

# Corporate Honesty and Business Education: A Behavioral Model

Herbert Gintis and Rakesh Khurana\*

July 25, 2006

## Abstract

Since the mid-1970's neoclassical economic theory has dominated business school thinking and teaching in dealing with the nature of human motivation. However valuable in understanding competitive product and financial markets, neoclassical economic theory employs an incorrect, *Homo economicus*, model of human behavior that treats managers as selfish maximizers of personal wealth and power. The *Homo economicus* model implies that a firm's board of directors can best further stockholders' interests by (a) selecting managerial personnel who are focussed virtually exclusively on personal financial gain, and (b) inducing them to act as agents of the stockholders by devising incentives that minimize the difference between the financial returns to stockholders and the firm's leading managers. Moreover, while neoclassical financial theory, in the form of the efficient markets hypothesis, is a generally insightful portrayal of financial markets, this theory implies that a firm's stock price is the best overall measure of the firm's long-term value. This implies that managerial incentives should be tied to stock market performance, since this will best align the interests of managers and stockholders. However, this implication is invalid when managers can manipulating information flows that influence short term stock price movements. Neoclassical economic theory thus fosters a corporate culture that ignores the personal rewards and social responsibilities associated with managing a modern enterprise, and encourages an ethic of greedy materialism in which managers are expected to care only about personal financial reward, and in which such human character virtues as honesty and decency are honored only when they contribute to personal material reward. However, a wide range of experiments based on

---

\*Gintis: Santa Fe Institute and Central European University; Khurana: Harvard Business School. We would like to thank Stephen Burks, Michael Jensen, and Richard Shreve for helpful comments, as well as the Gruter Foundation, the Templeton foundation, and the John D. and Catherine T. MacArthur Foundation for support.

---

behavioral game theory contradicts the *Homo economicus* model. In complex environments where complete contracts cannot be written or enforced, honesty, integrity, intrinsic job satisfaction, and peer recognition are powerful motivators, leading to better results for contracting parties than reliance on financial incentives alone. In particular, many individuals place high value on such character virtues as honesty and integrity for their own sake, and are more than willing to sacrifice material gain to maintain these virtues. We suggest that business schools develop and teach a professional code of ethics similar to those promoted in law, education, science, and medicine, that the staffing of managerial positions be guided by considerations of moral character and ethical performance, and that a corporate culture based on character virtues, together with stockholder-managerial relationships predicated in part on reciprocity and mutual regard, could improve both the moral character of business and the profitability of corporate enterprise.

## 1 Introduction

Since the mid-1970's neoclassical economic theory has dominated business school thinking and teaching in dealing with the nature of human motivation. However valuable in understanding competitive product and financial markets, neoclassical economic theory employs an incorrect, *Homo economicus*, model of human behavior that treats managers as selfish maximizers of personal wealth and power. The *Homo economicus* model implies that a firm's board of directors can best further stockholders' interests by (a) selecting managerial personnel who are focussed virtually exclusively on personal financial gain, and (b) inducing them to act as agents of the stockholders by devising incentives that minimize the difference between the financial returns to stockholders and the firm's leading managers. Moreover, while neoclassical financial theory, in the form of the efficient markets hypothesis, is a generally insightful portrayal of financial markets, this theory implies that a firm's stock price is the best overall measure of the firm's long-term value. An implication of this theory is that managerial incentives should be tied to stock market performance, since this will best align the interests of managers and stockholders. However, this implication is invalid when managers can manipulating information flows that influence short term stock price movements.

Neoclassical economic theory thus fosters a corporate culture that ignores the personal rewards and social responsibilities associated with managing a modern enterprise, and encourages an ethic of greedy materialism in which managers are expected to care only about personal financial reward, and in which such human character virtues as honesty and decency are deployed only contingently in the interests of personal material reward. Of course, we cannot say that the corporate environment encouraged by business school is responsible for the high level of

---

managerial misconduct witnessed in recent years. We can say, however, that business education is deeply complicit because it has failed to provide a consistent, and accurate, alternative to the *Homo economicus* model.

We are, of course, hardly the first to level this charge against business education in America. Business schools have widely responded to criticism by adding a course on “business ethics” to the MBA curriculum. While welcome, such a move cannot compensate for the generally incorrect and misleading characterization of human motivation, based on the neoclassical *Homo economicus* perspective, promulgated in courses on managerial behavior and corporate culture. This model must be directly attacked and replaced by a more accurate model of human motivation.

We know the *Homo economicus* model is incorrect because a wide range of experiments based on behavioral game theory shows that honesty, integrity, intrinsic job satisfaction, and peer recognition are powerful motivators, leading to better results for contracting parties than reliance on financial incentives alone. In particular, many individuals place high value on such character virtues as honesty and integrity for their own sake, and are more than willing to sacrifice material gain to maintain these virtues. We suggest that business schools develop and teach a professional code of ethics similar to those promoted in law, education, science, and medicine, that the staffing of managerial positions be guided by considerations of moral character and ethical performance, and that a corporate culture based on character virtues, together with stockholder-managerial relationships predicated in part on reciprocity and mutual regard, could improve both the moral character of business and the profitability of corporate enterprise.

Ironically, stockholders themselves are among the victims of a corporate culture centered on greed, run by managers who no more care about real interests of current and future stockholders than they about any other stakeholder in the operation of the enterprise.

Several scholars have recently voiced their concerns that the model of human behavior taught to business school students, a model whose roots can be directly traced to *Homo economicus*, has contributed to the diffusion of poor management practices. Goshal and Moran (1996), for example, argued that the prescriptions that flowed from the *Homo economicus* model have led managers and stockholders to favor purely *extrinsic* motivators, including pay and threats, over *intrinsic* motivators, like excellence or quality, for mobilizing employee behavior. Such choices, they argue, actually undermine long-term organizational performance. Recently, other researchers have maintained that current business school teachings socialize students into an ethic of selfishness and limited accountability (Ghoshal 2005, Khurana, Nohria and Penrice 2005, Ferraro, Pfeffer and Sutton 2005). Khurana et al. (2005), for example, has suggested that contemporary business education has legitimized the idea that managers should solely orient their action and behavior around

---

maximizing shareholder value without regard for the effects of this action on other constituents, including employees, community or customers. Shiller (2005) writes that much of the economic and finance theory that students learn in business schools is not only empirically flawed but impairs a student's ethical compass: "[T]he view of the world that one gets in a modern business curriculum can lead to an ethical disconnect. The courses often encourage a view of human nature that does not inspire high-mindedness." Ghoshal's (2005):75 most recent criticisms are even more pointed. He argues that the contemporary business school curriculum and professors have directly contributed to the institutionalization of corrupt managerial practices, including obtaining excessive executive compensation and routinely taking ethical short-cuts. Current models of economic relationships teach students that managerial and employee contracts cannot be based to any significant degree on trust or trustworthiness. This view, in turn, sets in motion a self-reinforcing cycle in which students come to see opportunistic behavior, including lying and cheating, as unavoidable and hence as morally acceptable. Ghoshal writes: "Business schools do not need to do a great deal more to help prevent future Enrons; they need only to stop doing a lot they currently do. They do not need to create new courses; they need to simply stop teaching some old ones....[B]usiness school faculty-need to own up to our own role in creating Enrons. Our theories and ideas have done much to strengthen the management practices that we are all now so loudly condemning."

Changing the dominant culture of an occupational group is of course a difficult task. But it is not impossible, and involves a joint deployment of new incentives and a novel cultural initiative. An example of a successful initiative of this type is the initiative begun in 1974 by the Hong Kong government, which had endured a culture of corruption going back hundreds of years. In that year the government pronounced yet another of its period anti-corruption campaigns. But, this time the new Independent Commission Against Corruption (ICAC) was surprisingly successful (Hauk and Saez-Marti 2002). The reason for this success was that new incentives were supplemented by an intense anti-corruption socialization campaign in the schools. The result was that in 1986, 75.1% of the 15–24 age group, the first generation to experience this socialization campaign, believed that corruption was a social problem, compared to only 54% of the 45–64 age group. Similarly, in 1977 32% believed that "tipping" government employees to encourage them to perform their prescribed duties was illegitimate, compared to 72% only nine years later in 1986. In the same vein, in 1977 38% believed that under-the-table kickbacks were legitimate in business dealings, while in 1986, only 7% held this view. Ten years further into the campaign, in 1998 and 1999 surveys, about 85% of respondents aged 15–24 said they would not tolerate corruption either government or business corruption.

Of course, "socialization" could possible take the simple form of altering the

---

expectations of self-regarding agents, but we do not think that this is what occurs when honesty and integrity are promoted in social affairs. Rather, such programs draw upon the innate desire of a majority of citizens to lead a moral life, and to contribute to the punishment of those who behave immorally. In the body of this paper, we present evidence supporting this alternative scenario.

