MORALITY, COMPETITION, AND THE FIRM

THE MARKET FAILURES APPROACH TO BUSINESS ETHICS

JOSEPH HEATH

Morality, Competition, and the Firm

Morality, Competition, and the Firm

THE MARKET FAILURES APPROACH TO BUSINESS ETHICS

Joseph Heath

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
Heath, Joseph, 1967–
Morality, competition, and the firm : the market failures approach to business ethics / Joseph Heath.
pages cm
ISBN 978–0–19–999048–1 (hardback)
1. Business ethics. 2. Profit—Moral and ethical aspects. 3. Competition.
4. Corporations—Moral and ethical aspects. I. Title.
HF5387.H435 2014
174′.4—dc23
2013050434

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

CONTENTS

ACKNOWLEDGMENTS

I owe many thanks to the various people who have played an important role in the development of this project. First of all, the book would never have seen the light of day without the interest, encouragement, support, and oversight of Peter Ohlin at Oxford University Press in New York. Academically, I owe the greatest debt to Wayne Norman, first for convincing me that this was an exciting and worthwhile field of study, and second for serving as a collaborator, interlocutor, and source of unflagging enthusiasm over the years. Philosophers are a somewhat truculent lot, and so I have never before had the experience of seeing exactly eye-to-eye with anyone on some complex set of philosophical issues. My exchanges with Wayne over the years have provided welcome relief from the usual solitude of philosophical inquiry.

I have learned an enormous amount from John Boatright, who always manages to live up to his reputation as "the smartest guy in the room" (in the non-pejorative sense of the term). My work in this area has also benefited a great deal from conversations over the years with Sareh Pouryousefi, Dominic Martin, Idil Boran, Pierre-Yves Néron, Chris MacDonald, and Thorsten Busch.

I have benefited from the financial support of the Social Sciences and Humanities Research Council of Canada as well as the Pierre Elliot Trudeau Foundation.

Some of these papers have been published previously. I have not made any substantive changes, other than correcting the occasional error, switching all spelling to the American, as well as changing the format of the references in some papers to create a consistent style throughout the book. Special thanks to Jessica Brown for assistance with this task. I was not able to eliminate the occasional repetitions that occur, due to the initially independent publication of the papers, and so the most I can do is apologize and ask for a measure of forbearance on the part of the reader. I would like to thank both the editors and referees of the journals for the work that they put into the improvement of these papers, as well as to the publishers for the permission to reproduce them here.

Chapter 1. "A Market Failures Approach to Business Ethics," in Bernard Hodgson, ed., *Studies in Economic Ethics and Philosophy*, Vol. 9 (Berlin: Springer, 2004).

Chapter 2. "Stakeholder Theory, Corporate Governance and Public Management," with Wayne Norman, *Journal of Business Ethics*, 53 (2004): 247–265.

Chapter 3. “Business Ethics without Stakeholders,” *Business Ethics Quarterly*, 16 (2006): 533–557. I would like to thank Wayne Norman and Alexei Marcoux for their input and advice with the writing of this paper.

Chapter 4. “An Adversarial Ethic for Business: or, When Sun-Tzu met the Stakeholder,” *Journal of Business Ethics*, 69 (2006): 359–374. Thanks to the Social Sciences and Humanities Research Council of Canada for financial support in the development of this project.

Chapter 5. “Business Ethics and the ‘End of History’ in Corporate Law,” *Journal of Business Ethics*, 102 (2011): 5–20. The idea for this paper arose during a discussion with John Boatright. It benefited from subsequent discussion with Margaret Blair, Wayne Norman, Alexei Marcoux, Eric Orts, Alan Strudler, members of the Legal Studies and Business Ethics program at the Wharton School of Business, as well as participants in the Trans-Atlantic Business Ethics Conference held at York University.

Chapter 6. “Contractualism: Micro and Macro.” This paper was written for the New York University Colloquium in Legal, Social and Political Philosophy, convened by Thomas Nagel and Ronald Dworkin. I benefited from many of the comments made by participants, but especially the careful attention given to it by Professors Nagel and Dworkin. Thanks to Shlomi Segall for comments on the manuscript as well.

Chapter 7. “Efficiency as the Implicit Morality of the Market.” This paper revises some material previously published as “Ideal Theory in an *n*-th Best World: The Case of Pauper Labor,” *Journal of Global Ethics*, 9 (2013): 159–172. My thanks to Kathryn Walker, Monique Deveaux, and participants in the “Thinking Beyond Distribution” workshop at the University of Toronto Centre for Ethics for valuable comments. This paper was also presented at the Justitia Amplificata conference on “Relational Injustice” at Goethe University; my thanks to James Gledhill in particular for the invitation to participate. Thanks to Dominic Martin for comments on the manuscript as well.

Chapter 8. “The History of the Invisible Hand.” This paper was presented at the summer school of the Netherlands School for Research in Practical Philosophy, organized by Ingrid Robeyns and Rutger Claassen on the theme of “Ethics and Economics.” My thanks to the participants for helpful feedback, as well as to Manuel Wörsdörfer for comments that led to a number of specific improvements.

Chapter 9. “The Benefits of Cooperation,” *Philosophy and Public Affairs*, 34 (2006): 313–351. I would like to thank Christopher Morris, Thomas Hurka, and Edward McClennen for encouragement at key points in the development of this article. Thanks also to audiences at the Canada School of Public Service, the University of Guelph, and the Université de Montréal, along with the Editors of *Philosophy & Public Affairs*, for helpful comments.

Chapter 10. “The Uses and Abuses of Agency Theory,” *Business Ethics Quarterly*, 19 (2009): 497–528.

Chapter 11. “Business Ethics and Moral Motivation: A Criminological Perspective,” *Journal of Business Ethics*, 83 (2008): 595–614.

Chapter 12. “Business Ethics after Virtue.” My thanks to Christine Tappolet for comments and advice on the improvement of this paper. I have also drawn some material from a paper co-authored with Wayne Norman and Jeffrey Moriarty, “Business Ethics and (or as) Political Philosophy,” *Business Ethics Quarterly*, 10 (2010): 427–452.

Chapter 13. “Reasonable Restrictions on Underwriting,” in Patrick Flanagan, Patrick Primeaux, and William Ferguson, eds., *Research in Ethical Issues in Organizations*, Vol. 7 (Amsterdam: Elsevier, 2006).

Introduction

This volume brings together a series of papers that I have written over the past ten years on the subject of business ethics, along with a few more general pieces that articulate the normative foundations of the project. Together they provide a basic outline of what I refer to as the "market failures" approach to business ethics. It was not my original intention to make a contribution to this particular literature. My objective in this introduction will therefore be to sketch out a bit of the intellectual history that led to the development of the project and to say something about how the various pieces fit together. Along the way, I hope to make clear some of the more general political motivations that inform the market failures approach, since these have occasionally been the subject of misunderstanding.

I.1. The Intellectual History of the Project

Traditional philosophical business ethics is done within an "applied ethics" paradigm, in which the theorist begins by establishing a commitment to a particular ethical theory, such as Kantianism or virtue theory, then (typically) goes on to discuss some "moral dilemma" that might arise in a business context. This leaves business ethics firmly anchored within a framework of personal ethics, or what Robert Solomon (1992) refers to as the "micro" level of institutional analysis. My own approach, by contrast, arose as something of a byproduct of a larger (or "macro") project in political economy. The best way to understand it, I believe, is to see it in the context of this more general position.

My initial interest was in the role that the state plays in a modern capitalist economy. This led me (*inter alia*) to write a popular book defending certain features of the Canadian welfare state against the ever-present pull of the American model (*The Efficient Society*, published by Penguin in 2001). One of my central preoccupations in that book was public health care, since at the time of writing the deficiencies of the American system were not as obvious as they are now, while the Canadian single-payer system was under significant strain, thanks to both cost escalation and government budget constraints. What I found frustrating about the public debate in Canada at the time was that when partisans of the public health care system were called upon to defend it, they inevitably appealed to its egalitarian qualities—the fact that it guaranteed roughly equal care to all citizens. While not having any particular objections to this argument, it also seemed to me that it should not be the first line of defense. Equality is important, but promoting equality is not the only thing that the state does. The most obvious argument in favor of single-payer health systems is that they are more efficient—the Canadian system, for instance, delivers approximately the same volume of health care services and achieves similar health outcomes as the American, but at something approaching half the cost. Furthermore, the basic argument in defense of single-payer appeals to efficiency in the Pareto sense, namely, that by resolving a market failure in the health insurance sector, it corrects a mispricing of insurance that lowers the welfare of everyone in the society.

Of course, there are also egalitarian arguments in favor of public provision (although these turn out to be far more complicated than they may at first appear). But it seemed to me that if there are both efficiency arguments and equality arguments to be made for a particular public program, it is always better to lead with the efficiency arguments, simply because they are inherently less controversial. Equality arguments are essentially about how to resolve distributive conflict, and so always have a win-lose structure. This means that regardless of how compelling they are, there will always be a constituency with an interest in opposing them. Efficiency arguments, on the other hand, appeal to mutual benefit, or win-win transformations, and so do not necessarily create an oppositional constituency (see chapter 6, figure 6.2). Solving collective action problems is not something that anyone should have a stake in preventing.

In developing this intuition, however, I came around to the view that efficiency arguments were not only rhetorically more effective, but actually did a better job of articulating the underlying logic of many welfare-state programs. While the tax system no doubt institutionalizes a set of egalitarian commitments, the social programs that the welfare state provides are primarily driven by efficiency concerns. This is often obscured by the fact that many of the "goods" that the welfare state provides are insurance products and are therefore superficially redistributive. A public pension scheme, for instance, looks like a system of redistribution—since it takes money from one group and transfers it to another—but is in fact a collective retirement insurance scheme. A better way to think of it is as a bundle of collectively purchased life annuities, delivered through the public sector because private insurance fails to price these products at an appropriate level (see chapter 9). Like all insurance arrangements, it is designed to redistribute from the lucky to the unlucky, not from the rich to the poor. Public health insurance, of course, has the same structure (chapter 13).

This insight led me to regard efficiency as one of the underappreciated virtues of the welfare state—and therefore to take more seriously the traditional "public economics" view that the welfare state is primarily in the business of correcting market failure, not achieving distributive justice, even in many cases where it does not look like it is promoting efficiency (see Heath 2011). This in turn suggested that one could make the case for many of the core features of the welfare state (specifically, public education, health care, and pensions) without appealing to controversial egalitarian commitments (Moss 2004). The only concept required is that of market failure, followed by some account of the resources that the state is able to deploy in order to resolve particular forms of such failure.

Once I had the position formulated in this way, it became clear to me that the conceptual resources required to mount a defense of the welfare state were in essence no different from the ones that underlay a standard transaction cost account of the firm. Transaction cost theory takes as its point of departure the observation that we have a toolkit of different institutional forms that we can use to organize economic cooperation, with the two most important being markets and administrative hierarchies. Depending on the nature of the "transaction" in question, each different institutional form will have a different profile of costs. These costs must be understood broadly, to include not only direct monetary costs (e.g., hiring a lawyer to draw up a contract), indirect costs (e.g., losses due to inadequate enforcement of contracts), but also purely invisible costs (e.g., deadweight losses, due to potentially advantageous exchanges that do not occur because of fear of fraud).

The crux of the theory is that neither markets nor hierarchies dominate the other as an organizational form, and so which one will impose greater costs depends upon the nature of the transaction in question. As a result, what tends to arise in a market economy—when the appropriate legal devices are made available—is a mixture of organizational forms, with some production being organized in a decentralized fashion among various individuals or firms using the market to coordinate their relations, and other production occurring "in house," within the administrative hierarchy of the firm and subject to the authority of management. Ronald Coase's (1937) great insight was that—granted certain idealizing assumptions—the boundary of the firm will be determined by the relative cost of organizing production using these different governance structures. The boundary will also be quite dynamic. Mergers and acquisitions are processes through which transactions that were once mediated by the market are brought within the scope of managerial authority, while outsourcing is a process through which an administered transaction is dissolved and replaced by market contracting.

Now a "market failure" is defined simply as a circumstance in which markets fail to achieve a Pareto optimum (which is to say, where they leave room for improving at least one person's position without worsening anyone else's [Bator 1958]). So what the transaction cost theory of the firm claims is that the organization of economic activity through administrative hierarchies is everywhere explained by the existence of market failure. Less intuitively, it also implies that the organization of economic activity through markets is everywhere explained by "administrative failure." All of this is to say that neither the concept of market failure nor that of administrative failure does much work, taken alone, what matters is simply the relative cost profile of these different modes of economic organization.

This is an extremely powerful theory, and a far more subtle one than most people realize when they first encounter it. It also has both explanatory and ideological virtues. Most importantly, it is very useful when it comes to helping people to unlearn some of the more extreme or Panglossian views of the market that are often inculcated through early exposure to the introductory economics curriculum. It is important to point out, for instance, that the invisible hand of the market, despite its many virtues, is not magical, and it solves no more than a fraction of our economic problems. If the market actually produced perfectly efficient outcomes, *then there would be no need for corporations*. And yet corporations exist. Therefore, there must be non-trivial limitations on the efficiency properties of markets.

Once all of this has been established, it is fairly easy to make the further point that the corporation possesses certain inherent limitations when it comes to its ability to correct market failure, whereas the state has two qualities that give it a different cost profile when organizing economic cooperation, namely, that within its territory membership is universal and it exercises a monopoly over the powers of compulsion (Stiglitz et al. 1989). For certain transactions, this can result in lower costs. For instance, the state has the power to control adverse selection in a way that no private insurer does, which results in the state being the lowest-cost provider of a variety of different forms of insurance (including, typically, health insurance). This means that a straightforward transaction cost analysis is going to suggest that certain goods and services should be funded by taxation and provided through the public sector.

This analysis provided the basic rhetorical strategy of *The Efficient Society*. The objective was to show that once one accepts both the need for and the legitimacy of corporations, then one cannot but accept both the need for and the legitimacy of the modern welfare state. The rhetorical aspect of the argument is important because at the time that I presented it I did not actually know very much about the theory of the firm or transaction cost theory. Like many philosophers, I had spent a lot of time reading "macro" arguments about capitalism, communism, and the state, but had spent very little time studying the "meso" level—the medium-sized institutions that do most of the organizational work in the operations of a capitalist economy, such as the corporation in its myriad forms, stock exchanges, financial institutions, insurance companies, government agencies, and so on. For example, one can read both John Rawls's *A Theory of Justice* and Robert Nozick's *Anarchy, State and Utopia* and learn absolutely nothing about any of these institutions. Nozick, despite providing the most sophisticated libertarian defense of the market economy, has essentially nothing to say about the corporation (and in fact mentions it only twice over the course of his book). Rawls writes in considerable detail about various institutional features of the state, but again has nothing to say about the corporation. (This is because he does not consider it part of the basic structure, and so treats it as outside the scope of his principles of justice [Rawls 1999: 126].)

Perhaps because of this shortage of literature, the fact that *The Efficient Society* contained a chapter on the subject of the corporation (and was written in a popular, accessible style), led to its widespread adoption in undergraduate business ethics courses. In particular, Richard Wellen at York University began teaching the book in his very large "business and society" course, and subsequently invited me to meet with a group of enthusiastic students to discuss my views—such as they were—on the corporation. At around the same time, the late Bernard Hodgson invited me to a conference at Trent University to discuss "the invisible hand and the common good." Together, these led me to develop a stand-alone version of my argument about the firm and its normative implications—the paper that became "A Market Failures Approach to Business Ethics" and forms the first chapter of this book.

One of the things about this paper that will undoubtedly strike many readers is that, apart from being rather polemical in tone, it contains practically no references to the business ethics literature. Indeed, the only two articles I cite are Milton Friedman's famous piece (1970), as well as Andrew Stark's "What's the Matter with Business Ethics?" (1993). The reason is that I had, at the time, essentially no knowledge of business ethics (I had read Stark's paper only because he is a colleague of mine at the University of Toronto). The position that I presented was what I took to be a fairly straightforward implication of the "political economy" perspective that I had been developing. I called it a "market failures" approach to business ethics, although strictly speaking it would be more accurate to have called it a "Paretian" approach, since my major claim is that the market is essentially a staged competition, designed to promote Pareto efficiency, and in cases where the explicit rules governing the competition are insufficient to secure the class of favored outcomes, economic actors should respect the spirit of these rules and refrain from pursuing strategies that run contrary to the point of the competition.

This is a very natural position to take if one approaches the basic question of corporate social responsibility from the perspective of modern economics (understood broadly, to include both the transaction cost theory of the firm and the standard "public economics" understanding of the welfare state; e.g., Barr 1998). In particular, in the wake of the "socialist calculation" debate of the early twentieth century, as well as Friedrich Hayek's information-theoretic reformulation of the classic "invisible hand" argument for the market (discussed in chapter 8), it has become common to regard marketplace competition as essentially a system designed to generate a set of prices, which can in turn be used to achieve a more efficient allocation of productive resources and consumer goods. The major thrust of regulatory interventions in the market—most obviously in the case of environmental and consumer protection—is to correct imperfections that are distorting price signals and thereby leading to the misallocation of resources (such as overproduction of environmental "bads"). It is quite natural—it seems to me—to think that the basic thrust of "business ethics" is the same as that of these regulatory interventions, namely, to discourage firms from taking advantage of market imperfections, even in cases where legal regulation is not feasible (see chapter 1, figure 1.1).

I was therefore unsurprised to discover, after having articulated the "market failures" perspective, that essentially the same thought had occurred to other people. Kenneth Arrow (1973), for example, had expressed very similar views. And I have since discovered a strong current of similar thinking about business ethics in the *Wirtschaftsethik* tradition in Germany, most obviously in the work of Peter Koslowski (2001: 26–30) and Karl Homann (1993). Again, I find all of this unsurprising, simply because the "market failures" perspective is a very natural consequence of taking a broadly liberal theory of justice (such as the "minimally controversial contractualism" that I articulate in chapter 6) and adopting a fairly standard economic perspective on the market.

What was surprising, at least to me, was the resistance that I encountered when I first presented my ideas to business ethicists. During several of my early presentations I discovered that a whole series of steps that I took to be self-evident were in fact highly controversial. For example, I was extremely surprised to find business ethicists resisting the suggestion that markets are competitive, or that competition is somehow important to their function. Of course, they were not exactly denying it, they were mainly resisting it, based largely on the intuition that if we want businesspeople to behave themselves, putting too much emphasis on the competitiveness of market interactions is counterproductive. So, for example, while it seemed natural to me to draw comparisons between business ethics and the ethics of sport and games—given that they are all competitively organized domains of interaction—it turns out there is a long tradition in business ethics of denouncing this exact comparison (the usual object of opprobrium is Carr [1968]).

Thus I resolved to write a series of papers, each one focused on a different stumbling block that I had encountered. I also set out to read more carefully the business ethics literature, in order to understand better the assumptions and motivations of those who were critical of my project. The first part of this book is essentially a record of these efforts. The second part then tries to articulate some of the broader "political economy" considerations that inform the project—in particular, my understanding of contractualism, the relationship between cooperation and social institutions, and the general justification for the market economy. The third part contains what I like to think of as "extensions" of the theory, primarily focused on how the framework can be used to address issues that arise in a management context, or within the firm.

I.1.1. PROFIT

The first stumbling block was the fact that I failed to take issue with the profit-orientation of the firm, and by extension, had no particular objection to shareholder primacy. This obviously generated a headlong conflict with what has arguably been the major trend in North American business ethics, the view that managers should be held accountable to a variety of different "stakeholder" groups, with investors enjoying no special privilege. Coming at things from a background in political economy, the stakeholder perspective seemed to me a strange view, simply because making peace with capitalism essentially involves acknowledging the value of the profit motive (since it is the quest for profit that generates the competitive dynamic that allows the price system to exhibit the desirable properties that it does). Arguing against the profit orientation of firms seemed to me equivalent to arguing for some kind of market socialism, which may be a perfectly respectable position to take, but probably should not be a position *in business ethics*. It seemed to me rather a case of changing the topic (away from the question of how economic actors should behave in a capitalist economy to whether we should have a capitalist economy at all).

Furthermore, there is a deep and sophisticated literature in socialist economics dealing with the problems that were encountered by managers in organizations that are not subject to the discipline of profit-maximization, both in the former communist countries under central planning and in state-owned enterprises under democratic welfare states (Kornai 1992; Nove 1983; Roemer 1994; Stiglitz 1994). For example, there is the well-known fact that investor ownership imposes a "hard budget constraint" that is very difficult to replicate under public ownership. It seemed to me likely that the reorganization of firms along the lines proposed by stakeholder theorists would encounter many of the same difficulties. Thus my first foray into the literature was the collaborative piece with Wayne Norman (chapter 2), the general purpose of which was to encourage business ethicists to confront the literature in public administration that discusses these challenges.

This paper, however, dealt essentially with implementation problems that stakeholder theorists were likely to encounter, it did not take issue with the normative core of the theory. Thus the following paper (chapter 3) represents my first attempt to explain the normative inadequacies of the theory. The central argument is that stakeholder theory, in its standard formulation, expands managerial obligations to encompass the interests of groups other than just shareholders, but does so in a way that is either too broad or else arbitrary from the moral point of view. Arguments of this sort have become increasingly familiar in recent years and are achieving much broader acceptance (Boatright 2006; Orts and Strudler 2009; Marcoux 2000). There is, however, an important complication, which I became aware of after having acquired greater familiarity with the literature. It is common to talk about "shareholder primacy" as the legal norm in much of the world (in contrast to, say, the "co-determination" arrangement in Germany that gives workers representation in the governance of certain firms). As a matter of fact, what the law actually provides is a menu of options, which permits all sorts of different organizational forms. In particular, as reading the work of Henry Hansmann (1992) impressed upon me, there is nothing to stop any constituency group from serving as owners of the firm, by forming a cooperative rather than a standard business corporation. Furthermore, legal "partnerships" are often de facto cooperatives, or multi-constituency ownership structures (that include, say, workers and investors). Furthermore, it is not always the case that in open marketplace competition, standard business corporations do better than cooperatives. The insurance industry, for example, was for decades dominated by consumer cooperatives (or "mutuals").

Thus it turns out to be incorrect to describe modern market economies as being governed by a norm of shareholder primacy within firms. What they are actually governed by is a norm of *owner primacy*. This, combined with the empirical regularity that most firms are owned by the providers of capital, generates the illusion that shareholder primacy is the legal norm. This raises a whole host of questions, which business ethicists have been slow to confront (with notable exceptions, e.g., Boatright 2002). For example, is the "market for control" unfairly biased against non-shareholder groups, or are there good reasons for the prevalence of investor ownership? (Miller 1989: 83–93; Dow 2003). If one accepts Hansmann's redescription of the shareholder-owned firm as essentially an "investor's cooperative," how does this affect our thinking about managerial responsibility? If we have no problem with "member primacy" in the case of a cooperative, why should we take issue with it when the "members" happen to be the lenders? These are some of the questions that I address in chapter 5. In general, what I find transformative in Hansmann's analysis is the suggestion that shareholder primacy and profit-maximization be understood as just special instances of owner primacy and maximization of the residual claim. I take issue, however, with some of the normative conclusions that he draws from this analysis.

I.1.2. COMPETITION

The second stumbling block stemmed from the fact that, from the very beginning, I regarded the market failures perspective as implying a system of *adversarial ethics* for all transactions mediated by the price mechanism. Adversarialism, in my view, involves *deontic weakening* with respect to everyday morality, where certain actions that would ordinarily be obligatory may become optional (and, equivalently, actions that are forbidden may become permissible). Use of the price mechanism implies an adversarial orientation because prices are competitively determined. A competition, in my view (articulated at length in *The Efficient Society*) is essentially an institutionalized collective action problem, where we are released from the everyday-morality obligation to act cooperatively and are actively encouraged to play free-rider strategies. This is a slightly counterintuitive view of competition, although the central idea is well captured by the old saying that "every free market is a failed cartel." The major objective of chapter 4 is to articulate and defend the close connection between the price system, marketplace competition, and adversarialism in business ethics.

This is why, despite the various disanalogies, I think that it is still illuminating to draw comparisons between business ethics and the ethics of sport (and why I find myself agreeing with Alfred Carr's unpopular view that business requires individuals to "discard the golden rule" [1968: 145]). Obviously sport is both voluntary and unserious, in a way that having to earn a living in a capitalist economy is not. This would be an issue if I thought that the voluntariness of transactions was important to the normative justification of the market (which I do not). What I find illuminating about the comparison to the ethics of competitive sport is that it focuses on the question of "how far do you go to win?" It seems to me that the structure of reasoning one must employ (in particular, the way that one must think about one's intentions and goals to resolve this question) can be usefully applied in the business context as well (as I argue in chapter 4).

I.1.3. EFFICIENCY

The third major stumbling block is the almost singular emphasis that I put on the principle of Pareto efficiency in providing normative foundations for the view. This has the potential to give rise to a number of misunderstandings, particularly among those who see the word "efficiency" and assume that it is merely an instrumental principle, or think that there is some sort of conceptual connection between efficiency and the pursuit of individual self-interest. One need only contemplate the structure of a prisoner's dilemma (as in figure 6.1) to see that efficient outcomes are not an automatic consequence of individual utility-maximization, and that in such interactions the Pareto principle serves as a genuine constraint on the pursuit of self-interest (Gauthier 1986).

In a slightly more sophisticated vein, efficiency is often conflated with utilitarian welfare-maximization, or else the wealth-maximization standard proposed by "law and economics" scholars (Posner 1973). I use the term "efficiency," by contrast, in the strict Pareto sense, to refer to the principle that, whenever it is possible to improve at least one person's condition without worsening anyone else's, it is better to do so than not. In practice, this simply commits one to promoting cooperation (in the game-theoretic sense), it says nothing about the specific modalities of cooperation, other than that the benefits should be maximized (i.e., that the set of available Pareto improvements should be exhausted). In economic terms, one way of thinking about the Pareto standard is to regard it as a general prohibition on waste, since if it is possible to rearrange the allocation of resources in such a way as to improve one person's welfare without worsening anyone else's, and yet this is not being done, it means that some resources are being wasted under the status quo.

It is worth emphasizing that I do not assign a central place to efficiency in the assessment of markets because I think it is a foundational value, or that it arises endogenously out of state-of-nature interactions. I think it is an irreducibly normative principle, which constitutes one element of a broader contractualist theory of justice (of the sort that I sketch out in chapter 6). It is only one element because it must to be supplemented with some conception of distributive justice in order to provide a persuasive standard for evaluating the overall system of social cooperation. This does not mean, however, that every domain of interaction is subject to assessment under the full theory of justice. There is, in my view, a division of moral labor within our institutions, with markets being essentially special-purpose institutions designed to promote efficiency (a view defended in greater detail in chapter 7). Thus it is only when embedded within the broader context of a welfare state, which engages in both market-complementing and redistributive policies (primarily through the tax system), that capitalism as a whole can claim to be just. At the same time, this does not mean that market actors are accountable to the same moral demands that the system as a whole must satisfy. Individuals are given license to maximize profits (and to associate in various ways to engage in joint action aimed at maximization of profits), for the narrow reason that, in a reasonably competitive market, this is the best way to get prices that reflect social cost. In order to achieve this, individuals must be given a fairly broad exemption from norms of equality or fairness in the organization of their interactions. To the extent that this is justifiable, it is because the compromises made in the equality dimension (due to the competitiveness of market interactions) are outweighed by the benefits that accrue in the efficiency dimension (due to the operations of the price system). Because of this moral compromise at the heart of capitalism, one cannot hold economic actors engaged in market transactions to a higher standard than that of efficiency promotion. But this is all the more reason to be rigorous in holding them to this standard. A competitive market only serves to promote efficiency under certain conditions, and there are various ways of acting that subvert it. Such actions are not just unethical, but egregiously so, because they fail to satisfy even the artificially low standard that is set for the evaluation of marketplace behavior.

Now to say that the market failures approach to business ethics is "guided" by the Pareto principle is not to say that individuals, when deciding what to do, should ensure that the outcomes of their own actions are always Pareto improving. This is obvious in the context of a market economy, because competitively structured interactions are designed to produce win-lose outcomes (not the win-win outcomes required by the Pareto principle). When a company lowers its asking price, in order to get rid of unsold merchandise, its actions harm its competitors, in a way that is collectively self-defeating when the competitors respond in kind. Our reason for allowing this sort of collective action problem to persist is that it generates, as a byproduct effect, a movement of prices in the direction that will clear the market, and therefore that will maximize the number of efficiency-promoting exchanges with purchasers. The market therefore institutionalizes an indirect strategy for promoting Pareto efficiency, in the form of rules that specify the terms of marketplace competition, in particular, which competitive strategies are permissible and which are not.

Many of the rules of the market are legally enforced, but it is impossible to imagine a circumstance in which they would all be. For example, an enormous amount of legal emphasis is put on the minimization of externalities. The underlying principle, with respect to negative externalities, is that if an action has consequences that are damaging for some other person then there should also be a cost for the person who is engaged in it, and that, to the degree possible, the cost to the person doing it should reflect the magnitude of the damage done. (The principle for positive externalities is just the reverse.) When prices reflect social cost in this way, it ensures that the overall production of costs and benefits will be Pareto efficient. The most important legal mechanism that we have, when striving to achieve this outcome, is the system of property rights itself, which can be thought of as an all-purpose mechanism for internalizing externalities. Through ownership, the individual is able to "capture" much of the value that she produces through her labor, thereby minimizing positive externalities (and thus, increasing the incentive to produce value in the first place). And by asserting her property rights (e.g., against trespass or unauthorized use), she is able to "deflect" many of the negative externalities that others might like to impose on her. The tort system represents an extension of this basic mechanism into areas that are less clearly structured (e.g., generating legal constraint on various forms of nuisance behavior that are not in direct violation of anyone's property rights, or that violate these rights in unanticipated ways). And finally there is regulation, which attempts to control the production of negative externalities in myriad ways (e.g., restricting the production of atmospheric pollution, excessive noise, toxic and dangerous substances, etc.), without requiring private individuals to step forward and assert their rights.

And yet, despite all of this effort, there are still many circumstances in which the legal system is powerless to stop the production of negative externalities. In some cases this is simply because of cost considerations, which are quite high for criminal, tort, and regulatory law. The resources consumed in the effort to stop someone from doing something will often be of greater value than the losses imposed upon society by the behavior. In many cases the behavior is undetectable, or the victims unidentifiable. In other cases, particularly ones with an international dimension, no state has the legal authority or power to effectively control the behavior. What the market failures perspective suggests is that in such cases, economic actors have moral obligations that extend beyond their legal obligations, but that these moral obligations are merely an extension of the basic rationale underlying the law—understood broadly, to include the system of property rights, the tort system, and body of regulatory law. So if it is possible to increase revenue by displacing costs, rather than creating value, this may be morally prohibited, regardless of whether it is legal. In this way, the ideal of promoting Pareto efficiency generates a more specific deontology that constrains the behavior of economic actors. Furthermore, this deontology is generated by a normative theory that provides a unified account of the foundations of the market economy, the purpose of regulatory intervention and state ownership, as well as the basic "beyond compliance" obligations of firms (Norman 2012: 48–49).

