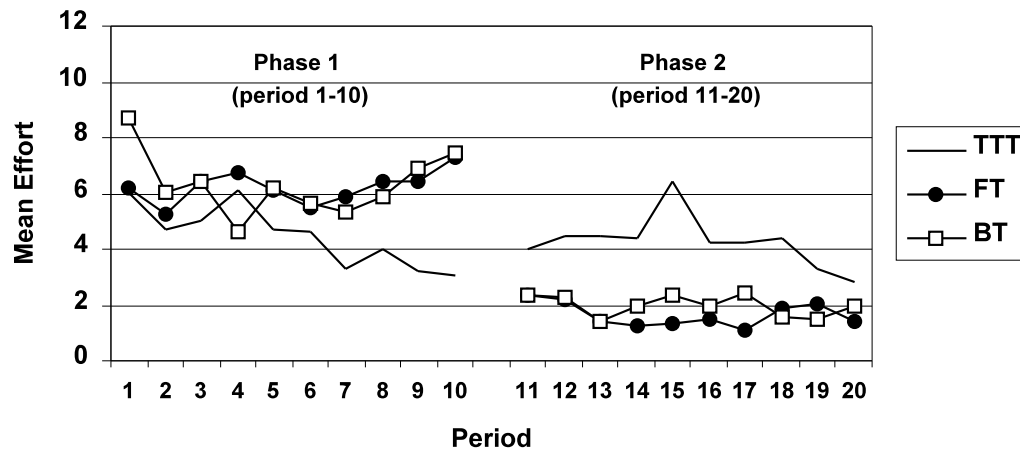


Figure 2.7 Effects of the prior experience of incentives on trust and reciprocity. Each treatment is a series of ten one shot interactions, each with a randomly selected partner. The difference in the second set of periods between the TT subjects and the FT or BT subjects is statistically significant. Source: Gaecher et al (xxxx)

CROWDING IN

Explicit incentives may enhance the salience of social preference by assuring people that those who conform to moral norms will not be exploited by their self-interested fellow citizens. Thus crowding in may also occur. Purely verbal messages of disapproval have a substantial positive effect on free riders' subsequent contributions (Barr (2001; Masclet, Noussair, Tucker, *et al.* (2003).) Incentives thus may recruit social preferences rather than dampening them. But other mechanisms are at work: social norms support the observance of traffic regulations, but these may unravel in the absence of state-imposed sanctions on flagrant violators. The rule of law and other institutional designs that limit the more extreme forms of anti-social behavior and facilitate mutually beneficial interactions on a large scale may enhance the salience of social preferences by assuring people that those who conform to moral norms will not be exploited by their self-interested fellow citizens.

This phenomenon may have been at work among the Hokkaido University subjects who cooperated more in a public goods experiment when assured that others (but not themselves) who did not cooperate would be punished (Shinada and Yamagishi (2007)) despite the fact that this had no effect on their own material incentives. They apparently wanted to be cooperative but wished even more to avoid being the sucker who is exploited by defectors

A different mechanism underlying crowding in was apparently at work in a public goods experiment by Galbiati and Vertova (2010). They found that the effect of a stated (non-binding) obligation to contribute a certain amount was greater when it was combined with a weak monetary incentive than when no incentives were offered. The monetary incentives 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.

CAN ONE GENERALIZE FROM EXPERIMENTAL EVIDENCE?

The experimental evidence for non-separability would not be very interesting if it did not reflect real-life behavior. Testing for separability in natural settings is difficult, but generalizing directly from experiments even for phenomena much simpler than separability is often unwarranted (Levitt and List (2007)). Consider, for example, the Dictator Game in which a one subject (the dictator) is assigned an endowment of money and asked to allocate some portion of it (including none) to a passive recipient. Typically more than 60% of subjects allocate a positive sum to the recipient, and the average given is about a fifth of the endowment. We would be sadly mistaken if we inferred from this that 60 percent of individuals would spontaneously transfer funds to an anonymous passer by, or that the same subjects would offer a fifth of the bills in their wallet to a homeless person asking for help. Subjects who reported that they had never given to a charity allocated 60 percent of their endowment to a named charity in a lab experiment (Benz and Meier (2006)).

Most individuals are strongly influenced by the cues of appropriate behavior offered by the situation in which an action is taken (Ross and Nisbett (1991)), and there is no reason to think that experiments are an exception to this context-dependent aspect of individual behavior. 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, interactions with other subjects are typically anonymous and without opportunities for ongoing face to face communication, unlike many social interactions. Third, subject pools may

be quite different from the real-world populations of interest, in part due to the process of recruitment and self selection. Finally, many of the experiments that provide evidence for the salience of social preferences are deliberately structured as strategic interactions like the Ultimatum Game that give scope for ethical or other-regarding behavior that may be absent in competitive markets and other important real world settings (Sobel (2007)). 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 other-regarding preferences subject to crowding out.

We can do more than speculate about these problems. Dean Karlan (2005) implemented a trust game among Peruvians participating in a micro-credit program; those who were least trustworthy (transferred less back to the “investor”) in the experiment were less likely to repay their real world loans. Jeff Carpenter and Erika Seki found that Japanese shrimp fishermen who contributed more in a public goods experiment were more likely to be members of 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. 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 while others cut large holes in their. Small trap holes are a form of defection, and just as in the public goods game it is the dominant strategy for a self-regarding individual. 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 lost opportunities that larger holes would entail. Fehr and Andreas Leibbrandt implemented both a public goods game and an experimental measure of impatience with the shrimpers. They found that both patience and cooperativeness in the public goods game predicted smaller trap holes (Fehr and Leibbrandt (2010))

While warranting caution in generalizing the details of experimental behavior to the real world, none of the external validity concerns is sufficient to dismiss the experimental evidence that social preferences are important behavioral motivations and that the salience of these preferences may be affected by explicit incentives. This is especially the case when experimental subjects exhibit motives such as reciprocity, generosity and trust that allow a consistent explanation of otherwise anomalous real world examples of crowding in or out, such as those mentioned at the outset.

CONCLUSION

These were precisely the motives that J.S. Mill wished to exclude when he narrowed the subject of political economy to the study of the individual “solely as a being who desires to possess wealth.” Evidence that social preferences are common, and that they underwrite mutually beneficial exchanges, but are often crowded out by incentives raises two distinct questions. Does the adverse effect of incentives on generosity, reciprocity, the work ethic and other motives essential to well functioning institutions, including markets, portend instability and dysfunction for any society in which explicit economic incentives are widely used? And, given that incentives play a central role in any market based economy and yet in many circumstances may prove disappointing or counter productive in their effects: what is the sophisticated Legislator to do?

I address these questions in the next two chapters.

III

IS LIBERAL SOCIETY A PARASITE ON TRADITION?

The corrosive effect of explicit economic incentives on the civic virtues and other social preferences may be conceded but held to be of little concern because, by comparison to other allocation mechanisms — gift exchange, consensual community governance, or central planning for example -- markets function tolerably well in their absence. Hayek (1948):11 for example held that the liberal market economy “is a system under which bad men can do least harm. It ...does not depend ...on our finding good men for running it, or on all men becoming better than they now are..” Thus, if market-based societies rely on incentives that reduce the supply of virtue, they may nonetheless be governed tolerably well if markets also reduce the demand for virtue. It is sometimes said that markets economize on virtue, meaning that “market-like arrangements reduce the need for compassion, patriotism, brotherly love, and cultural solidarity” (Schultze (1977):18).

Nonetheless the proper functioning of markets nonetheless depends critically on social and moral preferences (Arrow (1971)). For example, in the absence of a strong work ethic and feelings of reciprocity among employers and employees, an adequately functioning labor market would be impossible. If trust, truth-telling and other ethical behaviors were absent among borrowers and lenders, credit markets, likewise would collapse. If the “markets economize on virtue” reasoning is correct, the same is true with even greater force of other institutions, so that: “no social system can work ...in which everyone is ...guided by nothing except his own ...utilitarian ends” (Schumpeter (1950):448).

But if the explicit incentives that form the foundation of markets and public economics crowd out the virtues necessary for the functioning of a modern economy and polity, is the resulting society – liberal democratic capitalism – dynamically unstable in the very long run, in the sense that it does not sustain the individual values and motives necessary for its functioning?

The parasitic liberalism thesis, advanced in many variants over the past two centuries, holds that the proper functioning of markets and other institutions endorsed by liberals depends on family-based, religious and other traditional social norms that are endangered by these very institutions. Liberal society thus is said to fail Rawls’ test of stability: it does not “generate its own supportive moral attitudes.” (Rawls (1971):399).

If the self-interest based incentives that are intrinsic to markets also degrade the other-regarding and ethical motives on which the functioning of markets and other institutions depend, does this moral crowding out then lead eventually to economic dysfunction, instability and collapse of liberal society? An affirmative response – for example that “liberalism depends on virtues that it does not readily summon and which it may even stunt or stifle” (Berkowitz (1999):xiii) – was famously advanced by Daniel Bell (1976) in his *Cultural Contradictions of Capitalism* and earlier works (Bell (1973):48) “The historic justifications of bourgeois society -- in the realms of religion and character – are gone. ... The lack of a rooted moral belief system is the cultural contradiction of the society ...” Prominent exponents of related themes include Edmund Burke, Alexis de Tocqueville, Joseph Schumpeter, Frederick Hayek, and Jurgen Habermas.^a

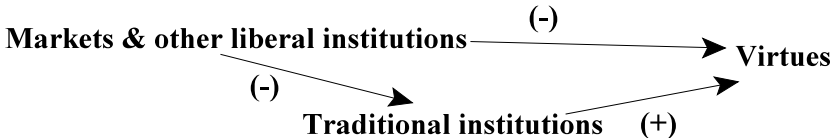


Figure 3.1. The causal structure of the parasitic liberalism thesis. “Virtues” represent the individual social norms and ethical commitments that are necessary for the proper functioning of markets and other liberal institutions. Arrows labeled (+) indicate a positive causal impact (variations in the source of the arrow result in variations in the same direction in the target of the arrow). Arrows labeled (-) are negative causal effects.

^a Burke (1791):64; Burke (1890[1790]):4-86; Tocqueville (1945):I 12; II 208, 334-337, 339; Hayek (1948); Polanyi (1957):76-77, 177; Habermas (1975):77, 79; Hirsch (1976):117-18; Schumpeter (1950) Some of the relevant passages appear in an appendix.

Simply put, the parasitic liberalism holds that markets crowd out the virtues necessary for the functioning of liberal institutions; and that this occurs both directly and by undermining religious, family, and other traditional institutions that would otherwise sustain these necessary virtues. Figure 1 illustrates this causal structure.

Surprisingly, as we will see, the “markets economize on virtue” response to the parasitic liberalism thesis not only fails to allay the concerns raised by these authors, it contributes to the instability of liberal institutions. The reason is that in tandem with moral crowding out, the comparative advantage of markets over other institutions in governing interactions among self-interested actors may set in motion a spiral of market-induced erosion of other regarding and ethical values, which in turn prompts greater reliance on markets, which in turn further erodes values, and so on.

The parasitic liberalism thesis is thus a claim about the mutual dependence of society-level institutions and individual preferences, and their joint dynamics in the very long run, one which ideally would be studied historically. To do this one would track over the centuries the scope and functioning of markets and other institutions along with various measures of the civic culture and individual values. Data, however, do not permit such a study. But we are able to clarify the causal structure of the thesis using evolutionary game theory, and test a key implication of the parasite thesis – that liberal societies would exhibit lesser levels of civic virtue – using recent experimental results from behavioral economics. That is what I will do in the pages that follow.

In the next section I model the joint dynamics of institutional and cultural change, showing the conditions under which the cultural dynamic of liberal society would confirm the parasitic liberalism thesis. Then (in section III) I present evidence that market-like incentives may crowd out ethical motivations, illustrating the parasitic liberalism thesis and the cultural and institutional processes by which it might work. The cross-cultural behavioral experiments presented in section IV, however, cast doubt on the thesis: liberal societies are distinctive in their civic cultures, exhibiting levels of generosity, fairmindedness, and civic involvement that distinguish them from non-liberal societies.

My interpretation of these seemingly conflicting experimental results (section V) is that the idealized view of tradition embodied in the parasitic liberalism thesis overlooks aspects of non-liberal social orders that are antithetical to a liberal civic culture. Thus while markets and other liberal institutions may indeed undermine traditional institutions as claimed, by attenuating familistic and other parochial norms and identities, this may enhance rather than erode the values necessary for a well functioning liberal order. And even if market incentives do crowd out values essential to the functioning of liberal institutions, these effects may be more than compensated by the cultural influence of non-market aspects of the liberal society such as the rule of law and social mobility, thereby sustaining the vibrant civic cultures observed in many liberal societies (section VI). The reader seeking schematic comparison of the parasitic liberalism thesis and my alternative explanation of the self-sustaining nature of liberal civic culture is in Figure 6.

When I refer to civic virtues I mean those social norms, ethical commitments and other-regarding preferences that facilitate the workings of the institutions advocated by liberals. Proponents of the parasitic liberalism thesis of course differ on which values are said to be essential in this regard, but the following are commonly held to be among the cultural foundations of a well functioning liberal order: willingness to help others at a cost to oneself (voluntarily paying taxes and contributing to public goods for example) and upholding social norms such as respect for private property, honesty, fair treatment, and political participation even when these do not enhance one's material benefits (Mill (1998):chapter 3; Rawls (1971):chapter 8).

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 attenuated ascriptive barriers to mobility (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.

II. PARASITIC LIBERALISM

Hume's often-cited "constitution for knaves" and Kant's "universal laws" for a "nation of devils" notwithstanding, liberal political theorists have never suggested that virtue is dispensable for the institutions that they endorsed (Kant (1970) :112-113 Hume (1898):116-117). For J. S. Mill, among the "causes and conditions of good government," the "principal of them ... is the qualities the human beings composing the society over which the government is exercised" (Mill (1919):11). The parasitic liberalism thesis thus does not hold that liberals have ignored the moral underpinnings of their favored social order, but rather that they have provided insufficient reason to think that these necessary virtues will flourish in a liberal environment.

Careful study of the works often said to provide such an account – those of Locke, Mill, and Rawls, for example – does not allay this concern. The great merit of these three authors is that they addressed the problem of the cultural dynamics that might underpin the institutions they advocated. But neither Locke's appeal to a gentlemanly home schooling (Locke (1968)), nor Mill's confidence that citizens in a liberal society will "spontaneously" adopt other-regarding preferences (Mill (1998):77), nor Rawls' belief that members of just associations will develop "bonds of friendship and trust" and through these "an attachment to the principles of justice" (Rawls (1971):461,470) provide reasons or evidence to think that these mechanisms entrusted with the perpetuation of liberal values would accomplish that end. In these and other works either the mechanism whereby a liberal culture could be sustained is not explained, or the reader is given no empirical evidence that the mechanism in question is up to the task. Rawls, perhaps, provides an exception. In addressing the danger of commitments to liberty being eclipsed by antagonistic and competitive status seeking, reasons that "social and economic differences" are "not likely to generate animosity" because "in a well-ordered society the need for status is met by the public recognition of just institutions, together with the full and diverse internal life of the many communities of interests that the equal liberties allow."(Rawls (1971)):1§82 He offers no evidence that this would be so. But evidence about the relationship between generalized trust and liberal political institutions presented in the last two sections below may be consistent with Rawls' claim.

The main conceptual challenge in investigating the claim that the liberal social order is not self-reproducing but rather depends on the vanishing cultural vestiges of a pre-liberal tradition, is that this requires an investigation of the joint dynamics of individual preferences and population-level institutions, one in which both institutions and individual preferences are endogenous, each providing conditions that influence the dynamics of the other. Modeling the evolution of institutions or of culture separately is difficult, and capturing the essentials of their joint evolution – the co-evolution of institutions and culture – is doubly so.

The two components of such a model must be a representation of the way that institutions affect the evolution of culture and the way cultures affect the evolution of institutions.

The first -- the idea that institutions affect culture -- is commonly illustrated by the role of families and religious and educational organizations in the socialization process; but it extends to institutions less transparently associated with the evolution of norms, tastes and the like, including economic institutions (Bowles (1998)). Supporting evidence comes from studies of parents' child-rearing values: parents value obedience more and independence less if at work they take rather than give orders (Kohn, Naoi, Schoenbach, *et al.* (1990)). We have also documented the influence of cooperative production (hunting large animals, for example, or the cooperative provision of local public goods) on values supporting cooperating in other settings (Gintis, Bowles, Boyd, *et al.* (2005).)

With respect to the second component, the effect of culture on institutions arises because the kinds of preferences that are prevalent in a population will influence the comparative advantage of particular institutions. By institutions I mean formal and informal formal rules governing social interactions, from the organization of families and firms to the structure of government. For example, where values such as reciprocity and fairness are prevalent, organizations based on partnerships may thrive, while in highly self-interested populations production may be carried out in organizations with close and punitive supervision.