## **2 Stockholder Resurgence: Agency Theory as a Tool for Controlling Managers**

For the better part of the Twentieth century, managers were treated in much the same manner as scientists, physicians, and other professionals. They were given a long leash and were expected to apply their craft in a professional and honorable manner. All this changed in the last quarter of the century. The attack on the professionalism of managers was broad based, but with respect to business schools two factors stand out. First, the economic decline of the 1970s and the accompanying lack of competitiveness of U.S. firms in the face of foreign competition created a context for a thorough critique of American management, including American business education, and gave rise to competing prescriptions about what could be done to improve America's corporations. Second, takeover firms and activist institutional investors began to present themselves forcefully as parties injured by incompetent corporate management, and as defenders of the individual shareholder. These newly empowered actors from outside of the firm sought greater control over the corporation.

Using the poor corporate performance of the 1970s as their backdrop, these takeover artists successfully recast the image of corporate managers and executives not as wise corporate statesmen trying to adjudicate the competing concerns of a variety of corporate constituents, but rather as a self-dealing, unaccountable elite whose primary interest was taking advantage of weak shareholders to promote a leisurely life-style and exaggerated material gain. The idea that American society could depend on managers to discipline themselves to achieve the desired outcome of corporations contributing to the well-being of society was increasingly replaced by the view that managerial effectiveness could only be attained through the overarching authority of the market. This was an ironic turn of events that could only be truly appreciated by Progressive reformers like Edwin Gay, Louis Brandeis, and Walter Lippmann, who saw enlightened managerial professionalism as society's best defense against an soulless and socially indifferent investor class.

The revisionism surrounding managerialism that took place during the 1980s had a profound impact on business education.<sup>1</sup> It represented an institutional shift

---

<sup>1</sup>Business schools were largely insulated from the student protests that deeply affect most of the

---

away from the basic managerialism framework that had defined and informed business school education and animated the managerial professionalization project from its start, eventually replacing it with a new conception that is never fully specified but whose broad outlines can be understood as a conception of management as an agent of shareholders, the corporation as a nexus of individual contracts, and the primary purpose of the corporation being to maximize shareholder value. The rhetoric and rationale undergirding this new conception of management with its emphasis on shareholder value was of a particular variant whose origins could be traced to the neoclassical model of *Homo economicus*, which played a particularly influential role in business education over the last two decades, especially in conceptualizing the role of management, the definition of the corporation, and the purpose of the corporation.

During this period, many leading business schools imported neoclassical economic theory practically wholecloth. There is little doubt but that, on the whole, the results were extremely effective. Business school students learned modern decision theory, risk assessment, portfolio management, efficient market theory, and were exposed to an array of analytical tools that are indispensable in modern corporate management.

Among these tools was the neoclassical *principal-agent model*, which was developed to explain the economic relationship between an employer (the principal) and an employee (the agent), an owner (the principal) and a manager (the agent), a lender (the principal) and a borrower (the agent), and other interactions where one individual (the principal) engages another (the agent) to act on his behalf, but cannot write a fully specified contract as to exactly what services the agent is to deliver.

The principal-agent model is built on the notion that individuals have a personal *preference ordering* that they attempt to maximize, subject to the informational, institutional, and financial *constraints* they face, and relying on their personal *beliefs* concerning the value of various outcomes and how their actions can produce outcomes. In this economics literature, this is known as the *rational actor model*. However, the term “rational” is so subject to flights of interpretive fancy, as we shall see below, that we prefer to call it more neutrally the *preferences, beliefs, and constraints* (PBC) model.

The PBC model is the very heart of modern economic theory, and in its absence, the explanatory success of modern economics would not have been possible. The principal-agent model flows logically from the PBC model, and suggests that the agents will maximize his personal payoffs, subject to whatever constraints are placed

---

university, partly because of the conservative attitudes and beliefs of the students and faculty drawn to teaching in business school.

---

upon him by the principal. The principal's task, therefore, is to devise a system of *incentives* that will lead a maximizing agent to behave in a manner favorable to the principal's own goals.

The principal-agent model was brought into business schools in the form of *agency theory* by Jensen and Meckling (1976) in a form directly particularized to elucidate the relationship between stockholders (the principals) and managers (the agents). It is important to note that the validity of the PBC model and of agency theory do not depend on the *Homo economicus* model of the individual. Indeed, Jensen and Meckling argued from the beginning that managers were not simply "money maximizers," and agency theory applies even where agents included altruistic and social purposes among their personal objectives (Jensen and Meckling 1994). However, neither these nor other agency theorists provided an alternative to *Homo economicus* that was sufficiently concrete and usable as to serve as an alternative to the ubiquitous *Homo economicus* paradigm. Agency theory in business school practice, by default, imported *Homo economicus* and became a theory of how to motivate selfish money maximizers. We suggest that agency theory be retained, but the underlying model of human agency be thoroughly overhauled.

### 3 Neoclassical Economic Theory

Following the implementation of the Ford and Carnegie Foundation reforms during the 1950's and 1960's, neoclassical economics asserted itself as the dominant discipline in business schools in the 1970's. A quick check of any of the major undergraduate and graduate microeconomics textbooks will convince those skeptical of our claims concerning ubiquity of the *Homo economicus* model: *never* is any assumption made concerning individual objects other than maximizing some combination of leisure and monetary reward. It is in this sense that neoclassical theory is so powerful an attack on the legitimacy of managerial authority as it had been constructed during the founding era of business schools and then revised in the post-World War II period. Its proponents acknowledge the difficulties that emerge with the dispersal of stockholdings and the rise of management, addressing by deploying the *Homo economicus* model, which conceptualizes the separation of ownership and control as one in which managers have no concern for shareholders' interest, except insofar as they are compelled to do so by their formal contractual conditions of employment.

In a series of papers following Jensen and Meckling (1976), Fama and Jensen (1983) and Jensen (1998) and other proposed a theoretical road map that asserted the purpose of the corporation to be that of maximizing shareholder value, and that since managers have interests that differ from those of stockholders, monitor-

---

ing these managers under conditions of wide stock dispersal is a major practical challenge. They argued that because their efforts are not easily observable, managers will perform fail to work towards stockholder goals. The challenge, they concluded, is to create an “alignment of incentives” in which managers’ personal financial interests will come into close correspondence with those of owners. While Jensen and Meckling often stressed the limitations of the *Homo economicus* model, their arguments were effectively drowned out in the mass adoption of traditional economic theory by the business school community.

Much of the discussion in these foundational papers focuses on the means by which owners can provide effectively align these interests. The research of Jensen, Meckling, Fama, and their colleagues emphasizes three mechanisms: monitoring managerial performance, providing comprehensive economic incentives, and promoting an active market for corporate control. Monitoring managerial behavior involves the deployment of complex accounting practices and the appointment of a professional board of directors whose members operate in the stockholders’ interest by virtue of their need to maintain their personal reputations. The alignment of incentives involves remunerating management in the form of company stock and stock options, so that managers and owners face exactly the same incentives, and hence self-interested managers will maximize shareholder value as a byproduct of maximizing their own material gain. The market for corporate control leads to stock prices reflecting firm fundamentals, and ensures that poorly performing “insiders” will be threatened and ultimately be replaced by efficiency and profit-oriented “outsiders.”

Neoclassical theory quickly created a unified approach to organizations and corporate governance in American business schools, catalyzing academic revolutions in corporate finance, organizational behavior, accounting, corporate governance, and the market for corporate control. Neoclassical theory spearheaded a new paradigm in to organizational theory. Unlike much of the earlier scholarship in business schools, many of the core ideas of neoclassical theory were not derived from inductive observation, but instead, through what seemed to be logically compelling or received wisdom. In the early 1970’s economists brought a high degree of theoretical rigor and analytically powerful modeling to business school education and research, the absence of which had concerned the Ford and Carnegie Foundations and haunted business education from the start. Drawing on the legitimacy of the economics discipline, the business school had the authority to redefine the character of managerial expertise.

The intimate bonding of disciplinary knowledge and its implications for professional identity is a fundamental postulate in the social sciences. Of course, people actively use knowledge to advance their influence and privilege. But, as Michel Foucault has noted, the process is more profound, since the classification of



---

knowledge, grounded in behavioral science, creates distinctions we come to see as *natural*, thus limiting our thinking by providing scripts and preconstituted habits of thought. The classifications we create to run our lives come to rule over us like alien beings. In particular, by trivializing culturally induced motivations and driving out of intellectual discourse such notions as honesty, trustworthiness, and fairness, as meaningful aspects of managerial motivation.

In particular, business economists often dismiss the arid mathematical formulations of their counterparts in the economic departments of academia, in favor of a thicker description of personal motivation and organizational behavior. However, the abstract formalization defines the conceptual basis for the business school rendition of microeconomic theory, and leaves no room for more than minor embellishments around the edges. As a result, even those who do not “believe” the neoclassical model, but simply use the model as a structural backdrop to a more nuanced description of the business world are in fact prisoners of it.