I.2. The Goal of Business Ethics

Many business ethicists, perhaps responding to the pressures of modesty, deny that they have any ambition to make people behave more ethically. I actually consider that objective to be central to my task. I strongly agree that the point of philosophy is not just to understand the world, but to change it. Furthermore, as I suggest in chapter 11, if persuading our students to behave more "ethically" in later life seems like too lofty an ambition, perhaps making them less likely to behave criminally would be a useful contribution, and a more achievable one. After all, most of the classic "cases" that business ethicists like to discuss, from the Ford Pinto to the Enron scandal to the Deepwater Horizon disaster, are not really "ethics" cases at all, but rather examples of occupational or corporate crime. In fact, one of the distinguishing features of business ethics, as a domain of applied ethics, is that it deals with an area of social life in which crime is a very serious problem.

This is one of the reasons why, in my view, explaining the complementarity of morality and law is very important. I take it to be one of the central tasks of business ethics to articulate the normative foundations of regulation and to explain why managers should adopt a moral attitude toward compliance. Indeed, if business ethicists were to forget about "moral dilemmas" entirely and just focus their energies on trying to articulate the moral reasons for firms to comply with existing laws, they would stand a much better chance of producing some benefit for society.

It seems to me that one of the useful tasks that business ethicists can perform, in the service of this ambition, is to combat two extraordinarily pernicious views (or, one is tempted to say, "ideologies") that are, unfortunately, quite widely held. The first is the idea that, in a market economy, corporations have no obligation beyond respect for the law. Perhaps the most high-profile proponent of this doctrine is *The Economist* magazine, where it is reiterated often enough to suggest that it is part of the magazine's official editorial stance. The following is a typical articulation of the view: "A company's job is to make money for its shareholders legally. Morality is the province of private individuals and of governments," and so if politicians want to change business behavior, "they should pass laws, not make speeches" (*The Economist* 2011: 18). One can easily find examples of businesspeople expressing variants of the same view (Carr 1968: 146–147).

Many will recognize this as a somewhat unnuanced rearticulation of the view that Milton Friedman (1970) expressed, when he claimed that the only social responsibility of business is to increase its profits. Friedman actually qualified this slightly, claiming that managers should typically try "to make as much money as possible while conforming to the basic rules of the society, both those embodied in law and those embodied in ethical custom" (1970). It is clear, however, that he did not imagine "ethical custom" imposing very significant constraints on business behavior. With respect to reducing pollution, for example, Friedman was clearly of the view that managers are obliged to do the minimum required by law—"ethical custom" generated no additional constraints that could be appealed to (or at least none that he mentions). A manager who voluntarily reduced emissions was, in effect, usurping a public power, and unjustly imposing a "tax" on shareholders.

According to this view, when faced with a demand to discontinue a particular business practice on grounds of "social responsibility," the appropriate response from the business manager is to say: "As long as it's legal, we are going to do it. If there's a problem with that, then the government should pass a law to make us stop." Yet although this claim is often made, it is difficult to take seriously when one stops to think about it. After all, who could possibly want that level of legal regulation of economic activity? Because law is the most intrusive and costly form of social control, it is typically appealed to as an intervention of the "last resort" (Simpson 2002: 112). Furthermore, there are well-known difficulties associated with trying to regulate the behavior of firms that adopt a broadly uncooperative orientation (i.e., that exhibit no moral constraint in their attitude toward compliance). For example, there are clear trade-offs involved in determining the level of specificity at which regulations should be framed: write the rules too broadly and it creates legal uncertainty, as well as higher enforcement costs; write the rules more narrowly and it encourages circumvention (or "gamesmanship"), as well the inefficiency caused by overly rigid specifications (Braithwaite 1981–82: 483–484).

Indeed, there would appear to be a strong element of bad faith in *The Economist*'s espousal of this doctrine, since the magazine is also a vocal critic of government "overregulation." This makes it difficult to believe that they actually support the dramatic extension of regulation that would be required if firms were to abandon all self-restraint in the pursuit of their objectives. Thus one begins to suspect that a shell game is being played, where moral constraints are rejected on the grounds that they should be juridified, but then legal constraint is rejected on the grounds that it is too costly, passing it back to moral constraint.

The position is also strangely unmotivated. Why would a particular institutional actor be exempt from moral constraint? The idea that we are all as individuals obliged to act morally, but when a few of us get together and sign articles of incorporation, we are suddenly licensed to do anything at all in pursuit of our interests, subject only to the constraint of law, lacks even prima facie plausibility. It is not clear that any social institution has the power simply to absolve people of moral responsibility for their actions. So to the extent that anyone thinks that firms are outside the scope of moral constraint, the most likely explanation is undoubtedly some form of muddled "invisible hand" reasoning. Rather than thinking that articles of incorporation offer *carte blanche* to act immorally, the view is more likely that morality becomes unnecessary in a business context, because the invisible hand of the market guarantees that individuals are only able to pursue their self-interest in ways that will also, as a byproduct effect, maximize social welfare. So the aims of morality are, as it were, guaranteed, without any need for constraint. This is, for example, the view articulated by David Gauthier, who argues that there is simply "no need for morality" in the competitive market: "Where earlier thinkers saw in the unbridled pursuit of individual interests, the ultimate source of conflict in human affairs, the defenders of laissez faire see in it rather the basis of the true harmony that results from the fullest compossible satisfaction of those interests. The traditional moralist is told that his/her services are not wanted" (1982: 47).

The problem with this view is not so much that it is bad ethics but that it is bad economics. It vastly overestimates the success of market institutions (in effect, the legal framework that structures economic activity) at achieving this reconciliation. One need only consider the "efficiency conditions" required for the first fundamental theorem of welfare economics to obtain (Schultz 2001). It is easy to see that the reconciliation of public and private interests is never guaranteed by any set of existing market institutions. And when the alignment of private and public interests is not achieved automatically by the market, some attempt must be made to do it consciously and explicitly, through moral constraint. Gauthier does not disagree with this: "Where the Invisible Hand fails to direct each person's actions to the public interest—or, as I shall prefer to say, to mutual benefit—the Visible Foot takes over. Hand and Foot share a common aim" (1982: 41).

The central disagreement between myself and Gauthier is over how much work actually gets done by "hand," and how much is left over for the "foot" of morality.[1] Here the progress of economic thinking over the past thirty years has done much to clarify the issues. At the time that Gauthier was dismissing the need for business ethics, it is worth recalling, there was still very little awareness that pollution could cause serious quality of life issues. This is why Gauthier, like Friedman before him, despite being aware of the problem of externalities, glibly dismisses them. Gauthier was also writing at a time when "information economics" was still in its infancy, and as a result simply ignored the possibility of market failure caused by information asymmetries. So he had nothing useful to say about the market for insurance, professional services, or even branded goods, where information dynamics play a dominant role in determining whether transactions will occur and at what price.

Thus the first pernicious view—that firms are obliged to respect the law, but have no "beyond compliance" obligations—is often put forward insincerely, but when it is sincerely held, it is usually based on an overestimation of the effectiveness of the invisible hand at promoting social welfare. The market failures approach to business ethics arises, then, as a natural consequence of combining a normative commitment to Paretianism with a modern economic understanding of the conditions under which mere compliance in a market is unlikely to promote social welfare.

The second pernicious view constitutes a further attempt to expunge ethics from the realm of the marketplace interaction. Many social scientists have, over the years, been uncomfortable with "ethics," on the grounds that moral motivation is thought to be something mysterious and obscure. External incentives, by contrast, seem much more solid, quantifiable, and verifiable in their effects. This general feeling of unease was, notoriously, elevated to the level of a strict methodological precept by economists (for discussion, see chapter 10). One of the consequences of this theory of action was that it made the law seem much more intelligible than morality. Indeed, many economists affected a certain cynicism toward morality, believing that it could not actually motivate or constrain agents, but was more likely a rationalization of self-interest. Law, by contrast, seemed like a more respectable social-scientific construct—you get punished if you break it, so there is no reason to doubt that legal rules actually constrain the way that economic agents behave. One can believe in both legal regulation and individual utility-maximization without cognitive dissonance.

---

[1] Indeed, if it were not irremediably obscure, the phrase "Hand and Foot share a common aim" could easily serve as the slogan for the market failures approach to business ethics.

The problem with this view is not just that it is dismissive of morality, but that it also generates a terrible misunderstanding of how compliance with legal rules is actually achieved. To the extent that this misunderstanding is encouraged by the adoption of an "economic" model of human behavior, it is important to emphasize that the specific thesis—"people obey the law because they fear punishment"—is an a priori deduction from the model, it is not a fact that anyone has discovered about the world. When one turns to the empirical literature, it turns out, as a generalization, to be false. While deterrence (i.e., the threatened application of sanctions) is an important element of compliance, it is very far from being the entire story. The most obvious reason is that the state lacks the resources and the information required to deter most crime. Given existing legal infrastructure, if people actually sat down and calculated, in a hard-headed way, how best to advance their own individual interests, then crime would be the rule rather than the exception. As a result, we (as a society) rely very heavily on internal controls (such as moral constraint) and informal social controls (such as stigmatization) to achieve conformity with legal rules. Furthermore, it is generally recognized among criminologists that these mechanisms do most of the heavy lifting when it comes to controlling crime. One of the most influential schemas in the literature is John Braithwaite's "enforcement pyramid," which features internal self-regulation at the bottom, followed by informal social controls, legal persuasion, and finally deterrence, in an ascending hierarchy (Ayres and Braithwaite 1992: 38–39). The reason it is a pyramid is that self-regulation is sufficient to control most of the population most of the time. It is only where this fails that informal social controls become active, and it is only when these fail that official legal intervention is required. Even then, legal intervention seldom starts out with punishment, but usually begins with some form of engagement aimed at correcting the behavior (such as an order, or a warning). Thus deterrence is at the top of the enforcement pyramid simply because it is only for a very small percentage of the population that the threat of punishment is required in order to achieve compliance.

It is sometimes assumed that deterrence is going to be more important when it comes to controlling corporate crime, compared to street crime, because corporations are more instrumentally rational, and hence more likely to be doing explicit cost-benefit calculations in deciding what to do. Yet surprisingly, the exact opposite is true; society relies more heavily on voluntary compliance in controlling white-collar crime than it does with street crime. This is because the state has much greater difficulty detecting, prosecuting, and securing convictions with white-collar criminals. There are a number of reasons for this (Simpson 2002: 45–60; Coleman 1989: 185–194). The first has to do with the nature of the victims, who are usually a very diffuse group and often unaware that they have been victimized. This makes the detection of criminal behavior much more difficult, since the state relies very heavily on victim complaints as a cue to initiate investigation. With street crime, the state also derives considerable advantage from the fact that most crimes are committed by individuals or loosely formed groups. This is why organized crime is so difficult to combat (and why significant enforcement effort is aimed at breaking up organized criminal groups, or preventing their formation). Many of the due process safeguards that have been put in place are aimed at protecting isolated individuals from the more organized power of the state. But when crimes are committed by a corporation, the state is dealing with a highly organized group (which, in the standard run of cases, the state has no desire to disrupt or break up), which nevertheless enjoys the same due process rights that protect individuals. Furthermore, corporations are often able to invest more resources in mounting a defense than the state can marshal for its enforcement efforts.

All of these factors, combined with the information asymmetries that make it very difficult to figure out when a crime has been committed, makes the probability of apprehension very low with most white-collar offences. Every so often a major scandal or prosecution will generate closer scrutiny, but when it does, it will often reveal systematically criminal behavior across an entire industry that has gone undetected for years. (To take just one example, under the 2012 mortgage fraud settlement in the United States, almost all the major banks operating in the American mortgage market admitted to using forged documents to illegally foreclose on homeowners [Dayen, 2013]. Commentators struggled to find adjectives that could fully convey the enormity, scale, and boldness of the crimes committed, especially in an industry that should have been expecting heightened scrutiny in the wake of the 2008 financial crisis.)

Simplistic cost-benefit reasoning suggests that these enforcement problems could be made up for by an increase in the associated penalties. There are, however, limits on this strategy in the case of corporate crime. Most obviously, limited liability puts a ceiling on how large fines and damage awards can be—if they become too large firms will simply declare bankruptcy. Furthermore, the threatened loss to managers who are making the decisions is often much lower than the potential loss to shareholders, so there are agency problems to be considered. This is in fact what underlies the trend toward criminal prosecution of individual managers, instead of monetary damages from the firm. But this creates all kinds of other problems, because of the internal complexity of firms and the difficulty of assigning responsibility. Again, the issue is that the criminal law is very strongly tailored toward dealing with individuals, not groups, and so when actions actually are planned and carried out by a group agent, like the firm, it can be very difficult to apply many of the essential categories of criminal law.

Thus the dominant view among criminologists is that it is often impossible to come up with a threatened punishment that is large enough, credible enough, and sufficiently well targeted to serve as an effective deterrent against corporate crime (Braithwaite 1981–82). So to the extent that corporations do respect the law, it must be due to a higher level of moral self-restraint than is operative in the area of street crime. This is, of course, not implausible when one looks at the profile of a typical businessperson. As Edwin Sutherland pithily observed, "businessmen are generally not poor, are not feebleminded, do not lack organized recreational facilities, and do not suffer from the other social and personal pathologies" (1968: 58). So there is certainly no reason to expect them to be immune to the force of moral constraint. Unfortunately, there is a certain line of thinking—again, strongly influenced by economics—which denies that corporations have any moral obligation to obey the law. Perhaps the most high-profile exponents of this view are Frank Easterbrook and Daniel Fischel, who argue that managers should adopt a purely instrumental orientation toward regulation, and comply with it only when it is in the firm's interests to do so. Any punishments the firm may incur as a result of breaking the law should be regarded as just another cost of doing business. According to Easterbrook and Fischel, "Managers do not have an ethical duty to obey economic regulatory laws just because the laws exist. They must determine the importance of these laws. The penalties Congress names for disobedience are a measure of how much it wants firms to sacrifice in order to adhere to the rules; the idea of optimal sanctions is based on the supposition that managers not only may but also should violate the rules when it is profitable to do so" (1982: 1177).

This latest refinement of "Chicago school" doctrine is far more audacious than what Friedman proposed. Friedman thought that firms should maximize profits using all legally available means. Easterbrook and Fischel claim that illegal means are perfectly acceptable as well, so long as the firm's actions are profit-maximizing ex ante, taking into consideration the probability of apprehension and severity of punishment. Indeed, they go further, arguing that managers are positively obliged to break the law when it is profitable to do so, anything less would be a dereliction of duty toward shareholders. They arrive at this alarming conclusion by taking the doctrine of proportionality—the idea that the severity of a punishment should reflect the severity of the offense—and interpreting it to mean that, by making a particular sort of behavior illegal, the legislature is not really prohibiting it, but just pricing it. This is complemented by the economistic view that the law can control behavior only by providing external incentives, and so it would be unreasonable to expect firms to do anything other than respond to the incentives provided. It follows that, when faced with a demand to discontinue a particular business practice on grounds that it is illegal, the appropriate response from the business manager is to say: "As long as it's profitable, we are going to do it. If there's a problem with that, then the government should increase the penalties for non-compliance."

It is difficult to express just how irresponsible this view is, particularly when promulgated by a sitting judge. And yet these are not ideas from the fringe. Easterbrook and Fischel are among the most influential theorists of corporate law of the twentieth century, and their position has become something close to Chicago school orthodoxy. Yet anyone acquainted with the empirical literature can see that what they are proposing is a recipe for utter chaos. Furthermore, there is good reason to think that very few people would be attracted to the idea of street criminals doing these sorts of calculations when deciding whether to mug a pedestrian or break into one's home. This suggests, in turn, that their analysis is not so much a distinctive normative position as it is an ideology (i.e., a system of motivated false belief), one that is based on a deep misunderstanding of how market economies actually function.

In this social context, it seems to me that the central role of business ethics is to provide an "immanent critique" of corporate conduct (Benhabib 1986). Its objective is not to bring in "outside" moral considerations to condemn the latest outrage, but to clarify and to correct the self-understanding of participants in the market economy, who are being bombarded—both by the business press and a certain segment of the academy, who appear not to have recovered from the epiphany they experienced in their first-year economics class—by a seductive but ultimately false suggestion that the institutions of the market free them from all forms of moral constraint. In order to do so, it has no need to appeal to normative standards beyond those that are already implicit in the institutions of a market economy.

I.3. Further Directions

By way of conclusion, I would like to mention a few areas in which the market failures approach to business ethics stands in need of much further theoretical development—issues that are touched on in the papers collected here, but are by no means given a satisfactory treatment. The doctrine as it is presented is, first and foremost, intended as an approach to the question of corporate social responsibility, in the sense that it deals with the obligations that managers have to individuals outside the firm, in relations that are mediated by the price mechanism (or should be mediated by the price mechanism, in the case of certain externalities). It says very little about relations inside the firm, among those engaged in administered, rather than market, transactions (e.g., between line managers and those who are subject to their authority). The central difference is that administered transactions are cooperative, in a way that competitive relations in the market are not. Because these relations are not intrinsically adversarial, the norms that govern behavior within the firm more closely resemble the norms of everyday morality. For example, wages are much more egalitarian within firms than across different firms (Frank 1985: 35–57). Although a number of different factors play into this, it is not difficult to imagine that concerns about fairness have much greater force within firms simply because people need to work together cooperatively.

However, the fact that norms within the firm more closely resemble everyday morality should not mislead one into thinking that they are the same. For instance, the firm can only go so far in satisfying people's intuitions about "fair wages," because it is constrained by the need to keep its internal transaction costs lower than what the market alternative would be. (For example, the more egalitarian the wage structure is within a firm, the more attractive it will be to "contract out" certain functions.) So while firms may cultivate their own internal values or ethos, as part of the corporate culture, which is much "thicker" than the extremely minimal framework that governs market transactions, it is still constrained by a general set of efficiency imperatives that arise out of the need to keep its administrative transaction costs lower than those of a corresponding set of market transactions. Because of this, while I think there is much to be said for Allen Buchanan's "agency risk minimization" (1996) analysis of the ethics of bureaucratic organizations, I also think there is a lot more to the story than what agency vocabulary is able to reveal. (To take just one example, reciprocity is extremely important when it comes to establishing cooperative relations between individuals in face-to-face interaction, yet the principal-agent framework is designed to handle only unilateral relationships of influence. It is therefore likely to overlook many of the devices that, say, managers use to motivate employees, devices that may in turn generate moral obligations.) Thus my discussion in chapter 4, particularly figure 4.3, provides too simplified a picture of these relations. At the moment, however, the most that I feel I can say about intrafirm relations is that they are complex and also extremely contextual. Analyzing them from an efficiency perspective may not provide the most perspicuous understanding; it may simply be the only thing that can be said about them at a high level of generality.

The second major issue that I would someday like to address more thoroughly concerns the non-ideal aspect of the theory (in the sense described in chapter 7). In particular, I am interested in trying to delimit more carefully the circumstances in which competitive pressure actually does offer economic actors a legitimate excuse for acting immorally, as well as the subsidiary obligations that might accompany that excuse. A related question arises about the circumstances in which firms might be justified in breaking the law. Both Wayne Norman and I have a tendency to talk about "regulation" as though it was mostly aimed at correcting market failure, which is why the "spirit of the law" is taken to be consonant with the general thrust of business ethics. But this is not always a fair characterization of regulation. There is, of course, the famous thesis advanced by "public choice" theorists, who see regulation almost in its entirety as a consequence of rent-seeking by various actors. In this context though, it is worth observing that this account of regulation is not based on empirical evidence, it is an a priori deduction from an economic model of action combined with the assumption that legislators always act in a self-interested fashion, assigning no value at all to the public interest. In other words, public choice theory and the attendant view of regulation is part and parcel of the economistic ideology that must be rejected in order for business ethics as an enterprise to make any sense at all. It cannot therefore be appealed to as an objection to any particular view in business ethics.

A more plausible challenge to the market failures approach would be one that points to specific regulations, such as a professional licensing requirement that serves an obvious cartelizing function, then highlights the conflict between the inefficiency of the law and the efficiency imperative arising from an ethical perspective. Should a firm then be prepared to break the law? It seems to me that two points are in order with respect to such a case. The first is the familiar one from the literature on civil disobedience, which observes that a law can be unjust or immoral, yet still be legitimate, and therefore worthy of obedience. The threshold at which civil disobedience becomes justified is typically set higher than the level at which a law is determined to be unjust. The second is the observation that, as long as corporate crime remains such a serious blight on society, the problem of excessive business compliance is likely to remain somewhat academic, in the pejorative sense of the term. Given a choice between the world we live in and some other possible world in which businesses automatically and unthinkingly obeyed every law, including those that are unjust, I would not hesitate to choose the latter.

The details of this, however, would need to be worked out, since there are a number of controversial issues—including the distinction between excuses and justifications, which I discuss briefly in some of the papers here, but make no attempt to defend systematically.

Finally, readers are sometimes a bit flummoxed by the fact that I present myself as a critic of "economism," while at the same time making extensive use of economic concepts and accepting the essential correctness of several contemporary economic theories. The reason for this is never directly articulated in any of these papers, but it is due to the fact that I reject the narrow instrumental conception of rationality at the heart of utility-maximization theory, while nevertheless accepting several "meso" level economic theories—such as the transaction cost theory of the firm, as well as something like the traditional supply-demand model of price determination. This is simply because I do not believe that the correctness of these theories depends on the correctness of the traditional utility-maximization model; they can be given other micro-foundations. (Those who are interested in my official views on action theory, and the way in which moral considerations can be integrated into a formal model of action can find some indications in chapter 10, but also are encouraged to consult my book *Following the Rules*, where the position is laid out in greater detail.)

PART } I

The Corporation and Society

1 }

A Market Failures Approach to Business Ethics

"Business ethics" is widely regarded as an oxymoron. The only way to be a good soldier in an unjust war is to disobey orders, or maybe even to desert. Many people believe, along similar lines, that the only way to maintain one's ethical integrity in business is not to go into business. The reasons for this are not hard to find. Students are still routinely taught in their introductory economics classes that in a market economy, when engaged in market transactions, individuals act out of self-interest—whether it be by maximizing profits as producers, or by maximizing satisfaction as consumers. This sets up an almost indissoluble link in people's minds between "profit-maximization" and "self-interest." As a result, anyone who thinks that the goal of business is to maximize profits will also tend to think that business is all about self-interest. And since morality is widely regarded as a type of constraint on the pursuit of individual self-interest, it seems to follow quite naturally that business is fundamentally amoral, if not immoral.

The problem is that the association between profit-maximization and self-interest so often taken for granted is based upon a naïve and inadequate theory of the firm. Profit-maximization and self-interest are not the same thing, and the failure to distinguish adequately between the two can be a source of enormous confusion. Business ethics, as a subject, is essentially concerned with the moral responsibilities of managers. Managers often find themselves placed in circumstances in which the imperative to "maximize shareholder value" conflicts with their self-interest. Thus there are many cases in which profit-maximization should be viewed as a managerial obligation, not as an expression of self-interest.

Because of this somewhat elementary confusion, there has been a marked tendency in the business ethics literature to dismiss out of hand views that take the profit motive seriously. In particular, Milton Friedman's classic article "The Social Responsibility of Business is to Increase its Profits," is more often treated as a piece of apologetic than as a serious piece of moral reasoning (Friedman 1970). This is unfortunate, since the moral laxity on display in Friedman's work is not so much a symptom of an inadequate normative framework as it is a consequence of specious economic reasoning. Or so I will attempt to show.

The more serious consequence of this confusion is the widespread perception that, in order for business ethics to be genuinely ethical, it must extend managerial responsibility to groups other than shareholders. This is, I believe, often the intuition underlying "stakeholder" theories of managerial responsibility. In this paper, I will argue that such efforts are misguided. Profit-maximization, understood as an obligation, rather than as an expression of self-interest, provides a perfectly legitimate platform for the development of a robust moral code. However, if profit-maximization is an obligation, the question naturally arises where this obligation stems from. It is in seeking to justify the profit motive that we discover that the appropriate form of managerial responsibility is not to maximize profits using any available strategy, but rather to take advantage of certain specific opportunities for profit. In many cases, the set of conditions under which profit-seeking is permissible is reflected in the legal environment in which firms operate. I will argue that business ethics is best understood as a set of additional constraints that preclude legally permissible, but not normatively justifiable, profit-maximization strategies.

1.1. The Profit Motive

Andrew Stark's controversial 1993 *Harvard Business Review* article, "What's the Matter with Business Ethics?" argued that conventional business ethics was "largely irrelevant for most managers," because it failed to offer them any "practical" advice (Stark 1993). "Moral philosophy," he argued, "tends to value *altruism*, the idea that an individual should do good because it is right or will benefit others, not because the individual will benefit from it" (Stark 1993: 40). As a result, business ethicists have had too little to say about "the potential conflict between ethics and interests," and in particular, how managers should handle such conflicts when they arise.

This article had many people nodding their heads in agreement. But to see just how peculiar the claim is, suppose that the subject had been medical ethics instead of business ethics. Substitute "doctors" for "managers" throughout. Now imagine criticizing medical ethics on the grounds that it fails to offer doctors any "practical" advice on what to do in cases where the imperatives of patient care conflict with their self-interest. Suppose the patient doesn't really need an operation, but the doctor could make a lot of money by performing it anyway. What to do, what to do?

I would suggest, *pace* Stark, that we do not need professional ethicists to tell us where our obligations lie in such cases. Everyone knows that when there is a straightforward conflict between our self-interest and our moral obligations, the moral obligations win, at least from the moral point of view. This is not "ethical absolutism," as Stark maintains, it is simply the logic of moral justification. The question of when we may be forgiven for disregarding our moral obligations (i.e., acting immorally) is a separate one and is in no way specific to the domain of business ethics.

So why does Stark's argument sound even remotely plausible, whereas a comparable argument in medical ethics would be dismissed out of hand? The confusion has two distinct sources. The first arises from the way that introductory economics is usually taught. The standard microeconomics textbook starts out with the assumption that individuals maximize utility. When it comes to particular goods, these utility functions can be represented as a set of indifference curves. These indifference curves are then taken to provide the supply and demand curves. The thesis that individuals maximize utility is interpreted to mean that consumers will seek to maximize satisfaction, and suppliers will seek to maximize profits. Finally, in order to make the model more "realistic" consumers get aggregated together into "households," and suppliers into "firms"—each of which is thought to maximize some joint utility function.

While everyone understands that "the firm" is something of a black box in this analysis, the result is still an unhelpful blurring of the boundaries between the pursuit of self-interest and the maximization of profits. Stark, for instance, variously describes the conflict that managers face as one between "self-interest and altruism," "ethics and interests," "ethical demands and economic realities," "moral and financial costs," "profit motives and ethical imperatives," and even "consumer's interests" versus the "obligation to provide shareholders with the healthiest dividend possible" (Stark 1993: 44). Here we see a clear blurring of the distinction between self-interest, profit-maximization, and the obligation to shareholders.

We understand implicitly that the professional conduct of doctors is to be entirely governed by their obligations to their patients, and thus that they are not permitted to let considerations of self-interest intrude. Profit-maximization has precisely the same status for managers. To my knowledge, no one has ever tried to defend the managers of RJR-Nabisco, or Enron, on the grounds that they were simply acting in their own self-interest. Of course, if the incentive systems have been properly designed, managers will find it to be in their interest to maximize shareholder value (in the same way that doctors generally find it to be in their interest to cure their patients). But this is accidental and irrelevant from the moral point of view. In the case of a conflict, the obligations simply trump the relevant set of interests. Where things get interesting is when multiple obligations conflict, as in the case of a doctor who can improve a patient's chances of survival by lying to him about his condition, or of a manager who finds herself able to please investors by initiating an unnecessarily severe downsizing.

The second major source of confusion stems from the moral status of the objective sought by managers—profit-maximization. The doctor's obligations to the patient flow quite naturally from the objective, which is to restore the patient to health. Health is widely regarded as a good thing, and thus the doctor's actions serve to promote a state of affairs that is morally desirable. This makes the doctor's actions directly justifiable, even intrinsically altruistic. Things are more complicated in the case of business. It is not clear that profits are intrinsically good. Furthermore, when a manager makes a decision that disadvantages workers in order to benefit owners, the profit-maximization imperative generates a distributive transfer that is by no means morally sanctioned. In fact, under the typical set of circumstances, the transfer will be regressive, and thus problematic from the moral point of view.

The asymmetry arises from the fact that profit-maximization is only indirectly justified. It is useful to note that this problem is one that business ethics shares with legal ethics. The adversarial trial system imposes upon lawyers an obligation to do whatever is in their power to defend or advance the interests of their client, even when these interests are highly refractory to the concerns of justice. Thus, the professional obligations of lawyers often conflict with the imperatives of everyday morality. What justifies their behavior is the fact that they operate in the context of an institution with differentiated roles. The desirable outcome is a product of the interaction between individuals acting in these roles, none of whom are actually seeking that outcome. Justice is best served when there is both vigorous prosecution and vigorous defense.

Thus the effective trial lawyer "promotes an end which is no part of his intention." The adversarial system may, for example, maximize acquittal of the innocent, even though neither the prosecution nor the defense adopts that as their objective. As a result, neither lawyer's conduct can be justified by the intended outcome. It is justifiable only through the consequences that the pursuit of this outcome leads to, when combined with the actions of the others.

The same can be applied to the case of managers. The manager should seek to maximize profits for the same reason that the defense lawyer should seek to have his client acquitted—not because the acquittal of his client would be a good thing, or even because his client wants to be acquitted and is paying the bill, but rather because the adversarial trial system as a whole is taken to be the best form of institutional arrangement to serve its appointed function. This is why one cannot do legal ethics without a broader appreciation of how the legal system as a whole functions, and what valuable tasks the various roles are thought to discharge. Similarly, one cannot do business ethics without some appreciation of what justifies the system of private enterprise.

Thus the straightforwardly moralizing critique of the profit motive is jejune (comparable to attacking lawyers for "defending rapists and murderers"). We need to understand why criminals should be entitled to the best possible defense, in order to understand the responsibilities of lawyers. Similarly, we need to understand why corporations should be entitled to pursue profits, in order to understand the responsibilities of managers.

1.2. What Justifies Profit?

The right of corporations to earn profits is sometimes regarded as self-evident. This conviction usually stems from a set of broadly Lockean convictions, which suggest that individuals come naturally equipped with a set of property rights prior to the institution of government. Profit-maximization is then understood as the attempt to augment these holdings through labor input or voluntary exchange—neither of which the state has any obvious authority to restrict.

The problem with this Lockean view—apart from the fact that the underlying conception of rights is deeply problematic—is that corporations are not individuals, they are highly artificial legal constructs. Furthermore, the corporate organizational form provides individuals with a number of very tangible advantages that they do not enjoy as private citizens. The most significant among these is limited liability—the ability to insulate their own private resources from those of the corporation, so that they cannot be pursued by creditors in the event of default. Because of this, creating a corporation is widely regarded as a privilege, not a right. This makes it legitimate for the state to impose certain obligations, in return for the privileges granted.

Many of the corporations chartered by the state are nonprofit. They are specifically prohibited from showing more than a modest revenue surplus. So why permit an exception for other firms? To put it in Marxian terms, why should society tolerate the private appropriation of the social product?

The answer to this question is somewhat complex. Basically, it is that society wants to encourage competition between suppliers. This competition, when combined with competition between purchasers, will affect the prices at which goods trade. Under the correct circumstances, competition will push prices toward the level at which markets clear (i.e., suppliers will not be left with unsold merchandise, and consumers will not be left with any unsatisfied demands). When this occurs, it means that society has succeeded in minimizing the overall amount of waste in the economy. It means that fewer resources will have been spent producing goods that no one wants, at the expense of goods that people do want.