Recently developed models of the co-evolution of cultures and institutions (Bowles (2004), Belloc and Bowles (2010)) allow a precise formalization of the parasitic liberalism thesis. I simplify by representing institutions by a measure of the extent to which markets (as opposed to other institutions) allocate resources (m), while representing preferences by a single-valued

measure of civic virtue (v), where the latter represents the prevalence of norms that contribute in essential ways to the functioning of liberal institutions. The objective of the model is to represent the mutual determination of m and v so as to characterize the pair or pairs $\{m, v\}$, such that both are unchanging when account is taken of the effects of each on the other. These so-called stationary pairs are termed cultural-institutional equilibria and they are subject to change only due to exogenous events. While obviously not representing the thinking of any particular variant of the parasitic liberalism thesis, the structure of the model captures two key ideas representing empirical claims about the nature of the two components mentioned above: institutions affect culture and culture affects institutions. (The following model is represented mathematically in the appendix).

The first, concerning the effect of institutions on culture is that markets crowd out virtues. This may occur by two mechanisms. In the first preferences are endogenous: social interactions typical of a society in which market institutions play a major role (and traditional institutions do not) favor a cultural learning process that is inimical to individuals acquiring and retaining the values needed for liberal institutions to function well. Proponents of the thesis have not specified the causal mechanisms by which this process might work, but it is not difficult to suggest a number of plausible candidates (Ben-Porath (1980)). One is that traditional institutions such as the patriarchal family and religious organizations are the main locus of socialization in the values necessary for the liberal social order. The other is that markets themselves (as well as market-like incentives used by public bodies) reward self-interest and penalize those with other-regarding or ethical values (Bowles (2004), Hwang and Bowles (2010)). An alternative mechanism whereby markets might crowd out virtues, occurs when the market framing of a decision situation makes the pursuit of individual self interest ethically permissible, and as markets become more extensive, this framing is extended to relations with family, neighbors, fellow citizens and work-mates. In this second mechanism preferences are situation-dependent rather than endogenous, and markets provide a frame that tends to be generalized.

The markets crowd out virtues relationship is illustrated in Panel A of Figure 2. In all four panels of this figure, each point a cultural-institutional state characterized by the indicated level of market extent (institutions) and virtue (culture). Each point on the downward sloping

“markets crowd out virtue” function gives the equilibrium level of virtue that results from the indicated level of market extent and some given extent of traditional institutions (the arrows indicate that from points above the line virtues tend to decline and conversely). For example the culture of a society with market extent m' (for the given level of tradition) would be v' . We label this function $v = v(m; \tau(m^{-1}))$ which says that virtue depends on both the extent of markets and of traditional institutions, where $\tau(m^{-1})$ represents the inverse relationship of the current extent of traditional institutions and markets in the past. I will return to the indirect effect of markets on virtues via the effect on traditional institutions presently.

The second key idea is an empirical claim about how culture affects institutions. It holds that because markets economize on virtue, they will be more widely used in societies in which virtue is less prevalent. Put differently, markets have a comparative advantage (over other allocation mechanisms) where virtue is scarce. This claim is not part of the parasitic liberalism thesis *per se* (though it is advanced independently by Hayek); but is necessary to capture the downward cultural-institutional spiral that some of its proponents suggest. This spiral occurs because the extent of the market in allocating resources is determined in a decentralized way by the choices of countless economic agents and it will vary with the cost advantages of markets relative to other institutions that may accomplish the same ends. Thus whether firms produce or purchase a particular component of the product they produce, for example, depends on the supervision and other costs of the direct command relations that distinguish firms from markets and that are entailed by production of the component, relative to the costs of search, bargaining over prices and other costs of using the market (Coase (1937).) The relative costs of the “build” versus “buy” options will depend on the ethical, self-interested and other motives of those involved.

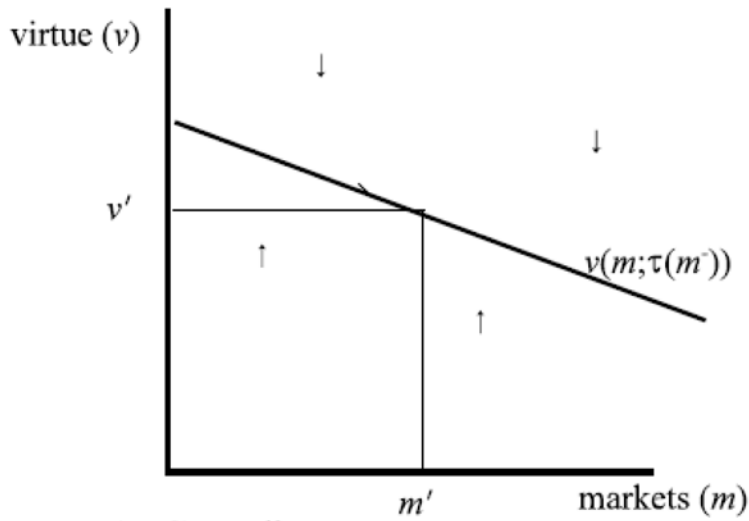
As a result, the level of virtues will influence the extent of the market; and because of the comparative advantage of markets in governing interactions among entirely self-interested individuals enjoyed by markets (relative to bureaucracies, families and other institutions), the relationship is inverse: higher levels of virtue being associated with a reduced extent of the market. This “markets economize on virtue” relationship is illustrated by the downward sloping line in Panel B of Figure 2. We label this function $m(v)$. Thus for any given level of virtue (say,

v) here is an equilibrium extent of the market (m) that is stationary, in the sense that no actor with the capacity to alter the extent of the market may benefit from doing so. As in Panel A, the arrows indicate the direction of change out of equilibrium (that is, points off of one or both of the functions), the extent of markets shrinking when they exceed the level indicated by the function and expanding when the reverse is true.

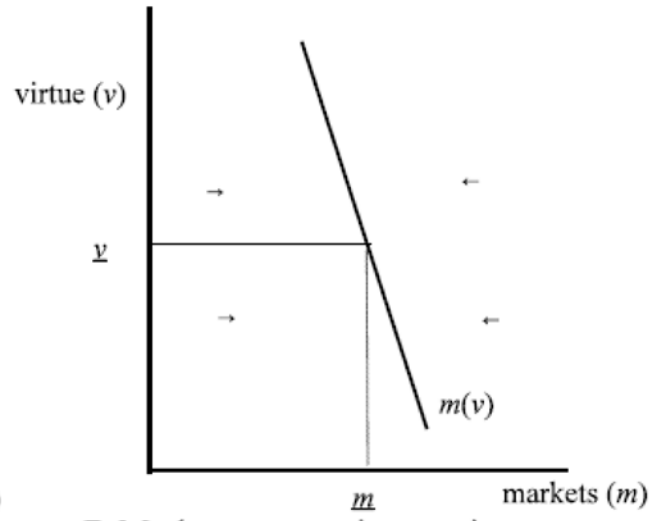
Because we want to know the conditions under which both culture and institutions will be stationary, we are interested in a state (that is, a $\{v, m\}$ pair) that is common to both functions, namely the intersection of the two lines representing relationships labeled “markets crowd out virtue” and “markets economize on virtue.” The joint influence of these two relationships shown in Panel C of Figure 2 gives us the equilibrium level of virtue and extent of the market, namely, the pair $\{v^*, m^*\}$ where these represent what is termed a temporary equilibrium, that is one defined for a given extent of traditional institutions.

The long-term effects of markets on tradition and thereby on virtue are shown by the dashed lines in the final panel of Figure 2. Recall that the line $v = v(m; \tau(m^-))$ – the crowding out function – says that the level of virtue in any period depends inversely on the current extent of the market as well as on traditional institutions, which in turn depend (also inversely) on the extent of the market in the past. This captures the indirect effect of markets on virtues: the cumulated effects of markets undermine traditional institutions, and as a result the temporary equilibrium level of virtue for any given extent of the market deteriorates over time, leading to a downward drift in the crowding out function. The result (in temporary equilibrium) is to increase the dependence on the market and diminish virtue, compromising institutional functioning and leading over time to the gradual displacement of the initial cultural-institutional temporary equilibrium (**a**) under the influence of the deleterious long term effects of markets on tradition. Note that a downward shift in the function of a given magnitude results in an even larger downward displacement of the cultural-institutional equilibrium due to the reciprocal effects of the markets economize on virtue and the resulting downward spiral.

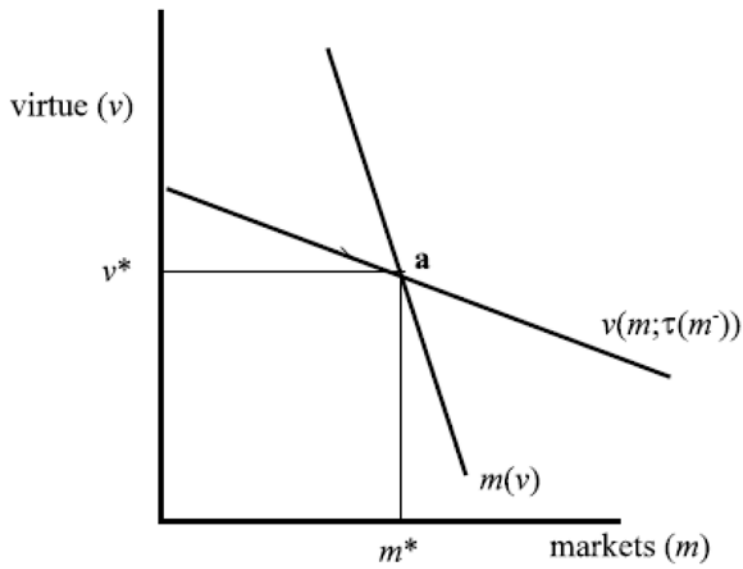
Figure 3.2: Parasitic liberalism: A temporary cultural-institutional equilibrium and the long-term effects of market-induced erosion of tradition (next page) Arrows indicate the direction of adjustment. Panel A: the effect of the extent of markets on virtue. Panel B. The effect



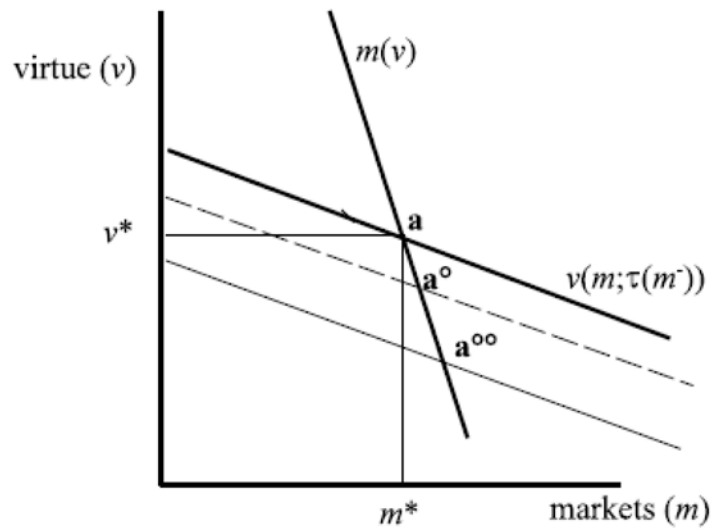
A: Crowding out



B: Markets economize on virtue



C: A temporary cultural-institutional equilibrium



D: The Long-run Erosion of tradition

of virtue on the extent of markets. Panel C: A temporary cultural-institutional equilibrium (for a given extent of traditional institutions). The state $\{v^*, m^*\}$ indicated by point **a** is a cultural-institutional equilibrium that is also stable (a chance displacement away from **a** is self-correcting, as the arrows show). Panel D. The long run effect of the induced demise of traditional institutions on the equilibrium levels of both virtue and market extent. Dashed lines indicate the effect of the current extent of the market in subsequent periods operating via the effect of markets on traditional institutions, displacing the cultural-institutional equilibrium to points a° , $a^{\circ\circ}$ and so on.

The dynamic illustrated by Panel D is a mathematical representation of the parasitic liberalism thesis, namely the existence of a configuration of virtues and market extent that erodes tradition leading to a displacement of the cultural-institutional equilibrium to one with a lesser levels of civic virtue and greater reliance on markets and characterized by a reduced level of economic output. There is some evidence in its favor.

III. EXPERIMENTAL EVIDENCE FOR THE PARASITIC LIBERALISM THESIS

Exploring the parasitic liberalism thesis empirically requires data on values across social systems. Measuring values is notoriously difficult, and the cross-cultural empirical study of civic virtues presents additional challenges. Differences across cultures in responses to widely-used survey self-reports confound differences in the responders' preferences with differences in self-presentation concerns or in the objective situation of the respondent, and moreover are sensitive to subtle differences in wording (which due to language differences are unavoidable in cross cultural research). Consider, for example, the standard survey question said to measure an individual's level of trust: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" An individual – one with some given amount of underlying trust of others -- would answer this question quite differently depending on where the individual lived.

However, one may infer values indirectly from behavior in experiments involving real material stakes and in which the decision structure facing individuals is identical across cultures. The fact that experimental subjects play anonymously is also valuable, as the civic virtues in question are not confined to behavior towards family members and friends, but must extend to unknown fellow citizens. The experimental evidence of the previous chapter is of course

consistent with the parasitic liberalism thesis, but it does not address the specific mechanisms said to be at work, namely the important role of pre-liberal institutions and the values they support in sustaining a liberal civic culture.

Direct evidence on the causes of crowding out is provided by a large team of anthropologists and economists who 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 U.S. undergraduates (Barr, Wallace, Ensminger, *et al.* (2009) , Henrich, Ensminger, McElreath, *et al.* (2009) .) In the Dictator Game an experimental subject (the dictator) is assigned an “endowment” of money by the experimenter and asked to allocate some, all, or none of it to a passive recipient. Then the game ends (the recipient taking home the dictator's offer and the dictator taking home the rest). 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 (also from an endowment provided by the experimenter) to impose a monetary fine on the dictator. The game then ends: the dictator keeps the part of the endowment that was not allocated to the respondent minus the fine imposed by the third party(if any), while the respondent keeps the amount allocated by the dictator. The third party “onlooker” keeps the initial endowment minus any amount spent fining the dictator.

The presence of a third party should induce dictators to adjust their allocations upwards (compared to the Dictator Game), the desire to avoid the material cost of the fine supplementing whatever generosity or fairmindedness motivated the dictator to share with the recipient in the absence of this incentive. Surprisingly, in only two of the 15 populations were the dictators' 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* ($t = -6.8$) 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.

Experimental design typically does not provide sufficient information to allow

investigation of the reasons why explicit incentives had the unintended effect. But in this case we can say something about the underlying causal mechanisms. Crowding out of specifically ethical motives is suggested by the following comparison. Pooling the 15 subject populations, in the standard Dictator Game, the dictator's adherence to one of the world religions (Islam or Christianity, including Russian Orthodoxy) raised allocations by 23 percent ($t = 3.5$), compared to those unaffiliated with a world religion. But in the Third Party Punishment Game with the very same individuals, this estimated "religion effect" was one tenth as large and was not significantly different from zero. In the Accra sample, the dictator's allocation in the standard Dictator Game was strongly correlated with his or her frequency of attendance at church or mosque; but this "religion effect" vanished in the Third Party Punishment Game. 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. Tellingly, the self-reported economic circumstances of the dictator (reflecting his or her own need for income) did not predict offers in the standard Dictator Game, but were very salient (and statistically significant) in the Third Party Punishment Game: needy dictators gave less. The presence of the economic incentive (the fine) apparently substituted economic interest for religious values.

While far from adequate, there is thus some empirical evidence consistent with the main causal claim of the parasitic liberalism thesis that market-like incentives may crowd out ethical motivations.

IV INDIVIDUALISM AND CIVIC VIRTUE

But the parasitic liberalism thesis does less well in a direct test: by most measures liberal societies appear to have more flourishing civic cultures. As the result of three large cross-cultural behavioral experiments we now have behavioral measures across a broad range of economic and political systems concerning individuals' cooperativeness, fair-mindedness and other predispositions commonly considered to among the civic virtues. 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 virtues flourish in liberal societies, though to varying degrees. The idea that pre-liberal tradition underpins the civic virtue essential to the functioning of liberal institutions finds little support in these data.

The cross cultural data are sufficient to reject the key inference of the parasite hypothesis, namely that liberal societies would exhibit a scarcity of civic virtues by comparison to non liberal societies. But they do not allow a test of the causal relationships accounting for the statistical patterns that I will presently report, for this would require cases in which differences in the extent of markets and other liberal institutions are exogenous with respect to the cultural norms under study, and it is difficult to imagine how such cases might be generated. It could well be, for example, that as the experiments presented above suggest, markets do crowd out virtue but that other liberal institutions more than compensate in sustaining a liberal civic culture. Indeed this is precisely the alternative model that I will propose.

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). In this game subjects are anonymously paired 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 rarely has been observed in literally hundreds of experiments in dozens of countries.

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

A second phase of this project studied primarily rural peoples in Africa, Oceania, and South America (Henrich, McElreath, Barr, *et al.* (2006), Henrich (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 might surprise a proponent of the parasitic liberalism thesis because it appears here that markets induce a kind of generosity by the proposer or anticipation of fairmindedness on the part of the respondent. But they are not inconsistent with the experimental evidence that I presented in the previous section in its support. 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) 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' fairmindedness in the Ultimatum Game by the minimum offer (fraction of the pie) that they reported (at the outset of the game) that would accept (this is also the amount the subject is willing to forgo in order not to accept an unfair offer.) This so called minimum acceptable offer (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 (the Hadza subjects offered a quarter, on average).