#### 4 Neoclassical Theory: A Simple Model

Before offering a critique of the *Homo economicus* model, it is useful to present a simple scenario that illustrates the major outlines of the model and its assumptions. Suppose the stockholders hire a manager to run the firm, and the firm’s expected profits  $\pi$  is 100 times the manager’s effort  $e$ , minus the salary  $s$  of the manager ( $\pi = 100e - s$ ). We may think of the manager’s effort  $e$  not simply as how hard he works, but also how careful he is to choose profit-maximizing strategies, as opposed to strategies that aggrandize his position, or favor other goals. Suppose the manager’s effort varies between a minimum of 0.1 and a maximum of 1 ( $0.1 \leq e \leq 1$ ), so the stockholders’ profit lies between 10 and 100 minus the salary  $s$  paid to the manager. Suppose further, for the sake of concreteness, that the subjective cost of effort to the manager is given by the schedule  $c(e)$ , where  $c(0.1) = 0$ ,  $c(0.2) = 1$ ,  $c(0.3) = 2$ ,  $c(0.4) = 4$ ,  $c(0.5) = 6$ ,  $c(0.6) = 8$ ,  $c(0.7) = 10$ ,  $c(0.8) = 12$ ,  $c(0.9) = 17$ , and  $c(1) = 21$ . No other effort levels can be chosen by the manager.

If the manager is self-regarding, caring only about his net salary, which is  $s - c(e)$ , then no matter how much the stockholders pay the manager, he will choose  $e = 0.1$  to maximize his return, and the firm will make profits  $\pi = 100 \times 0.1 - s = 10 - s$ . Supposing further that the manager must be paid at least  $s_o = 1$  to take the job, the stockholders will set  $s = s_o = 1$ , ending up with profit  $\pi = 10 - 1 = 9$ .

However, agency theory suggests that the manager’s incentives must be aligned with the owners, by paying the manager a salary that increases with the owners’ profits. For instance, suppose the manager is paid 10% of the firm’s profits. Then, the manager earns  $11e - c(e)$ , which is maximized when  $e = 0.3$ , giving the manager

---

the payoff  $11(0.3) - c(0.3) = 1.3$ , but the firm now earns  $\pi = 89(0.3) = 26.7$ . Paying the manager 46% of profits leads the manager to set  $e = 1.0$ , so his payoff is  $46(1) - c(1) = 46 - 21 = 25$ , and the firm earns  $\pi = 54(1) = 54$ . You can check that the owners maximize profits by paying the manager 24% of profits, leading the manager to set  $e = 0.8$ , so his payoff is  $24(0.8) - c(0.8) = 7.2$ , and the firm earns  $\pi = 76(0.8) = 60.8$ .

Of course, in the real world the information will never be remotely so clean. Profits will be more like  $\pi = ae + b$ , where  $a$  may have mean 100, but will be a random variable whose value depends on many complex economic conditions, and  $b$  may similarly be a random variable whose value is generally much larger than  $ae$ . Since stockholders know only  $\pi$  and not its breakdown into a managerial and non-managerial component, basing managerial compensation on  $\pi$  may not provide very effective incentives. Moreover, the manager is likely to be extremely risk-averse simply because his wealth is so much less than the firm's profits that variations in earnings may swamp his expected salary. Stockholders can deal with risk-aversion by diversification of their portfolios, but the manager cannot, without seriously weakening the alignment of incentives with that of the stockholders. Finally, there is no single, unambiguous measure of the firm's health for which managerial behavior is responsible. By emphasizing one aspect (e.g., stock prices) in managerial compensation, stockholders are inducing self-regarding managers to sacrifice the firm's overall health on the altar of those aspects upon which his compensation is based.

If managers are purely financially motivated, there is no alternative to neoclassical theory's quest for aligning stockholder and managerial incentives, no matter how imperfect the alignment turns out. But, why do we assume that managers are purely financially motivated? In other fields where the subtlety and complexity of agent expertise render explicit contracts manifestly incapable of aligning the interests of agents with that of their principals, very different employment conditions are generally deployed. For instance, we do not consider physicians, research scientists, or college professors as agents who will maximize their income subject to whatever incentives and constraints they face. Rather, we assume such agents have a *professional ethic* to which they subscribe, the maintenance of which is more valuable than material reward. Of course, we recognize that due safeguards must be in place to apprehend the occasional agent who violates professional norms, and professional norms themselves may not be perfectly aligned with social needs. Nevertheless, if the health system, scientific research, or higher education were run on the principle that the highest-level decision-makers are motivated solely by material reward, and if the training of professionals in these fields stressed that there are no binding ethical rules, and obeying laws should be subject to cost-benefit calculation, there is little doubt but that such systems would fail miserably.

---

It might be thought that the analogy of business leaders with their counterparts in health, science, and education is grossly overdrawn, if only because business is explicitly based on ruthless competition, whereas the fields depending on professionalism to align interests are generally service-oriented. However, a brief exposure to the competitive pressures in any of health care delivery or scientific research is generally sufficient to convince an open-minded student of their highly competitive nature, and of the massive gains that can accrue to those who violate their professional ethics. In any of these fields one can become massively rich before being ejected from the field, and even the latter contingency is relatively unlikely.

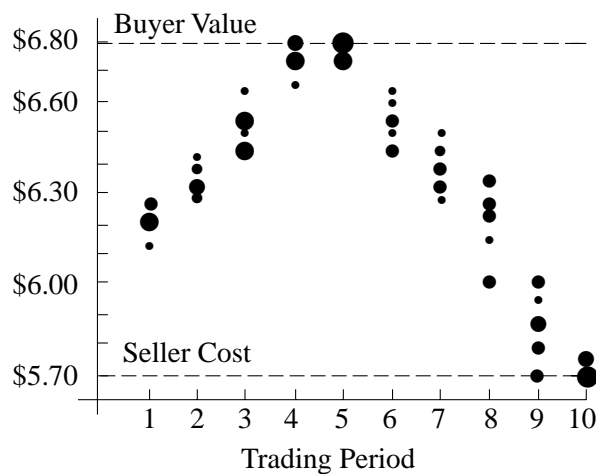
Perhaps it is not surprising that business schools would model their problems following economic theory, because both business and economic theory deal with “the economy” in a much more direct way than the service areas in which professionalism is duly recognized. But, why does economic theory model individuals as maximizing personal financial gain? We turn next to this question.

## **5 Why Does Economic Theory Model the Individual as *Homo economicus*?**

Economic theory models individuals as maximizing personal financial gain because in anonymous market exchange individuals generally act precisely in that way. The great strength of traditional economic theory lies precisely in its capacity to explain competitive markets. For this reason, our critique of the *Homo economicus* model should not be seen as a wholesale rejection of economic theory as relevant for business education. For instance, neoclassical economic theory holds that in a market for a product, the equilibrium price is at the intersection of the supply and demand curves for the good. Indeed, it is easy to see that at any other point a self-regarding seller could gain by asking a higher price, or a self-regarding buyer could gain by offering a lower price. This situation was among the first to be approached experimentally, *the neoclassical prediction virtually always receiving strong support* (Smith 1962, Smith 1982, Holt 1995). Here is a particularly dramatic example, provided by Holt, Langan and Villamil (1986) (reported by Charles Holt in Kagel and Roth, 1995).

In the Holt, Langan, and Villamil experiment there are four “buyers” and four “sellers.” The good is a chip that the seller can redeem for \$5.70 but the buyer can redeem for \$6.80 at the end of the game. In analyzing the game, we assume throughout that buyers and sellers maximize personal financial gain. In each of the first five rounds, each buyer was informed, privately, that he could redeem up to four chips, while eleven chips were distributed to sellers (three sellers were given

three chips each, and the fourth was given two chips). Clearly, buyers are willing to pay up to \$6.80 per chip for up to four chips each, and sellers are willing to sell their chip for any amount at or above \$5.70. Total demand is thus sixteen for all prices at or below \$6.80, and total supply is eleven chips at or above \$5.70. Since there is an excess demand for chips at every price between \$5.70 and \$6.80, the only point of intersection of demand and supply curves is at the price  $p = \$6.80$ . The subjects in the game, however, have absolutely no knowledge of aggregate demand and supply, since each knew only his own supply of or demand for chips.



**Figure 1:** Simulating a Market Equilibrium: The Double Auction. The size of the circle is proportional to the number of trades that occurred at the stated price.

The rules of the game are that at any time a seller can call out an asking price for a chip, and a buyer can call out an offer price for a chip. This price remains “on the table” until either it is accepted by another player, or a lower asking price is called out, or a higher offer price is called out. When a deal is made, the result is recorded and that chip is removed from the game. As seen in Figure 1, in the first period of play, actual prices were about midway between \$5.70 and \$6.80. Over the succeeding four rounds the average price increased, until in period 5, prices were very close to the equilibrium price predicted by neoclassical theory.