Thus the primary reason for introducing the profit motive into the economy is to secure the operation of the price mechanism. The price mechanism is in turn valued for its efficiency effects. It allows us to minimize waste. The formal proof of this is often referred to as "the first fundamental theory of welfare economics" (hereinafter FFT), or else, in a nod to Adam Smith, the "invisible hand theorem." The central conclusion is that the outcome of a perfectly competitive market economy will be Pareto-optimal—which means that it will not be possible to improve any one person's condition without worsening someone else's.

The importance of the price mechanism is often underestimated. Since the profit orientation of firms definitely has some adverse social consequences, this can sometimes make it difficult to see what the big gains are that justify our tolerance for the various abuses. In order to put things into perspective, it is helpful to consider the difficulties that we would face trying to make decisions in the absence of a set of prices. This is the situation that planners often confronted in the former Soviet Union. Imagine that one of your plants increases its production, so that you now have the capacity to produce an extra 500 tons of plastic. What to do with this material? You need to figure out where it is most needed. But how do you decide? Suppose, to simplify enormously, that there are two possible uses: to make toothbrushes or soup ladles. The question is: which do people need more of?

In a market economy, these needs will be expressed in the form of relative willingness to pay. If stores have too many ladles, and not enough toothbrushes, they will be willing to order more toothbrushes, and pay more for them. This in turn means that the toothbrush makers will be willing to pay more for the plastic. Thus, if all firms sell to the highest bidder, the resources will be channeled toward the use for which there is the greatest need. But if there is not a competitive market for all these goods, not only will firms not have the incentive to engage in the necessary transactions, but the absence of prices will make it difficult for anyone even to determine which transaction should be occurring. Planners in the former Soviet Union used to get around this problem by sometimes looking at commodity prices in Western Europe and North America, and using these figures to do calculations for their own economy. In fact, they used to joke that in the event of a global communist revolution, it might be worthwhile to keep Hong Kong capitalist, so that everyone else would know what prices their goods should be trading at.

The joke has a very serious underlying point. Without prices, you simply cannot organize a complex economy, whether it be capitalist, socialist, or communist. And not just any prices will do. There are an enormous number of price points at which exchanges can occur. In cases where there is only one supplier or one consumer, this gives one side considerable power to dictate terms. Under such conditions, there is no reason to expect that the price level chosen will be the price that clears the market. Thus the price system will not induce efficiency. But when there is more than one supplier, or more than one customer, each one is in a position to undermine the negotiating power of the other. If one supplier insists on a price that is too high, the customer can go to the competition. The competitor is then able to make a profit by undercutting the other one's price, making up for it through a larger volume of sales. The result is a race to the bottom among the suppliers, in which they competitively underbid one another until the market clears, and all profit disappears.

Thus the central rationale for having private profit-seeking firms is to establish competition among suppliers and consumers. This competition drives prices toward market-clearing levels, allowing society in turn to generate a more efficient allocation of its resources and labor time.

It should be noted that this concern with competitive markets, and market-clearing prices, is not simply an abstract philosophical theory about what might justify profit-maximization. The entire legal structure of the firm, along with the regulatory environment, has been organized in such a way as to promote not just competition, but the precise type of competition that is likely to generate market-clearing prices. This is true of everything from antitrust to consumer protection law. In the past decade in Russia, corporations have been known to maximize profit by blowing up each other's factories and assassinating each other's chief executives. Much of the massive legal apparatus that governs corporate behavior in more mature capitalist economies is designed to ensure that firms seek to maximize profits through a much more limited set of strategies—namely, those strategies that are likely to generate more efficient production, along with a more efficient allocation of goods and services in the economy.

Thus, if we ask what the obligations of managers are, the answer can be provided quite directly. The function of the market economy is to produce the most efficient use of our productive resources possible. This can be done, roughly speaking, by achieving the price level at which all markets clear. The role of the firm in that economy is to compete with other suppliers and purchasers for profits in order to drive prices to that level. Thus managers are obliged to do what is necessary in order for the firm to maximize profits in this way. Profits show that the balance of "needs satisfied" to "resources consumed" is positive, while losses show that the resources would have been put to better use elsewhere. Hence the old saying that if we penalize a man for making a profit, we should penalize him doubly for showing a loss.

1.3. Milton Friedman

The approach to business ethics that takes profit-maximization as a central concern is often viewed with suspicion, since it has traditionally been used more as an apologetic for irresponsible behavior than as a platform for a good-faith effort to develop a code of ethics.

As we have seen, in order to be plausible, the profit-maximization approach to business ethics cannot identify profit-maximization with individual utility-maximization on the part of managers. The naïve version of the "invisible hand" view, according to which markets miraculously transform private vices into public virtues, has clearly become obsolete in the era of professional management.

Thus when Milton Friedman argued that the social responsibility of business is to increase its profits, his primary emphasis was on the fiduciary relationship between managers and shareholders (Friedman 1962). The manager is in a similar position with respect to the shareholder that the lawyer is in with respect to a client—he is expected to advance the interests of the principal, not his own. This requires trust, and hence moral obligation, between the two parties. And, of course, there are many ways in which the lawyer can exploit this relationship for private gain, as can the manager.

This makes Friedman's view a genuine code of ethics, and not simply an apologia for self-interest. However, while Friedman is clear that managers are subject to genuine moral constraint, he is less than clear about the source of these obligations or constraints. At one point, he suggests that the manager is bound to assist the shareholder in the satisfaction of his or her desires, and that profits just happen to be what most shareholders want. This is clearly absurd—the manager is not the personal servant of the shareholder. The shareholder might like to have the manager do his laundry, and if he can supply appropriate incentives, he may even succeed in getting the manager to do it. But there is no sense in which the manager is morally obliged to do so, by the mere fact that the shareholder desires it. The manager's responsibility toward the shareholder is clearly restricted to the latter's investment returns. Or, as Friedman puts it when he is being careful, the responsibility of managers is "to make as much money for their stockholders as possible" (Friedman 1962: 133).

However, even this more restricted concept of managerial responsibility is not enough to explain the source of the obligation. Simply making a promise is not enough to generate an obligation, in cases where the end in view is itself not morally justifiable. Promising to help a friend rob a bank does not generate an obligation to rob the bank. Thus the manager's obligation to help the shareholder maximize profits must be derivative of the latter's entitlement to do so. And since it is the FFT that justifies this entitlement, Friedman's argument derives managerial responsibilities from the efficiency argument for capitalism on the whole.[1]

This implicit dependence upon the FFT is discernible in a seemingly innocuous caveat that Friedman tacks onto the formulation of his central thesis. Here is what he says:

> The view has been gaining widespread acceptance that corporate officials and labor leaders have a "social responsibility" that goes beyond serving the interests of their stockholders or their members. This view

[1] Friedman also has a parallel argument concerning the role of markets in promoting freedom. But this line of thinking is, in my view, so riddled with fallacies that it does not merit serious consideration. Furthermore, it seems fairly obvious that Friedman's preference for market solutions to almost every social problem came from his conviction that governments were inefficient and markets were efficient.

> shows a fundamental misconception of the character and nature of a free economy. In such an economy, there is one and only one social responsibility of business—to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game, which is to say, engages in open and free competition, without deception or fraud. (Friedman 1962: 133)

Thus he argues that managers must maximize profits, not *tout court*, but rather subject to the "rules of the game," and in particular, subject to the constraint that they do so "without deception or fraud." The fraud constraint is unexceptional and redundant, since it is illegal. (It goes without saying, for instance, that one should not profit through theft or murder.) But why not deception? One is allowed to win a chess game through deception. In fact, deception is a common feature of strategic interactions. What's wrong with making money through deception?

The answer cannot be that the general moral imperative against lying is binding upon managers in all contexts. Everyday morality compels us to treat others as we ourselves would like to be treated, and yet the last thing we want a manager thinking about, before declaring a giant year-end clearance sale, is how she would feel if the competition did the same to her. More generally, price competition is an interfirm prisoner's dilemma—the outcome is suboptimal for all the competitors. Many moral norms have as their primary function the elimination of such collectively self-defeating interaction patterns. Yet in the case of businesses, we want them to remain stuck in the prisoner's dilemma. In fact, any agreements designed to eliminate these outcomes are specifically prohibited by law. So we cannot simply appeal to the fact that an action is prohibited by everyday morality as grounds for imposing this same prohibition upon managers, unless we want to adopt the very strict universalist view that morality does not permit any institutional differentiation.

Thus the problem with deception, in Friedman's view, cannot arise from any strict deontic prohibition. The problem with deception is that it violates one of the conditions needed for the economy to achieve an efficient outcome. It is these conditions that Friedman is adverting to as well when he talks about an obligation to engage in "free and open" competition.

The relationship between honesty and efficiency in market transactions requires very little demonstration. If suppliers lie to consumers about the character of the goods that they are acquiring, then the prices at which their exchanges are concluded are not going to reflect the actual need for the goods in question. This will generate inefficiencies in the economy.

To take a very concrete case, consider the so-called "goulash capitalism" episode in Hungary. Shortly after the transition from communism to capitalism, Hungary was struck by a wave of lead poisoning. The source of the epidemic was eventually tracked down to paprika. After privatization, several paprika suppliers began adding ground-up paint—much of it lead-based—to the spice, in order to improve its color. In other words, a competition developed to produce the best-looking paprika, not the best quality paprika. Needless to say, if consumers had been properly informed as to the quality of the goods they were purchasing, they would not have bought any. Thus the deception perpetrated by these firms resulted in a huge loss of welfare to consumers. Health authorities eventually had to step in and destroy the entire paprika supply in the country, in order to eliminate all the contaminated goods.

This is a case of what economists call "market failure." In order for the FFT to obtain, a set of very restrictive conditions must be satisfied. These are referred to as the Pareto conditions. The state in which all the Pareto conditions are satisfied is often called, somewhat misleadingly, "perfect competition." When one or more of the Pareto conditions are not satisfied, the competitive equilibrium of a market economy will be less than Pareto-optimal. When a Pareto-inferior outcome is realized, this is referred to as a market failure.

One of the constraints that must be satisfied in order for the Pareto conditions to obtain is that information must be symmetric. Each party to the transaction must have the same information (not only about the prices and goods that are directly relevant to the exchange, but about all other prices and goods in the economy as well). Thus what Friedman is suggesting, in effect, is that managers have no right to take advantage of market imperfections in order to increase corporate profits. The set of permissible profit-maximizing strategies is limited to those strategies that would be permissible under conditions of perfect competition.

In this view, there is a natural complementarity between law and morality. As mentioned, the primary function of the legal regulation of the market is to prevent market failures—both by ensuring that firms do not collude to escape the prisoner's dilemma that competition imposes upon them, or by preventing them from displacing costs in a way that is not fully reflected in the price at which goods trade. In a perfect world, it would be possible to create perfect markets. However, in the actual world, the legal mechanism is a somewhat blunt instrument. In many cases, the state simply lacks the information needed to implement the necessary measures (sometimes because the information simply does not exist, but often because the state has no way of extracting it truthfully from the relevant parties). Even when the information can be obtained, there are significant administrative costs associated with record-keeping and compliance monitoring. Thus the deadweight losses imposed through the legal mechanism can easily outweigh whatever efficiency gains might have been achieved through the intervention. This makes legal regulation unfeasible.

Moral constraints, on the other hand, are subject to no such costs. Corporations, for instance, are often in a position where they can produce misleading advertising that stops short of outright falsity. In a perfect world, advertising would provide nothing more than truthful information about the qualities and prices of goods. However, the vagaries of interpretation make it impossible to prohibit anything but the most flagrant forms of misinformation. Thus misleading advertising stands to false advertising as deception does to fraud. It is something that would be illegal, were it not for practical (or perhaps even accidental) limitations on the scope of legal regulation. Profiting from such actions is therefore morally prohibited, because it runs contrary to the objectives that the market system was instituted to promote.

Friedman's view is often rejected on the grounds that it is morally lax. It basically lets business off the hook on the question of social responsibility. The above analysis shows, however, that Friedman's argument is not a Trojan Horse for naked self-interest. Despite some confusion, it is clear that Friedman's managers have genuine ethical responsibility to shareholders, and that this responsibility is derived from the FFT. The problem is that Friedman arbitrarily limits the set of obligations to those that support only some of the many Pareto conditions.

For example, Friedman argues that pollution reduction is one of the illegitimate responsibilities pressed upon managers in the name of "social responsibility." But pollution is a negative externality—a cost associated with some economic activity that is transferred to a third party without compensation. These externalities exist because the set of markets is incomplete. We cannot exercise property rights over the air that we breathe, for example. As a result, while we can charge people for dumping noxious substances on land that we own, we cannot do the same when they dump it in the air. For this reason, one of the Pareto conditions effectively requires that there be no externalities. Any corporation that pollutes is essentially profiting from a market imperfection. This means that there is no difference, from the moral point of view, between deception and pollution—both represent impermissible profit-maximization strategies. Friedman's decision to prohibit deception, while giving the wink to environmental degradation, is arbitrary and unmotivated.

Figure 1.1 shows the basic structure of Friedman's normative framework. The overall set of profit-maximizing strategies is partitioned into three categories, separating out the immoral and the illegal strategies from the normatively acceptable ones. The efficiency standard can be used to make both cuts. The "acceptable/unacceptable" distinction is imposed by the efficiency properties

Profit maximization strategies

Acceptable	**Immoral**	**Illegal**
e.g. lowering price, improving quality	e.g. pollution, deceptive advertising	e.g. fraud, theft, embezzlement, false advertising

FIGURE 1.1 Friedman's normative framework

of the market system as a whole. The set of unacceptable strategies can then be subdivided into "immoral/illegal" using a transaction cost or regulatory cost analysis.

1.4. A Market Failures Based Code

The above reflections suggest that there is no reason to think that a business ethics focused on profit-maximization cannot deal with the obligations that have traditionally been described under the heading of "social responsibility."[2] What so often annoys people about corporations—and what gives profit-seeking a bad name—is the exploitation of one or another form of market imperfection. People generally have no problem with companies that make money by providing good service, quality goods, low prices, and so forth. In my opinion, if all companies fully internalized all costs, and charged consumers the full price that production of their goods imposed upon society, it would be impossible to make the case for any further "social responsibility" with respect to, for example, the environment.

In fact, one of the major advantages of the market failures approach to business ethics is that it is the only one that is able to pick out the "right" level of pollution. There can be no ethical imperative to eliminate pollution completely, since without some pollution there would be no economy. Society as a whole must be willing to accept some degradation of the environment in exchange for the goods produced. What is important is that the level of pollution be determined by people's actual preferences, not simply the subset of those preferences that happens to be legally enforceable. In other words, the cost of production should be the same as the social cost. This is precisely the state that would obtain if businesses derived no profit from displacement of costs that markets do not internalize.

What other sort of constraints does this approach impose? Imagine for a moment a deontically perfect world, in which everyone could be counted on to comply with all moral requirements. How should an ethical corporation

[2] In referring to "social responsibility" I am implicitly drawing a somewhat rough-and-ready distinction between internal and external obligations. A corporation is usually an organizational hierarchy. Thus the manager is involved with two very different types of relationships—those with other members of the organizational hierarchy, and those with individuals outside of it. Employees of a firm are "inside," for example, whereas customers are generally "outside"—precisely because the latter are not under the control of the organization in the way that the former are. Thus environmental issues, consumer relations, business partnerships, relations with rival corporations and with government will all be classified as "external," and hence as falling under the rubric of "social responsibility." For the purposes of this paper, I will also classify shareholders as "external." This is an admittedly crude division of the conceptual terrain, but I think that it does separate out a set of quite distinct issues, which need to be addressed on their own terms. For example, I think that a number of fundamental norms of reciprocity that apply to internal relations do not apply to external ones.

behave in such a world? The answer is quite simple. The firm should behave as though market conditions were perfectly competitive, even though they may not in fact be. The following list of imperatives provides some examples of the restrictions that this would imply:

1. Minimize negative externalities.
2. Compete only through price and quality.
3. Reduce information asymmetries between firm and customers.
4. Do not exploit diffusion of ownership.
5. Avoid erecting barriers to entry.
6. Do not use cross-subsidization to eliminate competitors.
7. Do not oppose regulation aimed at correcting market imperfections.
8. Do not seek tariffs or other protectionist measures.
9. Treat price levels as exogenously determined.
10. Do not engage in opportunistic behavior toward customers or other firms.

I think it is clear from this list that, rather than being morally lax, the market failures approach is actually quite restrictive. In fact, in the real world, any firm that began to unilaterally respect these constraints would be quickly eliminated from the marketplace. For instance, the requirement that firms compete only through price and quality excludes the use of non-informative advertising as a way of building market share. Advertising, as a form of non-productive competition, imposes significant deadweight losses on the economy. For example, Molson and Labatt spend $200 million per year on advertising. Studies have shown, however, that this competition is zero-sum. The amount of beer consumed has actually fallen over the years—thus the two companies are, at best, simply stealing customers back and forth from one another. This drives up the price of beer, a situation that is only sustainable because of market imperfections—namely, the significant economies of scale in the brewing industry, which constitute an effective barrier to entry.

Assuming that the nuisance value of beer ads exceeds their entertainment value, this means that society as a whole would be better off if the breweries stopped advertising. But it would be suicide for either company to do so unilaterally. The situation is identical to that of a country hoping to escape from an arms race through unilateral disarmament. Such a situation provides an ideal occasion for the old "they are doing it, so we have to do it too" defense of noncooperation. (This is an argument used in favor of illegality as well, e.g., when foreign competitors are able to engage in business practices that would be considered corrupt in the home country.)

Of course, the fact that other people are not going to respect their moral obligations does not undo the obligation for everyone else. It may provide an excusing condition—a reason why one need not respect one's moral obligation in this case. At the same time, one is still obliged to do what is necessary in order to bring about the conditions under which the obligations could be fulfilled. And it cannot be argued that these demands are too onerous in principle, since the demands simply articulate the way that capitalist economies are supposed to function in the first place. Thus it is only the possibility of unethical behavior by others that could justify noncompliance.

There are a variety of different ways in which businesses might try to bring about the conditions under which they could satisfy these ethical demands. The first is that they might engage in "experiments in trust,"—build up cooperation through reciprocity over time. We are already familiar with this process from the dynamic of arms negotiations. Thus, for example, firms might all agree to scale back their advertising expenditures by a fixed percentage every year, until they are eliminated completely. Compliance in the first round of cuts would help to build confidence going into the second.

Firms might also enter into agreements to restrict unethical conduct outside the framework of formal law. Antitrust concerns create an environment in which legislators are very suspicious of such agreements—especially those that would limit competition. However, it is worth distinguishing between productive and nonproductive forms of competition. Firms governed by the profit motive, given the opportunity to collude, will eliminate the former, whereas firms governed by moral principles will eliminate the latter. One can imagine the development of an environment, through trust-building exercises, in which corporations could demonstrate their commitment to ethical conduct, and thus earn the trust of legislators. In such an environment, corporations could enter into binding agreements with one another to enforce ethical conduct.

Finally, there is the point sometimes made in the literature that firms which actively profit from market imperfections are, in effect, tempting legislators and regulators to intervene. And when the state does intervene, the costs associated with compliance usually leave all of the firms involved worse off than they had been prior to their exploitation of the imperfection. Thus companies may pressure one another to respect moral principles using the "stop it or you'll get us all caught" appeal. This sometimes provides an incentive structure that is able to secure the desired pattern of behavior even in the absence of regulation (although fans of "industry self-regulation" have a tendency to overestimate the number of circumstances in which such incentives are present).

1.5. Further Directions

The market failures approach to business ethics elaborated here shows that a very robust moral code can be developed out of the idea that the fundamental obligation of managers is to maximize shareholder value. It has always been accepted that managers must do so within the framework of the law. The suggestion here is simply that an ethical manager is one who does so while respecting not only the letter of the law but also its spirit—which is to create the conditions necessary for private enterprise to generate an efficient allocation of goods and services in the economy.

However, there is a significant complication with this view, one that merits further discussion. The problem arises from what is known as the "general theory of the second best," or the "second-best theorem" for short (Lipsey and Lancaster 1956). This theorem shows that in a situation in which one of the Pareto conditions is violated, respect for all of the other Pareto conditions will generate an outcome that is less efficient than some other outcome that could be obtained by violating one or more of the remaining conditions. In other words, while perfect competition generates a perfectly efficient outcome, a situation that is as close as possible to perfect competition will not generate an outcome that is as close as possible to perfect efficiency.

The second-best theorem blocks a line of analogical reasoning that has long appealed to economists. Everyone understands that Newtonian physics, for instance, employs a number of idealizations. We also understand that the more closely the real world resembles these idealizations, the more closely the objects at our disposal will respect these laws. So while we do not have access to a frictionless plane, we can often substitute a very smooth tabletop in order to illustrate a variety of principles. Furthermore, the smoother the tabletop, the more closely the objects on it will conform to the predictions of ideal theory.

People sometimes like to extend this sort of analogy to economics. Perfect competition, according to such a view, is like a frictionless plane. It is an idealization. But the more closely the real world resembles this idealization, the more closely the various predictions will obtain. Friedman is, like many others, tempted by this form of reasoning. He writes, for instance, that:

> Of course, competition is an ideal type, like a Euclidian line or point. No one has ever seen a Euclidian line—which has zero width and depth—yet we all find it useful to regard many a Euclidian volume—such as a surveyor's string—as a Euclidian line. Similarly, there is no such thing as "pure" competition. Every producer has some effect, however tiny, on the price of the product he purchases. The important issue for understanding and for policy is whether this effect is significant or can properly be neglected, as the surveyor can neglect the thickness of what he calls a "line." (Friedman 1962: 120)

On the basis of this analogy, we may be tempted to conclude that if perfect competition generates perfect efficiency, then near-perfect competition should generate something as close as possible to perfect efficiency. The second-best theorem shows that this line of reasoning is unsound. If one of the Pareto conditions is violated, then the closest approximation to perfect competition will produce an outcome that is less efficient—and thus worse for society—than some more distant alternative. This has massive consequences. It means, for example, that if there is even one trade barrier or tariff in place, then minimizing the number of tariffs will not necessarily produce the best outcome—we may be better off imposing some additional tariffs. Similarly, if one sector of the economy is subject to monopolistic pricing, then having prices in all the other sectors determined by competition will produce an outcome that is inferior to some other outcome that would result if these prices were not competitively determined.

Of course, the kind of information that would be required in order to figure out how to achieve the second-best outcome is almost always unobtainable. The second-best theorem is primarily a limitative result. It shows us that we cannot use the FFT to derive normative conclusions under real-world circumstances. Thus the second-best theorem basically blocks the line of reasoning that Friedman develops. It also presents a very fundamental challenge to the market-failures based approach to business ethics being mooted here. It suggests that ethical behavior, in the absence of complete reciprocity, may be bad not only for the firm that sticks its neck out, but for the rest of society as well.

Of course, this does not mean that the efficiency standard is deprived of all normative force. It simply means that we cannot make the big sweeping generalizations that were the stock-in-trade of economists of Friedman's generation. In particular, it means that the properties of general equilibrium models are not going to be relevant to the normative evaluation of actual economies. Moral reasoning in a business context must be a more contextual affair. We cannot simply adopt the best competitive strategy, then hope that the invisible hand will take care of the rest. Even if we are in perfect conformity with both the spirit and the letter of the law, profit-maximization may still generate an inferior outcome.

There are several responses that suggest themselves at this point. The first is that the FFT specifies the conditions under which a Pareto-optimum is attainable. But in day-to-day life, this optimum is irrelevant. Every voluntary exchange generates a Pareto improvement. It is through these tangible, incremental efficiency gains that the private market system has established its merit. Thus, instead of offering a "top-down" justification of profit-seeking—through appeal to the general equilibrium of the economy as a whole, one could adopt a more "bottom-up" strategy, which would appeal to the particular efficiency gains that the firm is able to realize among its shareholders, its employees, and its customers.

We can think of this approach as a "resource custodianship" perspective. The ultimate goal of the economy as a whole is to satisfy human needs. The demand for various goods is an expression, however imperfect, of the intensity of these needs. The function of the price system is to channel resources toward the satisfaction of the most important of these needs (not according to an objective measure, of course, but rather according to each individual's own assessment of his or her needs). Thus the firm purchases a bundle of productive inputs in order to satisfy these needs, and profit—when earned under the correct conditions—is the reward that is enjoyed for having done a better job at satisfying these needs than any of its rivals.

Thus we can think of all productive resources as being "earmarked" for the satisfaction of needs. The managers and shareholders are the custodians of these resources. Their job is to convert these resources into consumer welfare—and when they do, they are rewarded with a profit. As a result, whenever the firm uses these resources in a way that does not contribute to welfare, but rather imposes deadweight losses on the economy as a whole, it is acting as a poor custodian of these resources.

Using this sort of "bottom-up" reasoning, I believe that all of the constraints outlined in section 4 could be justified in some form. In this framework, the Pareto conditions would function as a set of heuristics, allowing us to determine what type of conduct, in general, is likely to constitute an illegitimate source of gain. However, actually making the case requires a more detailed analysis, one that examines the specific conditions of the market in question. These remarks are clearly unsatisfactory. The more general research program, however, is one that I believe has considerable promise.

2 }

Stakeholder Theory, Corporate Governance, and Public Management

(with Wayne Norman)

2.1. Introduction

For supporters of the "stakeholder theory" (SHT) of the firm, shareholders are but one of a number of important stakeholder groups. Like customers, suppliers, employees, and local communities, shareholders have a stake in, and are affected by, the firm's success or failure. According to one typical formulation of the claim, "*In the same way* that a business owes special and particular duties to its investors…it also has different duties to the various stakeholder groups" (Gibson 2000: 247).[1] The firm and its managers have special obligations to ensure that the shareholders receive a "fair" return on their investment; but the firm also has special obligations to other stakeholders, which go above and beyond those required by law. In cases where these interests conflict, the demands and interests of some stakeholders, including shareholders, must be moderated or sacrificed in order to fulfill basic obligations to other stakeholders.

Naturally, this idea of "shareholders as just another stakeholder group" is not one that underlies corporate law in most market economies. In corporate law, shareholders are given preeminent status as the owners of the firm. They are able to elect all or most of the members of Board of Directors, which in turn has the right to hire and fire senior executives and approve or reject important policies and strategies of the firm. In effect, the shareholders have the right to treat the firm as a vehicle to maximize the return on their investment. While the board is supposed to ensure that the firm respects its legal and contractual obligations to other stakeholder groups, it is also fully within its

[1] Emphasis added. Note that Gibson is articulating this claim, not necessarily defending it.

rights to instruct managers to consider the ultimate purpose of the firm to be the maximization of profits and shareholder value.[2]

Because of the extraordinary status and control that shareholders are given under corporate law, stakeholder theorists have tended to devote relatively little attention to defending shareholder rights. The assumption has been that shareholders already have the power to ensure that their interests are taken into account by the firm and its managers. Stakeholder theorists who have considered the basis for shareholders' rights have usually tried to demonstrate why these rights should be limited or circumscribed by the rights or interests of other stakeholder groups.

"Enron" should make us reconsider this assumption. We use "Enron" here as a symbolic stand-in for the wave of corporate scandals that rocked American business between late 2001 and throughout 2002 (involving leading firms like Arthur Andersen, WorldCom, General Electric, Tyco, Qwest, Adelphia, Halliburton, Global Crossing, AOL Time-Warner, Merrill Lynch, Health South, and, of course, Enron). As it turns out, shareholders in the Enron era did not have the power to assure that their interests were fully taken into account by senior management. While there is no common explanation of what went wrong in these companies, we can nevertheless trace the source of almost all of these scandals to a breakdown of the governance relation between shareholders, the board, and the senior executives. There are obvious lessons here for those with a vested interest in the system of shareholder-focused capitalism—i.e., investors, brokerages, auditors, financial regulators, legislators, and so on—and their reaction has been swift. Authorities tried quickly to identify the flaws in the governance relation that had facilitated the most egregious malfeasance, and then proposed "patches," often in the form of revised regulations or voluntary codes, to discourage or prevent similar scandals in the future.[3] The principal aim of virtually all of these post-Enron reforms has been to strengthen the accountability of corporate executives to their boards and their shareholders.

---

[2] There is, of course, considerable ambiguity concerning the meaning of profit, shareholder value, and so forth, as well as legitimate doubts as to whether any firm actually seeks to maximize, rather than simply satisfice, with respect to any of these objectives (see Boatright 1999: 190–191).These debates are not essential to our purposes—we use the term "profit-maximization" simply as a shorthand way of referring to the pecuniary interests of shareholders, however these may be specified. The discussion of principal–agent theory, below, contains further clarification in this regard. Anyone interested in the notion of sustainability for a business should be concerned about economic rather than merely accounting profits; that is, profits after all inputs, including the cost of capital, have been paid out. "If a company is unable—over the long-term—to earn a return on its capital that covers the cost of its capital, then ultimately, it will fail due to the inability to attract the capital needed to replace its assets" (Grant 1998: 33). This implies that to be sustainable, businesses must be highly profitable. Even in bull market years, more than half of the largest 1,000 non-financial corporations in the United States can fail to cover their costs of capital.

[3] The most prominent of these is the Sarbanes-Oxley Act of 2002, which tightens up corporate governance and auditing rules. The NYSE and other major stock exchanges have also raised their governance requirements for participating firms. There have been a number of voluntary measures, including an agreement by the "final four" accounting firms to avoid conflicts of interest involved in offering consulting and auditing services to the same clients; and the move by many firms to treat stock options as an expense.

In this paper we argue that "Enron" offers even more important lessons for stakeholder theorists who oppose the dominant shareholder-focused conception of the firm. First, stakeholder theorists have underestimated the extent to which shareholder interests and shareholder control are crucial to furthering the interests of other stakeholders of the firm. Every one of the stakeholders of Enron was harmed when its senior managers conspired against the interests of the shareholders and when investors lost confidence in the company. And second, issues of governance and corporate law have received insufficient attention among advocates of a radical departure from the shareholder-focused conception of the firm. Although we will in several places highlight the reasons for believing there should be a strong convergence of the interests of shareholders and other stakeholders, our focus will be on the relevance of agency problems to governance in general, and to the governance of "stakeholder-friendly" firms in particular.

The breakdown of the governance relation in the scandals of the Enron era was at heart a failure of these firms and their shareholders to protect themselves against agency problems. By exploiting information asymmetries and conflicts of interests on the board, the agents (senior executives) were able to act against the interests of the principals (the shareholders), and to do so with a reasonable expectation of evading punishment. The central question posed in this paper will be whether governance relations in firms that assume primary obligations not just to shareholders but to other stakeholder groups as well can be safeguarded from comparable agency problems. This question will be approached from two angles: first, by looking at some abstract, structural features of agency problems that are likely to pose a challenge for what we might call a stakeholder theory of governance; and second, by proposing what we believe to be a fruitful area of empirical and scholarly research for those interested in the viable governance of stakeholder-friendly firms in the private sector, namely, the study of governance failures in public sector firms with multi-stakeholder, or "social responsibility" mandates. The primary goal of this inquiry will be to make the case for bringing together two very extensive debates—within stakeholder theory, on the one hand, and within public management, on the other—that have hitherto been carried on in mutual isolation. The moral of this cautionary tale about agency theory and public management will be that any naïve restructuring of corporate law and corporate governance to encourage stakeholder management could result in firms that are prone to both the internal fraud of Enron and the colossal inefficiencies of, say, Ontario Hydro or British Steel.[4] Because of these potential problems,

[4] For a quick overview of the financial performance of public enterprises, see Ramanadham (1991: 117–120). A more detailed meta-analysis can be found in Boardman and Vining (1989). The leader of Ontario's socialist New Democratic Party once described Ontario Hydro as "a demonic, empire-building force unto-itself."

the basic normative intuitions behind stakeholder theory might best be met by strategies carried out within firms that retain a shareholder-focused governance structure.