More comprehensive evidence and (as we will see in the next section) an idea that may explain the empirical challenges to the parasitic liberalism thesis come 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). Cultural differences among these subject pools may be somewhat attenuated, however, because (unlike the previously-mentioned field experiment studies) 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, an n-player prisoners' dilemma thought to capture the structure of many so called social dilemmas – payment of taxes, participating in political activities, reducing one's environmental impact – in which individual and group interests conflict. 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 rounds of the game (often ten)

This game provides information on three behavioral dispositions that may be considered to be civic virtues: 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 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), Dnipropetrovsk (Ukraine), and Samara (Russia).

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 ($r = 0.53$), 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 ($r = 0.38$). These and the statistics reported below are calculated from data in Herrmann, Thoni, and Gaechter (2008b). (Definitions of the measures and a table of their values are in the appendix.)

Individually costly voluntary contribution to a public good to be shared with strangers is surely a measure of the civic virtues upon which a liberal social order is said to depend. That these contributions are greater in nations characterized by individualism, rule of law, social equality and democracy is puzzling, but whatever its explanation, it is not consistent with the parasitic liberalism thesis. Understanding why these correlations occur will cast further doubt on

the thesis.

V. ORDER IN LIBERAL AND LINEAGE-SEGMENTED 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 3.) 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 mentioned above, shirkers who were punished responded by significantly increasing their contributions in subsequent periods. In only one of the 5 low contributing subject pools did shirkers respond positively to punishment (in four the response was not significantly different from zero.)

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 it not in Muscat and Samara. The 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. 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.

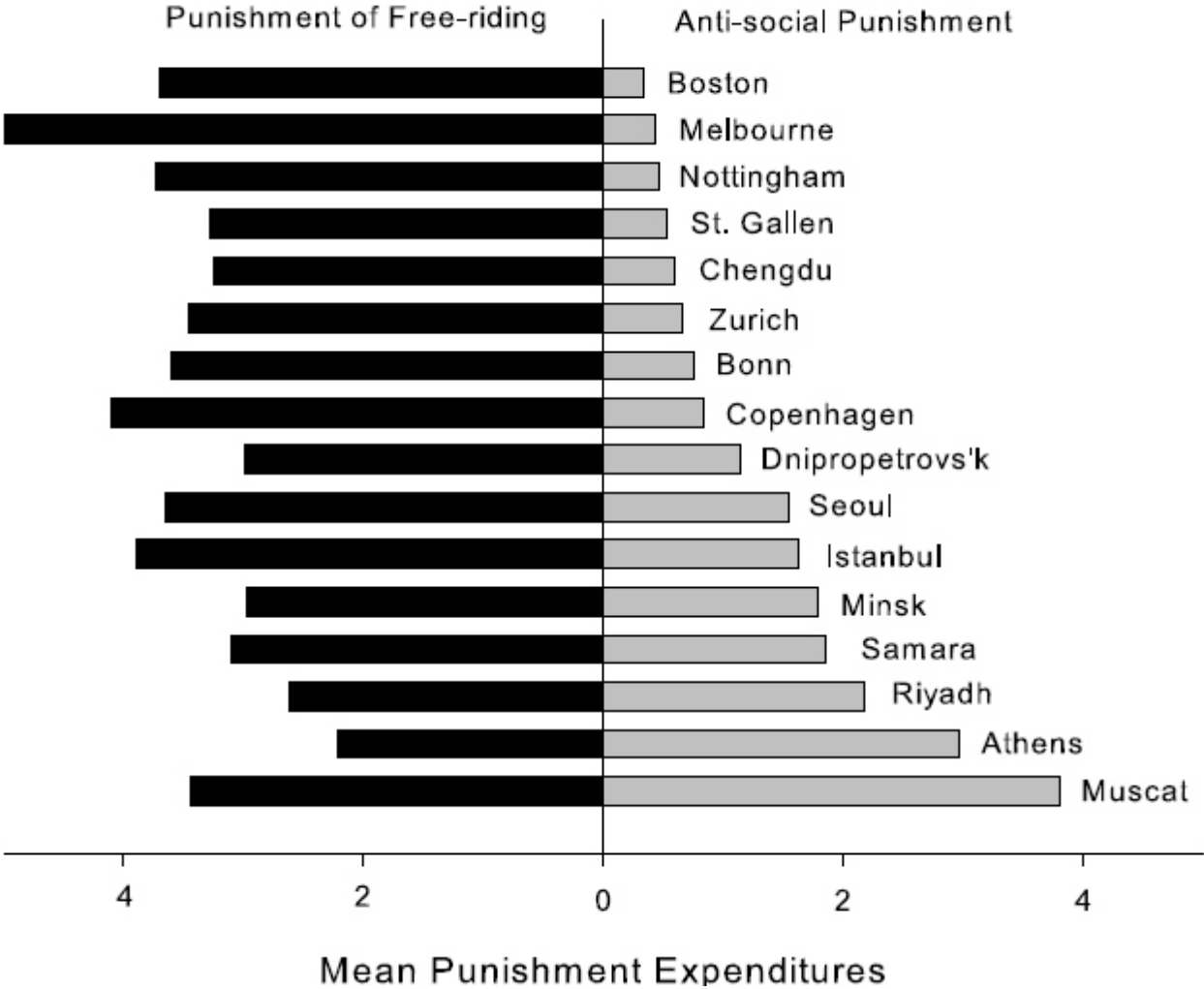


Figure 3.3 Anti-social punishment in a public goods game. Bars to the left of 0 indicate the extent of punishment of those who contributed less than the punisher. To the right are punishments targeted on those who contributed the same or more (darker shades) than the punisher. The very high level of anti-social punishment in Athens is remarkable, but not surprising in view of these correlations: by these measures Athens is very different from the top-contributing liberal locations and most similar to Seoul.

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. 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.

VI. A LIBERAL CIVIC CULTURE.

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 be 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.

A result is to reduce the value of those familial and parochial ties on which lineage-segments and other traditional identities are based, thereby creating a cultural environment favorable to the evolution of more universal norms that apply to strangers as well as family and clan, and may favor a greater interest in participating in democratic political activities, such as signing petitions or participating in demonstrations or boycotts. The strong inverse association between these indicators of democratic practice and 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 consistent with this view (Alesina and Giuliano (2009).)

Not surprisingly the emergence of the rule of law appears to be associated with a parallel shift from trust in kin and other particular individuals to generalized trust, consistent with 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. This process appears to have been at work in the 11th century Mediterranean trading system, which 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).) Because markets also flourish under these conditions (especially the protection of individual property under the rule of law), market-based societies may exhibit high levels of civic culture.

The relationship of markets to liberal civic culture may not be not entirely accidental, however, as a case can be made that the spread of markets did contribute to the emergence of

representative states with limited executive powers (which if the above argument is correct, favored the evolution of generalized trust), and to national systems of schooling-by-strangers, what Gellner termed exo-socialization (Gellner (1983)). Indeed Gellner argues convincingly 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 national standardization of language and culture facilitated occupational and geographical mobility, rendering individuals' income-earning assets less specific to place and craft and thereby complementing the other literal and defacto forms of insurance provided by liberal institutions (D'Antoni and Pagano (2002).)

The rule of law and other non-market aspects of liberal society that insure against worst-case outcomes not only undermine the value of familial and parochial loyalties; they 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. This is most probably the motivational mechanism underlying the few experiments in which material incentives and moral motives were complements rather than substitutes, the former enhancing the salience of the latter.

This crowding in effect of the rule of law 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 subjects' own material incentives (Shinada and Yamagishi (2007)). 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.

While this risk-reduction aspect of the liberal state effects the entire panoply of social interactions, I will illustrate it by the case of market exchange. 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 4. 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 Figure 4

	C	D
Cooperate	4,4	0,3
Defect	3,0	2,2

	C	D
Cooperate	4,4	1,2
Defect	2,1	2,2

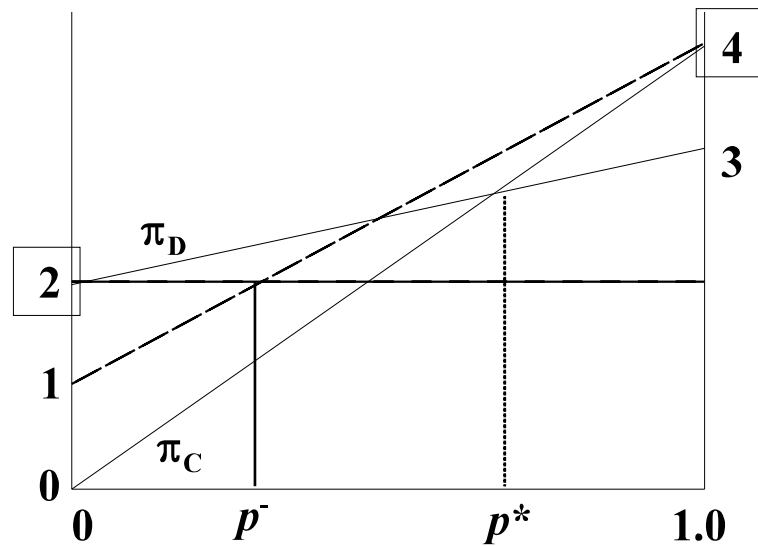


Figure 3.4. The rule of law and cooperative norms. Left panel: payoffs in the exchange game (upper without, lower with rule of law); right panel expected payoffs based on the type of one's partner (solid lines without, dashed lines with rule of law). The two Nash equilibria are mutual defect and mutual 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

The important feature of the payoff matrix is that a defector takes the goods of the cooperator, but at some cost, 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, due to the symmetry of the game, 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.

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 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.^b Thus the rule of law could promote the spread of trusting expectations and hence of trusting behavior in a population.

VII. CONCLUSION

If the interpretation offered here can be sustained by more adequate empirical investigation, the parasitic liberalism thesis fails not because it misunderstands the cultural consequences of markets or the tendency of liberal institutions to erode traditional institutions and cultures, but rather because it overrates the benign contribution of tradition to the moral underpinnings of liberal institutions, and underrates the contribution of the liberal state and other non-market aspects of liberal societies to the flourishing of these values.

If this reasoning and that of the previous sections is correct then we need to revise the model of parasitic liberalism in Figure 2. Instead of tradition being essential to liberal institutions yet endangered by their functioning, we need to account for the cultural and institutional effects of the non-market aspects of the liberal social order. By defining and enforcing property rights, the rule of law may increase the scope of markets at a given level of virtue (shifting the "markets economize on virtue" function to the right, as shown in Figure 5).

^b Rawls (1971) provides a different mechanism: "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 appeal to the would-be cooperator of pre-emptive defection as a risk minimizing strategy. But if public institutions are sufficient to deter defection directly, the result would be the same whatever the frequency of cooperative individuals in the population.

The effect, considered in isolation would be to displace the cultural-institutional equilibrium from a to a^- and thus to reduce the equilibrium level of virtue.

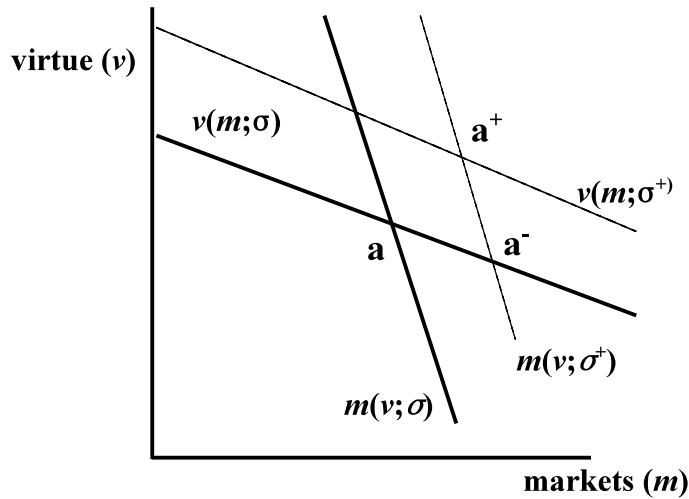
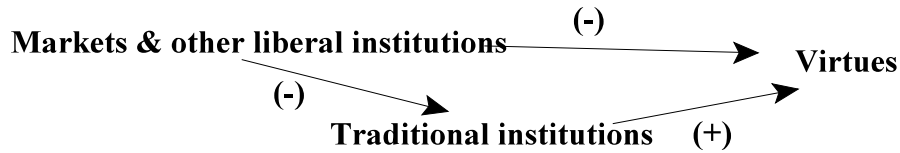


Figure 3.5. Markets, Liberal States, and Civic Virtue. The dashed lines indicate the effects of the liberal state (σ^+ , see text). In the case shown both v^* and m^* increase, but this need not be the case.

But if my interpretation of the evidence is correct, there are two compensating cultural effects. First, the rule of law, exo-socialization, cultural standardization and mobility enhance the level of virtue for any level of markets (shifting the “markets crowd out virtue” function upwards in Figure 5.) Second, the fact that traditional institutions are undermined may, on balance, contribute to rather than undermine the values on which the functioning of liberal institutions depend, thereby enhancing the upward shift in the same function. The result is displace the cultural-institutional equilibrium from a to a^+ and to increase either or both the equilibrium level of virtue and the extent of markets.

A schematic summary of both the parasitic liberalism thesis and the alternative theory of the liberal civic culture appears in Figure 6.

A. Parasitic liberalism



B. A liberal civic culture

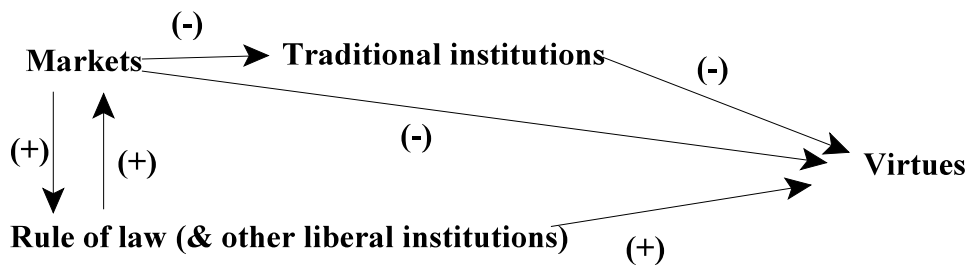


Figure 3.6. The causal structure of the parasitic liberalism thesis and an alternative.. The top panel , the variable definitions and the interpretation of the arrows are identical to Figure 3.1

This model, suitably extended, might account for the flourishing of civic culture found in some market-based societies. If true, on these grounds one would also expect that among liberal societies, the more market oriented societies (meaning those with greater values of m for any given v) and those with more limited reach of the other liberal institutions that may offset the market crowding out effects (the rule of law, exo-socialization, cultural standardization and mobility) will support lesser levels of civic virtue. While not an adequate test, this hypothesis finds some support in the substantial levels of generalized trust in Sweden compared to the U.S.(Rothstein and Uslaner (2005),Kumlin and Rothstein (2005)) and in the decline in measures of trust and civic engagement in the U.S. and contrast to continental Europe (Bartolini (2009) Sarracino (2009)). Indirect evidence consistent with the predicted inverse relationship between virtue and the extent of markets is found in the fact that the U.S., perhaps the most market-based of the advanced economies, also excels in the fraction of its labor force devoted to what Jayadev and I call guard labor, namely, that devoted to (or the consequence of) maintaining order (Jayadev and Bowles (2005) and Bowles and Jayadev (2007)).

While the parasitic liberalism thesis is thus not supported, the result is hardly an endorsement of laissez faire. Even under the idealized assumptions usually sufficient to vindicate unregulated markets (perfect competition and complete contracts in the exchange of all goods and services), the interaction of culture and institutions modeled here implies that once the joint dynamics of culture and institutions is taken into account, even these idealized conditions under which the Arrow-Debreu “invisible hand theorem” holds are insufficient to support the claim that

all competitive equilibria are efficient (in the Pareto sense). Remarkably, this is true even if the decentralized decisions that result in the extent of the market (described in section II above) perfectly reflect the relevant trade offs of using the market rather than alternative institutions so as to minimize the costs of doing business given the prevailing values in the society, in the manner described by Coase (1960) and Coase (1937). The reason is that in devising institutional solutions for the governance of economic transactions and other social interactions, individuals do not take account of the influence of their choices on society-wide long term cultural evolution. When markets crowd out the virtues that underpin effective governance they thus generate a cultural analogue to environmental spillovers. As a result, given a status quo of idealized “surplus maximizing” Coasean institutional choice, there will exist some restriction of the scope of markets that will increase economic output (as conventionally measured). (A proof of this proposition is in the appendix). Note that this “cultural market failure” places no normative weight whatsoever on virtues per se; the market failure occurs because markets under-provide virtues that contribute to economic output, much as the private economy under-provides public goods such as basic scientific knowledge and environmental amenities.