In period six and each of the succeeding four periods, buyers were given the right to redeem a total of eleven chips, and each seller was given four chips. In this new situation, it is clear (to us) that there is an excess supply of chips at each price between \$5.70 and \$6.80, so the only place supply and demand intersect is at \$5.70. While sellers, who previously made a profit of about \$1.10 per chip in each period, must have been delighted with their additional supply, succeeding periods witnessed a steady fall in price, until in the tenth period, the price is close to the neoclassical

---

prediction, and now the buyers are earning about \$1.10 per chip. A more remarkable vindication of the neoclassical model would be difficult to imagine.

## 6 Expanding the Rational Actor Model: Beliefs, Preferences, and Constraints

Perhaps because of the stunning success of the twin principles of rationality and self-interest in the analysis of market behavior, economists are wont to conflate the two concepts. The great economist Francis Ysidro Edgeworth (1881) proclaimed at the dawn of the neoclassical era that “The first principle of economics is that every agent is actuated only by self-interest.” Indeed, at least until recently, economists have considered maximizing material gain to be an essential aspect of rationality. In fact, however, the concepts are quite distinct.

The economist’s rational actor model can be shown to apply over any domain in which the individual has *consistent preferences*, in the sense that if he prefers A to B and he prefers B to C, then he prefers A to C. Given this assumption and a few technical conditions, it is easy to show that we can model the individual as *maximizing a preference function* subject to constraints. Moreover, if the individual’s choice space is sufficiently rich, a few simple additional assumptions allows us to conclude that the individual has *consistent beliefs* concerning the probability of the states of affairs that determine his payoffs (Kreps 1990). Because of the minimal requirements that “rationality” imposes on preferences and beliefs, and because the term “rational” has so many diverse meanings in philosophy, the humanities, and the behavioral sciences, we shall abandon the traditional term rational actor model in favor of the more descriptively accurate term, the beliefs, preferences and constraints (BPC) model.

The Nobel laureate economist Gary Becker pioneered the extension of the beliefs, preferences, and constraints model to such non-market areas as drug addiction, racial discrimination, crime and punishment, the family, education, and fertility (Becker 1976, Becker, Tommasi and Ierulli 1995). While Becker and his coworkers virtually identify rationality with self-interest, it is clear that consistent preferences are compatible with the application of ethical values to individual choice. This point has been made with great clarity by Andreoni and Miller (2002), which we will now describe.

In the *dictator game*, the experimenter gives subject 1 a certain amount of money and instructs subject 1 to give any portion of this he desires to an anonymous subject 2. Subject 1, the dictator, keeps whatever he does not choose to give to subject 2, the recipient. Obviously, a selfish dictator will give nothing to the recipient. Suppose the experimenter gives the dictator  $m$  dollars, and tells him that the price of giving

---

some of this to the recipient is  $p$ , meaning that each point the recipient gets costs the dictator  $p$  points. For instance, if  $p = 4$ , then it costs the dictator four points for each point that he transfers to the recipient. The dictator's choices must then satisfy the *budget constraint*  $\pi_s + p\pi_o = m$ , where  $\pi_s$  is the amount the dictator keeps and  $\pi_o$  is the amount the recipient gets. The question, then, is simply, is there a preference function  $u(\pi_s, \pi_o)$  that the dictator maximizes subject to the budget constraint  $\pi_s + p\pi_o = m$ ? If so, then it is just as rational, from a behavioral standpoint, to behave charitably toward the recipient as it is to behave selfishly.

Economist Hal Varian (1982) showed that the following Generalized Axiom of Revealed Preference (GARP) is sufficient to ensure not only rationality, but that individuals have nonsatiated, continuous, monotone, and concave utility functions—the sort found in traditional consumer demand theory. To define GARP, suppose the individual purchases bundle  $x_i$  when prices are  $p_i$ . We say consumption bundle  $x_s$  is *directly revealed preferred* to bundle  $x_t$  if  $p_s x_t \leq p_s x_s$ ; i.e.,  $x_t$  could have been purchased when  $x_s$  was purchased. We say  $x_s$  is *indirectly revealed preferred* to  $x_t$  if there is some sequence  $x_s = x_1, x_2, \dots, x_k = x_t$ , which each  $x_i$  is directly revealed preferred to  $x_{i+1}$  for  $i = 1, \dots, k - 1$ . GARP then is the following condition: If  $x_s$  is indirectly revealed preferred to  $x_t$ , then  $p_t x_t \leq p_t x_s$ ; i.e.  $x_s$  does not cost less than  $x_t$  when  $x_s$  is purchased.

The experimenters worked with 176 students in a elementary economics class, and had them play the dictator game eight times each, with the price  $p$  taking on the values  $p = 0.25, 0.33, 0.5, 1, 2, 3$ , and 4, with amounts of tokens equalling  $m = 40, 60, 75, 80$  and 100. They found that only 18 of the 176 subjects violated GARP at least once, and of these, only four were at all significant violations. By contrast, if choices were randomly generated, we would expect between 78% and 95% of subjects would violate GARP.

As to the degree of altruistic giving in this experiment, Andreoni and Miller found that 22.7% of subjects were perfectly selfish, 14.2% were perfectly egalitarian at all prices, while 6.2% always allocated all the money so as to maximize the total amount won (i.e., when  $p > 1$  they kept all the money, and when  $p < 1$ , they gave all the money to the recipient).

We conclude from this study that *we can treat altruistic preferences in a manner perfectly parallel to the way we treat money and private goods* in individual preference functions. There is thus nothing “irrational” about having other-regarding preferences when using the BPC model.

## 7 Strong Reciprocity in the Market for Managers

---

*Strong reciprocity* is the predisposition to cooperate in a group task, to respond to the cooperative behavior of others by maintaining or increasing one's level of cooperation, and to respond to noncooperative behavior of others by punishing the offenders, even at personal cost, and even when one cannot not reasonably expect future personal gains to flow from such punishment. When other forms of punishment are not available, the strong reciprocator responds to defection with defection.

The strong reciprocator is thus neither the selfless altruist of utopian theory, nor the selfish *Homo economicus* of neoclassical economics. The cooperative aspect of strong reciprocity is commonly known as *gift exchange* and the punitive side is *altruistic punishment*, the altruistic element coming from the fact that strong reciprocators promote cooperation even in groups with many selfish players, by making cooperation incentive compatible for such players (i.e., it costs less to cooperate than to defect and get punished).

Akerlof (1982) suggested that many puzzling facts about labor markets could be better understood if it were recognized that in many situations, employers pay their employees higher wages than necessary, in the expectation that workers will respond by providing higher effort than necessary. Of course, the relationship between employer and employee is of the same principal-agent type as that between stockholder and manager. Hence, if Akerlof is correct, we would equally expect trust and reciprocity to play a role in the market for managers. Fehr, Gächter and Kirchsteiger (1997) performed an experiment that validates this *gift exchange* model of the market for managers (see also Fehr and Gächter 1998).

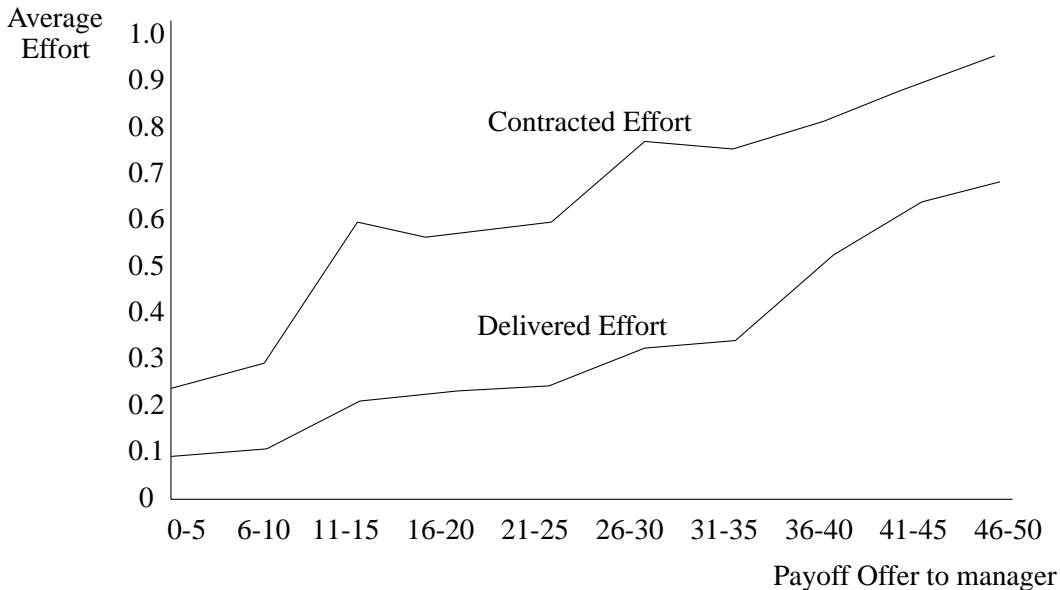
The experimenters divided a group of 141 subjects (college students who had agreed to participate in order to earn money) into “owners” and “managers.” The rules of the game are as follows. If an owner hires a manager who provides effort  $e$  and receives a salary  $s$ , his profit is  $\pi = 100e - s$ , just as in Section 4. The payoff to the manager is then  $u = s - c(e)$ , where  $c(e)$  is the same cost of effort function as in Section 4.