2.2. Corporate Social Responsibility and Stakeholder Theory

It is important to begin by clarifying which aspects of "stakeholder theory" are most relevant to the analysis of corporate governance. It is not the purpose of this paper to provide an overview of the vast literature on corporate social responsibility (CSR), stakeholder theory (SHT), and the so-called Triple Bottom Line (3BL). What matters is simply the core conviction of those committed to such models: that corporations have more extensive duties to key stakeholder groups like employees, communities, customers, suppliers, and so on, than is strictly required by law. All of these theories stand in opposition to a supposedly more classical conception of managerial obligation where, to quote Milton Friedman, the only "social responsibility of business is to maximize profits" (Friedman 1970), and where shareholders are the pre-eminent stakeholders. To get a clearer picture of both stakeholder theory and its classical alternatives, it is worth distinguishing several very different, if sometimes interrelated, theories[5]:

1. *Ontological SHT.* A theory about the fundamental nature and purpose of the corporation. A firm is essentially an "organizational entity through which many different individuals and groups attempt to achieve their ends" (Boatright 2000: 357). "The very purpose of the firm…is to serve as a vehicle for coordinating stakeholder interests" (Evan and Freeman 1988: 151). This stands in contrast to the shareholder-centered view of the firm as an economic entity that marshals resources for the purpose of making a profit for its owners.
2. *Explanatory SHT.* A theory that purports to describe and explain how corporations and their managers actually behave. "Managing stakeholder relations, rather than managing inputs and outputs, may provide a more adequate model for understanding what people in corporations actually do" (Boatright 2000: 391).
3. *Strategic SHT.* A theory about how devoting sufficient resources and managerial attention to stakeholder relations will tend to lead to positive (profitable) outcomes for the corporation.
4. *SHT of Branding and Corporate Culture.* A subset of strategic SHT, this is a theory about how a commitment to pay extraordinary

[5] The following set of distinctions expands upon the very influential four-part distinction described in Donaldson and Preston 1995.

attention to the interests of particular stakeholder groups (especially customers and/or employees, but also in some cases to "communities" concerned with the environment or with human rights) can be a fundamental aspect of a firm's basic branding and corporate culture. "Dolphin-friendly" or "The customer is king" can be profitable strategies.

5. *Deontic SHT.* A theory that determines the legitimate interests and rights of various stakeholders (presumably going above and beyond their legal rights), and uses these as a way of determining corporate and managerial duties.
6. *Managerial SHT.* A catch-all theory of management (incorporating theories of organizational behavior, HRM, CRM, leadership, operations research, accounting, and so on) that helps leaders and managers to realize the strategic benefits and satisfy the deontic requirements of SHT. "Stakeholder management requires, as its key attribute, simultaneous attention to the legitimate interests of all appropriate stakeholders, both in the establishment of organizational structures and general policies and in case-by-case decision making" (Donaldson and Preston 1995: 67).
7. *SHT of Governance.* A theory about how specific stakeholder groups should exercise oversight and control over management (e.g., which groups, in addition to shareholders, should be represented on the board, and how the board should function).
8. *Regulatory SHT.* The theory of which interests and rights of specific stakeholder groups ought to be protected by government regulation of business activities. In modern market societies, the dictate to "maximize profits while obeying the law" will necessarily involve fulfilling a vast body of obligations to suppliers, employees, customers, communities, and so on, since these obligations are legally binding.
9. *SHT of Corporate Law.* A theory about how traditional corporate law should be amended to reflect the principles and practices favored by Ontological, Deontic, and Governance approaches to SHT. Among other things, such an approach to corporate law would have to shield managers who favor non-profit-maximizing strategies of serving stakeholder interests from the wrath of shareholders and financial markets. Most importantly, it will have to give managers the ability to fend off hostile takeovers when other investors believe they could realize greater profits by changing managers and strategy.

There is a debate in the literature over whether it makes sense to talk about a unified stakeholder theory, or whether there are really many different kinds of theories that come into play. Without taking sides in this debate, one may conclude from the above list that thinking about the role of stakeholders in business involves a tremendous range of different theories, disciplines, and methodologies—from economics, law, ethics, political philosophy, and all of the social sciences underlying the managerial sub-disciplines, not to mention metaphysics (for Ontological SHT). Even when discussing any particular category of so-called SHT, the use of the term "theory" is often very loose indeed. Thus we follow the authors of a recent survey article in taking "stakeholder theory" to denote not a theory per se but "the body of research which has emerged in the last 15 years by scholars in management, business and society, and business ethics, in which the idea of 'stakeholders' plays a crucial role" (Jones et al. 2002: 19).

The form of SHT that will serve as the focus for the discussion that follows is Deontic SHT. The goal is not to explore the foundations of the theory, which claims that firms have an ethical duty to stakeholders above and beyond what is required by law—and, in particular, ethical duties that require the firm to operate in ways that will foreseeably reduce long-term profits. For the sake of argument, *we will consider the case of a firm that has assumed extensive extra-legal, profit-diminishing obligations to some of its stakeholder groups* and will then inquire about the implications of such a decision for managerial and governance processes. We will refer to this as a "Deontic stakeholder program" or a "strong CSR program."

For this reason, not much will be said about the way government regulation supports and enforces stakeholder rights and obligations. It is nevertheless extremely important to see these state functions as setting the context for almost any practical discussion of SHT; and it is astounding how seldom this is discussed in the literature. After all, if there is a sound moral argument for the claim that a particular corporation ought to assume extensive obligations to particular stakeholder groups, then there is prima facie a strong argument for the claim that all firms in its industry ought to assume these obligations; and therefore a strong argument for the regulation of this industry to ensure that these obligations are met by all firms on a level playing-field. Contrariwise, if there is a good argument against the state imposing a particular regulation to protect a certain stakeholder group of some industry (because, say, the costs of such a regulation would outstrip the benefits), then there may be an argument for the claim that no particular firm within that industry has a strong moral obligation to act as if there were a regulation.[6] This is not to deny that there are often moral (and of course self-interested) reasons to do certain things, even though it would not make sense for the state to require them. But the case has to be made.

[6] It is worth noting that even Chicago school economists and lawyers can be more receptive to well-designed regulation applying to all firms than to self-imposed "regulation" assumed by one firm. See, e.g., Easterbrook and Fischel (1991: 38), whose argument for some state regulation rather than a looser stakeholder-friendly governance structure emphasizes agency costs associated with the latter option.

When CSR theorists and stakeholder theorists ignore the role of state regulation, they are omitting some context. But if they ignore the role of corporate law in laying out governance structures, fiduciary duties, and stakeholder obligations, then their recommendations for socially responsible management are both incomplete and, quite possibly, incoherent.[7] Corporate law varies significantly from one country to another (and from one state to another in the United States), and regulates, among other things, the rights of shareholders and other stakeholders to determine membership on the board as well as certain fundamental transactions (such as mergers and sell-offs). It also regulates the duties of boards and board members, and the duties that managers have to the board and the shareholders. In effect, corporate law is what defines the legal bounds of the governance relationships between owners (and sometimes other stakeholders), the board, and managers. The most basic rights of shareholders that it regulates concern the information that the managers must disclose (about the financial health of the firm, but also potentially about its social and environmental record and policies), the rights to acquire and sell shares (including the tendering of offers to buy shares against the wishes of management), and about the voting rights of shareholders on board membership and basic corporate policies.

Why should the reform of corporate law matter to stakeholder theorists? There are at least two general reasons. The first is related to the argument for the prima facie preference for rules that are binding on all firms in a sector rather than self-imposed (and thus possibly disadvantageous) for one firm. If one is really committed to Ontological SHT—the idea that the firm exists essentially to serve the interests of all stakeholders—then why not build that into the governance structures by enabling certain stakeholders a fundamental role in governance, for example, by having representatives of these groups on the board (such as unions enjoy in certain European states)? The implications of such reformed governance structures will be discussed more extensively below, in the context of "multi-principal agency problems."

Perhaps more urgently, CSR and stakeholder theorists must be concerned about the justification and reform of corporate law, since many of the proposals they might recommend for socially responsible managers would be self-defeating under the current legal regime. Consider one simple illustration. In most market societies (Germany and Japan are exceptions), shareholders are given the exclusive right to elect the board, and the board is supposed to ensure that managers act in the shareholders' interests. Managers who forsake shareholders' interests may be fired by the board, and in some cases even sued by

[7] It should be noted that R. Edward Freeman has always taken seriously the implications of his theory for the reform of corporate law. Few other stakeholder theorists, however, have paid much attention to this challenge. See Marcoux (2000) for a concise discussion of the relevance of corporate law to CSR.

shareholders. Now one of the ways in which managers can fail to act in shareholders' interests is by following a strong CSR program, sacrificing a certain amount of profit to advance other stakeholder interests. (Other ways include, of course, lining their own pockets, or simply making bad decisions and managing ineffectively.) In such a case, under the governance structures laid out in corporate law in many countries, CSR managers could be fired. But even if managers are not fired by the current shareholders, they could very well be dismissed by future "corporate raiders." Corporate law governs the tendering of offers and the ways in which managers are able to resist hostile takeovers. One ever-present danger for a management team committed to a strong CSR program—and this holds even if they can convince the board to sanction such a program—is that CSR and stakeholder-friendly policies might fail to maximize profits and would, therefore, depress share prices. Investors who think that they could make more money with the resources of the firm under new management will then have an incentive to take the firm over and rid it of its CSR management team. In most market societies, such a takeover would be likely to succeed, because managers have limited freedom to create "poison pills" or "shark repellent," to make such acquisitions unpalatable for the raider. It would seem to follow that supporters of strong CSR and Deontic SHT must necessarily be in favor of reforming corporate law in ways that prevent the market from, in effect, swallowing up any stakeholder-friendly firm that failed to maximize profits.

This sort of situation illustrates a dilemma for stakeholder theorists that will be explored in the following two sections. On the one hand, if shareholder-centered corporate law is not reformed, then any CSR strategy that is not simultaneously profit-maximizing is likely to be snuffed out by free exchanges within financial markets. On the other hand, if corporate law is reformed to give managers the right to protect themselves and their CSR strategies from hostile takeovers, serious agency problems are likely to arise. Managers could use these protections simply to shield themselves from the market consequences of ineffective or even downright corrupt practices. Just as a feasible political theory cannot assume that leaders within the proposed system of government will be altruistic and public-spirited, a feasible theory of stakeholder management cannot assume that the managers will always have the stakeholders', rather than their own, interests at heart. It is possible that stakeholder management will give us both the worst of public management and the worst of Enron. The mere fact that Deontic SHT is a normative theory does not give us license to ignore this concern.

2.3. Governance and Principal–Agent Theory

Governance questions, along with questions about the nature and justification of the corporate law that sustains governance, are indispensable to any coherent theory of CSR. Questions of corporate governance, however, only become interesting when one refrains from thinking of firms as unified entities that make decisions and carry them out like individual agents. Notwithstanding its status as an artificial person under some articles of commercial law, the modern corporation is generally owned by thousands or millions of actual persons, directed by a dozen or so who are supposedly acting on the owners' behalf, and run by a deep hierarchy of managers—many of whom are also part-owners. Whatever obligations the corporation may have to outsiders can only be understood in the context of the vast and complex network of obligations between these owners, directors, managers and employees who think on behalf of the organization, but also on behalf of themselves. Principal–agent theory is one powerful tool for making sense of these obligations.

Principal–agent theory deals with situations in which one person, *the principal*, wants to induce another, *the agent*, to perform some task that it is in the principal's interest, but not necessarily the agent's. The principal can achieve this effect either through moral suasion (in effect, changing the agent's intentional states in order to make him more disposed toward performance of the task), or through the provision of incentives. Although the economics literature has tended to focus upon the latter mechanism, this is not an intrinsic feature of the model (Buchanan 1996; Campbell 1995). Almost any real-world principal–agent relationship will involve some combination of internal and external control. For an employee working under a piece-rate compensation scheme, the external incentives are largely sufficient to guarantee compliance with the principal's aims. In fiduciary relations, on the other hand, external incentives tend to be extremely weak, and so principals depend very heavily upon moral constraint on the part of the agent to secure compliance.[8]

In general, the employees, managers, and shareholders of a firm all have a common interest in ensuring the success of the enterprise. However, this common interest does not necessarily generate a natural harmony of individual interests. Individuals can often derive personal advantage from actions that are contrary to the common interest; in other words, they can "free ride." The most familiar example of such a strategy is *shirking*—investing less work effort in a task than possible (or than is expected), while enjoying the benefits of the higher effort levels of others. In effect, a productive, successful firm is a "public good" for its members (i.e., they all derive a benefit from it, but individually

[8] It should be clear so far, and below, that we fundamentally reject Neil Shankman's (1999) contrast between agency theory and stakeholder theory. Agency theory as such is neutral about who can be principals and agents; so a stakeholder theory can, and should, be concerned about agency problems that would arise if various stakeholders (and not just shareholders) act as principals. The fact (if it is one) that "most work in agency theory has focused on the relationship between owners and managers" (Shankman 1999: 322) does not tie agency theory fundamentally to a shareholder-focused conception of the firm. We thus reject most of Shankman's twenty-two points of contrast between agency theory and stakeholder theory (323–324).

self-interested action will fail to secure it). In order to produce this good, it is necessary to overcome a complex set of collective action problems. These collective action problems arise not only among co-workers, or between supervisors and employees. The separation of ownership and control in the modern corporation also generates the potential for significant free-rider problems between managers and shareholders. This potential divergence of interests is what makes it fruitful to conceive of the relationship between managers and shareholders as a principal–agent relationship.

The primary function of corporate governance structures is to mitigate or resolve these collective action problems. Of course, when actions are fully observable, and all other information is common knowledge, then the construction of such incentive systems is trivial. The principal–agent framework becomes interesting only when there is some information asymmetry between the principal and the agent. This is certainly the case between senior managers and shareholders or board members. Such asymmetries give rise to the potential for opportunistic behavior: managers can use their more intimate knowledge of what is going on within the firm to enrich themselves at shareholders' expense. Two situations have been the focus of particular interest in the literature:

- *Moral hazard* arises when the agent's action, or the outcome of that action, is only imperfectly observable to the principal. A manager, for example, may exercise a low level of effort, waste corporate resources, or take inappropriate risks.
- *Adverse selection* can arise when the agent has some private information, prior to entering into relations with the principal. Individuals with poor skills or aptitude will present themselves as having superior ones, people with low motivation will apply for the positions that involve the least supervision, and so forth.[9]

It has long been understood that extremely severe moral hazard problems may arise between senior management and shareholders. Experience has shown that managers may misappropriate or destroy not millions, but billions of dollars worth of corporate assets, when given the opportunity to do so. For example, the profligacy and waste that occurred at RJR Nabisco during the 1980s were due to a generalized failure to impose effective discipline upon management (Burrough and Heylar 1990). At one point, senior managers had a fleet of ten private jets and thirty-six pilots at their disposal, along with a private hangar at the Atlanta airport to service the fleet, complemented by a separate three-story facility to serve as a waiting lounge. The latter was built

[9] For an especially clear discussion of moral hazard and adverse selection, see Rasmusen (1989: 163–245).

and appointed under explicit instructions from the CEO that the budget was "unlimited" (Burrough and Heylar 1990: 93). In the Enron era, of course, the issue was not so much waste as it was the direct transfer of wealth from the corporation and its shareholders into the bank accounts and stock portfolios of senior executives. The root cause, however—an underlying moral hazard problem—was the same.

Thus despite the standard assumption in the business-ethics literature that management serves (at very least) the shareholders of the firm, in reality the alignment of incentives needed to obtain this result can be difficult to achieve. The exploitation of shareholders by management remains extremely common. Apart from the power to hire and fire managers, there are only two important levers that shareholders control and that serve as a check on management. The first is compensation. Firms often experiment with different compensation schemes, including performance pay, bonuses, stock ownership, and stock options, in order to give managers a personal interest in maximizing shareholder value. Managers can be given a bonus, for example, that is equivalent to some fraction of the output that can be achieved when they exercise high effort (an incentive structure similar to a sales commission). The second major control mechanism, as mentioned earlier, is the discipline imposed by the stock market through the threat of hostile takeover. Managerial waste and inefficiency tend to depress stock value, which makes the firm a more attractive target for a buyout. Such a change usually results in a consolidation of ownership, which gives the new shareholders both the power and the incentive to dislodge the old management, and then profit from the subsequent increase in the value of the firm.

These levers, however, are far from foolproof. Several factors contribute to the difficulties that shareholders have exercising effective discipline over management. The first is the magnitude of the information asymmetries that exist between the two groups, and the sheer cost associated with acquiring the information needed to assess managerial performance. There is also, as the "Enron" scandals reveal, ample opportunity for managers to conceal this information, or to frustrate the attempts of shareholders to gain access to it (not least by corrupting the auditing process that is supposed to give the board an independent assessment of crucial financial information).[10] Compounding the problem is the fact that shareholders often face their own collective action problem when it comes to oversight and discipline. Keeping an eye on management, and challenging certain decisions, requires an investment of time, energy, and resources. When there is a single dominant shareholder, that

[10] Enron's auditor, Arthur Andersen, arguably had a greater reason to be loyal to Enron's management—who, among other things, were able to extend or withdraw lucrative consulting contracts, include consulting on the accounting schemes supporting the notorious off-balance-sheet partnerships—than it did to Enron's board or its shareholders.

person will usually find it to be in his or her interest to take on these charges. Yet when ownership is extremely diffuse, the cost to each individual shareholder of managerial excess tends to be quite small, while the costs associated with disciplining management remain high. So absent some cost-sharing arrangement, no single shareholder will have an incentive to "mind the shop."

Understanding that the relationship between shareholders and managers is one fraught with agency risks helps to shed some light upon the importance that profitability plays in traditional corporate governance. Standard principal–agent models are one-dimensional: they assume that just one agent is relating to one principal and is performing only one task on his or her behalf. In reality, agency relationships are almost always multitask. Managers, for example, are expected not only to make profits but also to cut costs, maintain or improve product quality, increase market share, project a good corporate image, and so on. When there are multiple tasks, it becomes extremely difficult—often impossible—to design incentive schemes that will motivate the agent to produce an optimal performance (Laffont and Martimort 2002: 203–226). Sometimes tasks will be complementary: investing effort in one task will reduce the marginal cost of investing in some other (e.g., in manufacturing with economies of scale, increasing market share may lead to increases in productivity). Here there is no difficulty. Problems arise when the tasks are substitutes; when investing effort in one increases the marginal cost of investing in another (e.g., in retail, "cannibalization" of sales may mean that increasing market share leads to decreased productivity). In such cases trade-offs will be necessary.

To give just a sense of the problems that this may create, consider what will happen if the effort invested in one task is more observable than the effort invested in some other. In principle, it is possible to design a much "sharper" set of incentives for the task that is more easily observed. And yet if one were to do so, managers would tend to invest a disproportionate amount of energy into performance of that task. Thus it is necessary, in a multitask environment, to provide "dull" incentives across the board, even though the information conditions actually permit sharper incentives in certain domains. Several theorists have speculated that it is precisely because of this multitask problem that most middle managers simply receive a flat salary, with only slight variance for annual performance (Dixit 1997; Holmström and Milgrom 1991; Williamson 1985).

The incentive problem becomes even more acute if the principal lacks the information necessary to determine how the various tasks should be balanced against one another (in cases where they are substitutes). If the agent is given discretion in this regard, then accountability becomes almost impossible. The agent can always explain away poor outcomes in one task as a necessary consequence of better outcomes in some other. This is what explains the importance of the "bottom line" in traditional corporate governance. It provides the equivalent of a composite index, a common metric for evaluating the performance of management across all of the important dimensions. It also provides broad boundaries on the trade-offs that managers can make between shareholder return and other objectives, such as growth or product quality. It provides, in other words, a single "metatask" for which upper management can be held accountable.

The reason that it is possible to impose such a concern for profitability is that there is only one group of principals, the shareholders. The existence of multiple principals complicates matters further. Assuming that the various principals have different preferences over the set of possible managerial tasks, the overall effect will be to dull incentives yet again. Each principal will encourage the agent to perform best in the task that he or she (the principal) views as the most important, and to discourage the others. Thus any incentives provided by the various principals will have a tendency to cancel each other out, leaving the agent free to pursue his or her own interests (possibly to the detriment of them all).

2.4. Lessons from Public Management

It should be obvious from the above summary that strong CSR or 3BL proposals would significantly complicate the agency relationships that exist within the firm. The precise modalities vary, but in general one can say 3BL proposals—which ask managers to improve social and environmental "bottom lines" in addition to net income—would exacerbate the multitask incentive problem, while responsibilities to multiple stakeholder groups could generate multi-principal problems. Thus we are naturally led to inquire how corporate governance structures would need to be modified in order to reflect such a conception of managerial responsibility. Again, recent corporate scandals have shown that if such obligations cannot be effectively institutionalized, managerial malfeasance could easily undo any of the "social" gains achieved through the introduction of a broader concept of corporate social responsibility.

There is no reason in principle why these agency problems should be unmanageable. As we noted above, CSR does not create entirely new agency problems, it simply exacerbates existing ones. Managers are already obliged to grapple with multiple tasks. Furthermore, shareholders do not all have exactly the same interests (e.g., some are concerned with short-term profits, others with long-term growth, some are institutional investors, and others participate through "ethical" investments funds). Thus one can already think of managers as balancing the needs of multiple principals. "Enron" notwithstanding, existing governance structures seem to be capable of doing a tolerable job of keeping these agency problems under control. Could they not just be extended to handle a strong CSR program?

The prospects of this seem dim. The grounds for such an assessment stem from the experience of state-owned enterprises (SOEs) in the late 1960s and 1970s. The governance challenges that would arise out of the implementation of a strong CSR program in a private enterprise are structurally quite similar to the challenges that were faced by nationalized industries during this period. The experience in the public sector shows that it is extremely difficult to design governance structures under such conditions. The experience of SOEs shows that giving managers a "social responsibility" mandate, combined with the freedom required to carry it out, can lead not only to massive financial losses, but may not even result in improved social responsibility. SOEs have often done a worse job of serving the public interest than privately owned firms in the same industry. While there are a number of different factors that combine to produce this outcome, most analysts agree that agency problems *created by the social responsibility mandate itself* figure among the primary causes.

In the period following the Second World War, many firms were either nationalized or created under state ownership, not because of monopoly or market failure in the private sector, but out of a desire on the part of governments to have these enterprises serve the broader public interest. Consider the case of Canada, a country with a business culture and governance tradition similar to that in the United States, but with an interventionist state inspired by the European model. As in many countries, Canadian SOEs were (and in some cases continue to be) involved in the standard activities of electricity generation and distribution, telecommunications, postal services, water and sewage, ports and airports, and so on, primarily because it is (or was) difficult to organize a competitive market in these sectors. But the Canadian state has at various times also owned an airline (Air Canada), a railroad (Canadian National), and an oil company (Petro-Canada), not to mention numerous mining operations. It has been involved in shipbuilding, aerospace, forestry, oil and gas exploration, nuclear-reactor building, agricultural land ownership, interurban bus service, and automobile insurance. These SOEs competed directly against privately owned firms, either domestically or in international markets. The standard "public goods" rationale for state involvement is absent in these cases. The reason that the state was involved in these sectors followed primarily from the thought that, while privately owned firms pursued strictly private interests (i.e., profitability), public ownership would be able to ensure that these enterprises served the broader public interest. Thus managers in these SOEs were instructed, not just to provide a reasonable return on the capital invested, but to pursue other "social" objectives. Of course, this story was played out in just about every Western European country in the twentieth century—in many cases to an even greater extent than in Canada.

The social responsibilities that have often been imposed upon SOEs by the state can be summarized under five general categories[11]:

1. *Macroeconomic.* SOEs were at various times called upon to engage in counter-cyclical spending or to maintain employment during recessionary periods, in order to smooth out the business cycle; to promote full employment by creating excess capacity and engaging in "make work" projects; and to help control inflation by instituting wage and price controls. SOEs have also been called upon to assist the government in meeting specific fiscal objectives.
2. *National interest.* SOEs were often expected to bolster national industry by providing subsidized goods and services (especially energy) to domestic firms. They were expected to provide guaranteed markets for the product of these industries, by favoring domestic suppliers over foreign. They were often expected to serve the national interest by channeling investment into sectors that were deemed to be national priorities by the state, or to assist in the "incubation" of industries intended to bolster international competitiveness. They were also intended to keep under national ownership and control industries, information, and productive technology that were regarded as essential to national security.
3. *Redistribution.* SOEs played a significant role in helping the state to achieve redistributive goals. Most often, they were expected to abstain from any of the price discrimination that profit-maximizing private firms would engage in, and thus to provide the same services at the same price across the nation (e.g., postal service). In Canada, SOEs have also been heavily involved in regional development, either directly subsidizing across regions (e.g., rural passenger train service or flight service to relatively remote regions or small centers), or through region-specific investments.
4. *Model employer.* SOEs were expected to serve as model corporate citizens, in order to put pressure on private firms to follow suit. Thus they were often expected to pay higher wage rates, to offer superior benefits (e.g., on-site daycare) and better job security, or to hire more women or members of disadvantaged minorities.
5. *Reduction of externalities.* While most of the "social" responsibilities of SOEs could be described as the production of positive externalities, it is worth noting that certain SOEs are held in the public sector purely for the sake of controlling negative externalities. Most notably, liquor sales and gambling are often under state monopoly, out of concern that private enterprises in this domain would produce "too

[11] For a more detailed overview, see Ramanadham 1991: 76–81. Also Lewin 1982: 53–58.

> much" of the relevant good. Similarly, there is often a call for public ownership of industries that have the potential to create catastrophic environmental externalities (such as uranium mining and refinement, nuclear energy generation, etc.).

It should be clear from inspection that this set of objectives has many points of contact with the "wish list" of stakeholder obligations that proponents of CSR have been advancing over the years. This is no accident. The prevailing view among social-democratic political parties during the 1960s was that the ownership structure of private enterprise was responsible for failures of corporate citizenship. The solution was therefore to nationalize these firms, and then instruct the managers to behave in a more responsible fashion. Stakeholder groups could articulate their interests through the democratic political process, and SOEs could then be directly instructed to address these concerns. In a sense, an attempt was made to use the state (and public law) as a governance mechanism to institutionalize stakeholder capitalism.

This experiment, however, is now widely regarded as a failure, and not only on the right of the political spectrum. The "public interest" mandate of SOEs was abandoned by socialist, conservative, and Christian Democrat governments alike, long before the wave of privatization that swept through Europe and North America in the 1980s. The heady days of the 1960s, in which SOEs were encouraged to pursue a variety of social objectives, were followed by a long period of "commercialization," primarily during the 1970s, in which SOEs were instructed to abandon or curtail these activities, and to restructure their operations in accordance with more traditional business principles (Ferner 1988). In fact, the managers of firms in competitive industries were often instructed simply to maximize profits. Thus in 1974, for instance, a government directive instructed Canadian National Railroad to be profitable, and a new director was appointed with an explicit mandate to implement the necessary changes (Stevenson 1988). In 1978, the Air Canada Act instructed the airline (with comical understatement) to run its operation with "due regard to sound business principles and, in particular, the contemplation of profit" (Langford and Huffman 1988: 99). Both of these decisions were made by the left-of-center Liberal government of Pierre Trudeau, long before there was any discussion of privatization.[12]

Similar stories unfolded in France and Spain, where socialist parties imposed "commercializing" reforms upon the state sector. In fact, one of the reasons that it was so easy for subsequent right-wing governments to privatize state firms is that in most OECD countries they had already been restructured in such a way that their behavior was no different from that of private

---

[12] For an overview of the Canadian experience, which puts particular emphasis on the non-ideological character of many privatizations, see Tupper and Doern 1988.

enterprises. As Joseph Stiglitz has observed (1994: 173), by 1994 there was essentially no difference in the behavior of Texaco (private), Petrofina (public), and BP (mixed). In cases where SOEs operated in competitive sectors, commercialization relieved them of their social-responsibility mandate, and thus eliminated the primary reason for holding them in the public sector.

The basic reason for this commercialization of the SOEs was the realization that, not only were they consistently losing money, but they were often doing a worse job of promoting the public interest, under the explicit mandate to do so, than their privately owned counterparts. In several countries, governments suffered an almost total loss of control. In France, state oil companies freely speculated against the national currency, refused to divert deliveries to foreign customers in times of shortage, and engaged in predatory pricing policies toward domestic customers (Feigenbaum 1982: 109). In the United States, SOEs have been among the most vociferous opponents of enhanced pollution controls, and state-owned nuclear reactors are among the most unsafe (Stiglitz 1994: 250). Of course, these are rather dramatic examples. The more common problem was simply that the SOEs lost incredible amounts of money (Boardman and Vining 1989). These losses were enough, in several cases, to cast doubt upon the ongoing solvency of the state and to prompt currency devaluations. The reason that so much money was lost has a lot to do with a lack of accountability.[13]

The most widely accepted explanation for this perverse outcome is that the structure of public enterprise made it extremely difficult for the state to exercise effective discipline over its managers. Some of these agency problems are intrinsic features of public ownership, but some were produced by the specific character of the "social responsibility" mandate that managers were given during the 1960s. It was the latter that commercialization was intended to correct.

The idea that agency problems in the public sector are more acute than in the private is widely accepted. In some cases, this is due to the peculiar character of the state as an owner. For example, the public sector cannot give its managers an ownership stake in the operation that they run. The top end of the pay scale is also significantly lower than in the private sector, for a variety of reasons, and this may make it difficult for SOEs to attract or retain top managers. There is also the well-known problem of the "soft-budget constraint." If the managers of a privately owned firm cannot keep it in the black, shareholders will eventually withdraw their investment, regardless of the social

[13] For a fascinating and very careful analysis of one such case, see Palmer et al. 1983. They analyze two intercity bus companies in Canada, from 1969 to 1997, one private and the other public. They attempt to determine why the public firm had the highest fares, yet had an average rate of return on net worth of only 6.3 percent, compared to 20.6 percent for the private firm. They conclude that, although the public firm ran some unprofitable routes that otherwise would not have had service, the primary reason for its weak returns was overcapitalization, due to weak political oversight.

consequences. Because of this, private owners are able to issue much more credible threats to their managers. Politicians, on the other hand, would never allow a major public corporation to go bankrupt, and the managers know it. Thus public-sector managers have much less fear of losing money. They sometimes intentionally run deficits in order to secure budget increases.

These problems are all quite specific to the public sector, and of no particular interest to proponents of CSR in the private sector. Furthermore, because these problems are tied to structural features of the public sector, the commercialization of SOEs during the 1970s did nothing to correct them. The same can be said for SOEs that are monopolies, whether they are "natural" monopolies or artificial ones created through legislation. Obviously the state could not issue a directive to the managers of such firms, telling them to start maximizing profits, since doing so would defeat the purpose of having them in the public sector. The cases that are relevant are ones in which SOEs operated in competitive industries. These are the firms that were commercialized during the 1970s.