The shortcomings of the parasitic liberalism thesis did not arise because the classical economists' separability assumption was sustained; indeed, in exploring the thesis we found additional evidence that incentives and social preferences are less than additive, the former often crowding out the latter. The challenge, then, facing the sophisticated Legislator, who would “arrange a republic and order its laws” remains.

Thomas Schelling recalled the “exciting and stimulating times” he 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 real world economic incentives sometimes have surprisingly limited effects and may even be counterproductive. 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)) while 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).)

Since Richard Titmuss’ *The Gift Relationship: From Blood Donations to Social Policy* (1971), economists have been intrigued but for the most part unpersuaded by the claim that policies based on explicit economic incentives may be counter-productive when 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. (A Cornell 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)).

Moreover, where market failures required that the invisible hand get a helping hand from government policies, the then burgeoning field of mechanism design held out the promise that green taxes, training subsidies and other incentives could, along with markets, would implement an efficient allocation of resources even among entirely self-regarding citizens. Virtue was something that economists could safely ignore.

Research during intervening years has considerably sapped this optimism. Both the extent of incomplete contracts and the resulting market failures has grown, while the limited reach of incentives cleverly designed to induce self-regarding citizens to act in the common interest has become more evident. And in many settings, as we will see presently, private bargaining fails to implement efficient outcome even where contracts are complete. In part for this reason many contributors to the economic literature on optimal incentives – called mechanism design – have weakened the criterion for efficiency due to Pareto: that an allocation is efficient if there exists no feasible alternative such that at least one individual would be better off and non worse off. In place of this conventional Pareto-efficiency criterion they now favor “incentive-efficiency” which means simply the best that can be implemented given existing preferences. The example of market failure in the next section will dramatize the difference between these two criteria.

The recognition that the nature of citizens preferences do indeed matter coincided with the explosion of experimental evidence that other-regarding motives are common and that the separability assumption is routinely violated: incentives frequently crowd out civic virtue. The

result was to underline the policy maker's dilemma: the explicit economic incentives on which market allocations and the mechanism designer's interventions rely may reduce the salience of non-economic motives, crowding out ethical and other regarding preferences that would have motivated pro social actions in the absence of incentives..

This renewed attention to virtue has been further stimulated by the sheer enormity of the challenges now facing the world's populations including containing epidemic spread, addressing global climate change and governing the knowledge-based economy. In none of these are the complete contracts that underpin the invisible hand theorem even remotely feasible. As a result the conventional model of harnessing self interested motives to public ends now seems insufficient. Rousseau's seemingly prudent injunction that we should take “people as they are and laws as they might be” may now be a recipe for calamity if the people as they are would be entirely self interested (Rousseau (1762)).

Not surprisingly, 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?” and for women gives an affirmative answer for women (but not for men). A few economists, beginning with Albert Hirschman, turned to the design of public policy in a world in which 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).) Hirschman pointed out that economists 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

Economist, some of them at least, had not only rediscovered Titmuss; they had returned to Aristotle, asking what his sophisticated Legislator ought to do given this new knowledge about the possible anti-synergy between incentives and ethical motives. A sophisticated Legislator is one who knows that the separability assumption is likely to be violated. By contrast the naive Legislator assumes that incentives simply change the costs or benefits of some action

that the policy maker would like to influence. To the naive Legislator, a tax on smoking does not alter individual addiction or the commitment or social pressures to quit, or the social interactions among smokers and others; it simply makes the habit more expensive.

Not all economists, however, were on board. Even while accepting the evidence that separability often fails, an economist might still back the naive Legislator, reasoning that as long as properly designed incentives can implement socially desired outcomes irrespective of individual preferences, one need not worry about crowding out.

WHY NOT A CONSTITUTION FOR KNAVES?

When Hume proposed that under the right laws, the citizen's "avarice and ambition" could motivate him to "cooperate for the public good" the term "invisible hand" had not been coined. But Smith, and subsequent generations of economists showed how this could be accomplished through competitive markets, private bargaining, and, where these fail to support efficient outcomes, taxes, subsidies and other non-market incentives. These institutions are thought to implement efficient allocations of scarce resources while requiring neither mandatory participation in any activity (other than tax compliance and respect for property rights) nor that citizens be motivated by other-regarding or ethical preferences.

By the liberal trinity I mean the claim that under plausible conditions the following conditions may jointly obtain: a) individuals may have any preferences whatever, including entirely self-interested b) participation in economic activity is uncoerced, that is motivated by whatever these preferences happen to be, and c) the resulting allocations (possibly assisted by a benign but not omniscient Legislator's incentives) are Pareto efficient. Condition b) requires that any allocation must satisfy both of two conditions: incentive compatibility, requiring all individuals to do the best they can given their preferences, and a participation constraint requiring that participation in any activity be voluntary (preferable to withdrawal). Condition a) is an economist's rendition of what philosophers call liberal neutrality (the term is contested among philosophers, many of whom would consider the economists' version to be a caricature (Goodin and Reeve (1989))). Call the three elements of the trinity: liberal neutrality, liberty, and efficiency.

Demonstrating the conditions under which the trinity claim is true is thought to be the main contribution of economics to liberal theory, for if the reach of the claim is broad, this defeats objections that in order to promote economic efficiency a state might need to force citizens to engage in economic activities against their will or to adopt policies that favor some types of preferences over others. If true, this claim amounts to an elegant reconciliation of three liberal values: Pareto efficiency, voluntary participation, and neutrality among the ends that individuals wish to pursue. The claim underpins economists' "bring them on" approach to avarice, ambition and other individual motivations that some might consider to be character defects.

The claim is important for a less obvious reason. Were it true, then the fact that incentives sometimes compromise social preferences would be of less concern, at least from the standpoint of economic efficiency. The reason is that if efficient use of economic resources among entirely self interested citizens can be accomplished through well designed incentives, then the fact that social preferences may be crowded out as a result may be lamentable for other reasons, but would not constitute an impediment to economic efficiency.

The trilogy has been challenged in cases where the competitive markets or complete contracts assumptions of the invisible hand theorem does not obtain. But it is generally thought to be sustainable in the absence of familiar impediments to efficient bargaining (such as ill-defined property rights in cases of environmental spillovers or knowledge, or non-excludability of some aspect of the goods involved as in the case of public goods). In this case the Coase "theorem" (there is no theorem) showed that a large number of buyers or sellers knowing only their own preferences and the prices offered or asked by their trading counterparts would be able to bargain their way to an efficient allocation, eventually exhausting all potential gains from trade. Coase (1960) showed how the efficiency results of the invisible hand theorem can follow from much less stringent conditions, requiring only the absence of impediments to private bargaining rather than complete contracts and competition among large numbers of potential buyers and sellers. I once thought that this Coasean bargaining provides an empirically plausible model of exchange supporting Pareto-efficient allocations among self regarding actors as long as property rights are well defined (Bowles (2004)).

But this is not the case. 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. The result is that some mutually beneficial exchanges will not occur. In similar manner, the fact that individuals may benefit by misrepresenting their preferences prevents the Legislator planner from eliciting the information he needs to provide incentives for the efficient provision of a public good (Gibbard (1973), Laffont and Maskin (1979)). The Swedish Academy conveyed this unhappy news in a non technical summary of the field when it awarded the Nobel Memorial Prize in Economics to Eric Maskin, Roger Myerson, and Leonid Hurwicz in recognition of their contributions to mechanism design (Royal Swedish Academy of Sciences (2007)).

The problem is quite general, but it is best illustrated by a particular example provided by Chatterjee and Samuelson (1983) and Myerson and Satterthwaite (1983), The case does not involve public goods but rather is a textbook exchange in which Pareto-efficient outcomes are typically taken for granted: well defined private goods without any of the usual impediments to bargaining. Suppose a good is worth v_s to a particular seller who may sell this good to a buyer, for whom it is worth v_b , and these valuations are private. These valuations differ among the sellers and among the buyers. A large number of such buyers and sellers are randomly paired for a single interaction in which simultaneously the seller announces the minimum price at which she is willing to sell, s , and the buyer announces the maximum price at which he is willing to buy, b , and an exchange occurs if and only if $b \geq s$. This is called a double auction. See Figure 4.1

The price at which the exchange takes place depends on both b and s . Suppose the price is just a weighted average of the two announced valuations, so we have $p = kb + (1-k)s$ with k the relative weight given to the buyer's announcement taking any value from 0 to 1. The rule might be that the realized price is halfway between the two announced valuations – the so called split the difference rule – so that $k = \frac{1}{2}$. If instead $k = 1$, then the buyer sets the price, and because the seller has no influence on the price her only concern is to maximize the probability of a beneficial trade, thereby announcing her true value, ie. $s = v_s$. Analogous reasoning holds for the case of $k=0$ and the seller sets the price and the buyer truthfully announces $b = v_b$.

But assuming that the traders are happy to lie in order to maximize their payoffs, only in these polar cases where the price is entirely determined by either the buyer or the seller will they

report their true valuations. The reason why opportunities for beneficial trade will be foregone is that as long as the realized price is increasing in both traders' stated evaluations (that is as long as $0 < k < 1$) the seller has an incentive to overstate her valuation of the good and the buyer has an incentive to understate his valuation of the good. By misrepresenting their valuations both benefit from the resulting increase in the surplus they will garner should a transaction take place, but this comes the cost of reducing the probability of a transaction. As a result two traders among whom mutually beneficial trade is technically possible because $v_b > v_s$, may fail to transact because they report values such that $b < s$.

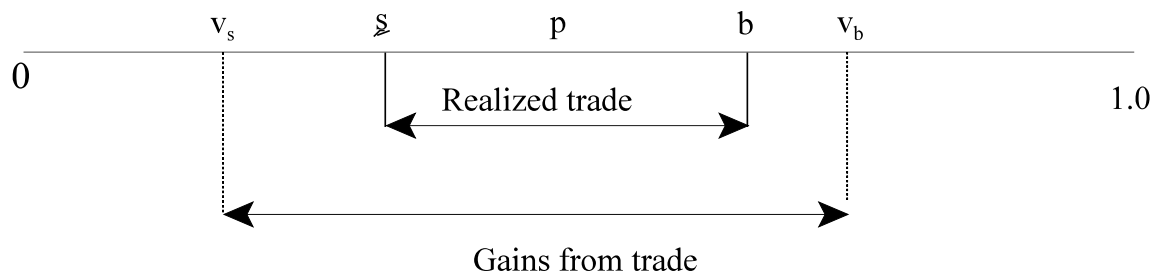


Figure 4.1: Buyers' and sellers' true and announced valuations of the good.

What Chatterjee and Samuelson (1983) and Myerson and Satterthwaite (1983) showed is not simply that a trader may benefit by misrepresenting the value of the good but that this will always be the case (as long as the realized price depends on the announced prices). The relationship between the real value and the stated value of the good is termed the revelation strategy of the two traders. The two functions $s = S(v_s)$ and $(b = B(v_b))$ give the valuation that the seller and buyer respectively will announce for every possible true valuation of the good in question. The expected payoff maximizing strategy for each trader pushes the level of misrepresentation to the point where the marginal gains in the share of the pie should a transaction take place are just offset by increased risk that the overstated selling price or understated buying price will preclude any transaction at all.

Of course the nature of this tradeoff for one trader depends on the strategy being followed by the other. So the payoff maximizing revelation strategy of each trader depends on the

revelation strategy adopted by the other. A strategy pair such that each is optimal given the other (the strategies are mutual best responses) constitutes a market equilibrium (technically, a Bayesian Nash equilibrium). Equilibrium strategy pairs result in a significant fraction of the set of all feasible mutually beneficial not taking place, as an example will show.

Suppose that traders settle on a price using the split the difference rule, and that both the buyers and sellers true valuations are uniformly distributed over the range of values shown in Figure 4.2 so that, for example, a seller is as likely to encounter a buyer who values the good at 57 cents as one willing to pay 38 cents. When traders are randomly paired, on the average in half the cases the seller will value the good less than the buyer (all of the pairings above the diagonal in Figure 4.2), and if they truthfully revealed their valuations, a mutually beneficial exchange would occur. But if traders use their payoff maximizing equilibrium strategy pairs, the announced valuations are such that no trade occurs unless the difference in the valuation of the good is at least $\frac{1}{4}$. As a result, the probability that a match will result in a trade is twenty-eight percent rather than one-half as it would be for truthful traders.

In the figure the problem can be seen to arise because many possible trades (pairs of true valuations) that satisfy the participation constraint --- $v_b \geq v_s$. -- so that both traders would benefit if they exchanged the good, do not satisfy the incentive compatibility constraint, so they would not be implemented by payoff maximizing traders.

The split the difference pricing rule is the best available: if $k \neq \frac{1}{2}$ the expected fraction of matches resulting in a trade is less than 28 percent. The unhappy result of this bargaining process is termed "incentive-efficient" as one can do no better under the circumstances (meaning: accepting that revelation strategies are adopted by payoff maximizing individuals who participate voluntarily in the market). But it clearly is not efficient in the usual Pareto sense because almost half of the mutually beneficial trades do not occur. (This statistic does overstate the losses, however, because the pairings for which the gains from trade are the greatest -- for which v_b exceeds v_s by a large amount, those occurring in the upper left portion of Figure 4.2 -- will in fact be implemented, as we have seen.)

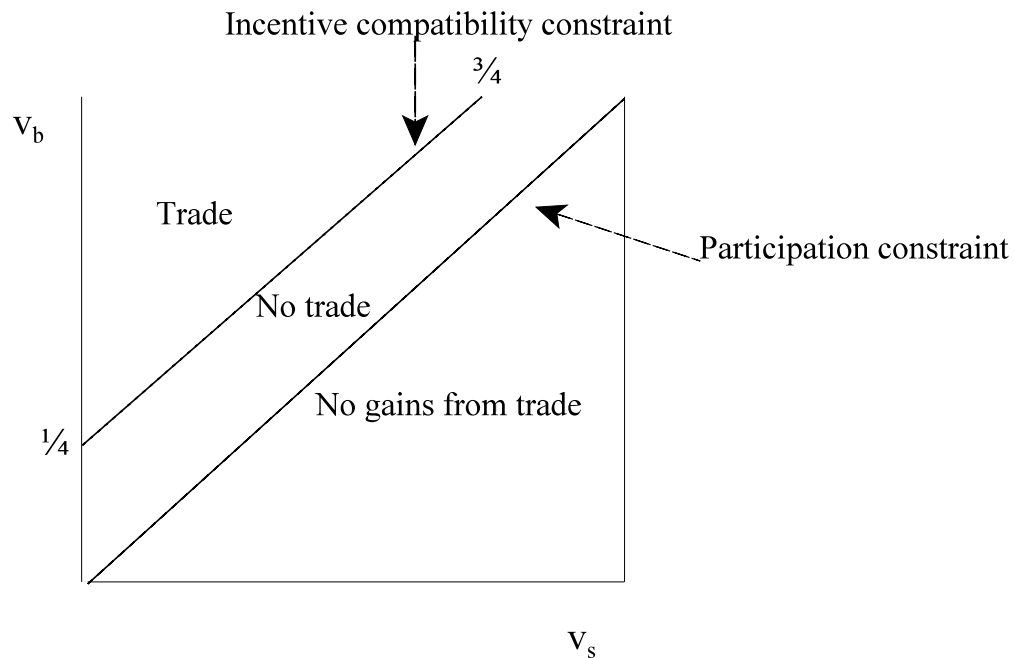


Figure 4.2. Incentive compatible trade and forgone opportunities for mutual beneficial trade in a Bayesian Nash equilibrium. Note: Valuation pairs above and to the left of the two constraints satisfy the constraints. For pairs of traders with valuations between the two constraints no transactions take place despite the traders valuations of the good being such that gains from trade are technically possible.

Chatterjee (1982) (extending Chatterjee, Pratt, and Zeckhauser (1978)) provided an incentive-compatible mechanism for this double auction such that each trader's best response is to truthfully report their valuations; as a result of which all mutually beneficial trades occur. The mechanism requires up front 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 payment is effectively a tax on false revelation that is just sufficient to make telling the truth a best response. Chatterjee, Pratt, and Zeckhauser (1978) describe this mechanism as “the payment to each player of the expected externality generated by his action” a kind of bargaining equivalent of "make the polluter pay" green taxes.

The hitch is that if traders know their own valuations, some would do better by withdrawing from the mechanism, which therefore works only if participation is involuntary

(violating the participation constraint). These 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 over a public good: under their mechanism truthful revelation is incentive compatible, but the mechanism requires that participation be mandatory.

These results show that the liberal trinity cannot be sustained even in a case deliberately cooked to make it work. Of the three legs of the liberal stool – neutrality, liberty, and efficiency – any two are consistent, but not three.

From this, the sophisticated Legislator will take away two uncomfortable facts: Incentives alone will not implement a fully efficient use of economic resources; and partly for this reason the Legislator must be concerned about the nature of individual preferences and the possibility that incentives may affect them adversely.

THE SOPHISTICATED LEGISLATOR'S DILEMMA: OVERUSE OF INCENTIVES?