The sequence of actions is as follows. The owner first offers a “contract” specifying a salary  $s$  and a desired amount of effort  $e^*$ . A contract is made with the first manager who agrees to these terms. An owner can make a contract  $(s, e^*)$  with at most one manager. The manager who agrees to these terms receives the salary  $s$  and supplies an effort level  $e$ , which *need not equal the contracted effort*,  $e^*$ . In effect, there is no penalty if the manager does not keep his promise, so the manager can choose any effort level,  $e \in [0.1, 1]$ , with impunity. Although subjects may play this game several times with different partners, each owner-manager interaction is a one-shot (non-repeated) event. Moreover, the identity of the interacting partners is never revealed.

If managers maximize personal financial gain, they will choose the zero-cost

effort level,  $e = 0.1$ , no matter what salary is offered them. Knowing this, owners will never pay more than the minimum necessary to get the manager to accept a contract, which is 1 (assuming only integral salary offers are permitted).<sup>2</sup> The manager will accept this offer, and will set  $e = 0.1$ . Since  $c(0.1) = 0$ , the manager's payoff is  $u = 1$ . The owner's payoff is  $\pi = 0.1 \times 100 - 1 = 9$ .

In fact, however, this self-regarding outcome rarely occurred in this experiment. The average net payoff to managers was  $u = 35$ , and the more generous the owner's salary offer to the manager, the higher the effort provided. In effect, owners presumed the strong reciprocity predispositions of the managers, making quite generous salary offers and receive higher effort, as a means to increase both their own and the manager's payoff, as depicted in Figure 2. Similar results have been observed in Fehr, Kirchsteiger and Riedl 1993, 1998.



**Figure 2:** Relation of Contracted and Delivered Effort to Worker Payoff (141 subjects). From Fehr, Gächter, and Kirchsteiger (1997).

Figure 2 also shows that, though most managers are strong reciprocators, at any salary rate there still is a significant gap between the amount of effort agreed upon and the amount actually delivered. This is not because there are a few “bad apples” among the set of managers, but because only 26% of managers delivered the level of effort they promised! We conclude that strong reciprocators are inclined

<sup>2</sup>This is because the experimenters created more managers than owners, thus ensuring an excess supply of managers.



---

to compromise their morality to some extent.

To see if owners are also strong reciprocators, the authors extended the game by allowing the owners to respond reciprocally to the *actual effort choices* of their workers. At a cost of 1, an owner could *increase* or *decrease* his manager's payoff by 2.5. If owners were self-regarding, they would of course do neither, since they would not (knowingly) interact with the same worker a second time. However, 68% of the time, owners punished managers that did not fulfill their contracts, and 70% of the time, owners rewarded managers who overfulfilled their contracts. Indeed, owners rewarded 41% of managers who *exactly* fulfilled their contracts. Moreover, managers *expected* this behavior on the part of their owners, as shown by the fact that their effort levels *increased significantly* when their bosses gained the power to punish and reward them. Underfulfilling contracts dropped from 83% to 26% of the exchanges, and overfulfilled contracts rose from 3% to 38% of the total. Finally, allowing owners to reward and punish led to a 40% increase in the net payoffs to all subjects, even when the payoff reductions resulting from owner punishment of managers are taken into account.

We conclude from this study that the subjects who assume the role of "manager" conform to internalized standards of reciprocity, even when they are certain there are no material repercussions from behaving in a self-regarding manner. Moreover, subjects who assume the role of owner expect this behavior and are rewarded for acting accordingly. Finally, owners reward good and punish bad behavior when they are allowed, and managers expect this behavior and adjust their own effort levels accordingly. In general, then subjects follow an internalized norm not only because it is prudent or useful to do so, or because they will suffer some material loss if they do not, but rather because they desire to do so *for its own sake*.

## 8 Gift Exchange and Reputation with Incomplete Contracts

"An economic transaction," says Abba Lerner (1972), "is a solved political problem. Economics has gained the title of queen of the social sciences by choosing solved political problems as its domain." Lerner's observation is correct, however, only insofar as economic transactions are indeed *solved* political problems. The assumption in the theory of competitive markets that gives this result is that *all economic transactions involved contractual agreements that are enforced by third parties (e.g., the judiciary) at no cost to the exchanging parties*. However, some of the most important economic transactions are characterized by the *absence of third-party enforcement*. Among these, of course, is the relationship between the stockholders and the managers of a large corporation.

In the previous section we showed that most experimental subjects are strong

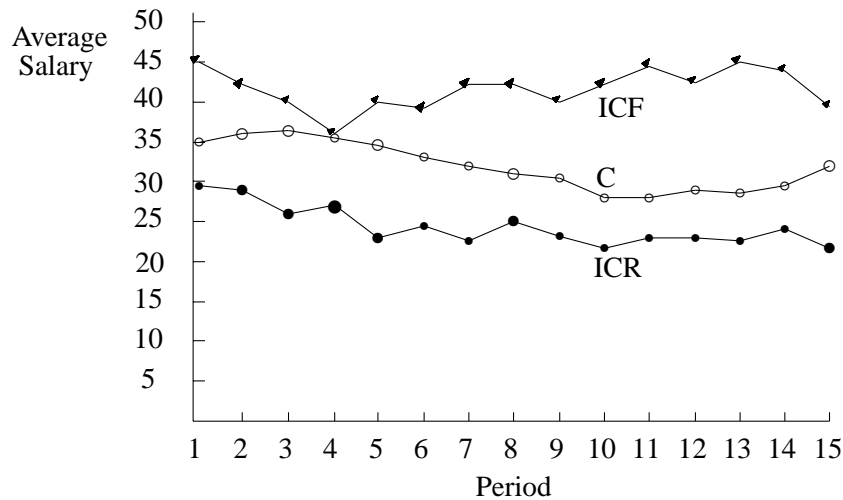
---

reciprocators who offer a high level of effort, provided they trust their employers to reward their successes, without the need for a complete, exogenously enforced contract. In this section we present another experiment, conducted by Brown, Falk and Fehr (2004), that shows clearly that when contracts are incomplete, owners prefer to establish long-term relationships with managers based on trust and gift exchange rather than contractually aligning the incentives of managers with their own. The result in this experiment is a market for managers dominated by long-term trust relationships.

Brown, Falk, and Fehr (BFF) used 15 trading periods with 238 subjects and three treatments. The first treatment was the standard complete contract condition (C condition) in which managerial effort is contractually specified. The second treatment was an incomplete contract condition (ICF condition) with exactly the same characteristics, including costs and payoffs to owner and manager, as in Section 7. In addition, however, managers were given a payment of 5 points in each period that they were out of work. In both conditions, subjects had identification numbers that allow long-term relationships to develop. The third treatment, which we call ICR, was identical to ICF, except that long-term relationships were ruled out (subjects received shuffled identification numbers in each experimental period). This treatment is thus identical to the gift exchange model in Section 7, except for the 5 point “unemployment compensation.”

All contracts formally lasted only one period, so even long-term relationships had to be explicitly renewed in each period. If managers are self-regarding, it is easy to see that in the ICR treatment, all managers will supply the lowest possible effort  $e = 1$ , and owners will offer salary  $s = 5$ . Each firm then has a profit of  $10e - 5 = 5$ , and each manager has payoff  $s - c(e) = 5 - c(0) = 5$ . This outcome will also occur in the last period of the ICF treatment, and hence by backward induction, will hold in all periods. In the C treatment with self-regarding agents, it is easy to show that the owner will set  $s = 23$  and require  $e = 10$ , so managers get  $s - c(e) = 23 - c(10) = 5$  and owners get  $10e - s = 100 - 23 = 77$  in each period. Managers are, in effect indifferent between being employed and unemployed in all cases.

The actual results were, not surprisingly, quite at variance with the self-regarding preferences assumption. Figure 3 shows the path of salaries over the fifteen periods under the three treatments. The ICR condition reproduces the result of Section 7, salaries being consistently well above the self-regarding level of  $s = 5$ . If the C condition were a two-sided double auction, we would expect salaries to converge to  $s = 23$ . The ICF condition gives the highest salaries after the fourth period, validating the claim that under conditions of incomplete contracting, long-term trust and gift exchange relationships will prevail, and the distribution of gains will be more equal between buyers and sellers.