The primary reason for commercializing these SOEs was to discontinue the practice of issuing multiple objectives to managers. Anthony Ferner summarizes the essential problem when he writes that, for SOEs:

> The way in which their objectives are defined through the political process and then 'transmitted' into the enterprise raises fundamental problems. First, the political demands on public enterprises lead to objectives that are confusing, changeable and often mutually at odds. Second, partly for this reason, but for others as well, the relationship between the state and public enterprises is dogged by difficult questions of enforcement: how can the political authorities ensure that the objectives set for state enterprises are effectively pursued? (1998: 30)

The reaction to this difficulty, in states throughout most of the Western world, was to give up on the goal of giving SOE managers multiple social responsibilities. This should be a cause of concern among proponents of CSR. In a sense, the history of nationalized industries in the twentieth century suggests that CSR was tried, and turned out to be a failure. At the very least, proponents of CSR must learn from this experience and think about how private corporations might institute governance structures that would allow them to avoid the problems that plagued the public sector. In this respect, it is helpful to look at these problems, and to divide them up into the categories of multitask and multi-principal problems.

2.5. Multitask Problems

The history of SOEs in the twentieth century makes it perfectly clear that firms cannot simply give managers multiple tasks, and then tell them to do "the best they can" in all dimensions. As Stiglitz argues, this sort of vagueness created serious agency problems in the public sector:

> [T]he ambiguity of objectives provides the managers further discretion to pursue their own interests. In the private sector, there is one over-riding concern: profits. In the public sector, there may be a multiplicity of objectives—economic (such as employment) as well as non-economic (national security). Managers can always claim that the reason they are losing money is not that they are inefficient or incompetent, but that they have been pursuing other goals. And it is virtually impossible for an outsider to judge the validity of those claims. (1989: 32)

Whenever there are trade-offs between different objectives, managers can explain the failure to meet one target as the "cost" imposed by their attempts to meet some other. Revenue shortfalls can be explained as a necessary consequence of maintaining employment. Layoffs can be justified as a necessary precondition for profitability. This makes it impossible for the principal to lay down any unambiguous performance criteria for the evaluation of management, which in turn leads to very serious agency problems. As long as the manager is determining how the various objectives should be balanced, assigning managers multiple objectives gives them something equivalent to a "get out of jail free" card—an automatic ticket to escape accountability for their own professional failings.

It should be noted that the multitask agency problem is not just an incentive problem, one that can be resolved through good will or more effective "internal" controls. Even managers who are willing to make a good-faith effort to do the best they can may find themselves lacking the information that they need in order to determine how well they are doing, or whether they could be doing better. Competition in the private sector not only creates incentives; it also provides important information about how firms are doing. In the early 1970s, for example, the big three automakers in the United States for the most part were simply unaware of how inefficient their operations had become. It was not until they were exposed to competition from the Japanese that they realized how much better they could be.

In order to make these sorts of comparisons across firms, however, managerial objectives must be commensurable. Having the single directive of profit-maximization permits comparisons across firms, because all managers are trying to do roughly the same thing, in a similar economic environment. But if managers have the freedom to balance objectives as they see fit, then the basis for comparison disappears, because any differences can be dismissed as a consequence of the opportunity cost of the specific type of balancing undertaken. A firm that puts more emphasis upon regional equality, or employment security, would simply not be comparable to a firm that put more emphasis on profitability. Thus the information needed for managers to assess even their own performances would, in general, be unobtainable.

Thus from a governance perspective, the only really feasible arrangement is for the principal to specify the balance that he or she would like to see obtain between the various objectives. Unfortunately, information asymmetries will often prevent the principal from doing so. It is generally impossible for an outsider to know what opportunities for profit are available, what internal efficiencies might be achieved, what level of risk-taking is appropriate, and so on. Even a very hands-on senior manager would be lucky to have such information.

There is considerable precedent on this issue in the history of public management. In France, it was decided early on that managers could not be given the discretion to balance objectives as they saw fit. The initial solution proposed was to explicitly calculate the cost that "social objectives" imposed upon the firm, then to measure performance on the basis of profitability after full compensation for these costs. The 1967 *Rapport sur les entreprises publiques* (or "Nora report") concluded that:

> Unless we can clearly distinguish the potential for profit specific to a particular economic activity from the costs imposed by the public interest constraints, there are no standards for these enterprises: no criteria of good management, no incentive to improve management, and no penalty for bad management. How then can we expect balanced finances from these enterprises, along with the innovative, autonomous and responsible action that constitutes its guarantee? (Nora 1967: 25)

Throughout the 1970s, the French state engaged in a process of "contracting" with the SOEs—developing elaborate arrangements that specified what the enterprise would be expected to achieve in each of the different categories of objectives. These contracts, however, proved difficult to negotiate, and even more difficult to enforce (Lewin 1982: 65–66).

This is an ongoing challenge for proponents of CSR and 3BL. In a sense, having three bottom lines is equivalent to having no bottom line. Thus, it is incumbent upon partisans of 3BL schemes to explain how they intend to handle the multitask incentive problem that their proposals create. At very least, such an effort should take as its point of departure the experiences of the public sector, since SOEs have at least three decades worth of experience in dealing with these issues. But the prospects are not encouraging. Norman and MacDonald (2004) have argued that there is no common metric that can be used in a 3BL context for evaluating social and environmental performance relevant to other stakeholders. If this is correct, then it is very difficult to see how any reform of corporate law designed to permit managers to pursue a 3BL agenda would not also open the door to rampant malfeasance.

2.6. Multi-Principal Problems

If 3BL approaches to corporate social responsibility involve assigning multiple tasks to managers, the Deontic SHT tradition foresees an arrangement under which managers would be accountable to multiple principals. Consider the following claim from Edward Freeman, who is widely credited with having introduced the concept of stakeholder theory into contemporary management theory:

> My thesis is that I can revitalize the concept of managerial capitalism by replacing the notion that managers have a duty to stockholders with the concept that managers bear a fiduciary relationship to stakeholders. Specifically I include suppliers, customers, employees, stockholders, and the local community, as well as management in its role as agent for these groups. . . . Each of these stakeholder groups has a right not to be treated as a means to some end, and therefore must participate in determining the future direction of the firm in which they have a stake. (1984: p. xx)

From the perspective of agency theory, this gives rise to an obvious objection. As Frank Easterbrook and Daniel Fischel write, "A manager told to serve two masters (a little for the equity holders, a little for the community) has been freed of both and is answerable to neither. Faced with a demand from either group, the manager can appeal to the interests of the other" (1991: 38).

The problem with treating managers as "agents" of all these groups is that there is often a straightforward divergence of interest among stakeholders. Union wage demands may directly impinge upon profitability, expansion of capacity may have a negative environmental impact, and so on. Consider, for example, an industrial union like the Canadian Auto Workers. Conflicts between this union and the "Big Three" automakers over wages and benefits are well publicized. Less well publicized have been conflicts engendered by the union's constant pressure to expand private automobile production and use, which has put it on a collision course with environmentalists, as well as public transit users and advocates. The union has also lobbied against the interests of Canadian-owned players in the automotive industry (mostly subcontractors making auto parts for "outsourced" production), not to mention the customers and employees of Japanese and European automotive firms. Under traditional corporate governance, shareholders may face a collective action problem when it comes to disciplining management, but at least they all share the same general interest with respect to the firm. The interest positions of stakeholders, on the other hand, often put them in zero-sum conflict with respect to other stakeholders or the decisions of the firm.

Thus holding managers accountable to the interests of all these different groups can create a serious multi-principal problem. In the best case scenario, accountability to multiple principals, in cases where each principal has control over some of the incentives that govern the agent, will result in a "dulling" of these incentives. Unions may reward managers for decisions that shareholders will punish. In the worst-case scenario, the incentives imposed by the two groups will cancel each other out completely, leaving the manager indifferent to the concerns of the principals.

The introduction of multiple principals also has a tendency to create a system of incentives that is dynamically unstable. A single principal is likely to have a stable set of preferences. With multiple principals, the system of incentives is likely to reflect a balance of power that may not be stable over time. This instability was a constant complaint of public-sector managers prior to commercialization: they were not only held accountable to multiple objectives, but these objectives would change from month to month, or day to day.

Managers governed by such an agency structure are likely to engage in strategic behavior in order to avoid accountability. They can often play one group or principal off against another. If the principals actually had independent power to sanction the agent, then it would be very unlikely that any workable governance structure could be established. In SOEs, the initial solution was to make the enterprise directly accountable to a single ministry, an industrial board, or a holding company. It was very rare for public-sector managers to be held directly accountable to stakeholder constituencies. Instead, stakeholder groups were given representation in some decision-making body or institution, which was charged with the task of reconciling the divergent interests, and issuing a coherent set of imperatives to management. This in itself was no easy task. In Spain, for example, the state holding company for SOEs initially had "representatives of the ministries of finance, commerce, industry, public works, agriculture, as well as the ministries of the army, navy and air force, on its board of directors" (Ferner 1988: 31). The result was almost completely unworkable.

Furthermore, the creation of a unified governance structure "on paper" does not mean that multi-principal agency problems go away in practice. Even though SOE managers were technically accountable to only a single agency, they could usually exercise considerable influence over the process of deliberation that informed the agency's decisions. Thus managers would routinely "play politics" with stakeholder groups, in order to change the balance of political power. Managers of public utilities, for example, would often appeal to large industrial clients, who had an interest in maintaining low rates, in order to help them lobby for expanded capacity, or to resist demands for profitability. The ability of management to selectively disseminate or leak information gives them a particularly powerful card to play in these affairs.

Is it plausible to think that such problems might become more tractable within a private enterprise system, with firms dedicated to CSR? It is difficult to see how. The primary problem with stakeholder groups is that, with the exception of trade unions and some environmental groups, they tend to be very poorly organized. Thus it is inconceivable that any such group should be able to exercise any direct control over management. From a governance perspective, proponents of CSR must be committed to having stakeholder groups represented on the board of directors of a firm.

Yet some stakeholders are so poorly organized that it is difficult to imagine them even coming together to elect a representative to the board. For example, when a firm accepts inflationary wage demands, which it then "passes along" through price increases, it creates a significant negative externality both for other workers and consumers. In certain economic climates, these wage-and-price decisions are far more deleterious to the public than any environmental externalities produced by the firm. Thus one of the primary mandates of SOEs was always to promote the "public interest" in price stability. Yet does one ever hear about firms refraining from contributing to inflationary pressures in the CSR literature? No, because the victims of inflation are simply too poorly organized to be described as a "stakeholder group." But is the suffering of the pensioner who has her pension wiped out by inflation any less real than that of the customer who winds up buying shoddy goods, or the supplier whose contract is abruptly terminated?

Thus we must be careful not to allow SHT to create an institutional bias that favors the better organized over the poorly organized. Yet what resources do the poorly organized have to press their interests? The traditional answer is that they have government. So thinking through the institutional implications of Deontic SHT leads quite easily to a SHT of governance that requires the state to take a leading role in appointing directors to the board of directors of firms, in order to make sure that all "stakeholder" groups are represented. Yet this is precisely what the failed nationalization strategy was intended to achieve. Thus it is absolutely incumbent upon proponents of SHT and CSR to explain why the solution that was eventually deemed superior in the private sector—arm's length regulation of profit-oriented firms—is not also the best suited to addressing their concerns.

2.7. Beyond the False Antagonism of Shareholders and Stakeholders

The argument developed above takes issue with the view that shareholders are "just another" stakeholder group. There are good reasons for according shareholders' interests priority in corporate governance. This is not because shareholders' interests are intrinsically more important, and certainly not because shareholders themselves, as individual persons, are more important than other persons. Nor has the argument been based on any strong conception of property rights, which accords shareholders priority because they are the owners of the firm. Shareholders do, of course, in some sense own the firm, and are treated as owners in many aspects of commercial law. But the argument presented here is consistent with the now-orthodox view in law and economics that shareholders (especially those with Class-A voting stock) are merely one of many providers of capital and financing for the firm.[14] The argument is based on governance considerations: because of the structure of risks and rewards that attracts shareholders to invest in the firm (or scares them away)—and in particular the fact that shareholders uniquely lack contractual guarantees of a return on their investment—they have the right incentives to be accorded a special role as the watchdogs (or the appointers of the watchdogs) over managers.[15] In effect, only by according shareholders a special role in protecting their stake in the firm can we expect managers to run the firm in a way that is in the long-term interests of other groups with a stake in the firm (such as employees, suppliers, customers, bondholders, lenders, etc). It is not greedy shareholders who are the enemies of other stakeholders; it is greedy (or lazy or unethical or unsupervised or simply unqualified) managers. We did not need "Enron" to teach us this; but the recent scandals have provided a textbook illustration of the agency problems that form the heart of the challenge of governance.

This paper takes issue quite specifically with the form of Deontic SHT that contemplates sacrificing profits and shareholder wealth in order to fulfill extra-legal, moral obligations to other stakeholder groups. The two central arguments both involve highlighting the agency risks that make such a corporate strategy likely to be self-defeating.

The first argument considered the prospects of a Deontic stakeholder project within the context of governance structures backed by corporate law. Under the current regime in most market societies, a management-sanctioned (or even board-sanctioned) CSR strategy that sacrificed shareholder wealth for benefits to other stakeholders would likely be self-defeating: in extreme cases, shareholders could sue the managers for neglecting their duties to increase profits; and under a more likely scenario, such a strategy would lead to a drop in share prices that would make the firm easy prey for corporate raiders. In order to permit such a CSR strategy, proponents would thus have to argue for reform of corporate law, and in particular, for reforms that would allow management to fend off hostile takeovers.

Such reform would be ill advised. On its own, it would significantly reduce the accountability of managers to anyone's judgment other than their own. The

---

[14] This is the so-called nexus of contracts theory or the contractarian model of the firm, which describes the firm as a set of explicit and implicit contracts. The firm is neither an entity nor a thing capable of being owned. "It is simply a legal fiction that encompasses a set of contractual relations" (Bainbridge 2002: 26). This theory of the firm is widely credited to Ronald Coase (1937).

[15] This rationale for shareholders' special role in governance is defended at length in Jensen (2001: especially chs. 4, 5, and 8). See also Boatright (1999: 176–194) for a more philosophical exploration of this model.

CEO could explain shortfalls in performance in terms of stakeholder commitments even if the real explanation involved managerial incompetence or even fraud. Of course, another consequence of a corporate law that allowed managers to shield themselves from shareholder wrath in this way might be that equity financing itself would ultimately dry up: people would be unlikely to buy securities with no fixed rate of return if they were not confident in management's ability to earn a profit. Firms would then be forced to offer shares at a discount or to seek financing at high fixed rates of interest from financial institutions that would generally be less than enthusiastic about financing a firm with a Deontic stakeholder strategy. In sum, strong CSR requires radical reform of corporate governance structures and corporate law, and there is no reason to think that some of the more obvious strategies for doing so would prove acceptable to society at large, or even to those most enthusiastic about promoting corporate social responsibility.

The preceding section provided a brief exploration of some further reforms of governance and corporate law that might be required. In addition, presumably, to allowing managers to shield themselves and the firm from hostile shareholders, such reforms would give other stakeholders a direct role in governance through representation on the board. This might be thought to be an improvement on the situation in which managers were "trusted on their own recognizance" to carry out CSR strategy to the detriment of share values, because at least it would make managers accountable to the stakeholders they are supposed to be benefiting. But as was shown, any confidence in this system of accountability should be undermined by the multi-principal problems that it creates.

These alternate governance scenarios suggest that we must be cautious about giving managers the means and discretion to carry out profit-consuming CSR strategies (although, again, not CSR strategies that enhance profitability), because such governance structures are open to abuse. This is the *post-Enron lesson* for stakeholder theorists. One cannot justify a system of stakeholder governance on the naïve assumption that managers will always be motivated to act in stakeholders' interests rather than their own. The other, *post-nationalization lesson*, emphasizes the agency risks produced by governance structures that give agents multiple objectives and the discretion to decide the appropriate trade-offs between them. Such scenarios are dangerous with both earnestly committed CSR managers and less-than-earnest managers willing to use this discretion and the favorable information asymmetry to advance their own interests. The basic structure of such a governance regime is a multitask agency problem, and any supporters of a robust CSR program would do well to study the history of largely unsuccessful attempts by democratic governments of all stripes to make multi-stakeholder-friendly SOEs viable.

The analysis in this paper has tried to weave together issues and theories from three fields—CSR and stakeholder theory; governance and agency theory;

and public management of SOEs—that, by and large, have been debated by different groups of theorists in mutual isolation. The central conclusion is simply that theorists interested in the flourishing of stakeholder-friendly, socially responsible firms would do well to explore the challenges raised in these other two fields. Of course, given the vast scope of all three of these fields, it has not been possible to present any knock-down arguments about the limits of a stakeholder theory of governance. This paper has not explored all of the permutations of CSR management, all of the possible reforms of corporate law that would favor stakeholders, nor all of the case studies of SOEs to find governance structures that facilitated the efficient pursuit of multiple objectives. Our hope is simply to have presented a case for why stakeholder theory should benefit from a much more thorough exploration of these issues.

The more specific conclusion is that there is a need for a fundamental reconsideration within CSR and SHT circles of the demotion of shareholders to the status of "just another stakeholder group"—at least when it comes to thinking about corporate governance structures and corporate law reform. This should not in any way be taken as a repudiation of CSR, or of business ethics, or of integrity-based management more generally. It is not difficult to make a business case for CSR, and there are many inspiring examples of corporations that have been financially successful over many years, and even generations, as a result of a thriving, values-based culture. Stakeholder theorists have as much or more to learn from the successes of these firms as they do from the failures of unsuccessful SOEs. But in so doing what they must try to understand is not merely the business case for CSR, but the CSR case for business.[16] This paper is meant as a modest contribution to the latter: a small part of the broader case for the claim that stakeholder theorists should take a second look at the governance advantages of a shareholder-focused, profit-maximizing corporation—or at least, at the pitfalls of certain naïve departures from this model.

[16] There is an extensive literature debating the business case for ethics and CSR. For a brief survey, see Gibson 2000 and Paine 2003. For a lengthy critique of the business case, see Henderson 2001. For a concise summary of the convergence of CSR and shareholder-value approaches to business strategy, see Grant (1998: ch. 2).

3 }

Business Ethics without Stakeholders

Over the past two decades, the "stakeholder paradigm" has served as the basis for one of the most powerful currents of thinking in the field of business ethics. Of course, stakeholder vocabulary is used even more widely, in areas where it is not necessarily intended to have any moral implications (e.g., in strategic management).[1] In business ethics, however, the stakeholder approach is associated with a very characteristic style of normative analysis, namely, one that interprets ethical conduct in a business context in terms of a set of moral obligations toward stakeholder groups (or one that helps "to broaden management's vision of its roles and responsibilities to include interests and claims of non-stockholding groups") (Mitchell et al. 1997: 855). Seen in this light, the primary moral dilemmas that arise in a business context involve reconciling these obligations in cases where stakeholder interests conflict. Thus, ethicists who are impressed by the stakeholder paradigm have become highly adept at translating any moral problem that arises in the workplace into the language of conflicting stakeholder claims (see, e.g., DesJardins and McCall 2005).[2]

The question that I would like to pose in this paper is whether the stakeholder paradigm represents the most fruitful approach to the study of business ethics. The vocabulary of stakeholder obligations has become so ubiquitous that in many contexts it is simply taken for granted. Yet the stakeholder approach is one that comes freighted with very substantive—and controversial—normative assumptions. Naturally, there are many who have criticized the stakeholder paradigm as part of a broader skeptical critique of business ethics in general, one which denies that firms have any "social responsibilities" beyond the maximization of profit.[3] This is not my intention here. I will

[1] For a discussion of the scope and impact of stakeholder theory, see Donaldson and Lee 1995. For an overview, see Harrison and Freeman 1999.

[2] Solomon and Martin (2004: 310) go so far as to introduce the environment as "the silent stakeholder."

[3] For a recent, high-profile example, see "A Survey of Corporate Social Responsibility" (*The Economist*, January 20, 2005).

argue that firms do have important social responsibilities, ones that extend far beyond mere conformity to the law. The question is whether the stakeholder paradigm represents the best framework for articulating the logic and structure of these obligations.

In order to serve as a point of contrast, I would like to provide an outline of two other possible approaches to the study of business ethics: one, a more minimal conception, anchored in the notion of fiduciary obligations toward shareholders, and the other, a broader conception, focused on the regulatory environment in which firms operate.[4] I will then attempt to show that the latter, which I refer to as a "market failures" approach, offers a more satisfactory framework for articulating the concerns that underlie traditional appeals for increased corporate social responsibility.

3.1. Business Ethics as Professional Ethics

There is one point that all three of the approaches that I will be presenting here have in common. All three conceive of business ethics as a species of professional ethics.[5] In the same way that medical ethics concerns, first and foremost, ethical questions that arise from the professional role of doctors, and legal ethics deals with questions that arise from the professional practice of lawyers, business ethics deals with questions that arise out of the professional role of managers. This is a narrower sense of the term "business ethics" than one sometimes encounters, but as we shall see, there are some advantages to be had from focusing on this somewhat constrained set of issues.

In each case, the assumption is that a professional role itself imposes its own set of obligations upon the person, which are not necessarily part of general morality (although they may be sanctioned by, or derived from, general morality). For example, both doctors and lawyers have a special obligation to protect client confidentiality, an obligation that arises out of their professional role. In other words, this obligation is one that is imposed upon each of them, not *qua* individual, but *qua* doctor, or *qua* lawyer. According to this conception, business ethics is concerned with the special obligations that arise out of the managerial role, and which are imposed upon the manager *qua* manager.

---

[4] The first two correspond well to the typology introduced by John Hasnas (1998: 19–42). On the third, my "market failures" model differs from the "social contract model," in that it provides more explicit recognition of the adversarial structure of market transactions.

[5] Thus, for example, Milton Friedman, the most influential proponent of the shareholder-focused view, criticizes the loose talk about "business" having social responsibility and argues that these responsibilities, should there be any, must fall upon the shoulders of managers (Friedman 1970). Similarly, R. Edward Freeman, in his classic work on stakeholder theory (Freeman 1984), identifies it quite explicitly as a set of obligations that fall upon managers, as part of their professional role.

The reason that it is helpful to conceive of business ethics as a set of moral obligations arising out of the professional role of the manager is that it serves to head off the commonly expressed accusation that business ethics is just blue sky dreaming, or a wish list of things that ethicists would like corporations to do, many of which will turn out to be unrealistic in practice. According to the "professional ethics" view, business ethics represents an attempt to articulate a code of conduct that is already implicit both in the structure of corporate law and in the best practices of working managers. This helps to allay the suspicion that business ethics is some alien code, which ethicists seek to impose upon corporations from the outside.

Not everyone accepts the "professional ethics" view. There is an influential strain of thinking in business ethics that treats moral obligations as perfectly invariant across persons. (This tendency is perhaps summed up best in the title of John C. Maxwell's recent book, *There is No Such Thing as "Business" Ethics: There's Only One Rule for Making Decisions* [2003].[6]) Thus some theorists begin by specifying an undifferentiated moral code (whether it be Kantian, utilitarian, Christian, Aristotelian, or what have you); they then treat business ethics as a subject concerned primarily with reconciling pressures that arise in a business context with the obligations that are imposed by this general morality (e.g., the Bible says "thou shalt not bear false witness," so what do you do when the boss asks you to lie to a client?).[7] From this perspective, the managerial role shows up, not as a source of positive moral obligations, but primarily as a source of social pressures that may conflict with morality.

Absent from this perspective is any clear conception of the role that the professions play in a modern economic system (or of the way that a professional "ethos" can give rise to a system of distinctive moral constraints[8]). The primary difference between having a job and practicing a profession involves the element of trust and fiduciary responsibility associated with the latter. In some situations, it is possible for parties in an employment relation to specify all the terms of the contract, to monitor performance completely, and to institute a system of incentives that guarantees perfect compliance. Stacking boxes in a warehouse is an example of an employment relation of this type. These are jobs, and in them, employees are not usually thought to have any special responsibilities beyond those specified in the contract (i.e., the terms of employment). Employees in these sorts of jobs are normally paid by the hour, and have a fixed workday, in recognition of the market-like structure of the transaction.

---

[6] One can find a considerably more sophisticated, but essentially similar, version of this idea in Norman Bowie (1995).

[7] This is the framework that is implicitly assumed by Andrew Stark, in his widely discussed paper (Stark 2003).

[8] The *locus classicus* is Émile Durkheim (1958).

Things become more complicated, however, when it is impossible to specify the terms of an employment contract completely, imperfect observability of effort makes monitoring difficult, or information asymmetries make the design of a perfect system of performance incentives impossible. In such cases it is impossible to eliminate moral hazard, and so the purchaser of labor services must rely in large measure upon the voluntary cooperation of the seller in order to secure adequate work effort.[9] Thus a certain amount of trust, or moral constraint, is required in these relationships. Contracts usually specify goals and obligations in very general terms, and the person supplying the services is expected to use his or her own judgment to decide how best these terms should be satisfied. The purchaser often lacks not only the information and skills to determine the best course on her own, but is often incapable of even verifying that the supplier has done so after the fact. This is the condition that Oliver Williamson refers to as "information impactedness," and it represents the primary force driving professionalization (Williamson 1973: 318).

In certain cases, reputation effects are enough to motivate good faith work effort for individuals in these roles. For example, most people have no ability to evaluate the claims and recommendations made by their auto mechanic, and the cost of getting a second opinion can be prohibitive (in both time and money). Thus they have no choice but to trust the mechanic. But as a result, reputation and "word-of-mouth" plays an important role in the market for automobile repairs. The market for contractors, plumbers, and hair stylists has a similar structure. These groups are not generally thought of as professionals, because the market still does a tolerable job of overcoming the important information asymmetries.

It is not an accident that these cases all involve purchases that consumers make frequently, where there is significant opportunity for repeat business. In markets where larger, more infrequent purchases are made, or where information asymmetries are even greater, it is much more difficult for purchasers of services to impose discipline upon suppliers through reputation mechanisms. As a result, suppliers who deploy highly specialized knowledge must work harder to secure the trust of potential clients, simply because the client may never have the opportunity to verify the quality or value of the services received. In some cases, the trust requirements are sufficiently high that these suppliers will form their own membership association, in order to impose an internal "code of conduct" more stringent than the requirements of general labor and contract law. The most well-known examples are the "bar" for lawyers, along with the various medical licensing boards for doctors. These sorts of associations are especially important in professions where the only people

[9] For an overview of moral hazard in this context, see Milgrom and Roberts 1992: 167–192.

competent to evaluate a particular individual's performance are other members of that same profession.

Economists sometimes suggest that the function of these organizations is merely to cartelize a particular segment of the labor market. This is a good example of the "naïve cynicism" often exhibited in this field—where the automatic identification of pecuniary incentives as the dominant motive leads to sociologically naïve analyses of particular institutions. These associations also play an important socializing role, helping to instill genuine respect for a set of moral obligations that are often specific to the profession.[10] For example, many engineers in Canada wear an iron ring on their little finger, which is conferred during a ceremony called "The Ritual of the Calling of an Engineer" (developed in 1925 by Rudyard Kipling). The ring is a symbol of the Pont de Québec Bridge, which collapsed in 1907 as it was nearing completion, killing seventy-six people. A subsequent Royal Commission declared that errors committed by the bridge's principal engineers were the primary cause of the tragedy. Initially, the rings were said to have been made with iron from the collapsed bridge. In the present day, the rings are intended simply to serve as a reminder to working engineers that the lives of many people depend upon their efforts. Engineers have more than just an obligation to put in a day's work for a day's pay, they must also consider the impact that their actions will have upon the eventual users of the structures or products they design. Many engineering students describe the ceremony as genuinely moving and find that the ring serves as a constant reminder of their professional ethical obligations.

The existence of a professional association, a certification system, a common body of accepted knowledge, and a shared ethics code, are sometimes treated as the distinguishing marks of a genuine profession (see Khurana, Nohria, and Penrice 2005). This involves some confusion of cause and effect. What makes the complex body of knowledge important is that it generates an information asymmetry, which creates a moral hazard problem that threatens to undermine any market transaction involving such specialists. Thus specialists must work hard to cultivate trust among potential purchasers of their services. A certification system, along with a professional association that imposes a stringent code of conduct, is one way of achieving this objective. There may be cases, however, in which a certification system is difficult to devise, or a professional association difficult to organize. Such is the case, traditionally, with managers (especially during the era when most were promoted up from the shop floor). Nevertheless, the economic role that managers occupy is a professional one, precisely because of the information impactedness in the domain of services

[10] This is why, as R. M. MacIver emphasizes, "Each profession tends to leave its distinctive stamp upon a man, so that it is easier in general to distinguish, say the doctor and the priest, the teacher and the judge, the writer and the man of science than it is to discern, outside their work, the electrician from the railwayman or the plumber from the machinist" (MacIver 1922: 11).

they provide. The nature of the managerial role is such that they need to be both trusted and trustworthy. This is reflected in the fact that most systems of corporate law treat senior managers as fiduciaries of the firm (Clark 1985). Thus the mere fact that managers do not belong to professional associations does not mean that they are not professionals, or more importantly, that there is not a distinctive set of ethical obligations that arise out of their occupational role. The fact that they are in a position of trust is what matters.[11]

Thinking of business ethics in terms of "professional ethics for managers" is an attractive perspective, insofar as it offers some relatively clear criteria for the evaluation of different "theories" or "paradigms" within the field. Managers who take social responsibility seriously already have some very firm intuitions about what constitutes ethical and unethical conduct. The question is whether the vocabulary and the principles that business ethicists develop offer a more or less perspicuous and coherent articulation of these intuitions—whether their theories help us to achieve greater clarity, or whether they sow confusion. This is the standard that I shall be employing in this paper. Thus my criticism of the stakeholder approach to business ethics is not that it is false or incoherent. I shall merely try to show that the vocabulary, and the theory that underlies it, is inherently misleading, and thus does not promote useful ways of thinking about corporate social responsibility.[12]

3.2. The Shareholder Model

The managerial role arises as a consequence of the so-called separation of ownership and control in the modern corporation. In the early stages of development, most corporations are run by the founders, who are also generally the principal owners. At a later point, the owners may choose to employ managers to assist them in running the firm, or to take over that role entirely. In the same way that individuals employ lawyers in order to advance their interests in a legal context, owners hire managers in order to advance their interests in a business context. Of course, as the firm becomes more mature, this relationship becomes significantly more complex (leading many to argue that the shareholders in a publicly traded corporation cannot be regarded as

[11] It is worth noting that there have been some moves afoot among business schools to start offering students some of the trappings of a professional association. One school in Canada, for instance, has begun offering a ring ceremony modeled on that of engineers, where students "make a public oath to behave honorably and, in return, receive an inscribed silver ring to wear as a reminder" (Gadd 2005). It seems to me that the question of whether we want to describe management as a profession should not depend upon the success or failure of such efforts.

[12] There are parallels between this aspect of my argument and that of Norman and MacDonald, who argue that so-called 3BL accounting is also "inherently misleading." See Norman and MacDonald 2004: 254.

its "owners" in any coherent sense). Nevertheless, the fact that shareholders are residual claimants in a standard business corporation means that their interests are not protected by an explicit contract. As a result, there is a set of fiduciary principles governing the relationship between managers and shareholders.[13] Because the fiduciary relationship imposes upon managers a very broad "duty of loyalty" and "duty of care" toward shareholders—concepts with explicit moral overtones—this particular relationship might be thought to serve as a natural point of departure for the development of a theory of business ethics (in the same way that duties toward the patient form the core of professional ethics for doctors, duties toward the client the core of professional ethics for lawyers, etc.)