A generation ago Robert Lucas (1976) rocked the standard paradigm in economic policy with a simple observation: taxes and other government interventions in the private economy affect not only the costs and benefits of actions that citizens may take, but also their beliefs about the actions of others (including the government). Here is an example (not one given by Lucas): announcing a campaign of stiffer penalties for non-payment of taxes provides an incentive to pony 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 beliefs to be endogenous, and then studies the joint equilibrium in which both citizens' beliefs and the targeted economic actions are stationary when account is taken of their mutual dependence. 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 was taken to be so important that it is now expressed in upper case letters, bestowing an honor on the Lucas Critique withheld even from the invisible hand. Here I take up Lucas' logic and apply it to cases in which incentives affect both beliefs and preferences, and may thus have unintended effects

The sophisticated Legislator faces a challenge that (with few exceptions) has yet to be addressed in the public economics literature: how to design optimal taxes, fines, or subsidies when 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 preferences which will determine the effects of the incentives.

This is not some impenetrable chicken and egg problem however. It just requires that for every policy under consideration the effects on preferences be considered and the effects of the policy be counted as the outcomes once the effects of preferences are accounted for. The Legislator is thus not simply selecting, say, a tax rate, but rather a tax rate and the distribution of preferences 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 of equilibrium preferences the Legislator can re-think the problem of optimal incentives. The economic intuition underlying Titmuss' claim that incentives are overused is that because crowding out reduces the effectiveness of incentives, they would be used less by a sophisticated Legislator cognizant of the crowding out problem, by comparison to his naive Legislator. If crowding out is so strong that the incentive has an effect the opposite of its intent, this is of course the case. But the effect of crowding out need not be literally counterproductive in this sense, and where the effectiveness of incentives is blunted but not reversed, the implications for the optimal use of incentives are far from obvious.

A modest way of addressing this problem would consider the problem of non- separability as exogenously given and simply determine the optimal level or mix of incentives taking account of their effects on preferences. Sung-Ha Hwang and I have studied a Public Goods game (the n- person Prisoners Dilemma described in Chapter II) in which citizens have other regarding preferences that motivate them to contribute more to the public good than would be the case in their absence (Hwang and Bowles (2010) and Bowles and Hwang (2008)). Related studies are

those of Bar-Gill and Fershtman (2005), Bar-Gill and Fershtman (2004) and Heifetz, Segev, and Talley (2007).

To make the problem interesting we assume that the citizens are not so altruistic that they fully take account of the benefits that their contributions confer on others, so that under-contribution will occur. The Legislator may provide a self-interested motivation to increase contributions beyond that motivated by the citizens' altruism by paying each citizen a subsidy proportional to their contribution. If the incentive works as intended this would implement an allocation closer to the social optimum. I take account of the cost of administering the subsidy that increases with the amount of the subsidy (collecting accurate information about contributions costs more if citizens have a larger incentive to misrepresent their contributions). The sophisticated Legislator knows that the incentives may compromise the citizens' pre-existing value of contributing (as in Figure 2.1, where the arrow from "incentive" to "value" is negative).

The optimal level of the subsidy is then determined as follows. If the negative effect of the incentive on the citizen values is so great that strong crowding out holds so that the effect of the subsidy would be to reduce contributions, the Legislator will of course abandon the use of subsidies as counter productive. But suppose the effect of the subsidy is positive even if attenuated by the crowding out of social preferences. Then to determine the optimal subsidy the Legislator must trade off the costs of raising the subsidy against its effect in raising contributions. Where marginal crowding out occurs (the negative effect of the subsidy on the citizens' values is greater, the greater is the subsidy), then the effectiveness of the subsidy is reduced and the tradeoff just indicated may lead the Legislator to adopt a lesser subsidy, just as Titmuss' logic would have it. But this is not the entire story.

The reason why the sophisticated Legislator may make greater use of incentives under these conditions is that if incentives work less well than would be the case under separability, then there are two offsetting influences on their optimal use. The one that forms the basis of the Titmuss critique is that crowding out reduces the marginal effect of the subsidy on the target's behavior. But there is a second often overlooked effect. Because the incentive is less effective (either categorically or marginally), the under-provision of the public good will be exacerbated and if there are decreasing returns to the level of provision of the public good, the marginal benefit of altering the target's behavior is therefore correspondingly greater. The intuition is transparent: the

doctor who discovers that a treatment he has been prescribing is less effective than he thought may opt for stronger doses rather than weaker, or for abandoning the treatment. This surprising case of under-use of incentives by the naive Legislator occurs when crowding out is categorical (and not too large), because in this case crowding out does not change the marginal effect of the incentive. But Hwang and I show that the sophisticated Legislator may make greater use of incentives even when only marginal crowding out occurs, if the marginal returns to the public good are diminishing at a sufficient rate.

I called this a modest approach because the Legislator simply took the crowding out phenomenon as a given. A less modest approach to the design of public incentives where separability may not hold is to recognize that the degree of non-separability is not given, but can be affected by the nature of the incentives and by the purposes for which they are deployed. Designing policies that can convert incentives from being substitutes of social preferences to their complements, however, requires an understanding of why crowding out occurs.

WHY DO INCENTIVES CROWD OUT SOCIAL PREFERENCES?

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 supports self interested action. We saw in Chapter II, for example, that the threat of a fine in a Trust Game (compared to the no-threat treatment) activated a brain region associated with the processing cost and benefit data (Li, Xiao, Houser, *et al.* (2008).) And there is some evidence that Ultimatum Game respondents who reject low offers exhibit heightened activation of the bilateral anterior insula, an area associated with negative emotional states such as anger and disgust (Sanfey *et al.* 2003) Camerer *et al.* (2005) comment:“It is irresistible to speculate that the insula is a neural locus of the distaste for inequality and unfair treatment. . . ”

But these are intriguing findings, there does not appear to be any simple mapping from the deliberation-affect distinction in cognitive processing to the self regarding --ethical and other regarding distinction in behavior. Negative and positive social emotions appear to be critically involved in both pro-social behavior such as cooperation (Singer and Steinbeis (2009)) as well as in entirely self regarding behavior; while deliberative processing may support a reflective Benthamite universalism or a calculating self-interest.(Table4.1 below). Not surprisingly, then

there is no simple correspondence between the behavioral distinction (self- vs other-regarding) and the prefrontal cortex - limbic system distinction. Thus while as Jonathan Cohen (2005) says, the prefrontal cortex “may be a critical substrate for *Homo economicus*” the deliberative decision maker, it is no more implicated in the self-interest assumed in most economic models than is the limbic system.

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

Table 4.1. No simple mapping between deliberative and emotional cognitive processing and pro social behavior.

What can the Legislator, then, conclude about why incentives sometimes crowd out social preferences? There are two quite different reasons, I think, why incentives and social preferences may be less than additive, and they operate on quite different time scales. The first concerns the fact that preferences are state-dependent (situation-dependent) and the presence and level of an incentive constitutes a particular state, while the second concerns the endogenous nature of preferences and the effects of incentives on how preferences are acquired.

According to the first, 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. This commonplace idea among psychologists has (with few exceptions, Akerlof and Kranton (2010)) been missed by economists. Incentives addressed to our acquisitive desires sometimes appear to 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. Individuals' constitutive objectives may override their acquisitive motives if a payoff maximizing response to an incentive would make them a chump or a victim, with crowding out the result. A negative 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

(2010)). Equally, responding in a self interested way an incentive may make the actor a good citizen or maybe just an intelligent shopper. Which of these it is may explain why incentives sometimes work exactly as economists predict on the basis of self interest and sometimes even surpass these expectations, but often they compete with the constitutive motives that Mill proposed to ignore, sometimes overriding them.

The importance of constitutive rather than simply acquisitive motives may be at work in the negative response to incentives that convey adverse information about the individual imposing the incentives. Recall that in the Trust Game implemented by Fehr and Rockenbach (2003) the investor's threat to fine the trustee if the back transfer was not sufficient reduced the level of reciprocity of the trustee: conditional on the investor's transfer to the trustee, back-transfers declined under the fine condition. This was especially the case when it appeared that the intent of the fine was to induce the trustee to grant most of the joint surplus to the investor. Where the investor announced modest levels of desired returns such that the investor and the trustee would both substantially share in the joint surplus, the use of the fines reduced back-transfers by an insignificant amount. But where the announced desired back-transfer would have allowed the investor to capture most of the surplus had the trustee complied, the reduction in back-transfers was 38 percent. It appears that the use of the fine in these conditions signaled more the unfair intent of the investor than simply his distrust of the trustee.

The fact that in this latter case incentives revealed that the principal is untrusting or self-aggrandizing helps explain the contrasting effect of incentives imposed by peers who do not stand to benefit personally. An example is the Public Goods with Punishment experiment in which fellow group members have the opportunity to reduce their own payoffs in order to punish (reduce the payoffs of) others in their group once each member's contributions are revealed. In this experiment group membership is sometimes shuffled after each period so that in subsequent periods a punisher will not be in the same group with the target of his punishment. Thus the punisher cannot benefit from the target's response. 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. In this setting there is a strong positive response by low contributors (Fehr and Gaechter (2002), Fehr and Gaechter (2000a) and Carpenter, Bowles, Gintis, *et al.* (2008).)

A plausible explanation of the effectiveness of incentives in this case is that, when punished by a peer who had nothing to gain by doing so, those who have contributed less than others interpret the punishment as a signal of public-spirited social disapproval by fellow group members seeking to uphold a social norm and willing to sacrifice payoffs to do so. As a result targeted free riders and even free riders who escaped punishment feel shame, which they redress by subsequently contributing more. In this case the incentive (prospect of peer imposed fines) has crowded in social preferences.

The fact that incentives designed to garner most of the joint surplus crowd out reciprocity and other social motives also provides a clue to the following puzzle: if incentives reduce the joint surplus why are they so widely used. Why then do we ever observe pie-shrinking incentives in practice? The pie metaphor gives away the answer: even if incentives reduce the total gains associated with a project, their use may give the principal a sufficiently larger slice of the smaller pie to motivate the principal to use them. 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 experiment above, was constructed so that had subjects responded optimally on the basis of self-regarding preferences, the total 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. But 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. Under these "optimal" contracts, profits were more than double than 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.

Thus one of the reasons why agents respond negatively to incentives -- that they benefit the principal at the agent's expense -- also explains why incentives may nonetheless be used by profit-maximizing principals, despite the fact that the result is a smaller pie. If a mutually acceptable division of the pie could be decided in advance (and enforced ex-post) this problem would not arise, but such ex-ante agreements are typically not feasible in real economies.

A second explanation of the sometimes anti-synergistic effects of incentives operates on a much longer time scale and concerns the effect of incentives on how we learn new preferences and discard old ones.

A plausible representation of this learning process is that is a Darwin-inspired model in which individuals consciously or more often unwittingly copy the behaviors (and the preferences motivating them) of successful people whom they meet or know about (Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1985)). Incentives and other aspects of economic organization affect this process because they influence both who meets whom and the set of behaviors that are feasible and rewarding given the kinds of tasks that people undertake (Bowles (2004).)

Here is an example on how incentives might impede the learning of pro-social preferences. It is based on two empirical regularities. The first is the powerful effect of exposure as a source of learned preferences, documented by Robert Zajonc (1968) and subsequent works (Murphy and Zajonc (1993) and Murphy, Monahan, and Zajonc (1995)). The exposure effect is one of the reasons that cultural transmission may favor the numerous over the rare, independently of their economic success: social pressures for uniformity are among the most convincingly documented human propensities.(Boyd and Richerson (1985), Ross and Nisbett (1991), and Bowles (1998) and the works cited there.) Following Boyd and Richerson, we assume a degree of conformist cultural transmission by which we mean that the likelihood that an individual will adopt a particular preference varies with prevalence of that behavior in the population (independently of other influences on learning such as relative payoffs.)

The second is that the presence and extent of incentives to contribute to a public project (or to engage in similar activities that benefit others) make the behavior (contribution) a noisier signal of an individual's preferences, resulting in observers interpreting some generous acts as merely self-interested. This is the key mechanism underlying the model of Benabou and Tirole (2006) showing how incentives may crowd out pro-social behavior. In similar fashion, Joel Sobel (2007) asks "do markets make people selfish?" with the reply, no, but they may make people appear to be selfish.

Taken together these two assumptions imply that the extensive use of incentives may reduce the perceived frequency of individuals with generous preferences in the population,

leading by the conformist effect to an evolutionary disadvantage of generous over self interested traits in the cultural evolutionary process.

A second example of how preferences might evolve under the influence of economic institutions is the well known repeated Prisoners' Dilemma studied by Axelrod and Hamilton (1981), refining and extending the earlier insights of Shubik (1959), Trivers (1971), and Taylor (1976). They showed 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 “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 will support a society of cooperators.

But the kinds of incentives that are said to implement efficient outcomes among self interested individuals often result in interactions being of short duration. As an illustration, consider a common pool resource like a forest or a fishery that is subject to overexploitation because nearby villagers do not own the resource. But suppose that sustainable use is accomplished because most villagers are tit for tat cooperators. 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 it, thereby providing 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 lake – provides each with an incentive to sustain the resource and to monitor those who over-exploit it. But this will also enhance exit options and reduce the expected duration of interactions, 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, anthropologists and In the village Palanpur (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). The development of modern labor markets in central highland Peru appears to have made the traditional contributions of communal labor to produce local public goods 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)). 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)).

The sophisticated Legislator now has a lot on his plate. Incentives that appeal to citizen's acquisitive motives may, first, collide with the citizen's constitutive motives and second, create a cultural environment in which social preferences are less likely to be learned and more likely to be abandoned. But perhaps he also has enough information to sketch the outlines of a new model of policy making in light of these inconvenient facts.

LAWS AS THEY MIGHT BE FOR CITIZENS AS THEY MIGHT BE

John Stuart Mill (whose definition of the boundaries of political economy we mentioned at the outset) and economists since have recognized that the purposes of individual economic action are constitutive as well as acquisitive. But what some have missed what our Legislator must now grapple: the fact that our acquisitive and constitutive motivations may not be separable. Jeremy Bentham's *Introduction to the Principles of Morals and Legislation* (1789), is arguably the first text in what we now call public economics. In it he explained how proper incentives should harness self-interested objectives for public ends by making "it each man's *interest* to observe ... that conduct which it is his *duty* to observe" But, like Hirschman, he also understood the constitutive side of action and the need to design incentives that 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.)

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 policy was highly effective, apparently by invoking the shame of the tax evaders at their transgression of a social norm.

The fact that punishments are “moral lessons” as well as incentives may help resolve one of the puzzles in the literature we have just surveyed. In a widely cited natural experiment, in Haifa, at six randomly chosen day care centers, a fine was imposed on parents who were late in picking up their children at the end of the day (in a control group of centers no fine was imposed). Parents responded to the fine by significantly greater tardiness: the fraction picking up their kids late more than doubled (Gneezy and Rustichini (2000)). When after 16 weeks the fine was revoked, their enhanced tardiness persisted, showing no tendency to return to the *status quo ante*. Over the entire 20 weeks of the experiment, there were no changes in the degree of lateness at the day care centers in the control group. The counter-productive imposition of the fines appears to illustrate strong crowding out: using a market mechanism (the fine) seems either to have signaled a low cost of their tardiness or to have undermined the parents’ sense of personal obligation to avoid inconveniencing the teachers (Gneezy (2003))

But the small tax on plastic grocery bags enacted in Ireland in 2002 had the opposite effect: in two weeks it resulted in a 94 percent reduction in their use and appeared to crowd in social preferences (Rosenthal (2008)). Carrying a plastic grocery bag joined wearing a fur coat in the closet of anti-social practices. The contrast is instructive. In the Haifa case, the experimenters (respecting standard experimental protocols) provided no justification for the introduction of the fine on the tardy parents. Moreover the parents’ occasional lateness could have occurred for reasons beyond their control rather than as the result of a deliberate disregard for the inconvenience it caused the teachers. Finally, lateness was not so common as to be widely broadcast to the other parents. By contrast, the introduction of the Irish plastic bag tax was preceded by a substantial publicity campaign dramatizing the threat posed by the bags to the beauty of the emerald isle. Moreover, the use of the bags required a deliberate choice made in a

highly public condition. In the Irish case, as in the experiments by Galbiati and Vertova (2010) and Galbiati and Vertova (2008) mentioned in Chapter II, 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.

Bentham's reasoning does not recommend abandoning Hume's objective of harnessing self-regarding preferences to public ends: self interest is a powerful motive. What Bentham's view of punishments as both incentives and moral signals instead recommends is that legislators and other policy makers find ways that incentives may serve “the purpose of a moral lesson” and thereby to turn the separability problem on its head, making incentives and morals complements rather than substitutes.