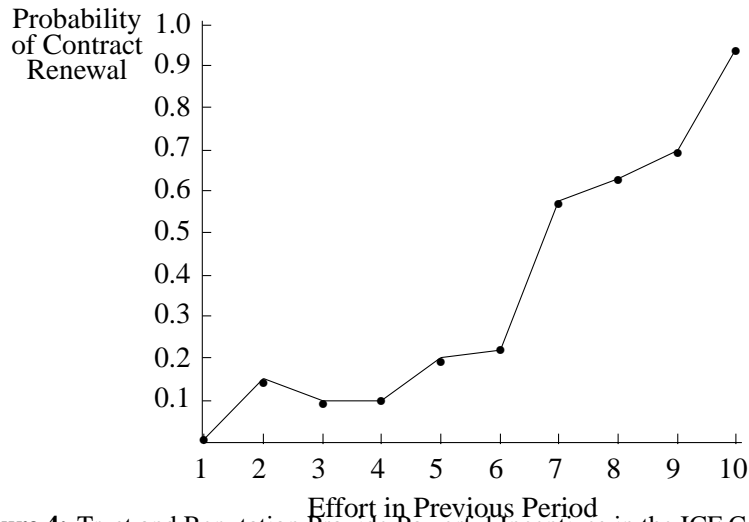


**Figure 3:** Salaries over Fifteen Periods (Brown et al. 2004). The C treatment is complete contracting, the ICF treatment is incomplete contracting with long-term relationships permitted, and the ICR treatment is incomplete contracting with no long-term relationships permitted.

By paying high salaries in the ICF condition, owners were capable of effectively threatening their managers with dismissal (non-renewal of contract) if they were dissatisfied with managerial performance. Figure 4 shows that this threat was in fact often exercised. Managers with effort close to  $e = 10$  were non-renewed only about 5% of the time, whereas managers with effort below  $e = 7$  were rarely renewed.

Figure 5 shows that the effect of different contracting availabilities strongly affects the level of productivity of the system, as measured by average effort levels. Under complete contracting, effort levels quickly attain near-efficiency ( $e = 10$ ), and remain there. Contingent renewal of long-term relationships achieves between 80% and 90% efficiency, with a significant end-game effect, as the threat of non-renewal is not very effective on the last few rounds. The gift exchange treatment (ICR), while supporting effort levels considerably above the self-regarding level, is considerably less efficient than either of the others, although it predictably suffers a smaller end-game effect than the ICF condition.

One extremely interesting pattern emerging from this study is the interaction of gift exchange and threat in the owner-manager relationship. One might think that they would be mutually exclusive, on the grounds that one cannot both feel charitable towards one's employer while at the same time being threatened by him. Yet, many of us will recall from personal experience this ambiguous co-presence of good will and fear. In this study, the importance of gift exchange in the long-term



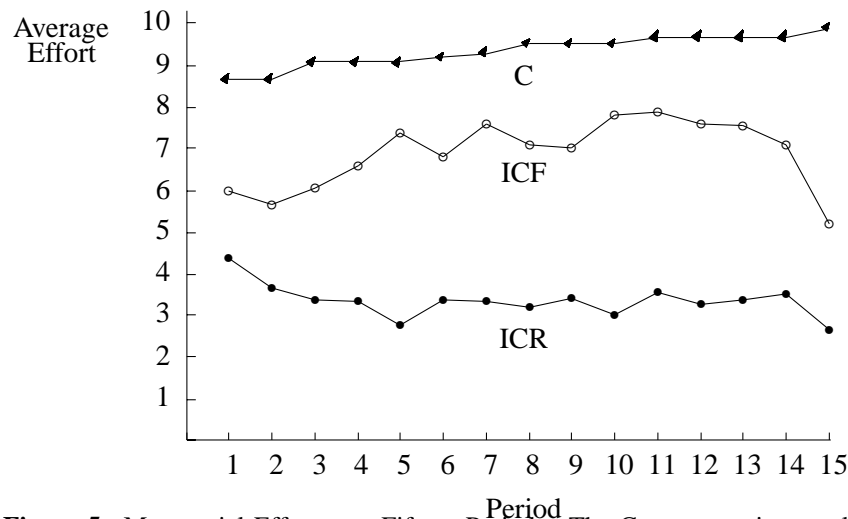
**Figure 4:** Trust and Reputation Provide Powerful Incentives in the ICF Condition

relationship is exhibited by the fact that even in the last two periods, where the threat of dismissal is weak or absent, effort levels are considerably above those of the pure gift exchange condition. Thus, gift exchange appears to be stronger when accompanied by the capacity of the owner to harm, as though the fact that the owner has not exercised this capacity increases the willingness to supply effort.

## 9 The Character Virtues

*Character virtues* are ethically desirable behavioral regularities that individuals value for their own sake, while having the property of facilitating cooperation and enhancing social efficiency. The character virtues include *honesty*, *trustworthiness*, *promise-keeping*, and *fairness*. Unlike such other-regarding preferences as strong reciprocity and empathy, these character virtues operate without concern for the individuals with whom one interacts. An individual is honest in his transactions because this is a desired state of being, not because he has any particular regard for those with whom he transacts. Of course, the sociopath *Homo economicus* is honest only when it serves his material interests to be so, whereas the rest of us are at times honest even when it is costly to be so, and even when no one but ourselves could possibly detect a breach.

Common sense, as well as the experiments described below, indicate that honesty, fairness, and promise-keeping are not absolutes. If the cost of virtue is sufficiently high, and the probability of detection of a breach of virtue is sufficiently small, some if not most individuals will behave dishonestly. When one is aware that



**Figure 5:** Managerial Effort over Fifteen Periods. The C treatment is complete contracting, the ICF treatment is incomplete contracting with long-term trust and reputation relationships permitted, and the ICR treatment is incomplete contracting with no long-term relationships permitted.

others are not virtuous in a particular region of their lives (e.g., marriage, tax paying, obeying traffic rules, accepting bribes), one is more likely to allow one’s own virtue to lapse. Finally, the more easily one can delude oneself into inaccurately classifying a non-virtuous act as virtuous, the more likely will one allow oneself to carry out such an act.

One might be tempted to model honesty and other character virtues as *self-constituted constraints* on one’s set of available actions in a game, but a more fruitful approach is to include the state of being virtuous in a certain way as an argument in one’s preference function, to be traded off against other valuable objects of desire and personal goals. In this respect, the character virtues are in the same category as ethical and religious preferences, and are often considered subcategories of the latter.

The importance of character virtues, which appear to be confined to our species, is ignored in neoclassical economics, which prefers to ground social cooperation in *enlightened self-interest*. This predilection goes back to Bernard Mandeville’s “private vices, public virtues” (1924[1705]) and Adam Smith’s “invisible hand” (2000[1759]). The Darwinian struggle for existence may explain why the concept of virtue does not add to our understanding of animal behavior in general, but by all available evidence, it is a central element in human behavior. The reasons for this are the subject of some speculation (Gintis 2003a,2006b), but they come

---

down to the plausible insight that human social life is so complex, and the rewards for prosocial behavior so distant and indistinct, that adherence to general rules of propriety, including the strict control over of such deadly sins of anger, avarice, gluttony, and lust is individually fitness-enhancing (Simon 1990).

Numerous experiments indicate that most subjects are willing to sacrifice material reward to maintain a virtuous character, even when outsiders cannot penetrate the soul to investigate its contents. Sally (1995) undertook a meta-analysis of 137 experimental treatments, finding that face-to-face communication, in which subjects are capable of making verbal agreements and promises, was the strongest predictor of cooperation. Of course, face-to-face interaction violates anonymity and has other effects besides the ability to make promises. However, both Bochet, Page and Putterman (2006) and Brosig, Ockenfels and Weimann (2003) report that only the ability to exchange verbal information accounts for the increased cooperation.

For example, consider the following *trust game*, first studied by Berg, Dickhaut and McCabe (1995). In this game subjects are each given a certain endowment, say \$10. Subjects are then randomly paired, and one subject in each pair, player A, is told he can transfer any number of dollars, from zero to ten, to his (anonymous) partner, player B, and keep the remainder. The amount transferred will be *tripled* by the experimenter and given to player B, who can then give any number of dollars back to player A (this amount is not tripled). A player A who transfers a lot is called *trusting*, and a player B who returns a lot to player A is called *trustworthy*. In the terminology of this paper, a trustworthy player B is a strong reciprocator, and a trusting player A is an individual who expects his partner to be a strong reciprocator.

Clearly, if all individuals have self-regarding preferences, and if each player A believes his partner has self-regarding preferences, then player A will give nothing to player B. On the other hand, if player A believes player B is inequality averse, he will transfer all \$10 to player B, who will then have \$40. To avoid inequality, player B will give \$20 back to player A. A similar result would obtain if player A believes player B is a strong reciprocator. On the other hand, if player A is altruistic, he may transfer something to player B, on the grounds that the money is worth more to player B (since it is tripled) than it is to himself, even if player A does not expect anything back. It follows that several distinct motivations can lead to a positive transfer of money from A to B and then back to A.