Yet despite the fact that moral obligations toward shareholders are such a striking feature of the managerial role, in the business ethics literature they are the subject of considerable controversy and are often downplayed or dismissed. (Marjorie Kelly, the editor of *Business Ethics* magazine, set the tone for one end of this discussion with the title of her article, "Why All the Fuss about Stockholders?") (Kelly 2001). There are several reasons for this relative neglect of the shareholder, some worse than others. In popular debates, there is a tendency when talking about "the corporation" simply to conflate to the two groups (managers and owners), or to assume that there is a greater identity of interests between them than is usually the case. The standard microeconomics curriculum encourages this, by starting out with the assumption that individuals maximize utility, but then aggregating consumers together into "households" and suppliers into "firms"—each of which is thought to maximize some joint utility function—without explaining the transition (this gets reserved for more advanced courses). Even though it is understood that "the firm" is something of a black box in this analysis, the result is still an unhelpful blurring of the distinction between the pursuit of self-interest on the part of individuals and the maximization of profit on the part of firms, and thus a tendency to overestimate the extent to which the latter flows naturally from the former. As a result, it is easy to underestimate the potential for moral hazard in the relationship between managers and shareholders.

The recent scandals at Enron, Parmalat, Tyco, WorldCom, Hollinger, and elsewhere, have shown that shareholders neglect these difficulties at their own peril. In each of the major scandals, managers were able to enrich themselves primarily at the expense of shareholders. (It may be helpful to recall that at its peak, Enron had 19,000 employees and a market capitalization of $77 billion. Thus for each employee who had to look for a new job as a result of the

---

[13] This should be interpreted as a positive (i.e., factual) claim about the structure of corporate law. See Easterbrook and Fischel 1991: 90–91. Whether managers should be fiduciaries of shareholders, or just shareholders, is of course the subject of considerable controversy among business ethicists. For a defense of the claim that they should be, see Marcoux 2003: 1–25.

subsequent bankruptcy of the firm, shareholders lost at least $4 million.) The fact that most of these scandals involved illegal conduct should not distract us from the fact that each illegal act was surrounded by a very broad penumbral region of unethical conduct. For example, it was never decided specifically whether the $2.1 million dollar party thrown by Tyco CEO Dennis Kozlowski for his wife's birthday, half paid out of company funds, constituted fraud or theft, but it most certainly represented a violation of his moral obligation to shareholders.

It is a mistake to believe that self-interest alone, combined with a few performance incentives, is able to achieve a harmony of interest between managers and shareholders. In this respect, a lot of the work done by economists (and game theorists) on the "theory of the firm" has been quite misleading. The overriding objective of many economists has been to extend the methodological tools—and in particular, the action theory—used in the analysis of markets to model the internal structure of organizations.[14] Thus "principal–agent" theory has focused almost entirely upon the use of external incentives as a mechanism for overcoming collective action and control problems within the firm. In so doing, economists have dramatically underplayed the role that trust, values, social norms, and other aspects of "corporate culture" play in determining organizational behavior.[15] Thus they have wasted considerable time and energy devising increasingly baroque performance pay schemes, while neglecting more obvious managerial strategies, such as encouraging employee loyalty to the firm or cultivating a direct concern for customer satisfaction.[16]

It is precisely because of the importance of these internal (i.e., moral) incentives, along with the enormous potential for abuse, that US corporate law essentially imposes a fiduciary relationship between senior managers and shareholders. It is helpful to recall, for example, the words of an influential US court judgment, concerning the obligations of managers:

> He who is in such a fiduciary position cannot serve himself and his cestuis second. He cannot manipulate the affairs of his corporation to their

[14] The paper that really set economists off in the wrong direction was Armen A. Alcian and Harold Demsetz's "Production, Information Costs, and Economic Organization" (1972: 777–795), with their suggestion that the firm is really just a "privately owned market" (795). It should be noted, however, that subsequent work by incentive theorists has been considerably less sanguine about the efficiency properties of such "markets."

[15] For a critique of these and other "framing assumptions" in agency theory, see Dees 1995.

[16] For example, the chapter in Milgrom and Roberts (1992) on moral hazard has a section entitled "Controlling Moral Hazard" (185–192), which discusses, among other things, employee monitoring, supervision, incentive contracts, performance pay, bonding, and ownership changes as managerial strategies for preventing shirking. At no point is it mentioned that employees may respond to changes in "internal" motives (such as whether they love or hate the company they work for). It also exhibits a lack of concern for the fact that external performance incentives, such as pecuniary compensation, have the potential to "crowd out" moral incentives, and thus in some cases generate collective action problems rather than resolve them. See Frey et al. 1996: 1297–1313.

> detriment and in disregard of the standards of common decency and honesty. He cannot by the intervention of a corporate entity violate the ancient precept against serving two masters. He cannot by the use of the corporate device avail himself of privileges normally permitted outsiders in a race of creditors. He cannot utilize his inside information and his strategic position for his own preferment. He cannot violate rules of fair play by doing indirectly through the corporation what he could not do directly. He cannot use his power for his personal advantage and to the detriment of the stockholders and creditors, no matter how absolute in terms that power may be and no matter how meticulous he is to satisfy technical requirements, for that power is at all times subject to the equitable limitation that it may not be exercised for the aggrandizement, preference, or advantage of the fiduciary to the exclusion or detriment of the cestuis. Where there is a violation of those principles, equity will undo the wrong or intervene to prevent its consummation.[17]

The obligations enumerated here are sufficiently broad that one could only imagine legal prosecution in cases of the most egregious violation. Thus a very robust theory of business ethics could be developed based simply on the injunction to respect the spirit of this judgment, along with the fiduciary obligations that it outlines toward shareholders. Yet despite this fact, far too little has been said on this subject. The dominant assumption has been that shareholders are able to take care of themselves. Many introductory business ethics textbooks cover topics like whistle blowing, truth in advertising, pollution, discrimination, and health and safety issues, yet neglect to discuss more common ethical challenges that employees encounter in their day-to-day affairs, such as the temptation to abuse expense accounts.[18] Strictly speaking, society should be no more willing to tolerate such abuses when carried out by business executives (wasting shareholders' money) than when carried out by politicians or civil servants (wasting taxpayers' money). The reality, needless to say, is quite different. Thus a simple duty of loyalty toward shareholders precludes a lot of the everyday immorality that goes on in firms (but which attracts attention only when it reaches spectacular proportions, as with the recent spate of corporate scandals).

[17] *Pepper v. Litton* 308 U.S. 295 (1939) at 311, cited in Clark 1990: 76. As Clark observes, the use of moral rhetoric in cases involving breach of managerial duty is highly significant, because as a general rule "our society is reluctant to allow or encourage organs of the state to try to instill moral feelings about commercial relationships in its citizens" (75).

[18] Although admittedly an unscientific survey, I have in my office fifteen different introductory business ethics textbooks, many of which discuss insider trading, but only one of which (Richardson 2004) makes any mention of the issue of employee expense account abuse or employee theft. Even then, the discussion focuses upon falsification of expenses and does not mention the issue of mere profligacy.

Thus the tendency to overestimate the degree of alignment of managerial and shareholder interests leads to more general failure to appreciate the extent to which shareholders are vulnerable in their relations with managers (just as patients are vulnerable in their relations with doctors or clients are vulnerable in their dealings with lawyers). There is, however, also a more principled reason that obligations toward shareholders tend to get downplayed. There is a widespread perception that the fiduciary relationship between the manager and the shareholder cannot serve as a source of genuine moral obligation. Even though I am morally obliged to keep my promises, if I promise my friend that I will rob a bank that does not mean that I am then morally obliged to rob a bank.[19] The same applies to fiduciary relations. Consider the following argument, due to Arthur Applbaum (2000). Imagine a Hobbesian state of nature, in which everyone treats everyone else abysmally. Such conduct is immoral. Now imagine that, in this state of nature, each person solemnly swears to stop pursing his own interests and to begin pursuing the interests of the person next to him. What changes? From the moral point of view, nothing much. It is still the war of all against all, except that now it is being carried out by proxy. Certainly the mere fact that each person is acting "altruistically"—advancing the interests of her neighbor, rather than her own—is not enough to transform this into a morally acceptable state of affairs. If it could, then the simple act of promising would permit unlimited "laundering" of immoral acts into moral ones.

Thus the discussion of the fiduciary responsibilities of managers quickly turns into a discussion of the moral legitimacy of the goals being pursued by shareholders. This in turn must lead to a discussion of the moral status of *profit* (since this is the interest of shareholders that managers are generally understood to be advancing). It is here that the "ethical" status of business ethics begins to seem problematic. Indeed, Milton Friedman's well-known article "The Social Responsibility of Business Is to Increase its Profits," which presents the ethical obligation to maximize the returns of shareholders as the cornerstone of a conception of business ethics, usually shows up in business ethics textbooks, not as the point of departure for further development of the theory, but rather as an example of an instructively mistaken point of view (see Beauchamp and Bowie 2001; Poff 2005; White 1993). The problem is that "profit" is associated, in many people's minds, with "self-interest."[20] "Ethics," on the other hand, is usually associated with

[19] See Alex C. Michalos's critique of "the loyal agent's argument," in Michalos 1995: 50–52. Also DeGeorge 1992: 65–66.

[20] Khurana, Nohria, and Penrice (2005), for example, argue that a bona fide profession requires of its members "a renunciation of the profit motive." They then blame "the doctrine of shareholder primacy" for recent corporate ethics scandals, on the grounds that it "has legitimized the idea that the benefits of managerial expertise may be offered for purely private gain." This "led directly to many of the worst profit-maximizing abuses unmasked in the recent wave of corporate scandals." Such an analysis is almost exactly backwards. The problems at Enron (for example) were not due to managers maximizing profits; they were due to managers failing to maximize profits, then creating special-purpose

behavior that is "altruistic," in some sense of the term. More precisely, morality can be understood as a "principled constraint on the pursuit of self-interest" (Gauthier 1986). If this is the case, then substituting "profit" for "self-interest" yields the conclusion that business ethics must represent some sort of principled constraint on the pursuit of profit—not an injunction to maximize it.[21]

In the case of doctors, who must do everything in their power to promote the health of their patients, it is easy to see that health is a good thing, and so efforts to promote it in others must also be good. This is more difficult to see in the case of managers and wealth, especially in cases when increasing the wealth of shareholders can only be achieved at the expense of others. Yet managers who take their responsibilities toward shareholders seriously are often put in a situation where they must effect pure distributive transfers—often regressive ones between workers and shareholders. Here it becomes difficult to see what is so ethical about business ethics.

Thus in order to see managerial obligations toward shareholders as genuine moral obligations, one cannot merely point to their fiduciary status, one must also come up with some justification for the role that profit-taking plays in a capitalist economy. There are two general strategies for doing so. The first, which might be thought of as broadly Lockean, defends profits as the product of a legitimate exercise of the shareholder's property rights, under conditions of freedom of contract. According to this view, the shareholder is entitled to these profits for the same reason that the creditor is entitled to repayment with interest, or that the worker is entitled to her wages. This is not very compelling, however, because the Lockean theory is one that defines the individual's legal rights, but makes no pretence of accounting for her moral obligations. Thus, for example, the Lockean thinks that we have no legal obligation to give anything to charity, and our property rights protect us from any seizure of our assets for such purposes. But this does not mean that we have no moral obligation to give to charity. Ordinary morality tells us that wealth is not an overriding value, and so there would appear to be many cases where the profit motive is trumped by other considerations. This makes it unethical for shareholders to pursue profits in particular ways, and thus unethical for managers to assist them in carrying out such strategies.

---

entities to keep more than $26 billion worth of debt off the balance sheet, precisely to generate the illusion of profitability. The fact that they were able to line their own pockets in the process demonstrates the extent to which the goal of maximizing one's own personal earnings and maximizing the profits of a firm can diverge. Professional conduct requires setting aside the goal of maximizing one's own earnings, but that does not preclude one from earning money for others. Divorce lawyers seek to secure the largest settlement for their clients, without that compromising their status as professionals.

[21] For an especially clear example of confusion on this score, see Duska 2007: 157–159. He talks about the "self-interested pursuit of profit" and argues that in order to diminish the level of self-interested behavior on the part of individuals within a firm it will be necessary to challenge the orientation toward profit-making on the part of the business as a whole.

The more promising defense of profit is the Paretian one, which points to the efficiency properties of the market economy as a way of justifying the profit orientation of firms. According to this view, the point of the market economy is not to respect individual property rights, but rather to ensure the smooth operation of the price system. The profit orientation is valued, not because individuals have a right to pursue certain interests, but rather because it generates the competition necessary to push prices toward the levels at which markets clear.[22] When markets clear, it means that all resources will have been put to their best use, by flowing to the individuals who derive the most relative satisfaction from their consumption. The spirit of the Paretian approach is best expressed in the "invisible hand" theorem of welfare economics, which shows that the equilibrium of a perfectly competitive market will be Pareto-optimal (i.e., it will be impossible to improve anyone's conditions without worsening someone else's) (see Barr 1998: 70–85).

Yet this framework still seems to be, in many ways, not "ethical enough" to satisfy many people's intuitions (Goodpaster 1991: 60). It offers a seal of approval, for instance, to a wide range of so-called sharp practices in market transactions (which, despite being legal, nevertheless offend our intuitive moral sensibilities). And while it has been pointed out many times that firms seldom profit in the long run from abusing employees, cheating customers, or taking advantage of suppliers, it nevertheless remains true that in certain cases it can be profitable to do so. In other words, it is simply not the case that the interests of shareholders always line up with those of workers, customers, suppliers, and other groups with an interest in the firm's decisions. There are genuine conflicts that arise, and it is not obvious that the ethical course of action for managers in every instance is to take the side of shareholders, respecting no constraints beyond those imposed by law. But if this is so, the question becomes how far one should go, as a manager, in advancing the interests of the principal, and when one should start showing more concern for others who are affected by one's actions. Yet even to pose the question in this way is to reveal the limitations of any theoretical approach to business ethics that takes obligations to shareholders as the sole criterion of ethical conduct in business.

3.3. The Stakeholder Model

The shareholder approach to business ethics suffers, first and foremost, from the taint of moral laxity. It does not seem to impose enough obligations upon

[22] John Kay writes, "It is not true that profit is the purpose of the market economy, and the production of goods and services the means to it: the purpose is the production of goods and services, profit the means" (2003: 351).

managers to satisfy the moral intuitions of many people. In particular, it suggests that, as R. Edward Freeman puts it, "management can pursue market transactions with suppliers and customers in an unconstrained manner" (Freeman 1998: 126). Thus the suggestion has been made that managers have moral obligations, not just to shareholders, but to other groups as well. Freeman introduced the term "stakeholders" as a "generalization of the notion of stockholders," in order to refer to "groups and individuals who benefit from or are harmed by, and whose rights are violated or respected by, corporate actions" (Freeman 1998: 129). He went on to make the suggestion that managers have fiduciary obligations toward multiple stakeholder groups.

This overall approach has proven to be remarkably influential, and it is not difficult to see why. After all, we understand quite clearly what it means for managers to have fiduciary obligations toward shareholders. By construing relations with "stakeholders" on analogy, Freeman provided an intuitively accessible framework for articulating the sorts of moral obligations that the shareholder model elides. (In the same way, the term "social capital" has become popular, precisely because people understand what capital is, and so construing social capital on analogy with real capital provides an intuitively accessible framework for thinking about collective action.)

Of course, the term "stakeholder" has been picked up and used quite widely, even by those who do not share Freeman's views on the structure of managerial obligations. For example, so-called strategic stakeholder theory argues that managers must exercise moral restraint in stakeholder relations *as a way of discharging their fiduciary obligations toward shareholders* (i.e., "ethics pays"). Freeman, on the other hand, claims that managers must exercise moral restraint in dealings with stakeholders *because managers have direct fiduciary obligations toward those stakeholders*. Shareholders, according to this view, are just one stakeholder group among many. Managers have fiduciary obligations toward shareholders only because shareholders are stakeholders, and managers have fiduciary obligations toward all stakeholders (Freeman 1998: 132).[23]

Thus Kenneth Goodpaster identifies the key characteristic of Freeman's theory when he refers to it as the "multi-fiduciary stakeholder" theory (Goodpaster 1991). What matters is the idea that managers have fiduciary obligations toward multiple groups—regardless of whether these groups are called stakeholders or something else. Thus the two components of the theory are separable—one need not conceive of stakeholder relations as fiduciary relations. Nevertheless, stakeholder vocabulary is often used as a way of expressing tacit commitment to the multi-fiduciary view. As a result, some of the obvious weaknesses of the position tend to be overlooked. As Goodpaster observes, the

[23] For an example of this view, further developed, see the list of "principles of an ethical firm" in Norman Bowie 1995: 90.

fact that managers have moral obligations with respect to customers, employees, and other groups, does not mean that these obligations must take a fiduciary form. There is some danger of being seduced by the metaphor, leading one to think that the status of stakeholders is much closer to that of shareholders than it in fact is. For example, the manager might have an obligation to respect certain rights of customers, without also having a fiduciary duty to advance their interests.

If managers really are to be regarded as fiduciaries of stakeholder groups, it raises immediate difficulties with respect to questions of corporate governance. Freeman suggests that the manager must become like "King Solomon," adjudicating the rival claims of various stakeholder groups. Yet giving managers the legal freedom to balance these claims as they see fit would create extraordinary agency risks. On the one hand, managers would need to be protected from being fired by shareholders upset over the performance of their investments.[24] But even more significantly, it would become almost impossible for members of any stakeholder group to evaluate the performance of management. It is difficult enough for shareholders to determine whether managers are actually maximizing profits, given available resources. But when profits can be traded off against myriad other objectives, such as maintaining employment, sustaining supplier relationships, and protecting the environment, while managers have the discretion to balance these objectives as they see fit, then there is really no alternative but to trust the word of managers when they say that they are doing the best they can. The history of state-owned enterprises shows that the "multiple objectives" problem can completely undermine managerial discipline, and lead to firms behaving in a less socially responsible manner than those that are explicitly committed to maximizing shareholder value (Heath and Norman 2004).

Setting aside these practical difficulties, the plausibility of multi-fiduciary stakeholder theory also depends quite heavily upon how broadly the term "stakeholder" is understood. This so-called identification problem has attracted considerable attention (see Mitchell, Agle, and Wood 1997: 856–858.) Freeman distinguishes between a "narrow definition" of the term, which refers to groups that are "vital to the success and survival of the firm," and a "wide definition," which refers to any group "who can affect or is affected by the achievement of the organization's objectives."[25] The former includes employees, customers, suppliers, but also, in most formulations of the theory, the local community. The wide definition, on the other hand, is so wide that it becomes equivalent to "all of society." (For example, every pricing decision made by the firm contributes to the national inflation rate, which in turn affects every

[24] Some US states have been moving in this direction, see discussion of "other constituency statutes" below.

[25] The narrow definition is from Freeman 1998: 129; the wide is from Freeman 1984: 46.

member of society. So if a stakeholder is anyone affected by the corporation, then everyone is a stakeholder in everything.) Yet the idea that managers are fiduciaries for "all of society" simply collapses business ethics into general ethics (i.e., general utilitarianism, Kantianism, Christian ethics, or what have you). Thus theorists who believe that the managerial role imposes special obligations upon the individual have tended to stick to the narrower definition of the stakeholder.

From the moral point of view, however, there seems to be no reason for the firm to pay special attention to stakeholders in the narrow sense of the term. There are plenty of good strategic reasons for managers to worry most about those whose contribution is vital to the success of the firm, but it is difficult to see what moral ones there could be. The groups that are conventionally classified as stakeholders in the narrow sense are not necessarily those with the most at stake in a particular decision, in terms of their potential welfare losses. In fact, if one looks at the standard list of stakeholder groups (customers, suppliers, employees, and the local community), it tends rather to be those who are the best organized, or who have the most immediate relationship to the firm, or who are best positioned to make their voices heard. Thus stakeholder theory often has a "squeaky wheel" bias.[26] For example, when General Motors considers closing down a plant in Detroit and moving it to Mexico, a standard multi-fiduciary stakeholder theory would insist that managers take into account the impact of their decision, not just upon their workers in Detroit but also upon other members of the community whose livelihood depends upon their wages. Thus the "local community" in Detroit where the plant is located would normally be counted as a "stakeholder." But what about the "local community" in Mexico, where the plant would be located? And what about the people there who would be getting jobs? (See Langtry 1994: 431–443.) Presumably they also have a lot at stake (possibly even more, in terms of welfare, given the relative poverty of the society in which they live). The fact that General Motors has built up a relationship over time with the people in Detroit may well count for something, but it cannot justify ignoring the interests of the people in Mexico. From the moral point of view, a potential relationship can be just as important as an actual one (Mitchell et al. 1997). The only real difference between the groups is that potential employees do not know who they are, and so are unable to organize themselves to articulate their interests or

[26] Mitchell, Agle, and Wood (1997) propose a very nuanced analysis of stakeholder groups, classifying them in a way that reflects their relative "salience" to managers. They go on to observe that "if the stakeholder is particularly clever, for example, at coalition-building, political action, or social construction of reality, that stakeholder can move into the 'definitive stakeholder' category (characterized by high salience to managers)" (879). This sort of observation shows how stakeholder analysis may be useful for strategic management, but when employed without further ado as the normative foundation of business ethics tends to favor the squeaky wheel.

express grievances. But it is difficult to see why—from a moral, rather than a strategic point of view—this should give managers the freedom to leave potential employees, or potential "local communities," off the list of groups that the firm has an obligation to.

Because stakeholder theory focuses on the relationship between the manager and different "groups" within society, it tends to privilege the interests of those who are well-organized over those who are poorly organized, simply because it is the former who are able to present themselves as a coherent body with a common set of interests. To see this bias in action, one need only look at the difference in the way that various stakeholder theorists conceive of "social responsibility" and the way that governments have traditionally approached it (Heath and Norman 2004). In this context, it is useful to recall that the widespread nationalization of industry that occurred in Western Europe after the Second World War was motivated, in large part, by the desire of democratic governments to make corporations behave in a more socially responsible manner. The thought was that corporations behaved irresponsibly because owners put their private interests ahead of the public good. By transferring ownership to the state, the people as a whole would become the owners, and so the corporation would no longer have an incentive to pursue anything other than the public good.

Needless to say, this initiative did not have precisely the results that were anticipated. The interesting point, however, lies in the agenda that various governments initially laid out for these firms. First and foremost, state-owned enterprises were expected to play an important role in assisting the state to implement macroeconomic stabilization policies: attenuating the business cycle by making counter-cyclical investments; maintaining excess employment during recessionary periods; and following self-imposed wage and price controls when necessary, in order to control inflation. Similarly, state-owned enterprises were expected to serve the national interest in various ways, either by providing goods at discounted prices when supplying domestic industry, serving as a guaranteed market for domestically produced goods, or by assisting in the "incubation" of industries intended to bolster international competitiveness. They were, of course, also expected to act as model employers with respect to their workers, to refrain from polluting, to promote regional development, and so forth. While there is significant overlap between the latter set of objectives and the traditional concerns of many stakeholder theorists, there are also some striking differences. In particular, one can search the stakeholder literature long and hard without finding any mention of the way that firms can contribute to macroeconomic stability. The reason, I would suggest, is that there are no organized or clearly identifiable "stakeholder" groups in this case. After all, how does one identify those who are harmed by inflation? It is, by and large, an extremely diffuse group of individuals. As a result, business ethicists working within the stakeholder paradigm have had a tendency simply to ignore them. For example, I am not aware of anyone having suggested that managers should refrain from granting inflationary wage increases to workers (i.e., increases that are not funded by productivity gains). Governments, on the other hand, have traditionally been concerned with these questions, precisely because they do have a mandate to defend the welfare of all citizens and to promote the public interest.

As a result, if one interprets the term "stakeholder" in the narrow sense, it introduces an unacceptable element of arbitrariness into business ethics. If one expands the definition, such that anyone affected by the firm's actions will be considered a stakeholder, multi-fiduciary stakeholder theory amounts to the claim that the manager should be motivated by general considerations of social justice. This risks rendering the stakeholder vocabulary nonsensical, since the concept of a "fiduciary" relation is inherently contrastive. Being a loyal fiduciary involves showing partiality toward the interests of one group, not an impartial concern for the interests of all. Furthermore, if the manager is obliged to show impartial concern, the question then becomes, is he or she the person best equipped, or best positioned, to be making these judgments? As Friedman pointed out long ago, normative issues at this level of generality seem to be a more appropriate topic for public policy and democratic deliberation (Friedman 1970). It is simply not obvious that the manager's obligations should be determined by these concerns.

Part of the unwillingness to accept this line of reasoning stems from a rejection of the idea that there might be an institutional "division of moral labor," such that not everyone is morally responsible for everything at all times. Many of the most subtle and difficult questions in professional ethics involve dealing with the way that obligations are divided up and parceled out to different individuals occupying different institutional roles. This is especially tricky in cases where the institution has an adversarial structure (Applbaum 1999). For example, the role of a defense attorney in a criminal trial is to advance the interests of her client by mounting a vigorous defense. Naturally, the overall goal of the procedure is to see that "justice" is served. But that does not make the defense attorney directly accountable to what she thinks is "just" in any particular case. Her job is to defend her client (and in fact, mounting a less-than-vigorous defense, because she happens to believe that her client is guilty, constitutes a serious violation of professional ethics). The victim of the crime is no doubt a "stakeholder" in these proceedings, but that does not mean that the defense attorney has a fiduciary obligation toward this individual. Both as a human being and as an officer of the court, she no doubt has ethical obligations toward victims of crime. But *qua* defense attorney, her obligation in many cases will be to disregard this everyday moral constraint. Justice arises through the interaction of her role-specific obligations with those of the crown prosecutor (or district attorney) and the judge. Of course, this is not to say that defense attorneys should do anything to secure the acquittal of their clients, or should not respect certain constraints in dealing with victims. There are clearly ethical and unethical ways to proceed. The point is that the vocabulary of fiduciary obligation does not provide a useful way of formulating these constraints. Furthermore, the idea that attorneys should seek to promote justice by balancing the interests of all affected parties is in tension with the role-differentiation that is a central component of the adversarial trial procedure.

Turning to business ethics, the first thing to note is that market transactions also have an adversarial structure (insofar as prices are competitively determined). One can see the problems that this creates for multi-fiduciary stakeholder theory by considering the attempts that have been made to classify "competitors" amongst the relevant stakeholder groups (or more often, the way that "competitors" are tacitly excluded without discussion).[27] After all, competitors are clearly affected by many of the decisions taken by the firm. Furthermore, since competitors have the power to drive the firm into bankruptcy, their behavior is often vital to its success or failure. Yet it seems obvious that managers do not have any fiduciary obligations toward rival corporations. After all, the price mechanism functions only because of an unresolved collective action problem between firms. No company sets out with the intention of selling goods at a price that clears the market. Often no one even knows what that price is. It is only when firms compete with one another, undercutting each other's prices in order to increase their market share, that the selling price will be driven down to market-clearing levels. This is a classic form of non-cooperative behavior, since it is not normally profit-maximizing overall for firms to sell at this price level. They do it only because they are stuck in a collective action problem.

Thus there is a significant difference between market transactions and the administered transactions that occur within the organizational hierarchy of the firm. The former, because they are mediated through the price system, have an intrinsically adversarial element, since prices are supposed to be determined through competition (and considerable legal effort is invested in the task of keeping things that way). Since many of the socially desirable outcomes of the market economy are a consequence of the operation of the price mechanism, it is not clear that individual firms, much less managers, should be held directly accountable to them. Yet the possibility of such differentiated roles is tacitly denied by the wide version of stakeholder theory, which demands that the manager be ethically responsible for balancing the interests of everyone who is affected by the firm's actions, regardless of whether they are in a competitive or a cooperative relationship.

---

[27] For an example of the former, see Freeman 1998: 132; for an example of the latter, see Mitchell, Agle, and Wood 1997.

3.4. The Market Failures Model

Despite these difficulties, the stakeholder paradigm still exercises an extraordinary grip over the imagination of many business ethicists (Buchholz and Rosenthal 2005: 137–148). It is all too often assumed that the stakeholder theory and the shareholder theory exhaust the logical space of alternatives. As a result, theorists like Marjorie Kelly and Max Clarkson have sought to defend stakeholder theory by mounting increasingly spirited attacks on the idea that managers have any particular obligations to shareholders. The cornerstone of this "nothing special about shareholders" defense is the claim that shareholders are not really "owners" of the firm in any meaningful sense (Kelly 1997).[28] Thus Clarkson cites with approval the fact that "serious questions are being raised about the belief, widely held in North America, that the purpose of the corporation in society is to maximize profits and financial value for the primary benefit of its shareholders, who are also assumed, mistakenly, to be the corporation's owners" (1998). (For a clear antidote to these sorts of views, see Hansmann 1992.)

It is perhaps worth noting that this particular strategy for defending the stakeholder paradigm has the unhelpful effect of making business ethics extremely unintuitive for those who actually work in a standard corporate environment, where the understanding that shareholders own the firm is still widespread. In particular, the downgrading of shareholder claims creates an enormous tension with corporate law, which remains very much committed to the idea that shareholders have a special status within the firm, and that managers owe them fiduciary duties (Easterbrook and Fischel 1991: 90–91). Of course, it is always possible for the law to be unethical. Nevertheless, this problem is more serious than it would at first appear. If one could produce a sound argument for the conclusion that managers have fiduciary obligations toward various stakeholder groups, one would also have produced a strong prima facie argument for the legal enforcement of these obligations. Thus stakeholder theorists have invested some effort in attempting to show that corporate law has in fact been evolving in the direction of increased recognition of stakeholder claims (see, e.g., Orts 1992: 134–135; Donaldson and Preston 1995: 75–76). And it is here, I think, that one can see where the most instructive misunderstanding arises.

There can be no doubt that the development of the welfare state in the twentieth century has coincided with increased regulation of the market. Health and safety in the workplace, the minimum wage, unionization procedures, product warranties, "truth in advertising" and product labeling, toxic emission

[28] See also Max Clarkson's introduction to *The Corporation and its Stakeholders* (1998:1–6). Bowie (1995:144–145) offers an approving survey of such strategies.

controls, environmental impact studies, even the size and location of commercial signage—have all become subject to increasingly strict controls. Furthermore, it is clear that all of these regulations respond, in one way or another, to the type of issues that have traditionally been of concern to business ethicists. Each regulation amounts to a legal prohibition of a form of corporate conduct that was at one time merely unethical. The question is how we should understand these developments. Freeman argues that the growth in regulation *constitutes an increased legal recognition of stakeholder claims (*Freeman 1998). This is, I will argue, a serious misunderstanding. The growth of regulation over the course of the twentieth century goes hand-in-hand with the increased positive economic role of the state in supplying public goods. Both represent strategies aimed at *correcting market failure*. As a result, I think that the concept of market failure provides a much more satisfactory framework for understanding the growth of regulation—and thus the increased legal entrenchment of the social responsibilities of business—than that of stakeholder claim recognition.

Setting aside Germany's "co-determination" arrangements, the closest one can find to an explicit recognition of stakeholder claims is the spread of statutes that allow boards of directors to consider the impact that a hostile takeover would have on non-shareholder groups in determining whether resistance to such takeovers would be "reasonable." These so-called other constituency statutes adopted in many US states (although not Delaware), typically permit (and occasionally require) "officers and directors to consider the impact of their decisions on constituencies besides shareholders" (Boatright 1994: 402; see also Hanks 1991: 97–120). Thomas Donaldson and Lee Preston describe this as a "trend toward stakeholder law" (1995: 76). It is significant, however, that these statutes do not impose fiduciary duties and were largely motivated by a desire on the part of legislators to make hostile control transactions more difficult, based upon a perception that takeovers generate significant social costs. Thus "other constituency" statutes have a lot in common with enabling statutes for "poison pill" and "shark repellent" defenses. I would argue that they are therefore better understood as an attempt to curtail a (perceived) market failure in the stock market than as a legal recognition of stakeholder claims.