There is nothing about mechanism design (or economics) that would preclude embracing more realistic psychological assumptions, and then rising to Bentham's challenge. But a consequence will be a more cautious and conditional embrace of the usual incentives. The caution that conventional incentive-based interventions may be worse than ineffective can be formalized as a preference-related analogue to has come to be 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 monopoly in one sector leading to price exceeding marginal cost, and some sector (perhaps the same one) contributing to environmental degradation so that the private marginal cost of production to the owner of the firm in this sector is less than the social marginal cost (taking the environmental external dis-economy into account). Then the correction of one of these market failures (making the monopolized industry competitive, for example) may take the economy farther away from an efficient outcome. 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 countering distortions induced by other violations. An example: if a producer generates environmental external dis-economies (and therefore produces more than the Pareto optimum level of output) this distortion can be countered if the same producer were a monopoly (and thus chooses an output at which price exceeds marginal cost ($p = mc$), thereby restricting

output). A competition policy which induced this producer to choose the competitive output level such that $p=mc$ could be welfare-reducing rather than welfare-enhancing. 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 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 equilibrium allocation. Considered dynamically in this way, the problem is more difficult than Hume's dictum that the legislator should harness knaves to the public benefit. Social preferences 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 selfish preferences, but also by evoking, cultivating and empowering public-spirited motives.

CONCLUSION

Is there a simple lesson for the Legislator? I think there is. The literature that followed Titmuss targeted explicit incentives as the cause of crowding out and recommended a reduced role for incentives in the governance of social interactions. Both the diagnosis and the policy implication are wrong.

Crowding out, as we have seen, may require greater, not lesser use of incentives. And, perhaps more important: fines, subsidies, and other monetary incentives *per se* are not the culprit. What accounts for the adverse effects are two things, I believe.

The first is the effect of incentives on the evolution of preferences, as we have just seen. But there is no presumption that incentives *per se* have deleterious effects in this regard. I explained in chapter III how the incentives provided by the rule of law under a liberal state could enhance trust and other social motives in a population. The fact that the ephemerality or

anonymity of market interactions may impede the evolution of some important variants of cooperative preferences (like tit for tat cooperation) is hardly an indictment of incentives per se.

The second reason why incentives may have adverse effects concerns the meaning of the fines or subsidies to the target of these incentives; and this depends on the social relationships among the actors, the information the incentive provides, and the pre-existing normative frameworks of the actors. This is the message of the contrast between the Irish grocery bag tax and the Haifa fines for tardiness, along with the fact that fines imposed on low contributors by peers in Public Goods Games have positive effects while fines imposed by principals on agents sometimes backfire. The same message is evident in the successes and failures of incentives and other mechanisms for procuring human blood and organs for medical uses (Healy (2006).)

Fines deployed to exploit or to control the target (or which give this appearance) are likely to be less effective than they would under separability, and may even be counterproductive. The reason is that they activate the target's desire to constitute himself or herself as a dignified and autonomous individual who is treated fairly by others. It is this constitutive aspiration that sometimes trumps the acquisitive motive addressed by the incentive, and leads to a contrary response. The same incentives deployed by individuals who do not stand to benefit personally, and which are intended to foster pro-social behavior are likely to crowd in pro social preferences. They do this by activating rather than diminishing the target's constitutive motives such as the desire to be a good member of a community, and a feeling of shame when others regard one as having failed in this.

V.

CODA: GOOD GOVERNMENT AND THE BEHAVIORAL SCIENCES. [to be written]

Over the last two decades behavioral scientists have established a number of empirical regularities with direct bearing on the enduring questions motivating political theorists and philosophers. Prominent among these are the seemingly ubiquitous human propensities for altruistic and reciprocal behavior motivated by feelings of sympathy, solidarity and moral commitment, and the no less ubiquitous limitations in human capacities for and proclivities to engage in higher order cognitive reasoning as the basis of ethical judgement and decision making. In light of empirical evidence now available, it appears that preferences and cognitive functioning alike are heterogeneous (differing substantially from person to person and among cultures), versatile (responding the cues provided by situations) -- and plastic (developing over the life cycle and evolving over longer periods under the influence not only of deliberate tutelage but also unwitting social effects).

Debates in political philosophy on the merits of markets and other institutions are not reducible to questions of fact about human behavior, but claims about these facts sometimes do enter in central ways. The preceding chapters are an initial attempt to bring modern behavioral science to bear on a essential questions about good government and the evolution of culture and society that has been central to political philosophy, yet curiously ignored by less philosophically inclined behavioral scientists. Here I address the surprising contrasts between the representation of human behavior in political theory and philosophy on the one hand and the behavioral sciences on the other.

[to be written]

Appendices

1 Experimental measurement of social preferences

2 The parasite hypothesis

3 Cultural-institutional market failures

4 Measures of institutional and cultural differences

Appendix 1. Experimental measurement of social preferences

Table A1: Seven experimental games useful for measuring social preferences

Table A1. Seven experimental games useful for measuring social preferences (from Camerer and Fehr, 2004)

<i>Game</i>	<i>Definition of the Game</i>	<i>Real life Example</i>	<i>Predictions with selfish players</i>	<i>Experimental regularities, References</i>	<i>Interpretation</i>
Prisoners' Dilemma	Two players may either Cooperate (C) or Defect (D) with $r(J,K)$ the payoffs to playing J with a K-playing partner as follows: $r(D,C) > r(C,C) > r(D,D) > r(C,D)$ and $2r(C,C) > r(C,D) + r(D,C)$. Mutual C maximizes the sum of payoffs; D is the individual payoff maximizing strategy irrespective of the strategy of the other.	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
Public Goods	n players simultaneously decide about their contribution g_i , $(0 \leq g_i \leq y)$ where y is player i's endowment; each player i receives $G_i = y - g_i + mG$ 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	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 $x = \zeta$; where ζ 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	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

Table A1
Cont'd

Trust	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	"Employer" offers a wage w to the "worker" and announces a desired effort level \hat{e} . If worker rejects (w, \hat{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 .	Non contractibility or non enforceability 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	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.

Note: ** denotes survey papers, * denotes papers that introduced the respective games.

Table A2: Are student experimental subjects more pro-social than the general public?

<i>Study</i>	<i>Result</i>
Bellemare, Kroger, and van Soest (2008)	“ extending the subject pool from students only to a more representative population [of Dutch citizens] ... generates a distribution with much greater levels of inequity aversion
Baran, Sapienza, and Zingales (2009)	Individual's sensitivity to social pressure (Crowne Marlowe (1960) social desirability scale) influenced behavior in a natural setting (contributions to the University of Chicago GSB), but not in a lab (trustee back transfer in a trust game).
Carpenter, Burks, and Verhoogen (2005)	In a DG, Middlebury College and Kansas City Kansas Community College students gave less than employees at a Kansas City distribution center (warehouse). KCKCC students gave more in the UG than warehouse workers who in turn gave more than MC students. Workers gave the same in the DG and the UG, while students gave much less in the DG (consistent with the view that in the UG. workers were not giving strategically while students were)
Burks, Carpenter, and Goette (2009)	In a sequential PD game bicycle messengers in Switzerland and U.S. were more cooperative than students.
List (2004)	Public goods contributions among participants at a sports card show closely approximated those in lab experiments (e.g. Fehr Gaechter 2000), Isaac et al (1994) and Andreoni (1988). Older (> 49) participants contributed more than college age, while middle aged approximated college age. Controlling for income older Florida residents also gave more to a university fund raising appeal. In the TV game show Friend or Foe (a PD like game for big stakes) older players were more likely to cooperate
Carpenter, Connolly, and Myers (2008)	In a dictator game with subject-named charity recipients, students gave 25 percent less than non students. Students also gave significantly less controlling for age and other demographics, and older respondents gave more controlling for student status. (Vermont). Forty eight percent of non students contributed the entire endowment; only 16 percent of students did.
Cardenas (2005)	In a common pool resource game students were less cooperative (extracted significantly more) than villagers (Colombia)

Cleave, Nikiforakis, and Slonim (2009)	Students who later volunteered for an experiment were as trustworthy but less trusting than the student population from which they were drawn (who were administered the trust game as a captive audience)
Falk, Meier, and Zehnder (2010)	Swiss students who exhibit stronger pro-social behavior in an unrelated field donation (contribution to a social fund) are not more likely to participate in experiments. Students and general population behavior is very similar in a trust game except that the back transfers conditional on investor allocation were significantly lower for students.
Gaechter and Hermann (2010)	Among Russian rural and urban experimental subjects non students contributed significantly more (18 per cent) than students in a public goods game; 23 per cent of the non students contributed their entire endowment while only 12 of students did so.

Appendix 2: Parasitic liberalism

Smith (1776)

The division of labor is limited by the extent of the market.

In the progress of the division of labor, the employment of the far greater part of those who live by labor, that is, of the great body of the people, comes to be confined to a few very simple operations, frequently to one or two. But the understandings of the greater part of men are necessarily formed by their ordinary employments. The man whose whole life is spent in performing a few simple operations, of which the effects are perhaps always the same, or very nearly the same, has no occasion to exert his understanding or to exercise his invention in finding out expedients for removing difficulties which never occur. He naturally loses, therefore, the habit of such exertion, and generally becomes as stupid and ignorant as it is possible for a human creature to become. The torpor of his mind renders him not only incapable of relishing or bearing a part in any rational conversation, but of conceiving any generous, noble, or tender sentiment, and consequently of forming any just judgment concerning many even of the ordinary duties of private life. Of the great and extensive interests of his country he is altogether incapable of judging. (Smith (1937), *Wealth...*) Book Five, Chapter I, Part 3, Article II.

Burke (1790, 1791):

Men are qualified for civil liberty, in exact proportion to their disposition to put moral chains upon their appetites, in proportion as their love of justice is above their rapacity; in proportion as their soundness and sobriety is above their vanity and presumption..” (Burke (1791):64, *Letter...*)

When I see the spirit of liberty in action, I see a strong principle at work; ... I should therefore suspend my congratulations on the new liberty of France, until I was informed how it had been combined with government; with public force; with the discipline and obedience of armies; with the collection of an effective and well-distributed revenue; with the solidity for property; with peace in order; with civil and social manners... without them, liberty is not a benefit while it lasts, and is not likely to continue long. ...the age of chivalry is gone. That of Sophisters, economists, and calculators has succeeded ...Nothing is left which engages the affection on the part of the commonwealth...so as to create in us love, veneration, admiration or attachment.(Burke (1890[1790]): 84-86 *Reflections...*)

Tocqueville (1834, 1840):

...religion has been entangled with those institutions which democracy destroys...Liberty cannot be established without morality, nor morality without faith. ... No free communities ever existed without morals. (Tocqueville (1945):I,12;II 208, *Democracy...*)

A democratic state of society, similar to that of the Americans, might offer singular facilities for the establishment of despotism; ...an innumerable multitude of men, all equal and alike, incessantly endeavoring to procure the petty and paltry pleasures with which they glut their lives Each of them, living apart, is a stranger to the fate of all the rest...his children and his private friends constitute to him the whole of mankind; as for the rest of his fellow citizens, he is close to them but he sees them not...he touches them but he feels them not; he exists but in himself and for himself alone...Above this race of men stands an immense and tutelary power, which takes upon itself alone to secure their gratifications and to watch over their fate. ...what remains, but to spare them all the care of thinking and all the trouble of living? Such a power does not tyrannize, but compresses, enervates, extinguishes and stupefies a people, till each nation is reduced to nothing better than a flock of timid and industrious animals, of which the government is the shepherd. ... servitude of the regular, quite and gentle kind I have just described might be combined more easily than is commonly believed with some of the outward forms of freedom.(II, 334-337)

...the manufacturing aristocracy which is growing up under our eyes is one of the harshest that ever existed in the world. (II,170-71.

... it is difficult indeed to conceive how men who have entirely given up the habit of self government should succeed in making a proper choice of those by whom they are to be governed; and no one will ever believe that a liberal, wise and energetic government can spring from the suffrages of a subservient people. (II, 339.

Marx and Engels (1847-8)

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 Marx (1956):32 (*Poverty..*)

The bourgeoisie, wherever it has got the upper hand, has put an end to all feudal, patriarchal, idyllic relations ...and left no other nexus between man and man than naked self-interest than callous “cash payment.” It has drowned the most heavenly ecstasies of religious fervor...in the icy waters of egotistical calculation. Marx and Engels Marx and Engels (1972) (*...Manifesto*)

Polanyi (1944)

Our thesis is that the idea of a self-adjusting market implied a stark utopia. Such an institution could not exist for any length of time without annihilating the human and natural substance of society p. 3 ... [The] market for labor implied no less than the wholesale destruction of the traditional fabric of society p. 77 Social history in the 19th century was thus the result of a double movement: the extension of the market organization in respect to genuine commodities was accompanied by its restriction in respect to the fictitious ones. p. 76. [T]he labor market was allowed to retain its main function only on... conditions] that ...would safeguard the human character of the alleged commodity, labor. p. 177 Polanyi (1957) (...*Transformation*)

Bell (1973, 1976)

The historic justifications of bourgeois society -- in the realms of religion and character -- are gone... ... The lack of a rooted moral belief system is the cultural contradiction of the society ... (Bell (1973)48, ...*Post-industrial society*)

.the problem of virtue arose because of the dual and necessarily contradictory role of the individual as *citoyen* and *bourgeois*. As the first, he had the obligation to the polity of which he was a part; as the second he had private concerns which he pursued for his own self-interest. (Bell (1976)21, *Cultural contradictions of capitalism*)

In historical retrospect, bourgeois society had a double source and a double fate. The one current was a Puritan, Whig capitalism in which the emphasis was not just on economic activity but on the formation of character (sobriety, probity, work as a calling). The other was a secular Hobbesianism, a radical individualism which saw man as unlimited in his appetite, which was restrained in politics by a sovereign but ran fully free in economics and culture. (Bell (1976):80)

American capitalism...has lost its traditional legitimacy, which was based on a moral system or reward rooted in the Protestant sanctification of work. (Bell (1976):84)

The major consequence of this crisis .. is the loss of *civitas*, that spontaneous willingness to obey the law, to respect the rights of others, to forgo the temptations of private enrichment at the expense of the public weal (Bell (1976):245)

Habermas (1975)

The "Protestant ethic" with its emphasis on self-discipline, secularized vocational ethos, and renunciation of immediate gratification, is no less based on tradition than its traditionalist counterpart of uncoerced obedience, fatalism and orientation to immediate gratification. These traditions cannot be renewed on the basis of bourgeois society alone.

Bourgeois culture as a whole was never able to reproduce itself from itself. It was always dependent on motivationally effective supplementation by traditional world-views. 77

...the remains of pre-bourgeois traditions, in which civil and familial-vocational privatism are embedded, are being non-renewably dismantled ..[they are] softened and increasingly dissolved in the course of capitalist development. (Habermas (1975):79, *Legitimation..*)

Hirsch (1976)

This legacy [of pre-capitalist moral codes] has diminished with time and with the corrosive contact of the active capitalist values. As individual behavior has been increasingly directed to individual advantage, habits and instincts based on communal attitudes and objectives have lost out (Hirsch (1976):117-18, *Social limits..*)

Appendix 3

A. Multiple cultural-institutional equilibria and cultural collapse

The model in the text is an extension of work in Belloc and Bowles (2010), Bowles (2004) and Bowles (2009).

There is a stationary level of virtue expressed by the function $v = v(m; \tau(m^-))$, where m^- represents past values of m and τ represents the extent of traditional institutions with $v_m < 0$ and $v_\tau > 0$ (v_x is the derivative of v with respect the variable x .) Thus when $v = v(m; \tau(m^-))$ the process of cultural updating is such that the level of virtue in the population does not change (i.e. is stationary, unless τ or m change). The $v(m; \tau(m^-))$ function is based on a process of cultural transmission in which an individual's values are periodically updated taking account of the relative payoffs of bearers of different values and the frequency of types in the population.

Likewise, the function $m(v)$ gives the stationary values of m for given values of v based on individuals structuring their interactions with others (choosing among, say, contractual or friendship, or familial ways of interacting in some particular activity) based on the relative payoffs of these various structures.

The intersections of these two functions are temporary equilibria (Grandmont (2007).)

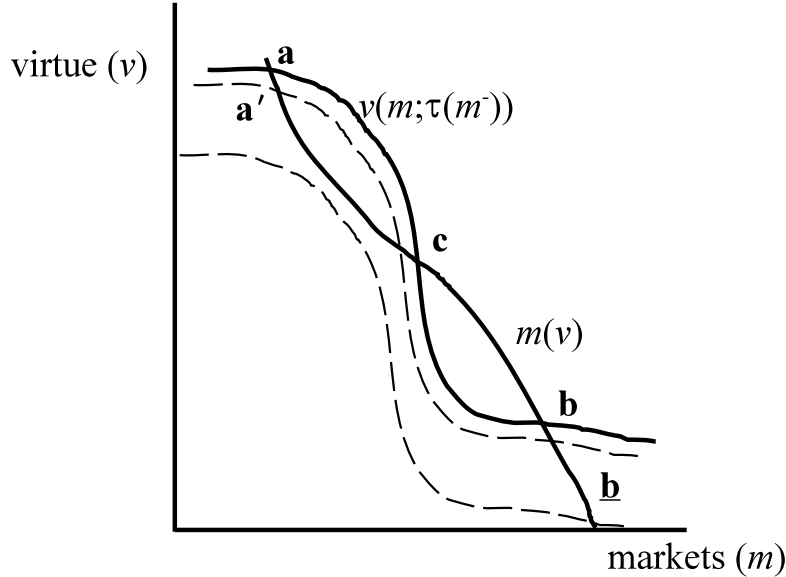


Figure 3A.1. Multiple cultural-institutional equilibria with punctuated-equilibrium dynamics. This figure is similar to Panel D of Figure 1 in the text except that the functions are non-linear. States **a** and **b** are asymptotically stable (self-correcting), while **c** is unstable. The erosion of traditional institutions given by the lower dashed line eliminates the upper stable equilibrium (as well as the unstable equilibrium) so that the only temporary equilibrium is **b**.