Berg et al. (1995) found that on average, \$5.16 was transferred from A to B, and on average, \$4.66 was transferred back from B's to A's. Furthermore, when the experimenters revealed this result to the subjects and had them play the game a second time, \$5.36 was transferred from A's to B's, and \$6.46 was transferred back from B's to A's. In both sets of games there was a great deal of variability, some player A's transferring everything, some nothing, and some player B's more than fully repaying their partner, and some giving back nothing.

---

As noted above, the trust game does not really draw on character virtues because Player B has made no explicit promises to Player A. To add a dose of character virtue to the mix, Ben-Ner and Putterman (2005) first implemented a standard trust game with 194 college student subjects. Each player was given \$10, and Player A was allowed to pass as much of his \$10 to Player B, the amount being first tripled by experimenter. Player B was then permitted to return to Player A whatever fraction of his newly-acquired money he pleased. Just as reported above, even without communication, most subjects exhibited some level of cooperative behavior, and indeed, even a strictly selfish Player A stood to gain from passing all of his money to Player B. The experimenters then ran the game a second time, now allowing newly-paired partners to communicate proposals and counterproposals *via* their computer screens prior to playing the one-shot, anonymous trust game. The results was much more cooperation, the average amount sent increasing from \$5.50 to \$8.10, despite that fact that there was no way a Player B could be punished for breaking a promise. Player A's were justified in trusting, since in fact 74% of Player B's actually kept their promises, and most of the other at least returned something to Player A.

One striking finding in this study is that when Player B is allowed to make a promise, the modal exchange (i.e., the most common) was that in which Player A sent all \$10 to Player B, who then kept \$10 of the \$30 given to him by the experimenter, returning \$20 to Player A, so that both parties ended up with \$20.

## **10 The Role of Community: Altruistic Third Party Reward and Punishment**

Prosocial behavior in human society occurs not only because those directly helped and harmed by an individual's actions are likely to reciprocate in kind but also because there are general *social norms* that foster prosocial behavior and many people are willing to bestow favors on someone who conforms to social norms, and to punish someone who does not, even if they are not personally helped or hurt by the individual's actions. In everyday life, third parties who are not the beneficiaries of an individual's prosocial act, will help the individual and his family in times of need, will preferentially trade favors with the individual, and otherwise will reward the individual in ways that are not costly but are nonetheless of great benefit to the cooperator. Similarly, individuals who have not been personally harmed by the selfish behavior of an another individual will refuse to aid the offender, even when it is not costly to do so, and will shun the offender and approve of the offender's ostracism from beneficial group activities.

By *community culture* we mean the sharing of cultural norms among community members and the willingness of community members to reward and punish com-

---

munity members based on their adherence to and violation of these norms, even if they are not personally the object of the resulting beneficence or harm. Professional standards are the cultural norms of occupational groups, and once established, lead members to reward and punish each other informally, without the need for costly litigation, legislation, and incarceration. By abjuring professional standards for managers in favor of a culture of greed, it is likely that business schools that have promoted the neoclassical model of stockholder-manager relations have so undercut the culture of professional honor among managerial personnel that the mechanism of informal third-party punishment and reward has sunk to dramatically low levels, thus contributing to a deficit in moral behavior on the part of contemporary managerial personnel.

There is strong experimental evidence supporting this line of reasoning. While self-regarding actors will never engage in third party reward or punishment if such behavior is at all costly, experimental subjects routinely punish and reward others who have impacted upon the welfare of other group members, although not themselves. An experiment conducted by Fehr and Fischbacher (2004) addresses this issue by conducting a series of third party punishment games using prisoner's dilemma and dictator games.<sup>3</sup> The experimenters implemented four experimental treatments, in each of which subjects were grouped into threes. In each group, in stage one, subject A played a prisoner's dilemma or dictator game with subject B as the recipient, and subject C was an outsider whose payoff was not affected by A's decision. Then, in stage two, subject C was endowed with 50 points and allowed to deduct points from subject A, such that every three points deducted from A's score cost C one point. In the first treatment (TP-DG) the game was the dictator game, in which A was endowed with 100 points, and could give 0, 10, 20, 30, 40, or 50 points to B, who had no endowment.

The second treatment (TP-PD) was the same, except that the game was the prisoner's dilemma. Subjects A and B were each endowed with ten points, and each could either keep the ten points, or transfer the ten points to the other subject, in which case it was tripled by the experimenter. Thus, if both cooperated, each earned 30 points, and if both defected, each earned ten points. If one cooperated and one defected, however, the cooperator earned zero and the defector 40 points. In the second stage, C was given an endowment of 40 points, and was allowed to deduct points from A and/or B, just as in the TP-DG treatment.

To compare the relative strengths of second and third party punishment in the dictator game, the experimenters implemented a third treatment, S&P-DG. In this

---

<sup>3</sup>The prisoner's dilemma is a two-player game in which both players gain if they cooperate, but each player has an incentive to defect no matter what the other player does. Thus, self-regarding players never cooperate in a prisoner's dilemma, although experimental evidence indicates that a majority of experimental subjects actively prefer to cooperate, provided their partners do so as well.



---

treatment, subjects were randomly assigned to player A and player B, and A-B pairs were randomly formed. In the first stage of this treatment, each A was endowed with 100 points and each B with none, and the As played the dictator game as before. In the second stage of each treatment, each player was given an additional 50 points, and the B players were permitted to deduct points from A players on the same terms as in the first two treatments. S&P-DG also had two conditions. In the S condition, a B player could only punish his *own* dictator, whereas in the T condition, a B player could only punish an A player *from another pair*, to which he was randomly assigned by the experimenters. In the T condition, each B player was informed of the behavior of the player A' to which he was assigned.

To compare the relative strengths of second and third party punishment in the prisoner's dilemma, the experimenters implemented a fourth treatment, S&P-PG. This was similar to the S&P-DG treatment, except that now in the S condition, an AB pair could only punish each other, whereas in the T condition, each agent could punish only a randomly assigned subject from another pair.<sup>4</sup>

In the first two treatments, since subjects were randomly assigned to positions A, B, and C, the obvious fairness norm is that all should have equal payoffs (an "equality norm"). For instance, if A gave 50 points to B, and C deducted no points from A, each subject would end up with 50 points. In the dictator game treatment (TP-DG), 60% of third parties (C's) punish dictators (A's) who give less than 50% of the endowment to recipients (B's). Statistical analysis (ordinary least squares regression) showed that for every point an A kept for himself above the 50-50 split, he was punished an average 0.28 points by C's, leading to a total punishment of  $3 \times 0.28 = 0.84$  points. Thus, a dictator who kept the whole 100 points would have  $0.84 \times 50 = 42$  points deducted by C's, leaving a meager gain of 8 points over equal sharing.

The results for the prisoner's dilemma treatment (TP-PD) were similar, with an interesting twist. If one partner in the AB pair defected and the other cooperated, the defector would have on average 10.05 points deducted by C's, but if both defected, the punished player lost only an average of 1.75 points. This shows that third parties (C's) care not only about the intentions of defectors, but how much harm they caused and/or how unfair they turned out to be. Overall, 45.8% of third parties punished defectors whose partners cooperated, whereas only 20.8% of third parties punished defectors whose partners defected.

Turning to the third treatment (T&SP-DG), second party sanctions of selfish dictators are found to be considerably stronger than third party sanctions, although both were highly significant. On average, in the first condition, where recipients

---

<sup>4</sup>It is worth repeating that the experimenters never use value-laden terms such as "punish," but rather neutral terms, such as "deduct points."

---

could punish their own dictators, they imposed a deduction of 1.36 points for each point the dictator kept above the 50-50 split, whereas they imposed a deduction of only 0.62 points per point kept on third party dictators. In the final treatment (T&SP-PD), defectors are severely punished by both second and third parties, but second party punishment is again found to be much more severe than third. Thus, cooperating subjects deducted on average 8.4 points from a defecting partner, but only 3.09 points from a defecting third party.

This study confirms the general principle that punishing norm violators is very common but not universal, and individuals are prone to be more harsh in punishing those who hurt them personally, as opposed to violating a social norm that hurts others than themselves.

## 11 Re-professionalizing Business Education

Neoclassical economic theory forms the central discourse and behavioral model of contemporary management education. Drawing on research and insights from game theory and behavioral economics we have argued that many of the core assumptions underlying this model are flawed. While we cannot say that the widespread reliance on the *Homo economicus* model has caused the highly level of observed managerial malfeasance, it may well have, and it surely does not act as a healthy influence on managerial morality. Students have learned this flawed model and in their capacity as corporate managers, doubtless act daily in conformance with it. This, in turn, may have contributed to the weakening of socially functional values and norms like honesty, integrity, self-restraint, reciprocity and fairness, to the detriment of the health of the enterprise. Simultaneously, this perspective has legitimized, or at least not delegitimized, such behaviors as material greed and optimizing with guile. We noted that this model has become highly institutionalized in business education. Fortunately, we believe that the potential for moving away from this flawed model is significant and thus can end this chapter on a more optimistic note for the future of business education.