The politics of "other constituency" statutes is a complex issue, however, which I do not want to get into here. My primary concern is to illustrate the style of analysis suggested by the market failures perspective. A market failure represents a situation in which the competitive market fails to produce a Pareto-efficient outcome (or for our purposes, let us say, fails egregiously to produce an efficient outcome). There are two primary institutional responses to market failure. The first involves the creation of the corporation itself, which is based upon the substitution of an organizational hierarchy and a set of administered transactions for a competitive market. The central characteristic of the firm, as Ronald Coase observed in his classic work, is the internal elimination of market transactions and the "supersession of the price mechanism" (Coase 1937: 389; Williamson 1973: 316). In more contemporary terms, we would say that the corporation substitutes a set of principal–agent relations for the non-cooperative relations of marketplace competition. However, because of the limitations of external incentive schemes, these agency relations can often be organized only through some combination of moral and prudential constraint (Noreen 1988). Thus the central focus of business ethics, in an intrafirm context, involves promoting cooperative behavior within these agency relationships (as Allen Buchanan [1996] has argued, in my view persuasively). First and foremost among these obligations will be the fiduciary duty that managers have as the agents of shareholders. Thus when dealing with relationships or transactions "inside" the organizational hierarchy of the firm, the market failures approach to business ethics follows the shareholder-focused view quite closely. With respect to individuals who are "outside" the firm, on the other hand, it is quite different.

The second primary institutional response to market failure is less drastic than the first; it involves preservation of the market transaction, but subject to some more extensive set of legal, typically regulatory, constraints. To see the rationale for this strategy, it is helpful to recall that the point of permitting profit-maximizing behavior among firms in the first place is to promote price competition, along with all the beneficial "upstream" and "downstream" effects of such competition, such as technical innovation, quality improvement, and so on. Under conditions of "perfect competition," lower price, improved quality, and product innovation would be the only way that firms could compete with one another. We can refer to these as the set of *preferred* competitive strategies. Unfortunately, in the real world, the so-called Pareto conditions that specify the terms of perfect competition are never met. In order for competition to generate an efficient allocation of goods and services, there must be an absence of externalities (e.g., a complete set of property rights), symmetric information between buyers and sellers, a complete set of insurance markets, and rational, utility-maximizing agents with dynamically consistent preferences. Because of the practical impossibility of satisfying these constraints, firms are often able to make a profit using *non-preferred* competitive strategies, such as producing pollution or selling products with hidden quality defects.[29] This is what generates market failure. The basic rules for marketplace competition laid down by the state—including the system of property rights—are designed to limit these possibilities, in order to bring real-world competition closer to the ideal (or to bring outcomes closer to those that would be achieved under the ideal, in cases where a functional competition cannot be organized). This is the motivation

[29] Kenneth Arrow (1973: 303–317) puts particular emphasis on the consequences of firms maximizing profits in cases where there are pollution externalities and information asymmetries that favor the firm. "The classical efficiency arguments for profit maximization do not apply here," he writes, "and it is wrong to obfuscate the issue by invoking them" (308).

that underlies not only direct state provision of public goods, such as roads, but also state regulation of negative externalities, such as pollution.[30]

Unfortunately, the law is a somewhat blunt instrument. In many cases, the state simply lacks the information needed to implement the measures needed to improve upon a marketplace outcome (sometimes because the information does not exist, but often because the state has no way of extracting it truthfully from the relevant parties). Even when the information can be obtained, there are significant administrative costs associated with record-keeping and compliance monitoring, not to mention the costs incurred by firms in an effort to evade compliance. Thus the deadweight losses imposed through use of the legal mechanism can easily outweigh whatever efficiency gains might have been achieved through the intervention. This often makes legal regulation unfeasible or unwise.

It is at this point that ethical constraints become germane. As we have seen, profit is not intrinsically good. The profit-seeking orientation of the private firm is valued only because of the role that it plays in sustaining the price system, and thus the contribution that it makes to the efficiency properties of the market economy as a whole. Ideally, the only way that a firm could make a profit would be by employing one of the preferred strategies. However, for strictly practical reasons, it is often impossible to create a system of laws that prohibits the non-preferred ones. Thus according to the market failures perspective, specifically ethical conduct in an extrafirm business context (i.e., when dealing with external parties) consists in refraining from using non-preferred strategies to maximize profit, even when doing so would be legally permissible. Put more simply, the ethical firm does not seek to profit from market failure. In many cases, doing so will be illegal—precisely because the state has tried, through increased regulation, to eliminate the use of non-preferred competitive strategies. Ethical constraint becomes relevant in the rather large penumbral region of strategies that are not illegal, and yet at the same time are not among the preferred.

Corporations, for instance, are often in a position where they can produce advertising that will quite likely mislead the consumer, but which stops short of outright falsity. In a perfect world, advertising would provide nothing more than truthful information about the qualities and prices of goods. However, the vagaries of interpretation make it impossible to prohibit anything but the most flagrant forms of misinformation. Thus misleading advertising stands to false advertising as deception does to fraud. It is something that would be illegal, were it not for practical limitations on the scope of the legal mechanism. Profiting from such actions is therefore morally objectionable, not because it violates some duty of loyalty to the customer (as stakeholder theory would

---

[30] For more extensive discussion, see Heath 2001b.

have it), but because it undermines the social benefits that justify the profit orientation in the first place. (In a sense, the invisible hand no longer works to transform private vice into public virtue in this case, and so we are left merely with vice.)

In this respect, the market failures approach to business ethics is a version of what Bruce Langtry calls "tinged stockholder theory," which holds that "firms ought to be run to maximize the interests of stockholders, subject not only to legal constraints but also to moral or social obligations" (1994: 434–435). Indeed, it has been well understood for a long time that a shareholder-focused model with a set of deontic constraints (or "side constraints") on the set of permissible profit-maximizing strategies represents a plausible alternative to the stakeholder model.[31] What distinguishes the market failures approach from other such proposals is the specific account of how these constraints should be derived. Rather than trying to derive them from general morality (as Langtry does by focusing on the "moral rights" of individuals affected by the firm, or as Goodpaster does even more explicitly through appeal to the "moral obligations owed by any member of society to others"), the market failures approach takes its guidance from the policy objectives that underlie the regulatory environment in which firms compete, and more generally, from the conditions that must be satisfied in order for the market economy as a whole to achieve efficiency in the production and allocation of goods and services. Furthermore, by focusing on the distinction between administered transactions and market transactions, it is able to offer a principled basis for the difference in structure between the intrafirm obligations owed to shareholders and the extrafirm obligations owed to other groups affected by the actions of the corporation.

When one adopts this market failures perspective, there is no reason to think that a conception of business ethics that continues to place primary emphasis upon the fiduciary responsibility toward shareholders cannot deal with the ethical obligations that have traditionally been described under the heading of "corporate social responsibility." What so often upsets people about corporate behavior—and what gives profit-seeking a bad name—is the exploitation of one or another form of market imperfection. People generally have no problem with companies that make money by providing good service, quality goods, low prices, and so forth. For example, if all companies fully internalized all costs, and charged consumers the full price that the production of their goods imposed upon society, I believe it would be impossible to make the case for any further "social responsibility" with respect to the environment. Thus the market failures approach to business ethics is able to retain the intuitively familiar idea that managers have fiduciary duties toward shareholders and that

---

[31] Goodpaster (1991: 67–68), for example, moots such a proposal. The term "side constraint" is from Nozick (1974: 28–32), whose discussion of the issue is also quite helpful.

the primary goal of corporations is to make a profit. Yet it is able to avoid the charge of moral laxity often leveled against the shareholder model of business ethics, because it imposes strict moral constraints on the range of permissible profit-maximization strategies.

There is a close analogy, from this perspective, between "corporate social responsibility" and the concept of "good sportsmanship" in competitive team sports. In the case of sports, the goal is clearly to win—but not by any means available. Every sport has an official set of rules, which constrain the set of admissible strategies. Yet it will generally be impossible to exclude strategies that respect the letter of the law, while nevertheless violating its spirit (e.g., taking performance-enhancing drugs that have other legitimate uses, and therefore have not been banned). "Good sportsmanship" consists in a willingness to refrain from exploiting these loopholes, while nevertheless retaining an adversarial orientation. In other words, the obligation is to be a team player and to compete fairly, but not necessarily to let the other side win. The fundamental problem with stakeholder theory is that it tries to eliminate the adversarialism of the managerial role, rather than merely imposing constraints upon it.

3.5. Conclusion

One of the charges that hostile critics frequently make against business ethicists is that they are implicitly, if not explicitly, anticapitalist. Insofar as one equates business ethics with the stakeholder paradigm, there is more than a grain of truth in this accusation. Goodpaster was certainly not wrong to observe that the multi-fiduciary stakeholder theory "blurs traditional goals in terms of entrepreneurial risk-taking, pushes decision-making toward paralysis because of the dilemmas posed by divided loyalties and, in the final analysis, represents nothing less than the conversion of the modern private corporation into a public institution and probably calls for a corresponding restructuring of corporate governance (e.g., representatives of each stakeholder group on the board of directors)" (1991: 66). There is, of course, nothing wrong in principle with arguing for institutional reforms of this kind. But a theory that has this as its consequence is unlikely to provide much guidance when it comes to dealing with the ethical challenges that arise in the day-to-day operations of firms in an unreformed capitalist economy.

One of the central advantages of the market failures approach to business ethics is that, far from being antithetical to the spirit of capitalism, it can plausibly claim to be providing a more rigorous articulation of the central principles that structure the capitalist economy. If firms were to behave more ethically, according to this conception, the result would be an enhancement of the benefits that the market provides to society, and the elimination of many of its persistent weaknesses. It would help to perfect the private enterprise system, rather than destroy it.

Of course, none of this is intended to show that one cannot continue to talk about corporate social responsibility in terms of stakeholder interests. The question is simply whether this vocabulary encourages a more or less perspicuous articulation of the important moral issues. In this respect, it is important to remember that the term “stakeholder” was coined precisely in order to suggest an analogy between the relationship that managers have with shareholders and the relationship that they have with other interested parties. But as we have seen, the moral obligations that managers have toward these disparate groups are not analogous; in fact they are quite dissimilar. So while the term “stakeholder” may remain a useful piece of shop-talk in strategic management circles, as a piece of ethical vocabulary, for use in a theory that tries to articulate the central moral obligations of managers, it is inherently misleading. It creates considerable mischief in business ethics, while offering no real conceptual gain.

4 }

An Adversarial Ethic for Business: or, When Sun-Tzu Met the Stakeholder

Some of the most serious confusions to arise in the business ethics literature stem from a failure to distinguish adequately between the moral obligations that managers have toward individuals who are "outside" and those who are "inside" the corporation. In the economic literature on the firm, especially in the transaction cost tradition, a sharp distinction is drawn between so-called "market transactions" (which involve buying and selling in the market) and "administered transactions" (which are governed by the rules that structure the bureaucratic hierarchy of the firm) (Shipman 1999: 267; Williamson 1975). The reason this distinction is so important for business ethics is that market transactions are governed by the *competitive* logic of the market environment in which the firm operates, whereas administered transactions are subject to the *cooperative* norms that govern collective action in a bureaucracy.

Generally speaking, the norms that structure systems of cooperation are significantly more exigent, from the moral point of view, than those that govern competitive behavior. Indeed, one of the hallmarks of competition—and one of the reasons that many people feel such unease with it—is that it appears to offer individuals temporary and partial exemption from some of the norms that ordinarily structure interpersonal relations. Thus, competition permits forms of behavior that would, in other contexts, typically be regarded as antisocial. There are many examples from the field of competitive sport that could be drawn upon to illustrate this principle. There is a special branch of ethics, referred to as *adversarial ethics*, that deals with the problem of determining appropriate standards of conduct in such contexts.[1] So far, however, there has been little or no recognition of the fact that a significant portion of the issues traditionally dealt with by business ethicists, namely, those that pertain to market transactions, fall into the domain of adversarial ethics.

[1] See, most importantly, Applbaum 1999.

The widespread failure to distinguish between market transactions and administered transactions, and thus to distinguish between adversarial and non-adversarial relations, has led many business ethicists to develop a "uniform" moral code, suggesting that the same test, or that the same standards be applied in all circumstances and in every transaction. This leads to a problem that has become endemic in the business ethics literature. In order to set a uniform standard "high enough" to govern cooperative relations within the firm (and thus to handle issues like employee health and safety, personnel management, and so on), it must be set so high that it essentially precludes adversarial behavior. This makes marketplace competition impossible, from the moral point of view, and so makes business ethics implicitly anticapitalist (Goodpaster 1991: 66; *The Economist* 2005). Yet if one turns around and lowers the standard, in order to permit adversarial behavior toward competitors, the moral code winds up licensing all sorts of sharp practices within the firm that are not only unethical, but that are even incompatible with the imperatives of good management (dependent, as the firm is, upon norms of reciprocity and a climate of trust in order to secure the good will and loyalty of its employees). This does an enormous amount to heighten the perception that, while ethics in business are all well and good, they represent niceties that, when push comes to shove, may need to be set aside.

The solution to this problem lies in the recognition that moral obligations in business are not uniform. There is, rather, an institutional division of moral labor. In market transactions, the checks and balances built into the system of commercial exchange are such as to permit more instrumental (or "self-interested") forms of behavior. In administered transactions, by contrast, these checks and balances are absent (indeed, managers often wield great power over the lives of subordinates), and thus the institutional context calls for much greater exercise of moral restraint. This is a very old idea—that the "invisible hand" of the market transforms certain vices into virtues, in a way that the "visible hand" of management does not. Unfortunately, those who take this line of reasoning seriously have had a tendency to overstate their case (e.g., by claiming that markets obviate the need for any sort of moral restraint [Gauthier 1982]). This has created, in turn, a rather hypertrophied aversion among business ethicists to any discussion of the ways in which markets might license a selective exemption from everyday moral norms.

In this paper, I would like to begin the task of developing an adversarial ethics for business. I do so by, first, analyzing the structure of competitive behavior, along with the specific forms of competition that constitute the economic environment in which firms operate. I then go on to show how this competitive environment licenses certain forms of "self-interested" behavior, but also imposes its own limits on the strategies that firms may adopt in the pursuit of their interests. This constitutes the core of an adversarial ethic for market transactions, one that is clearly distinct from the norms that govern administered transactions.

4.1. The Nature of Competition

Morality arises in response to the fact that human affairs, when left to their own devices, have a tendency to go very badly. Thomas Hobbes summed it up best with his observation that the unbridled pursuit of individual self-interest generates a "natural condition" in which life is "solitary, poor, nasty, brutish and short." This is because individuals who refuse to exercise any restraint in the pursuit of their self-interest rapidly become embroiled in collective action problems—interactions in which, despite acting in a self-interested fashion, each individual winds up with an outcome that is much worse than some other feasible outcome, which might have been achieved had they all chosen to act differently. Furthermore, a collective action problem can easily degenerate into a race to the bottom, in which each individual, responding to the actions of the others, generates an outcome that is successively worse, but where each iteration of the interaction only intensifies their incentive to act in the same way. An arms race is the most clear-cut example.

One of the primary functions of morality (and of social institutions more generally) has always been to impose constraints that prevent individuals from falling into these kinds of collectively self-defeating patterns of behavior (see Gauthier 1986; Schotter 1981). A simple golden rule, for example, which asks individuals to consider, before embarking upon a particular course of action, how they would feel if others acted the same way, has the potential to resolve the overwhelming majority of collective action problems, and thus to promote mutually beneficial forms of cooperation. Consider, for example, the rule against littering. When leaving a subway car, it is tempting to leave one's newspaper behind, rather than carry it along in search of a trash can. At the same time, people generally do not like riding in messy subway cars—the only reason they are tempted to leave the newspaper behind is that they are exiting the train. This creates a collective action problem (or a prisoner's dilemma). Figure 4.1 shows a simplified version of this game involving two riders, along with a graph of the payoffs (representing the level of satisfaction that the riders get from their morning commute).

The norm that prohibits littering takes the riders of the subway away from the strategic equilibrium, which is (1,1), and allows them to achieve the cooperative outcome (2,2). Of course, it is still in the interest of each rider to defect from the cooperative arrangement by littering. The social norm, insofar as it does constrain the conduct of the two riders, represents a genuine constraint; it is not merely their self-interest correctly understood. What makes the norm advantageous is the fact that general compliance generates a win–win outcome. This is the hallmark of moral action. (The philosopher Kurt Baier has written, with considerable plausibility, that being moral simply means "following rules designed to overrule self-interest whenever it is in the interest of everyone alike that everyone should set aside his interest" [1958: 314]. Even if this is not all of morality, it is certainly a sizable chunk of it.) Immoral action, on the other hand, tends to generate win–lose outcomes (and when everyone does it, lose–lose outcomes).

This analysis makes it somewhat easier to see why competition often appears to be so puzzling, and for many people, so morally problematic. While cooperation is designed to deliver win–win outcomes, competitions are specifically designed to produce win–lose ones (Skillen 1998: 171). Furthermore, the structure of a competition is designed to induce all of the competitors to defect rather than to cooperate (Heath 2001b: 93–97). Take the example of an athletic competition, such as long-distance running. If you took a randomly selected group of people and told them to run a race, promising to give a prize to the fastest, then generally speaking the prize would go to the person with the most natural ability (the right sort of frame and musculature, the best cardiovascular system, etc.). On the other hand, if you announce the contest well in advance, it is possible for those with less natural ability to improve their chances of winning by training for the race (thereby improving their musculature, cardiovascular system, etc.). Yet when the less talented begin to train, this just forces those with more natural ability to train as well, so that they can retain position. At the end of the day, when everyone trains equally, the person with the most natural ability still wins. Yet, everyone involved in the competition now is expending much greater time and effort to achieve this result, and thus the outcome is suboptimal from the standpoint of the competitors. In other words, training for an athletic competition is a form of defection (equivalent to littering the subway car, in figure 4.1). In fact, it is one that generates a race to the bottom. If everyone is training three hours a day, it gives those with less natural ability an incentive to train four hours a day. When those with more talent start to match that, and train four hours a day, it simply gives those with less talent an incentive to train five hours a day, and so on.

The fact that training has this structure has not escaped the attention of athletes. As Robert Frank and Philip Cook observe:

> The Academy Award-winning film *Chariots of Fire* portrays British collegiate track-and-field competitors who have developed an implicit norm that limits their training and practice time. Their apparent understanding is that since the most talented runner will win whether all train arduously or none does, the sensible thing is for no one to train very hard. This arrangement is challenged by an outsider with a rigorous training regimen. In response the incumbents bring considerable social pressure to bear upon the maverick. In the face of such pressure, most normal challengers might have succumbed. But this particular runner is tough, and he goes on to win in the end. (Frank and Cook 1995: 142)

		Player 2: Litter	Player 2: Don't Litter
Player 1	Litter	(1,1)	(3,0)
	Don't Litter	(0,3)	(2,2)

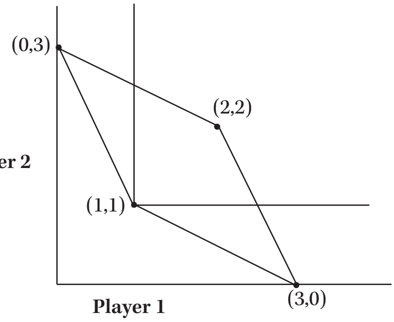


FIGURE 4.1 Prisoner's dilemma

Of course, when it comes to competitions our sympathies lie with those who "break ranks" and adopt the non-cooperative strategy of training. Indeed, the point of a competition is to encourage precisely this sort of "one-upmanship." Yet, why would society want to inflict this peculiar type of collective action problem upon people? The answer is that desirable competitions also generate positive externalities—benefits to people other than those directly involved. The competition is precisely how society induces those involved to produce these benefits, despite the personal inconvenience that it entails. Olympic athletes, for instance, might prefer not to have to give up their entire lives to train, but the intensity of competition generates a riveting display, in which spectators can see the frontier of human achievement being pushed back year after year.

Thus, the reason that "society" favors competition in certain areas of life has everything to do with the externalities that are generated. The difference between healthy and unhealthy forms of competition is that, in the former case, the external benefits outweigh the losses incurred by the competitors, while in the latter case they do not. Compare the case of training to that of performance-enhancing drugs (see Simon 1988). Both have the structure of a defection strategy. When one person starts training, everyone else is forced to train as well, in order to have any chance of winning. In the same way, when one person starts taking steroids, everyone else has to take steroids as well, in order to have any chance of winning. The difference is that training, although it represents an inconvenience to many people, usually improves the athlete's overall health, whereas performance-enhancing drugs have serious adverse health effects in the long run. (Indeed, it is a testament to the intensity of the race to the bottom among athletes that so many are willing to take them, and so many more would be willing to do so, in the absence of regulations prohibiting them and testing to monitor compliance.)

This is why competitions need to be so carefully monitored and regulated. In general, the participants are motivated by the incentive to defect (i.e., the desire to win) and not by the overall "social" objectives of the competition.[2] If this were not the case, then there would be no need to test for performance-enhancing drugs; athletes would simply refrain from taking them on the grounds that they are not "good for the sport." Yet, the logic of the collective action problem at the heart of athletic competition generally precludes this sort of high-mindedness. Thus healthy competitions are always in danger of degenerating into unhealthy ones. There was no better reminder of this than the scandal that erupted in American figure-skating in 1994, when skater Tonya Harding sent a member of her entourage out to kneecap her primary rival, Nancy Kerrigan. Needless to say, the point of a figure-skating competition is not to see who will be left standing at the end of the day, but rather to see who can perform the most impressive on-ice maneuvers. Practicing is a legitimate way of besting one's rivals; sending out thugs to handicap them is not. The former generates positive externalities that make the competition a "race to the top," while the latter clearly transforms it into a "race to the bottom." Thus, the difference between healthy and unhealthy competition lies not in the intentions of the competitors, but rather in the rules that constrain them and keep them from employing strategies other than those that generate positive externalities. There is nothing intrinsically right or wrong about any particular competitive strategy (after all, they are all forms of non-cooperative behavior), the question is simply whether the strategies chosen promote healthy or unhealthy forms of competition.

One can see already how this peculiar structure makes the moral evaluation of competitive behavior rather tricky. The problem is that the beneficial consequences of a competition arise necessarily as a byproduct of the competitive activity, while the objectives that the participants themselves seek often seem morally objectionable prima facie. The virtues of the competition, such as they are, are associated with the institutional structure (i.e., the set of rules) that constrains the participants' behavior, and not necessarily the intentions of the participants. Indeed, insofar as a competition does produce beneficial consequences, it is almost as though the participants were guided, by an invisible hand, to promote an end which was no part of their intention.

4.2. Competition in Business

Everyone knows that businesses operate in a competitive environment. However, the way that market exchange is presented in the standard microeconomics curriculum sometimes obscures the fact that marketplace competition also has at its core an unresolved collective action problem (indeed, it

---

[2] Thus Warren Fraleigh is careful to distinguish between the "intended end" of participants and the "purpose of the sport contest" (1984: 37–42).

is not just an unresolved collective action problem, but an institutionalized collective action problem, since attempts to resolve it are widely prohibited by antitrust law). Thus, it is worth reviewing briefly the structure of marketplace competition.

Familiarity with so-called “general equilibrium” models has conditioned many people to think of the point at which supply and demand curves intersect as the equilibrium of an exchange, and the price level at that point as the equilibrium price. Under certain conditions this may be true of aggregate supply and demand, but it is not true of individual supply and demand curves. When there is only one buyer and one seller, every price level at which some positive quantity of goods would be exchanged is the Nash equilibrium of a marketplace interaction in which either the buyer or the seller makes a “take-it-or-leave-it” offer to the other. Consider figure 4.2. The seller may find it advantageous to sell quantity x_1 at price level p_1, rather than x_2 at p_2. If he makes a “take-it-or-leave-it” offer to the buyer at that price, and the buyer believes that the price is firm, then it is in the buyer’s interest to accept.

This is why buying and selling in one-on-one interactions often involves so much posturing. Both parties know that if the other believes that the “final offer” is indeed a final offer, then he or she will accept, so long as the price is within the zone of exchanges that generate a mutual benefit (i.e., a “gain from trade”). However, there is no guarantee that the exchange will maximize the mutual benefit. Thus, the seller may wind up with unsold goods at the end of the day, simply because it was best to sell a smaller quantity at a higher price.

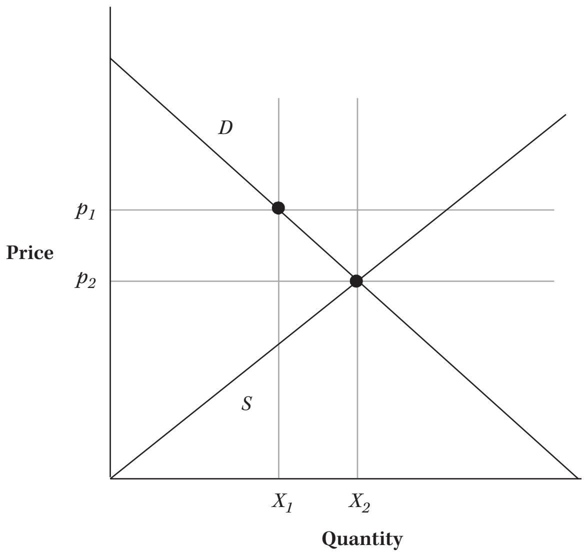


FIGURE 4.2 Market exchange

In other words, there is no expectation that markets will clear in exchanges between only one buyer and one seller.

As soon as another buyer or seller enters the market, however, the strategic situation changes completely. The presence of multiple buyers and sellers dramatically reduces the ability of any one buyer or seller to make a credible "take-it-or-leave-it" offer. If the price that the sellers are charging is above the price at the point where supply and demand curves intersect, then they will wind up with unsold goods at the end of the day. If they are both charging the same price, then one can assume that they will split the sales between them, and so both wind up with unsold goods. Yet this creates a temptation for both sellers. By dropping the asking price somewhat, it should be possible to sell one's entire inventory. The loss of revenue caused by the lower price will then be made up for by the increased volume of sales. Of course, if one seller does this, then the other has no choice but to respond in kind. The result is lower profits for both of them. This competition will continue until the volume of sales at a given price level leaves neither of them with unsold goods. This is the point at which supply and demand curves intersect (which is why the price at that point is known as the "market-clearing" price). The same sort of competition develops among buyers in cases where the price is lower than the market-clearing price—some buyers will be left with unsatisfied demand at the end of the day, and so will have an incentive to defect, by paying more than the going rate, in order to guarantee that they secure enough of the good.[3]

Clearly, it is not in the joint interest of either suppliers or buyers to compete with one another in this way. Thus, the reason that price competition is desirable is not that it benefits the people involved, but rather that it generates external benefits for society at large. In this respect, it is quite similar to athletic competition. But what are these external benefits, in the case of the competitive market? When suppliers compete with one another, it benefits buyers, and vice versa. Thus the competitive market works to eliminate "deadweight losses" from the economy, ensuring that the maximum number of mutually beneficial economic exchanges take place. But more importantly, a competitive market also gives rise to a set of prices, which provide crucial information to everyone else in society about the relative scarcity of the various resources,

[3] The model of marketplace competition presented here is similar to the neoclassical economic one, in that it posits two collective actions problems, one on the supply side and one on the demand side. (It differs in that it treats pricing decisions as the primary competitive strategy, whereas the standard neoclassical model represents individuals as "price-takers" who react to market conditions only by adjusting the quantity that they supply or that they purchase.) However, even in cases where there is very little competition among firms on the supply side, or among households on the demand side, one may see the emergence of what Galbraith (1952) called "countervailing power." In this case, a similar sort of competitive dynamic could be diagnosed, involving collectively self-defeating rent-seeking behavior on the part of increasingly oligopolistic agencies on both the supply and the demand side. In this case, the collective action problem exists between those on the supply and those on the demand side.

skills, goods, and services being exchanged. In the same way that an infrared camera takes invisible light and converts it to a wavelength that the human eye can see, the competitive market takes people's invisible preferences regarding both production and consumption and converts them to something that can be observed with the naked eye, namely, prices. This is what makes economically rational decision-making even roughly possible in every sector of the economy, including the public sector. The operation of the price system therefore allows for a more efficient (i.e., less wasteful) use of resources and labor.

Furthermore, the failure on the part of either producers or buyers to compete with one another can cause considerable mischief, insofar as it sends the wrong "signals," via the price mechanism, to other economic actors. When suppliers, through collusion or cartelization, are able to maintain prices for some good at above-market-clearing rates, it suggests that there is "not enough" of that good, and so encourages a shift of resources away from other economic activities toward increased production of that good, combined with a shift among consumers toward goods that serve as substitutes (assuming such are available). Similarly, when buyers form a "consumer co-op," or some similar organization, in order to hold out for lower prices, it sends the signal to suppliers that there is "too much" of the relevant good, and so encourages them to shift investment out of that sector.

This is, of course, the substance of "invisible hand" arguments for the market since Adam Smith. It is why David Gauthier, in his article "No Need for Morality: The Case of the Competitive Market," argues that in market transactions, moral constraints "would be not merely pointless, but positively harmful" (1982: 54). One is not merely encouraged to act non-cooperatively in a competitive market, social welfare considerations require one to do so, because the price mechanism requires competition in order to generate the right information about the relative scarcity or need for different goods.

Of course, it is important to recognize that there is nothing magical about the ability of markets to transform private vices into public virtues. This sort of laundering is a general feature of all competitively structured social interactions. And like all other forms of competition, market competition must be governed by a set of rules, restricting the range of strategies that individuals may employ, in order to ensure that it remains healthy. For suppliers, offering to sell at a lower price—and making the necessary changes in the production process that will enable one to do so—is the most important permissible strategy. Adjusting the quantity that is supplied and making improvements in product quality are also permissible.

But like every other form of competition, market competition also has a tendency to go off the rails when improperly regulated. In principle, there is no reason why firms could not compete with one another by blowing up each others' factories and hiring assassins to kill each others' CEOs. Such a scenario is no less implausible than figure skaters sending out thugs to kneecap their opponents. In fact, one need only look at the experiences of the various "transition economies" in the former communist bloc to see the sort of outrageous behavior that improperly regulated marketplace competition may generate. For example, in 1994, shortly after the privatization of agriculture and food production in Hungary, the country was swept by an epidemic of lead poisoning. After searching far and wide for the cause, doctors and scientists finally tracked down the source of the problem. Manufacturers of paprika—a staple of Hungarian cuisine—had been grinding up old paint, much of it lead-based, and adding it to the spice in order to improve its color. The practice was so widespread that officials in Hungary were forced to order all the paprika in the country removed from store shelves and destroyed. This is a clear example of firms using an impermissible strategy—exploiting an information asymmetry—in order to compete, and other firms being forced to do the same, in order to retain position. The race to the top of the competitive market is thereby transformed into a race to the bottom, one that can have devastating consequences for the society at large.