Figure 3A.1 illustrates a case in which two stable equilibria may exist, a result of the negative effect of markets on virtue being attenuated at both low and high levels of the extent of the market (so that the $v(m)$ function is non-linear and is flatter for high or low m .) In this case, for a society initially at the high virtue and limited market extent equilibrium (**a**) the effect of a modest downward shift in the values function resulting from the decay of tradition is to displace the cultural-institutional equilibrium to a nearby equilibrium (**a'**) with less virtue and more markets. A further decline in tradition however, may eliminate the upper equilibrium entirely (the lower of the two dashed lines), inducing a precipitous collapse of virtue and increased market dependence.

B. The Markets Economize on Virtue Function & Cultural-Institutional Market Failures

In Figure 3A2 (Panel B of Figure 3.1 with additional information) the loci labeled y and y^+ are isoquants, namely loci of pairs of m and v that yield a total income (of the society in question) of y and y^+ respectively with $y < y^+$. The position of the isoquants indicates that virtue contributes to the productivity of the society (its total income). Suppose, for illustration, that (as Ronald Coase hypothesized) the extent of the market is determined by an implicit transaction-cost minimizing process that maximizes income net of these costs for a given level of values. Then the $m(v)$ function is as shown, namely the locus of all points such that the isoquant is tangent to the horizontal dotted line indicating the given level of v . (See appendix 3). This gives the effect of values on the equilibrium extent of markets. If the extent of the market exceeded (or fell short of) that which maximized income for a given v , then the $m(v)$ function would be to the right (or the left) of that shown and aggregate income would be lower than that shown (for any given level of v). The idea that entirely decentralized contracting and other interactions would implement an efficient set of institutions in the Coasean sense is of course unrealistic; the key point is that markets will be used more were virtue is less. I adopt the Coasean framework simply because it makes clear that the parasitical liberalism thesis does not require any departures from conventional liberal economic models other than the fact that markets have cultural consequences.

The cultural-institutional equilibrium described in the text is not efficient because the process by which market extent is determined, whether in the spirit of Coase, that is efficiently, or (except accidentally) by any other decentralized means, does not take account of the cultural consequences of markets. Suppose that the extent of markets $m(v)$ is that which maximizes aggregate income $y(v,m)$ for each value of v . Figure A2 shows the isoquants based on $y(v,m)$ with slope $-y_m/y_v$. Because $y_v > 0$ (all institutions work better at higher levels of v) higher isoquants are associated with greater income (namely $y^+ > y$). The locus of points for which $y_m = 0$ (the isoquant is horizontal) give the $m(v)$ function.

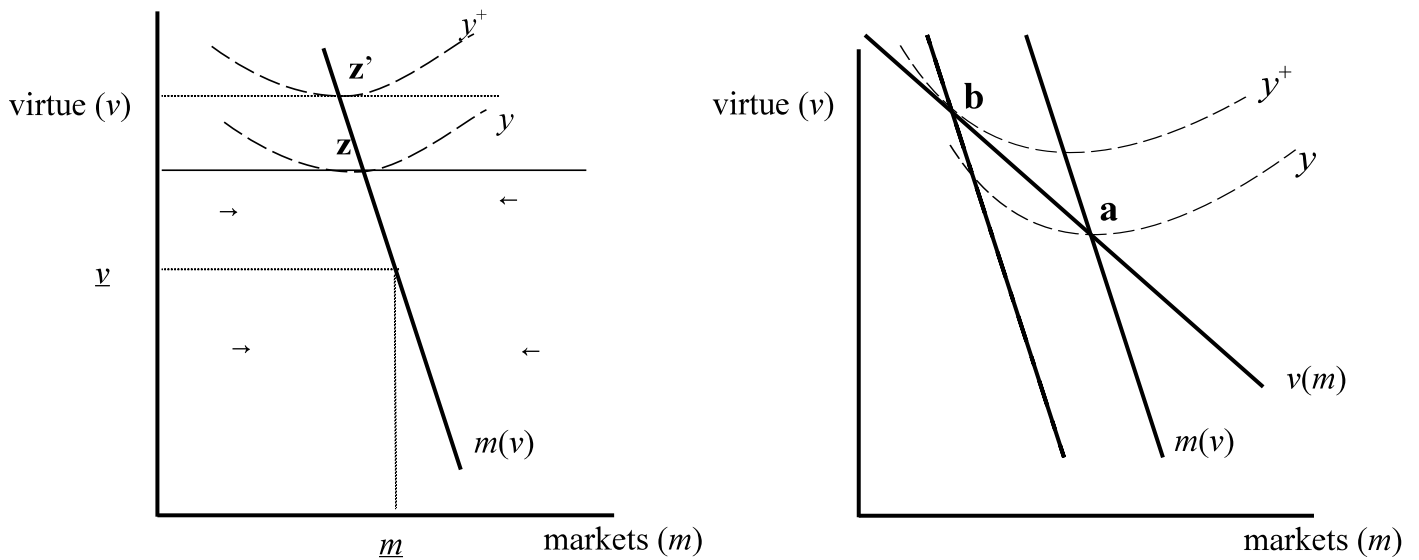


Figure 3A2. Derivation of the $m(v)$ function. 3A3. Cultural-institutional market failure

Figure 3 A2. The cultural-institutional equilibrium is a state such that both market extent and the level of virtue are stationary or $dm/dt = 0 = dv/dt$, namely point **a**. There exists an exogenous restriction of market extent (given by the dashed line) by a social planner that would displace the cultural institutional equilibrium to point **b**, resulting in a larger aggregate income. The social planner varies m to maximize y subject to the constraint that culture adjusts to the extent of the market according to $v = v(m)$. This income-maximizing level of market use balances the losses entailed by the use of non-market institutions (in cases for which, conditional on a given v , markets would do better) against the cultural benefits made possible by attenuating the deleterious market effects on culture.

A3. Described in text.

Appendix 4. Measures of Institutional and Cultural Differences

Definitions and Sources of Measures used in Sections IV and V

Rule of Law: measures the extent to which agents have confidence in and abide by the rules of society, in particular the quality of contract enforcement, the police, and the courts, as well as the likelihood of crime and violence. It is measured on a scale of -2.5 (being the weakest) to 2.5 (being the strongest). The data used is from the year 2002-2006. From the World Bank Worldwide Governance Indicators.

<http://info.worldbank.org/governance/wgi/index.asp>
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=999979

Democracy: the concept has been defined to measure: state of public corruption; current practice in human rights; political rights; free speech; and the overall state of the rule of law in 150 nations. The research was conducted by the World Audit, surveys 150 countries, and the measurements are updated each year (this article used the measures from the year 2008). The lower numbers signify a higher level of democracy and the high numbers a lower level of democracy as the world audit defines it according to the above-stated measures. The reported correlations in the text are for the negative of the measure, so that ‘democracy’ varies positively with the democratic traits above.

<http://www.worldaudit.org/publisher.htm>

Social inequality This is the “power distance” measure that the Hofstede defines it as [a] “dimension of national cultures It reflects the range of answers found in various countries to the basic question of how to handle the fact that people are unequal. It derives its name from research by a Dutch experimental social psychologist, Mauk Mulder, into the emotional distance that separates subordinates from their bosses. Scores for 50 countries have been calculated” The indicator increases with the score.(Hofstede and Hofstede (2005):41-42)

Individualism: Also from Hofstede, countries with low scores are considered collectivist countries, and high scores correspond with individualist societies: “Individualism pertains to societies in which the ties between individuals are loose: everyone is expected to look after himself or herself and his or her immediate family. Collectivism as its opposite pertains to societies in which people from birth onward are integrated into strong, cohesive in-groups, which throughout people’s lifetimes continue to protect them in exchange for unquestioning loyalty”(Hofstede and Hofstede (2005):75-76.)

Site	Trust	Law	Dem	Ineq	Indiv	Anti-soc P	Cont.
Boston	0.36	1.54	13	40	91	-8.117	18
Nottingham	0.29	1.72	10	35	89	-6.87	15
Copenhagen	0.67	1.94	2	18	74	-8.927	17.7
Bonn	0.38	1.73	11	35	67	-6.349	14.5
Zurich	0.37	1.96	5	26	69		16.2
St. Gallen	0.37	1.96	5	26	69	-5.876	16.7
Minsk	0.42	-1.23	137			-3.606	12.9
Dnipropetrovs'k	0.27	-0.74	129			-4.302	10.9
Samara	0.24	-0.88	119	93	39	-3.055	11.9
Athens	0.24	0.71	34	60	35	-2.38	5.7
Istanbul	0.16	0.02	69	66	37	-4.682	7.1
Riyadh	0.53	0.22	129	80	38	-3.273	6.9
Muscat		0.75	99			-0.486	9.9
Seoul	0.27	0.73	33	60	18	-4.634	14.7
Chengdu	0.55	-0.41	129	80	20	-6.004	13.9
Melbourne	0.4	1.79	8	36	90	-5.161	14.1

Table 4A1. Cultural-institutional measures. Anti-social punishment is the estimated dummy variable for the site in question (measuring the estimated difference between that subject pool's behavior and a predicted amount. Contribution (Cont.) is the average in the public goods with punishment experiment for the site indicated. Empty cells indicate absence of data (data on the controls in the estimating equation for Zurich were absent).

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, pp. 1384-414.
- Aristotle. 1962. *Nicomachean ethics*. Indianapolis: Bobbs-Merrill.
- Arrow, Kenneth J. 1971. "Political and Economic Evaluation of Social Effects and Externalities," in *Frontiers of Quantitative Economics*. M. D. Intriligator ed. Amsterdam: North Holland, pp. 3-23.
- Arrow, Kenneth J. 1972. "Gifts and Exchanges." *Philosophy and Public Affairs*, 1:4, pp. 343-62.
- Axelrod, Robert and William D. Hamilton. 1981. "The Evolution of Cooperation." *Science*, 211, pp. 1390-96.
- Baran, Nicole, Paola Sapienza, and Luigi Zingales. 2009. "Can we infer social preferences from the lab? Evidence from the trust game." *University of Chicago*.
- Bar-Gill, Oren and Chaim Fershtman. 2004. "Law and Preferences." *Journal of Law, Economics and Organization*, 20:2, pp. 331-53.
- Bar-Gill, Oren and Chaim Fershtman. 2005. "Public policy: with endogenous preferences." *Journal of Public Economic Theory*, 7:5, pp. 841-57.
- Barr, Abigail. 2001. "Social dilemmas, shame-based sanctions, and shamelessness: experimental results from rural Zimbabwe." Centre for the Study of African Economies Working Paper WPS/2001.11: Oxford University.
- Barr, Abigail, Chris Wallace, Jean Ensminger, Joseph Henrich, *et al.* 2009. "Homo Aequalis: A Cross-Society Experimental Analysis of Three Bargaining Games."
- Barry, Herbert III, Irvin L. Child, and Margaret K. Bacon. 1959. "Relation of Child Training to Subsistence Economy." *American Anthropologist*, 61, pp. 51-63.

- Bartolini, Stefano. 2009. *Did the decline in social capital depress Americans' happiness?:* University of Siena.
- Becker, Gary S. 1996. *Accounting for Tastes*. Cambridge, MA: Harvard University Press.
- Becker, Gary S. and George J. Stigler. 1977. "De Gustibus Non Est Disputandum." *American Economic Review*, 67:2, pp. 76-90.
- Belkin, Douglas. 2002. "Boston Firefighters Sick - Or Tired of Working." *Boston Globe*, 18 January, Third ed.: B1: Boston.
- Bell, Daniel. 1973. *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. New York: Basic Books, Inc.
- Bell, Daniel. 1976. *The cultural contradictions of capitalism*. New York: Basic Books.
- Bellemare, Charles, S. Kroger, and A van Soest. 2008. "Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities." *Econometrica*, 76:4, pp. 815-39.
- Belloc, Marianna and Samuel Bowles. 2010. "International Trade and the Persistence of Cultural-Institutional Diversity." *Santa Fe Institute Working Paper 09-03-005*.
- Benabou, Roland and Jean Tirole. 2003. "Intrinsic and extrinsic motivation." *Review of Economic Studies*, 70, pp. 489-520.
- Benabou, Roland and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96:5, pp. 1652-78.
- Benner, Erica. 2009. *Machiavelli's Ethics*. Princeton: Princeton University Press.
- Ben-Porath, Yoram. 1980. "The F-Connection: Families, Friends, and Firms and the Organization of Exchange." *Population and Development Review*, 6:1, pp. 1-30.
- Bentham, Jeremy. 1970 [1789]. *An Introduction to the Principles of Morals and Legislation*: Athlone Press.
- Benz, Matthias and Stephan Meier. 2006. "Do people behave in experiments as in real life?" *Institute for Empirical Research in Economics, University of Zurich*.
- Berkowitz, Peter. 1999. *Virtue and the Making of Modern Liberalism*. Princeton: Princeton University Press.
- 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*.

- Bisin, Alberto and Thierry Verdier. 2010. "The Economics of Cultural Transmission and Socialization," in *Handbook of Social Economics*. Jess Benhabib, Alberto Bisin and Matthew Jackson eds: Elsevier Science.
- Bliss, Christopher J. 1972. "Review of R.M. Titmuss, *The Gift Relationship: from human blood to social policy*." *Journal of Public Economics*, 1, pp. 162-65.
- Boehm, Christopher. 1984. *Blood Revenge: The Enactment and Management of Conflict in Montenegro and Other Tribal Societies*. Lawrence: University Press of Kansas.
- Bowles, Samuel. 1998. "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions." *Journal of Economic Literature*, 36:1, pp. 75-111.
- Bowles, Samuel. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press.
- Bowles, Samuel. 2009. "The Coevolution of Institutions and Preferences," in *Institutional and Social Dynamics of Growth and Distribution*. Neri Salvadori ed. London.
- 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, pp. 1811-20.
- Bowles, Samuel and Arjun Jayadev. 2007. "Garrison America." *The Economists' Voice*, 4:2, pp. Article 3.
- Bowles, Samuel and Sandra Polanía Reyes. 2010. "Economic Incentives and Pro-social behavior." *Santa Fe Institute*.
- Bowring, John ed. 1962. *The Works of Jeremy Bentham: Volume 8*. New York: Russell and Russell.
- Boyd, Robert and Peter J. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Buchanan, James. 1975. *The Limits of Liberty*. Chicago: University of Chicago Press.
- Burke, Edmund. 1791. *A letter from Mr. Burke to a member of the National Assembly in answer to some objections to his Book on French Affairs*. London: Dodsley, Pall-Mall.
- Burke, Edmund. 1890[1790]. *Reflections on the Revolution in France*. New York: Macmillan.

- Burks, Stephen, Jeffrey Carpenter, and Lorenz Goette. 2009. "Performance Pay and Worker Cooperation: Evidence from an artefactual field experiment." *Journal of Economic Behavior & Organization*.
- Camerer, Colin. 2003. *Behavioral Game Theory: Experimental Studies of Strategic Interaction*. Princeton: Princeton University Press.
- Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics*, 112:2, pp. 407-41.
- Camerer, Colin and Ernst Fehr. 2004. "Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists," in *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Joe Henrich, Samuel Bowles, Robert Boyd, Colin Camerer, Ernst Fehr and Herbert Gintis eds. Oxford: Oxford University Press.
- Cardenas, Juan Camilo. 2005. "Groups, Commons, and Regulations: Experiments with Villagers and Students in Colombia," in *Psychology, Rationality, and Economic Behavior: Challenging the Standard Assumptions*. Bina Agarwal and Alessandro Vercelli eds: International Economics Association.
- Cardenas, Juan Camilo, John K. Stranlund, and Cleve E. Willis. 2000. "Local Environmental Control and Institutional Crowding-out." *World Development*, 28:10, pp. 1719-33.
- Carpenter, Jeffrey, Samuel Bowles, Herbert Gintis, and Sung-Ha Hwang. 2008. "Strong Reciprocity and Team Production: Theory and Evidence." *Journal of Economic Behavior & Organization*, in press.
- Carpenter, Jeffrey, S. Burks, and E. Verhoogen. 2005. "Comparing students to workers: The effects of social framing on behavior in distribution games." *Research in Experimental Economics*, 1, pp. 261-90.
- Carpenter, Jeffrey, Cristina Connolly, and Caitlin Myers. 2008. "Altruistic behavior in a representative dictator experiment." *Experimental Economics*.
- Carpenter, Jeffrey P. and Erika Seki. 2010. "Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay." *Economic Inquiry*, in press.
- Cavalli-Sforza, L. L. and Marcus W. Feldman. 1981. *Cultural transmission and evolution : a quantitative approach*. Princeton, N.J.: Princeton University Press.
- Chatterjee, K. 1982. "Incentive compatibility in bargaining under uncertainty." *Quarterly Journal of Economics*, 97:1, pp. 717-26.