Recently business school scholars like Jensen (2006) have given thought about how to introduce the notion of the character virtue as a central element of economic value creation. Jensen has recently proposed a framework for value creation that resonates with one of the key character virtues associated with professionalism. The authors argue that *integrity* is a necessary condition to the maximizing of value. An economic entity has integrity when it is “whole and complete and stable.” Jensen then defines the behavioral characteristics of economic entities (agents inside organizations, groups inside organizations, and organizations themselves) that are in integrity mode: “nothing is hidden, no deception, no untruths, no violation of

---

contracts or property rights, etc.”For those who chose not to play by these rules, integrity requires you to make this clear to all others. This framework is then applied to financial and capital markets and suggests that many of the recent corporate scandals can be traced to the institutions, like investment banks, analysts, and auditors, purposefully setting aside the integrity imperative.

Recent research has demonstrated that business schools have a rich treasure of wisdom, idealism and vision on which to draw to move away from this dominant model (Khurana et al. 2005). It is an inheritance whose roots can be traced to the mission of the larger university as social institution charged with advancing the public good and one rooted in the original professionalizing mission of the university-based business school charter. The notion that those who lead and manage our society’s major private economic institutions might provide, or be responsible for providing, a public good is quite foreign to our current way of thinking about management. Yet, as alluded to in the introduction of this chapter, this idea was often voiced by those who led American business schools in the early decades of their existence. For example, in a speech titled “The Social Significance of Business” that he delivered at Stanford University’s School of Business shortly after its founding in 1925 (subsequently published as an article in Harvard Business Review of the same title), Wallace B. Donham, the second dean of the Harvard Business School, declared that the “ development, strengthening, and multiplication of socially minded business men is the central problem of business.” As Donham went on to say:

The socializing of industry from within on a higher ethical plane, not socialism nor communism, not government operation nor the exercise of the police power, but rather the development from within the business group of effective social control of those mechanisms which have been placed in the hands of the race through all the recent extraordinary revolutionizing of material things, is greatly needed. The business group largely controls these mechanisms and is therefore in a strategic position to solve these problems. Our objective therefore, should be the multiplication of men who will handle their current business problems in socially constructive ways.

Business schools did not begin with neo-classical theory. The founders of business schools never envisioned the notion that the sole purpose of the corporation was to serve only one master, the shareholder. Nor could they have ever imagined that the model that students would be trained into a world view that would conceive managers as self-interested agents with no consideration of any other values or imperatives but their own wallets. Those of us who study society, coach soccer, help our neighbors, raise children, and look within ourselves, all know such a view of human behavior is not true, and more important, if such a vision were fully realized,

---

would not offer the foundation for creating a sustainable society. Business schools need to recover what the sociologist Ferdinand Tönnies called *Gemeinschaft* or *community*. To this end, business school faculties and deans have an institutional responsibility to socialize students to a model of behavior that inspires them to respect other institutions in society, especially basic units like the family and community, and to inspire students to accept the responsibilities and obligations that come with occupying society's most powerful positions.

#### REFERENCES

- Akerlof, George A., "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics* 97,4 (November 1982):543–569.
- Andreoni, James and John H. Miller, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica* 70,2 (2002):737–753.
- Becker, Gary, *The Economic Approach to Human Behavior* (Chicago: The University of Chicago Press, 1976).
- , Mariano Tommasi, and Kathryn Ierulli, *The New Economics of Human Behaviour* (Cambridge: Cambridge University Press, 1995).
- Ben-Ner, Avner and Louis Putterman, "Trust Versus Contracting: Who Takes the Plunge, and When?," 2005. University of Minnesota.
- Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social History," *Games and Economic Behavior* 10 (1995):122–142.
- Bochet, Olivier, Talbot Page, and Louis Putterman, "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior and Organization* (2006).
- Brosig, J., A. Ockenfels, and J. Weimann, "The Effect of Communication Media on Cooperation," *German Economic Review* 4 (2003):217–242.
- Brown, Martin, Armin Falk, and Ernst Fehr, "Relational Contracts and the Nature of Market Interactions," *Econometrica* 72,3 (May 2004):747–780.
- Edgeworth, Francis Ysidro, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London: Kegan Paul, 1881).
- Fama, Eugene F. and Michael Jensen, "Separation of Ownership and Control," *Journal of Law and Economics* 26 (1983):301–326.
- Fehr, Ernst and Simon Gächter, "How Effective are Trust- and Reciprocity-based Incentives?," in Louis Putterman and Avner Ben-Ner (eds.) *Economics, Values and Organizations* (New York: Cambridge University Press, 1998) pp. 337–363.

- 
- and Urs Fischbacher, “Third Party Punishment and Social Norms,” *Evolution & Human Behavior* 25 (2004):63–87.
- , Georg Kirchsteiger, and Arno Riedl, “Does Fairness Prevent Market Clearing?,” *Quarterly Journal of Economics* 108,2 (1993):437–459.
- , — , and — , “Gift Exchange and Reciprocity in Competitive Experimental Markets,” *European Economic Review* 42,1 (1998):1–34.
- , Simon Gächter, and Georg Kirchsteiger, “Reciprocity as a Contract Enforcement Device: Experimental Evidence,” *Econometrica* 65,4 (July 1997):833–860.
- Ferraro, Fabrizio, Jeffrey Pfeffer, and Robert I Sutton, “Economics Language and Assumptions: How Theories become Self-fulfilling,” *Academy of Management Review* 30,1 (2005):8–24.
- Ghoshal, Sumantra, “Bad Management Theories are Destroying Good Management Practices,” *Academy of Management Learning & Education* 4,1 (2005):75–91.
- Gintis, Herbert, “The Hitchhiker’s Guide to Altruism: Genes, Culture, and the Internalization of Norms,” *Journal of Theoretical Biology* 220,4 (2003):407–418.
- , “A Framework for the Unification of the Behavioral Sciences,” *Behavioral and Brain Sciences (In Press)* (2006).
- Goshal, Sumantra and Peter Moran, “Bad for Practice: A Critique of the Transaction Cost Theory,” *Academy of Management Review* 21 (1996):13–47.
- Hauk, Esther and Maria Saez-Marti, “On the Cultural Transmission of Corruption,” *Journal of Economic Theory* 107,2 (2002):311–335.
- Holt, Charles A., *Industrial Organization: A Survey of Laboratory Research* (Princeton, NJ: Princeton University Press, 1995).
- , Loren Langan, and Anne Villamil, “Market Power in an Oral Double Auction,” *Economic Inquiry* 24 (1986):107–123.
- Jensen, Michael, *Foundations of Organizational Strategy* (Cambridge, MA: Harvard University Press, 1998).
- Jensen, Michael C., “Putting Integrity into Finance Theory and Practice: A Positive Approach,” mar 2006. Harvard Business School.
- and W. H. Meckling, “The Nature of Man,” *Journal of Applied Corporate Finance* 7,2 (Summer 1994):4–19.
- and William H. Meckling, “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure,” *Journal of Financial Economics* 3 (1976):305–360.
- Kagel, J. H. and A. E. Roth, *Handbook of Experimental Economics* (Princeton, NJ: Princeton University Press, 1995).

- 
- Khurana, Rakesh, Nitin Nohria, and Daniel Penrice, "Management as a Profession," in Jay W. Lorch, Leslie Berlowitz, and Andy Zelleke (eds.) *Restoring Trust in American Business* (Cambridge: The MIT Press, 2005).
- Kreps, David M., *A Course in Microeconomic Theory* (Princeton, NJ: Princeton University Press, 1990).
- Lerner, Abba, "The Economics and Politics of Consumer Sovereignty," *American Economic Review* 62,2 (May 1972):258–266.
- Mandeville, Bernard, *The Fable of the Bees: Private Vices, Publick Benefits* (Oxford: Clarendon, 1924[1705]).
- Sally, David, "Conversation and Cooperation in Social Dilemmas," *Rationality and Society* 7,1 (January 1995):58–92.
- Shiller, Robert J., "How Wall Street Learns to Look the Other Way," *New York Times* (February 8 2005).
- Simon, Herbert, "A Mechanism for Social Selection and Successful Altruism," *Science* 250 (1990):1665–1668.
- Smith, Adam, *The Theory of Moral Sentiments* (New York: Prometheus, 2000[1759]).
- Smith, Vernon, "An Experimental Study of Competitive Market Behavior," *Journal of Political Economy* 70 (1962):111–137.
- , "Microeconomic Systems as an Experimental Science," *American Economic Review* 72 (December 1982):923–955.
- Varian, Hal R., "The Nonparametric Approach to Demand Analysis," *Econometrica* 50 (1982):945–972.

c:\Papers\Corruption\Corporate Honesty-A Behavioral Model July 25, 2006