- Chatterjee, K and W Samuelson. 1983. "Bargaining under incomplete information." *Operations Research*, 31:5, pp. 23-39.
- Chatterjee, Kalyan, John W. Pratt, and Richard J. Zeckhauser. 1978. "Paying the expected externality for a price quote achieves bargaining efficiency." *Economics Letters*, 1, pp. 311-13.
- Cleave, Blair, Nikos Nikiforakis, and Robert Slonim. 2009. "Is there selection bias in laboratory experiments?" *University of Sydney*.
- Coase, R. H. 1937. "The Nature of the Firm." *Economica*, 4, pp. 386-405.
- Coase, R. H. 1960. "The Problem of Social Cost." *Journal of Law and Economics*, 3:1, pp. 1-44.
- Cohen, Jonathan. 2005. "The vulcanization of the human brain: A neural perspective on interactions between cognition and emotion." *Journal of Economic Perspectives*, 19:4, pp. 3-24.
- D'Antoni, M and Ugo Pagano. 2002. "National Cultures and Social Protection as Alternative Insurance Devices." *Structural Change and Economic Dynamics*, 13, pp. 367-86.
- d'Aspremont, Claude and Louis-Andre Gerard-Varet. 1979. "On Bayesian incentive compatible mechanisms," in *Aggregation and Revelation of Preferences*. Jean Jacques Laffont ed. Amsterdam: North Holland, pp. 269-88.
- Deci, Edward L. 1975. *Intrinsic Motivation*. New York: Plenum.
- Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, 125:6, pp. 627-68.
- 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, pp. 495-511.
- Falk, Armin and Michael Kosfeld. 2006. "The Hidden Costs of Control." *American Economic Review*, 96:5, pp. 1611-30.
- Falk, Armin, Stephan Meier, and Christian Zehnder. 2010. "Did we overestimate the role of social preferences? The case of self-selected student samples." *University of Bonn, Department of Economics*.
- 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, pp. 247-64.

Farooq, Omer. 2005. "Drumming tax sense into evaders." BBC News.

Fehr, E. and Andreas Leibbrandt. 2010. "Cooperativeness and impatience in the tragedy of the commons." *University of Zurich*.

Fehr, E. and Bettina Rockenbach. 2003. "Detrimental effects of sanctions on human altruism." *Nature*, 422:13 March, pp. 137-40.

Fehr, Ernst and Armin Falk. 2002. "Psychological Foundations of Incentives." *European Economic Review*, 46:4 - 5, pp. 687-724.

Fehr, Ernst and Urs Fischbacher. 2001. "Why Social Preferences Matter." *Nobel Symposium on Behavioral and Experimental Economics*: Stockholm.

Fehr, Ernst and Simon Gächter. 2000a. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review*, 90:4, pp. 980-94.

Fehr, Ernst and Simon Gächter. 2000b. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14:3, pp. 159-81.

Fehr, Ernst and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature*, 415, pp. 137-40.

Frey, B. and R Jegen. 2001. "Motivation Crowding Theory: A Survey of Empirical Evidence." *Journal of Economic Surveys*, 15:5, pp. 589 - 611.

Fryer, Roland. 2010. "Financial incentives and student achievement: Evidence from randomized trials." *NBER*.

Gächter, Simon and Benedikt Herrmann. 2010. "The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia." *European Economic Review*, in press.

Galbiati, Roberto and Pietro Vertova. 2008. "Obligations and cooperative behavior in public good games." *Games and Economic Behavior*, 64:1, pp. 146-70.

Galbiati, Roberto and Pietro Vertova. 2010. "How laws affect behaviour: Obligations, incentives and cooperative behavior." *Università Bocconi*.

Gambetta, Diego. 2008. *Do strong family ties inhibit trust?*: University of Oxford.

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," in *Trust: Making and Breaking Cooperative Relations*. Diego Gambetta ed. Oxford: Basil Blackwell, pp. 142-57.
- Gibbard, Allan. 1973. "Manipulation of voting schemes: A general result." *Journal of Economic Theory*, 41:4, pp. 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, pp. 7357-60.
- Gintis, Herbert. 1972. "A Radical Analysis of Welfare Economics and Individual Development." *Quarterly Journal of Economics*, 86:4, pp. 572-99.
- Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr eds. 2005. *Moral sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge: MIT Press.
- Gneezy, Uri. 2003. "The W effect of incentives." University of Chicago Graduate School of Business.
- Gneezy, Uri and Aldo Rustichini. 2000. "A Fine is a Price." *Journal of Legal Studies*, 29:1, pp. 1-17.
- Goodin, Robert E. and Andrew Reeve eds. 1989. *Liberal Neutrality*. London: Routledge.
- Grandmont, J.M. 2007. "Temporary Equilibrium," in *New Palgrave Dictionary of Economics*. Lawrence Blume and Stephen Durlauf ed. New York: MacMillian.
- Grant, Ruth. 2010. "Ethics and Incentives." *Duke University, Department of Political Science*.
- Greenberger, Scott. 2003. "Sick day abuses focus of fire talks." *Boston Globe*, 17 September, Third ed.: B7.
- 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, pp. 912-50.
- Greif, Avner. 2002. "Institutions & Impersonal Exchange: From Communal to Individual Responsibility." *Journal of Institutional and Theoretical Economics*, 158:1, pp. 168-204.
- Habermas, Jurgen. 1975. *Legitimation Crisis*. Boston: Beacon Press.
- Hayek, Friedrich. 1948. *Individualism and Economic Order*. Chicago: University of Chicago

Press.

Healy, Kieran. 2006. *Best Last Gifts*. Chicago: University of Chicago.

Heifetz, A., E. Segev, and E. Talley. 2007. "Market design with endogenous preferences." *Games and Economic Behavior*, 58, pp. 121-53.

Henrich, Joe, Robert Boyd, Samuel Bowles, Colin Camerer, *et al.* 2005. "Economic Man' in Cross-Cultural Perspective: Behavioral experiments in 15 small-scale societies." *Behavioral and Brain Sciences*, 28, pp. 795-855.

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, *et al.* 2009. "Markets, Religion, Community Size and the Evolution of Fairness and Punishment."

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Herich, Carolyn Lesorogol, Frank Marlowe, David Tracer, and John Ziker. 2010. "Markets, Religion, Community Size and the Evolution of Fairness and Punishment." *Science*, 327, pp. 1480-84.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, *et al.* 2006. "Costly punishment across human societies." *Science*, 312, pp. 1767-70.

Herrmann, Benedikt, Christian Thoni, and Simon Gaechter. 2008a. "Antisocial Punishment Across Societies." *Science*, 319: 7 March 2008, pp. 1362-67.

Herrmann, Benedikt, Christian Thoni, and Simon Gaechter. 2008b. "Supporting Online Material for 'Antisocial Punishment Across Societies'." *Science*, 319: 7 March 2008, pp. 1362-67.

Heyman, James and Dan Ariely. 2004. "Effort for Payment: A tale of two markets." *Psychological Science*, 15:11, pp. 787-93.

Hirsch, Fred. 1976. *Social Limits to Growth*. Cambridge, MA: Harvard University Press.

Hirschman, Albert O. 1985. "Against parsimony: three ways of complicating some categories of economic discourse." *Economics and Philosophy*, 1:1, pp. 7-21.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, 7:3, pp. 346-80.

Hofstede, G. and G.J. Hofstede. 2005. *Cultures and Organizations: Software of the Mind*. New York: McGraw-Hill.

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.

Hume, David. 1898. *David Hume, The Philosophical Works*. London: Longmans, Green, and Co.

Hume, David. 1964. *David Hume, Th Philosophical Works*. Darmstadt: Scientia Verlag Aalen.

Hurwicz, Leonid. 1975. "The Design of Mechanisms for Resource Allocation," in *Frontiers of Quantitative Economics II*. M. D. Intrilligator and D. A. Kendrick eds. Amsterdam: North Holland Press.

Hwang, Sung-Ha and Samuel Bowles. 2010. "The sophisticated planner's dilemma: optimal fines and subsidies when incentives affect preferences." *Santa Fe Institute*.

Irlenbusch, Bernd and G.K. Ruchala. 2008. "Relative Rewards within team-based compensation." *Labour Economics*, 15, pp. 141-67.

Jayadev, Arjun and Samuel Bowles. 2005. "Guard Labor." *Journal of Development Economics*, 79, pp. 328-48.

Kant, Emmanuel. 1970. "Perpetual Peace," in *Kant's Political Writings*. Hans Reiss ed. Cambridge: Cambridge University Press [1795].

Karlan, Dean. 2005. "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions." *American Economic Review*.

Kohn, Melvin. 1969. *Class and Conformity*. Homewood, IL: Dorsey Press.

Kohn, Melvin L. 1990. "Unresolved Issues in the Relationship Between Work and Personality," in *The Nature of Work: Sociological Perspectives*. Kai Erikson and Steven Peter Vallas eds. New Haven: Yale University Press, pp. 36-68.

Kohn, Melvin L. and Carmi 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, Carmi Schooler, et al. 1990. "Position in the Class Structure and Psychological Functioning in the U.S., Japan, and Poland." *American Journal of Sociology*, 95:4, pp. 964-1008.

Kumlin, Staffan and Bo Rothstein. 2005. "Making and breaking social capital: The impact of welfare state institutions." *Comparative Political Studies*, 38:339-362.

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," in *Aggregation and Revelation of Preferences*. Jean Jacques Laffont ed. Amsterdam: North Holland, pp. 289-308.
- Lanjouw, Peter and Nicholas Stern eds. 1998. *Economic Development in Palanpur Over Five Decades*. Delhi: Oxford University Press.
- Leibbrandt, Andreas, Uri Gneezy, and John List. 2010. "Ode to the Sea: The socio-ecological underpinnings of social norms." *University of Chicago, Department of Economics*.
- Levitt, Steven D. and John List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world." *Journal of Economic Perspectives*, 21:1, pp. 153-74.
- Li, Jian, Erte Xiao, Daniel Houser, and P Read Montague. 2008. "Neural responses to sanction threats in two-party exchanges." *Baylor College of Medicine*.
- Lipsey, R. and K. Lancaster. 1956-1957. "The General Theory of the Second Best." *Review of Economic Studies*, 24:1, pp. 11-32.
- List, John. 2004. "Young, Selfish, and Male: Field evidence on social preferences." *Economic Journal*, 114, pp. 121-49.
- Locke, John. 1968. *The educational writings of John Locke*. Cambridge: Cambridge University Press.
- 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, pp. 426-41.
- Machiavelli, Nicolò. 1984. *Discorsi sopra la preme 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, pp. 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 Vices, Publick Benefits*. Oxford: Clarendon Press.
- Mandeville, Bernard. 1988. "A Search into the Nature of Society," in *The Fable of the Bees*. F.B Kaye ed. Indianapolis: Liberty Fund, pp. 323-70.

- Marx, Karl. 1956. *The poverty of philosophy*. Moscow: Foreign Language Publishing House.
- Marx, Karl and Friedrich Engels. 1972. "The Communist Manifesto," in *The Marx-Engels Reader, 2nd Edition*. Robert Tucker ed. New York: W. W. Norton & Company (first published in 1848).
- Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. 2003. "Monetary and Non-monetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 93:1, pp. 366-80.
- Maskin, Eric. 1985. "The Theory of Implementation in Nash Equilibrium: A Survey," in *Social Goals and Social Organization; Essays in Memory of Elisha Pazner*. Leonid Hurwicz, David Schmeidler and Hugo Sonnenschein eds. Cambridge: Cambridge University Press, pp. 173-341.
- Mellstrom, Carl and Magnus Johannesson. 2008. "Crowding out in blood donation: Was Titmuss right?" *Journal of The European Economic Association*, 6:4, pp. 845-63.
- Mill, John Stuart. 1844. *Essays on some unsettled questions of political economy*. London: John W. Parker.
- Mill, John Stuart. 1919. *Considerations on Representative Government*. London: Longmans, Green, and Company.
- Mill, John Stuart. 1998. *Utilitarianism*. New York: Oxford University Press (originally published in 1861).
- Murphy, Sheila T. and R.B. Zajonc. 1993. "Affect, Cognition, and Awareness: Affective Priming With Optimal and Suboptimal Stimulus Exposures." *Journal of Personality and Social Psychology*, 64:5, pp. 723-39.
- Murphy, Sheila T., Jennifer L. Monahan, and R.B. Zajonc. 1995. "Additivity of Nonconscious Affect: Combined Effects of Priming and Exposure." *Journal of Personality and Social Psychology*, 69:4, pp. 589-602.
- Myerson, Roger and Mark Satterthwaite. 1983. "Efficient Mechanisms for Bilateral Trading." *Journal of Economic Theory*, 29, pp. 265-81.
- Pocock, J.G.A. 1975. *The Machiavelian Moment: Florentine Political Thought and the Atlantic Republican Tradition*. Princeton: Princeton University Press.
- Polanyi, Karl. 1957. *The Great Transformation: the Political and Economic Origins of our Time*. Beacon Hill: Beacon Press.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.

- Rosenthal, Elisabeth. 2008. "Motivated by a Tax, Irish Spurn Plastic Bags." *New York Times*: New York.
- Ross, Lee and Richard E. Nisbett. 1991. *The Person and the Situation: Perspectives of Social Psychology*. Philadelphia: Temple University Press.
- Rothstein, Bo and Eric Uslaner. 2005. "All for all: Equality, corruption and social trust." *World Politics*, 58, pp. 41-72.
- Rousseau, Jean-Jacques. 1762. *The Social Contract: or principles of political right*.
- Royal Swedish Academy of Sciences. 2007. "Mechanism Design Theory." Stockholm.
- Sarracino, Francesco. 2009. *Social capital and subjective well-being trends: comparing 11 western European countries*: University of Firenze.
- Schmitz, Hubert. 1999. "From ascribed to earned trust in exporting clusters." *Journal of International Economics*, 48, pp. 138-50.
- Schotter, Andrew, Avi Weiss, and Inigo Zapater. 1996. "Fairness and Survival in Ultimatum and Dictatorship Games." *Journal of Economic Behavior and Organization*, 31:1, pp. 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, pp. 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, pp. 330-39.
- Shubik, Martin. 1959. *Strategy and Market Structure: Competition, Oligopoly, and the Theory of Games*. New York: Wiley.
- Singer, Tania and Nikolaus Steinbeis. 2009. "Differential roles of fairness- and compassion-based motivations for cooperation, defection, and punishment." *Annals of the New York Academy of Sciences*, 1167, pp. 41-50.
- Sliwka, Dirk. 2007. "Trust as a Signal of a Social Norm and the Hidden Costs of Incentive Schemes." *American Economic Review*, 97:3, pp. 999-1012.
- Smith, Adam. 1937. *The Wealth of Nations*. New York: Modern Library (originally published in

1776).

Smith, Adam. 1976 [1759]. *Theory of Moral Sentiments*. Oxford: Clarendon Press.

Smith, Adam ed. 1976 [1776]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press.

Sobel, Joel. 2007. "Do markets make people selfish?"

Solow, Robert. 1971. "Blood and Thunder." *Yale Law Journal*, 80:8, pp. 1696-711.

Strauss, Leo. 1988. *What is political philosophy*. Chicago: University of Chicago Press.

Taylor, Michael. 1976. *Anarchy and Cooperation*. London: John Wiley and Sons.

Tilly, Charles. 1981. "Charivaris, Repertoires and Urban Politics," in *French Cities in the Nineteenth Century*. John M. Merriman ed. New York: Holmes and Meier, pp. 73-91.

Titmuss, Richard M. 1971. *The Gift Relationship: From Human Blood to Social Policy*. New York: Pantheon Books.

Tocqueville, Alexis de. 1945. *Democracy in America*: Vintage.

Trivers, R. L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology*, 46, pp. 35-57.

Tversky, A. and D. Kahneman. 1981. "The framing of decisions and the psychology of choice." *Science*, 211:4481, pp. 453-58.

Upton, William Edward III. 1974. "Altruism, attribution, and intrinsic motivation in the recruitment of blood donors." *Dissertation Abstracts International*, 34:12, pp. 6260-B.

Wilkinson-Ryan, Tess. 2010. "Do liquidated damages encourage efficient breach: A psychological experiment." *Michigan Law Review*, 108.

Woodburn, James. 1982. "Egalitarian Societies." *Man*, 17, pp. 431-51.

Woodruff, Christopher. 1998. "Contract enforcement and trade liberalization in Mexico's footwear industry." *World Development*, 26:6, pp. 979-91.

Yamagishi, T., K. S. Cook, and M Watabe. 1998. "Uncertainty, trust, and commitment formation in the U.S. and Japan." *American Journal of Sociology*, 104, pp. 165-94.

Yamagishi, Toshio and M Yamagishi. 1994. "Trust and commitment in the United States and Japan." *Motivation and Emotion*, 18, pp. 9-66.

Zajonc, Robert B. 1968. "Attitudinal Effects of Mere Exposure." *Journal of Personality and Social Psychology Monograph Supplement*, 9:2, Part 2, pp. 1-27.

INDEX