4.3. The Morality of Competition

Much of everyday morality has as its goal the prevention of collective action problems. It is possible to secure certain advantages by lying, but if everyone did it, no one would believe what anyone said, and everyone would be worse off. It is possible to advance one's interests by stealing from others, but if everyone did it, everyone would have to make costly investments in security and protection, etc. This is why the various formulations of the Golden Rule capture much of the spirit of everyday morality. But because the central mechanism in a competition is an unresolved collective action problem, there are bound to be numerous prima facie conflicts between competitive imperatives and those imposed by everyday morality. This is reflected in the fact that a naïve or mechanical application of the Golden Rule in a competitive situation is likely to generate the wrong results. Before kicking in the winning field goal, we do not want football players to be thinking, "How would I like it if the other team did that to me?" Similarly, before lowering prices, we do not want gas-station owners to be thinking "How would I like it if the station across the street did that to me?"

There is some debate among ethicists as to whether this conflict with everyday morality is real or apparent. Arthur Applbaum has offered a critical survey of arguments that "have been offered to back up the claim that the rules of [competitive] games provide moral permission to use tactics that would otherwise be wrong" (1999: 115). He argues that this conclusion, which seeks to dissolve the tension between adversarial practices and everyday morality, is in fact much more difficult to sustain than many have imagined. In some cases, participants sign waivers, whereby they explicitly consent to be treated by others in the way that the game rules dictate. But more often, whatever consent is present is merely implicit, and generalizing from this type of consent to the moral permissibility of prevailing practices is fraught with difficulty. For example, Applbaum observes that "when alternatives to participation in a game are poor, expectation of an adversary game does not imply consent to its rules. In buying a used car, you may fully expect to be deceived about its defects" (1999: 117)—this does not mean that the dealer is morally entitled to deceive you.

Thus, Applbaum argues that, in the majority of cases, adversarial institutions generate behavior that is morally wrong *pro tanto*, but perhaps permissible all things considered (i.e., when the systemic consequences of that behavior within that institution are brought into the picture). In the case of competitive behavior, this means that the consequences of defecting from the cooperative arrangement constitutes a genuine harm for the other competitors, but that the wrongness of this harm is outweighed by the positive externalities generated by the competition as a whole (e.g., the "ratcheting up" of effort and skill in a sporting competition), and thus the action in its context is morally permissible. This is, of course, still a somewhat tricky position to defend, since it involves a certain instrumentalization of the other competitors. The general point, however, is sound. Adversarial institutions do not provide individuals with a moral "get out of jail free" card, such that categories of moral evaluation no longer apply to their conduct (leaving them free to pursue whatever course of antisocial behavior happens to suit their fancy). In other words, these institutions do not dissolve morality. What they provide is, at best, a set of highly specific exemptions from particular moral obligations.

One can see this clearly reflected in the morality of sport. In fact, we can learn a great deal about the morality of adversarial relations by examining sports, both because games are highly artificial constructs, and so are governed by an unusually explicit set of rules and regulations, but also because sports play an important educational role in the socialization of the young, and so the underlying moral ideals tend to be quite well articulated. One need only look at what parents and coaches say to children after a game has gone poorly. The central moral ideal here is known as "sportsmanship" (Feezell 1988) or "being a good sport." This is a complex ideal, one that involves a number of different characteristics.[4]

> *Constrained competitiveness:* The good sport is one who maintains a zealously adversarial stance within the designated context of the game, but then drops this orientation and adopts a more cooperative demeanor when the game is over. Thus a classic way to demonstrate good

[4] For an excellent empirical survey, see Commission for Fair Play (1993: 34–38).

sportsmanship in a contact sport is for a player, after having knocked an opponent down, to offer him a hand up after the whistle is blown. The whistle that stops the play effectively signals a switch from adversarial to cooperative relations; a good sport is one who is able to make this switch without allowing residual ill will from the competitive segment to poison relations in the cooperative. (Indeed, one of the reasons that competitive sports are often thought to "build character" is that they force children to develop this more advanced form of self-control.) For similar reasons, a good sport does not "rub it in" after having won or behave sullenly after losing, but is rather "courteous in victory, gracious in defeat." Again this is a way of emphasizing the point that the win–lose structure of the interaction is confined to specific actions taken in the game; it does not extend to general participation in the sport.

No cheating: This almost goes without saying, but a good sport is one who respects the rules of the competition, even when the referee is not looking, or the chances of detection are slight. That having been said, it should be noted that the temptation to cheat is perhaps greater in adversarial relations than in everyday cooperative ones, precisely because the competition is already structured as a race to the bottom among competitors. Thus, the temptation to cheat may require greater force of character to resist in sport (another reason that it is felt to build character in the young). As we have already seen in the case of anabolic steroids and other banned performance-enhancing drugs, cheating can be a serious problem in sport and has the potential to undermine all of the beneficial side effects that make the competition "healthy" in the first place.

No gaming: "Gaming" the rules involves taking actions that are technically not prohibited, but are not intended to be permissible strategies. Such actions violate the spirit, rather than the letter, of the rules, and are prohibited by the ideal of sportsmanship.[5] Such strategies are sometimes referred to as "exploits" (precisely because they exploit an unintended feature of the structure of the competition). They involve actions that would be against the rules, but for some oversight (e.g., it never occurred to anyone that players would do it) or impracticality (e.g., it is impossible to enforce a rule against it). An example would be the use of bronchodilators among athletes to enhance their cardiovascular efficiency prior to a competition. The problem is that there is no real way to distinguish between those who genuinely have asthma, and so need the medication, and those who use it in the hopes of enhancing their performance.

[5] See Leaman 1988. He gives the example of a tennis player constantly stopping to retie her shoelaces, in order to unnerve her opponent (1998: 278). See also Steenbergen et al. 2001: 141–142.

Thus these substances are not officially banned, even though their use in many cases is clearly contrary to the spirit of the regulations that prohibit performance-enhancing drugs.

Taking the high road: Finally, and most fundamentally, the good sport is one who considers respect for the principles of good sportsmanship to be more important than winning. Faced with an opponent who has decided to "play dirty," the good sport does not take this as license to start playing dirty herself.[6] The consequence is that she may often suffer defeat, rather than stoop to the level of an unscrupulous opponent. This requires the greatest self-control of all, since it requires not just overcoming the desire to win, but also suppression of our disposition to punish, through reciprocation, those who violate moral norms.

The function of the rules that govern a sport is to promote healthy competition. The morality of sport is clearly structured by the same interest. In many cases, it simply complements the official rules, by mandating respect for the spirit, as well as the letter, of the rule. A competition is socially beneficial when players exercise restraint in the strategies that they employ, when they confine their adversarial behavior to certain specific contexts, and when they refrain from allowing moral lapses on the part of other competitors to transform the entire contest into a race to the bottom. Moral judgment, in this case, is always guided by a sense of what the overall "point" of the competition is, what the beneficial consequences of the activity are, and how the competition serves to generate them.

4.4. Implications for Business Ethics

There can be little doubt that the core element of any plausible conception of business ethics is going to be a system of principles that mandates cooperative behavior with regard to the various agency relationships that exist within the firm, first and foremost, the principal–agent relationship between senior management and shareholders (Buchanan 1996). These moral obligations are deeply entrenched, both in terms of institutional practices and in corporate law—most obviously, in the fact that courts treat senior managers as fiduciaries of the firm and directors as fiduciaries of shareholders (Clark 1985). The problem with this conception, however, is that it generates a system of moral obligations that tracks the agency relationships, and thus directly mirrors the

[6] "Taking the low road" is sometimes referred to, euphemistically, as "evening things up" (see Commission for Fair Play 1993). This suggests that violation of the rules by others generates a moral permission (perhaps even an obligation) for others to do so. According to the view developed here, it provides at best an excuse for doing so, never a justification (see Baron 2005).

organizational hierarchy of the firm. Individuals have duties toward those who are, in some sense, their superiors; employees toward their supervisors, managers toward executives, executives toward the board of directors, and via the board of directors, the shareholders. But what about other individuals who may be affected by the actions of the firm? What about customers, creditors, suppliers, or local communities? A conception of business ethics that focuses too narrowly upon obligations toward shareholders appears to give individuals free reign to engage in "sharp practices" in dealings with the latter groups.

Faced with this difficulty, one of the most influential impulses among business ethicists has been to take the fiduciary relationship that exists between managers and shareholders and use it as a model for positing additional fiduciary responsibilities between managers and so-called "stakeholder" groups. The claim, in effect, is that managers are agents with multiple principals, who must therefore exercise a duty of care and loyalty toward all of these different stakeholder groups.[7] Of course, many others have felt that this is the wrong way to proceed. Unfortunately, those who are opposed to this kind of "multi-fiduciary" stakeholder analysis have not done a very good job of formulating their objections. Several have suggested that managers should retain a fiduciary orientation toward owners, but their relations with other "patron" groups should be subject to deontic constraints (Goodpaster 1991; Langtry 1994). The standard argument has been that the relationship between managers and shareholders should be privileged because the latter are residual claimants, and are therefore much more dependent upon the good faith of management (Boatright 2002: 47–48). The interests of all the other major stakeholder groups—with some notable exceptions—are protected by contract. Since the agency risks in such relationships are low, the imposition of fiduciary duties would be otiose (Easterbrooke and Fischel 1991: 90–92).

The problem with this response, which defenders of stakeholder theory have emphasized, is that the mere fact that shareholders, as residual claimants, are more in need of protection from exploitation by managers than other stakeholder groups does not explain why there should be any sort of qualitative distinction in the nature of the moral obligations that are owed to them (Boatright 2002: 50–51). It may explain why they are owed a greater duty of care, but it cannot explain why *only* they should be owed a duty of care.

A more persuasive response would build upon the distinction between administered transactions and market transactions. As Ronald Coase put

[7] Thus, stakeholder approaches to business ethics often involve a commitment to what Goodpaster 1991 refers to as a "multi-fiduciary" view.

it, the most important organizational feature of the firm is the internal supersession of the price mechanism, along with the type of competitive behavior that it requires to function correctly. "Outside the firm, price movements direct production, which is coordinated through a series of exchange transactions on the market. Within a firm, these market transactions are eliminated and in place of the complicated market structure with exchange transactions is substituted the entrepreneur-coordinator, who directs production" (1937: 388). Thus, the difference in character of the moral obligations that managers owe to different individuals who are affected by the actions of the firm depends upon the nature of the transactions that occur between them, and in particular, whether these transactions are mediated through the price mechanism. Administered transactions—within the hierarchy of the firm, which includes both employees and shareholders (via the board of directors)—are organized as principal–agent relations, and are therefore governed by an essentially cooperative logic. This is why moral obligations in this case take on a fiduciary or quasi-fiduciary form, and are aimed at reducing agency risks. These obligations are, as Allen Buchanan has emphasized, obligations to *advance the legitimate interests* of the principal (1996: 424). Market transactions, on the other hand, are mediated by the price mechanism, and are therefore governed by an essentially competitive logic. Thus, moral obligations in this context have an adversarial character, because the market requires non-cooperative behavior in order to move prices toward the level that promotes the socially optimal use of resources. It follows quite naturally that these moral obligations cannot be fiduciary in nature, because one does not have an obligation to advance the interests of one's opponent in an adversarial context (if one did, then it would no longer be an adversarial context). It does not follow that these obligations may be any less strict; it just means that they must have a different form (see figure 4.3).

It should go without saying that there are also significant competitive aspects to relations within the firm. Indeed, most firms use internal competitions of various sorts (e.g., for bonuses and promotion) as a way of motivating work effort. In the same way, there are significant cooperative elements in market transactions, especially in cases where long-term contracts are in place. But this sort of complexity does not change the fundamental structural distinction, which has to do with the dominant mode of social integration in these domains. Intense personal rivalries may develop among players on a sports team, just as players from different teams may develop tacit norms of cooperation that limit the scope of competition. Yet there is still a fundamental distinction between what you owe to players on your own team and what you owe to those on a rival team. The same is true in business.

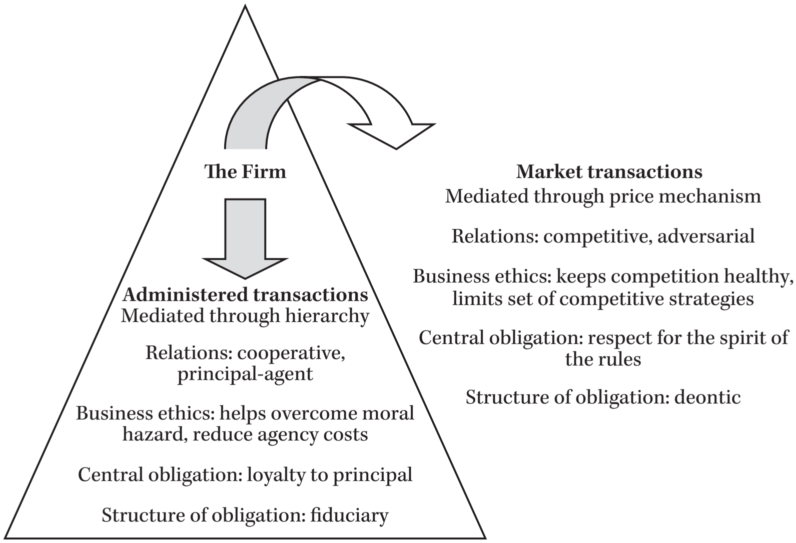


FIGURE 4.3 Administered and market transactions

Unfortunately, many theorists who are attentive to the difference between administered and market transactions have been misled by "invisible hand" arguments, which purport to show that nothing is owed to those on a rival team. Gauthier, for example, argues that because the perfectly competitive market reconciles the pursuit of self-interest with the production of socially beneficial outcomes, there is simply no call for moral evaluation: "The traditional moralist is told that his/her services are not wanted" (1982: 47). Thus, what he calls the "visible foot" of morality ("to be applied firmly to our backsides in order to redirect our concerns when individual gain and mutual benefit diverge" [1982: 41]), may be required whenever the "visible hand" of management is present, but wherever the "invisible hand" does the work of integrating our actions there is no need for it. Thus, markets represent "freedom from morality" (1986: 83).

Gauthier does mention one important exception to this claim. In order to get the perfect coincidence of self-interest and mutual benefit, the market must be perfectly competitive, and in order to be perfectly competitive, the market must satisfy certain conditions (usually referred to as the "Pareto conditions"). What Gauthier fails to emphasize is that it is impossible to satisfy these conditions in the real world. For example, perfect competition requires that there be no externalities and no information asymmetries anywhere in the economy. But there are always externalities and information asymmetries. Furthermore, it is not generally the case that the closest possible approximation of perfectly competitive conditions will yield the closest possible coincidence of self-interest and mutual benefit. Generally speaking, once one of the

Pareto conditions has been violated anywhere in the economy, there can be no presumption that satisfaction of the other Pareto conditions will lead to a more efficient outcome.[8]

Thus the invocation of the ideal of perfect competition as grounds for ignoring morality in the marketplace is, at best, the result of a weak grasp of the underlying economics, and at worst, positively misleading. This point has been emphasized by the economist Kenneth Arrow, who commands particular authority in this context, since it was he, in collaboration with Gerard Debreu, who finally proved the "invisible hand theorem," i.e. demonstrated that the equilibrium of a perfectly competitive market would be Pareto-optimal (Arrow and Debreu 1954). Arrow's argument for "ethical codes" to constrain the conduct of business emphasizes that when the Pareto conditions are violated, "the classical efficiency arguments for profit maximization do not apply . . . and it is wrong to obfuscate the issue by invoking them" (Arrow 1973: 308).

The problem with Gauthier's view (and those who share it, like Milton Friedman[9]) is that it confuses the adversarialism of market transactions with freedom from all moral constraint. Thinking that the invisible hand of the market eliminates the need for ethical conduct in business is like thinking that the competitive structure of sport eliminates the need for good sportsmanship. The market is not a free-for-all, any more than a competitive team sport is. Making a profit is the goal of business in the same way that winning is the goal of competitive sport. But the point is not to achieve this goal by any means possible; it is to achieve it in a fair and honest way.

The reason that such obvious truths have so often been ignored is that the law already prohibits firms from employing excessively antisocial competitive strategies. Thus some have been tempted by the view that it is redundant to constrain competition by adding on a moral prohibition, above and beyond the obligation to obey the law. But the law is a blunt instrument. If it is impossible to design a set of rules to create a perfect competition in sport, it is even more difficult to design a set of rules to perfect our system of markets. Thus there may be cases in which it is possible to employ competitive strategies in business that, while not technically illegal, nevertheless defeat the purpose of the market system. It is here that moral constraint is required. Arrow, for instance, identifies the problem of externalities and of asymmetric information as two cases in which "the simple rule of maximizing profits is socially inefficient." In such situations, "it is clearly desirable to have some idea of social responsibility" (1973: 309).

---

[8] This is a consequence of the so-called "second-best theorem." See Lipsey and Lancaster 1956.

[9] See Friedman 1962, 1970. The former, incidentally, contains a glaring example of the economic fallacy described above, in which "as close as possible to perfect competition" is assumed to generate "as close as possible to perfect efficiency" (1962: 120).

Parenthetically, it is important to distinguish between this view and one that regards the relevant moral constraints as simply an application of everyday morality to the role of the manager. Goodpaster, for instance, argues (plausibly) that managers have a fiduciary obligation toward shareholders, yet non-fiduciary obligations toward other "stakeholder" groups. However, when pressed to identify the source of these non-fiduciary obligations, he denies that they arise from the managerial role itself and suggests (implausibly) that they are simply a reflection of moral constraints that the principal is subject to. Thus, he argues that "the conscience of the corporation is a logical and moral extension of the consciences of its principals" (1991: 68). He criticizes the "invisible hand" view for suggesting that the agent has " 'moral immunity' from the basic obligations that would apply to any human being toward other members of the community"(1991: 68).

The problem with this analysis is that the competitive structure of the marketplace, insofar as it demands certain types of non-cooperative behavior, does in fact offer agents limited "moral immunity" from the norms of everyday morality. Managers are expected to be tough negotiators, to act strategically in the interests of the firm, to fire unproductive employees, to refrain from nepotistic practices, etc. Similarly, investors are entitled to withdraw their money from an unprofitable firm, regardless of the broader "social consequences" of their doing so. (This is essential to maintaining the "hard budget constraint" under which the private sector generally operates, with salutary consequences for the economy as a whole.) Thus, the moral constraints that the manager faces when dealing with various "stakeholders" are not merely the constraints of everyday morality, inherited from the firm's principals. There are a number of *sui generis* constraints that arise out of the managerial role, that are specific to the context of a competitive market economy. Indeed, their primary function is to specify the permissible means by which this competition can be pursued.

Take the example of advertising. Almost all advertising is false advertising by the standards of everyday morality. But from the standpoint of business ethics, this is neither here nor there. What is morally significant, with respect to the role-specific obligations of the manager, is that advertising has the potential to exacerbate information asymmetries in the market. Insofar as these information asymmetries undermine efficiency, such advertising runs contrary to the intended consequences of marketplace competition. In other words, it threatens to generate unhealthy forms of competition. The standard response on the part of the state has been to institute a set of "truth in advertising" laws, to prohibit advertising that makes deceptive claims, claims that are likely to mislead the consumer "in material respect" (J. W. Coleman 1989: 16). Yet there are many cases in which claims can be made that are misleading, and yet not strictly speaking false (e.g., food that is advertised as "now fat free" even though the product in question had never contained fat), or that are false without being materially misleading. These types of marketing claims are difficult, if not impossible, to exclude through regulation. But insofar as this sort of advertising works only by exploiting a market imperfection, in this case an information asymmetry, it is unethical. It remains legal only because it would be too costly or cumbersome to eliminate through regulation (or in some cases, simply because legislators have not yet gotten around to prohibiting it).

A similar situation arises when firms are given the opportunity to externalize costs (whether it be in the form of pollution, congestion, threats to safety, etc.). The presence of a pollution externality, for instance, means that the firm will be able to charge prices that are "too low," relative to the true social cost of producing the good. Rather than actually reducing the cost structure of its operations, the firm is simply displacing these costs onto others through an extra-market mechanism. As a result, an excessive quantity of resources will tend to flow to employments that generate negative externalities, while too little will flow to the production of goods that generate positive externalities. Even worse, when one firm takes advantage of the opportunity to externalize some of its costs of production (e.g., by "cutting corners"), it puts competitive pressure on all rival firms to follow suit. Thus, the exploitation of market failures can quickly transform the "race to the top" of the competitive market into a "race to the bottom."

The central ideal of an adversarial ethic for business should be the preservation of healthy competition, even when the law fails to offer sufficient guarantees. Looking at the specific ways in which markets can fail to promote healthy economic rivalry, and considering the analogy with the ethics of sport, we can suggest the following as a set of general conceptual templates for thinking about the conduct of business with respect to market transactions.

> *Do not exploit market failure:* This is the form that the principle of *constrained competitiveness* takes in an economic context. As Applbaum has observed, many books on competitive strategy are essentially "how-to" guides for creating and profiting from market failures (1999: 194–195). Taking advantage of externalities, information asymmetries, and market power represent the primary forms of unethical conduct in this regard (for more detail, see Heath 2004a: 84). The "Pareto conditions" that define the structure of a perfectly competitive market provide the chief guidelines for determining what counts as a market failure (see Schultz 2001: 99–104), although it is important to note that these are only guidelines. Managers themselves, for instance, are usually best placed to determine whether a particular competitive strategy generates gains for the firm by a genuine lowering of costs, or rather by an uncompensated displacement of costs.

Do not cheat: In many cases, efforts on the part of the state to correct market failure generate bodies of regulation that are unenforceable (J. W. Coleman 1989: 185–194). Other times, the penalties associated with violation of the law are so minor, relative to the gains that might be achieved, that the desire to maximize profits winds up favoring violation (Braithwaite 1981). Nevertheless, managers are morally obliged to respect both the letter and the spirit of the law, regardless of the fact that, from a cost–benefit perspective, it is in their interest to cheat. Otherwise put, the regulatory environment in which businesses operate should be regarded as a system of moral constraints and not merely as a set of incentives.

Do not game the rules: Any complex system of rules—such as a body of government regulation—will tend to have loopholes. There may have been oversights in the way that the rules were formulated, or the rules may simply interact with one another in unintended ways, generating potential "exploits." Clever people in business sometimes amuse themselves by searching for such exploits, in order to give their firm a competitive edge.[10] The problem of "creative" accounting typically falls into this category as well (Blake et al. 1998: 25). Such "gamesmanship" is unethical.

Take the high road: One of the major problems with approaches to business ethics that ignore the adversarial nature of market relations is that they also tend to ignore the single most important excuse for unethical conduct in business. In a non-adversarial context, the fact that one person acts unethically does not in itself create any additional pressure on others to do so. For example, if one surgeon performs some unnecessary procedures, it does not necessarily give other surgeons a reason to do so. In a competition, however, the fact that one person is deriving an advantage from unethical conduct necessarily generates a disadvantage for everyone else, and therefore creates pressure for everyone to follow suit. Once one athlete starts taking steroids, it is very difficult for the others to stand by and do nothing. Acting ethically, in this context, means losing the competition. In an economic context, the consequences of "losing" can be quite severe. Of course, the mere fact that one is embroiled in a competition does not give one *carte blanche* to do anything whatsoever, just because the other person "started it." One's ethical obligation is always to take the high road and refrain from adopting any unhealthy competitive strategies. Nevertheless, it is important for business ethicists

[10] Consider, for example, the actions of Enron traders gaming the California electricity market. See McLean and Elkind 2003: 264–283. An internal review of the practice generated the now-famous legal counsel that "this strategy appears not to present any problems, other than a public-relations risk arising from the fact that such exports may have contributed to California's declaration of a Stage 2 Emergency yesterday" (2003: 277).

> to recognize that managers, because of the competitive structure of the market economy, are systematically subjected to external pressure to engage in unethical conduct in a way that doctors, for example, are not. While these competitive conditions do not make it permissible to violate ethical constraints, they may provide a legitimate excuse for doing so (Austin 1979; Baron 2005).

There is one more general imperative that should be mentioned, which does not have a precise analog in sport. One of the more troubling features of the way businesses conduct themselves in the public sphere is that they consistently lobby against regulations that are designed to correct market imperfections (Baumol 1974). For example, the petroleum industry fought vociferously against the ban on leaded gasoline, just as American automakers lobbied against mandatory seat belts, safety glass, catalytic converters, fuel economy standards, etc. This is, in a sense, doubly unethical—not only did these firms exploit market failures, but they dedicated considerable resources to entrenching these failures (even when there was only a marginal business case to be made for doing so). Thus, the fifth imperative might be, "Don't oppose rule changes that have as their goal the correction of a market failure."

Warren Fraleigh, in *Right Actions in Sport*, defines the "good sports contest" as "one in which the personal intended ends of actions are congruent with or consistent with the purpose of the sports contest" (1984: 49). The central claim here is somewhat subtle: the participants need not actually intend the larger purpose, but their intentions must be consistent with it. The same can be said with regard to competitive strategies in business. Managers need not intend the greater social good; they may adopt competitive strategies with an eye only toward the maximization of profit. However, the strategies that they adopt in order to obtain profit must be consistent with the greater social good that serves as the "purpose" of the market economy, namely, efficiency in the production and allocation of goods and services. The imperatives outlined above represent an attempt to articulate the type of constraints that this sort of consistency imposes.

Naturally, the task of taking these very general conceptual templates and developing from them a set of more concrete moral norms exceeds the scope of this paper. I have sought to provide only a few suggestions. My primary goal has been to show that an adversarial approach to the ethics of market transactions—and, in particular, an approach that preserves the old-fashioned idea that managers bear fiduciary obligations only to the owners of a firm—need not exhibit any moral laxity, or provide an excuse for corporate misconduct. It should be obvious that the imperatives outlined above are extremely demanding, so much so that competitive pressures would probably prevent any corporation from respecting all of them in the near term. Thus, the adversarial approach presents an ethical ideal. The important point is that this ethical ideal is one that is consistent with the economic ideal of the free market, and thus, far from being antithetical to the spirit of capitalism, can rightly claim to be articulating its true essence.

4.5. Conclusion

There is a reason why Sun-Tzu's *The Art of War* is a popular read among management and law students, but not among medical interns and engineers. The former are both preparing for professional roles within institutions that have important adversarial features, while the latter are not. Unfortunately, among management students, reading *The Art of War* is far too often seen as an alternative to the study of business ethics, one that offers more "realistic" advice for dealing with the challenges that will arise in the corporate world. In part, this is the fault of business ethicists, for having systematically failed to acknowledge the adversarial structure of the market economy. In their effort to stave off facile appeals to the "invisible hand," and to condemn the moral laxity that such appeals usually encourage, too many have chosen to deny the reality of competition, or to resist the suggestion that this competition offers individuals "immunity" from any of the norms of everyday morality. In so doing, they have failed to articulate the implicit morality of the market (or the implicit logic of corporate law), which is organized around the goal of promoting healthy over unhealthy forms of competition.

This has had a number of unfortunate consequences. First and foremost, it has encouraged the idea that when the market is producing bad outcomes, the way to improve it is to change the objectives of the participants. According to this view, corporations do bad things because they are too greedy in their pursuit of profit, so the way to correct this problem is for them to be less greedy, or to pursue other objectives besides profit. The adversarial perspective, by contrast, displaces attention from the objectives of the participants to the rules that structure the interaction. It suggests that rather than demonizing profit, ethicists should be encouraging firms to respect the "spirit" of the regulatory structure that governs marketplace competition. People who get hung up on the unethical nature of profit are essentially allowing the *pro tanto* immortality of a competitive strategy to obscure the overall point of the institution. In this respect, they are like those who condemn lawyers for "defending rapists and murderers" without looking at the role that a vigorous defense plays in an adversarial trial procedure.

The second unfortunate effect of the failure to acknowledge the adversarial structure of market transactions has been an inability to counter the widespread perception that business ethics is too "touchy-feely" to be of any use in the hard-nosed world of business. The adversarial approach to business ethics outlined here, by contrast, is able to distinguish between "playing hardball"—hard bargaining, nickel-and-diming, aggressive pricing, etc.—all permissible in a market context, and "sharp practices" or "dirty pool"—deception, cost externalization, creative accounting, etc.—which exploit market imperfections, and thus violate the spirit, if not the letter, of the rules under which marketplace competition is conducted. Business ethics, according to this conception, is not an alternative to *The Art of War*; it is more like a Geneva Convention or a code of honor, a pact aimed at guarding against the almost universal tendency of competitive interaction, when left unsupervised, to degenerate into a race to the bottom.

5 }

Business Ethics and the "End of History" in Corporate Law

Henry Hansmann and Reiner Kraakman have argued that, after more than a century of extensive experimentation and debate, the "end of history" has now been reached in corporate law (Hansmann and Kraakman 2003). In the same way that the combination of a market economy and a democratic welfare state has emerged as the only attractive way of organizing society at the national level, the shareholder-owned, profit-oriented business corporation has emerged as the standard institutional arrangement for organizing production and investment. Underlying this increasingly standard arrangement is what they describe as a "widespread normative consensus that corporate managers should act exclusively in the economic interests of shareholders" (Hansmann and Kraakman 2000). Although intended primarily as an empirical observation, Hansmann himself has made a not-inconsiderable contribution to this emerging consensus through his work on corporate ownership—presented most systematically in his book, *The Ownership of Enterprise* (2000). The centerpiece of Hansmann's view is the claim that ownership of the firm is most naturally exercised by the group able to achieve the lowest costs of ownership, and that homogeneity of interest within the ownership group is a crucial factor in achieving lower costs. He defends this claim through a study of cooperatives, attempting to show that homogeneity is the source of the competitive advantage most often enjoyed by lenders over other constituency groups, such as workers, suppliers, and customers, when it comes to exercising control over the firm (and, in particular, the firm's managers).

This argument has a number of important implications for business ethics. After all, the majority of business ethicists have spent the past three decades swimming against what, according to Hansmann, is the current of history. While shareholder primacy may be the default view in corporate law, business ethicists have resolutely resisted this doctrine, arguing against the idea that shareholders should have any sort of privileged position within the firm. The most mainstream expression of this can be found in multi-fiduciary stakeholder theory (Goodpaster 1991: 60), which argues that members of the board of directors of a firm should act as agents for a variety of different "stakeholder" groups—not just shareholders— and that corporate law should be adjusted, where necessary, in order to accommodate or even enforce these obligations (Greenfield 2006). Hansmann's argument is particularly devastating to this doctrine, because of the emphasis that he puts on homogeneity of interest as the central characteristic that makes a group suitable for assuming ownership. If the failure of most cooperatives is due to the inner tensions and rivalries that exist amongst workers, customers, or suppliers, taken singly, then trying to put together a coalition between two or more of these groups, in order to exercise ownership, is going to be a non-starter. Yet that is essentially what stakeholder theorists are proposing.[1]

While I am largely sympathetic to this critique, my objective in this paper is not to provide any detailed argument in support of it. The pros and cons of stakeholder theory have been extensively discussed elsewhere.[2] Instead, I would like to consider what consequences Hansmann's argument would have for business ethics if its central empirical claim were correct—that the reason for the prevalence of the standard shareholder-owned firm is that it minimizes ownership costs. Some business ethicists—most notably Boatright (2002)—have argued that this amounts to a vindication of Milton Friedman's view that the only "social responsibility" of business is to increase its profits within the framework established by law. Non-shareholder groups, according to this perspective, have essentially chosen the protections offered by contract over the advantages that they could have obtained by assuming ownership (such as the open-ended "duty of loyalty" that is part of the fiduciary responsibility of senior managers). Because of this, they cannot reasonably turn around and expect managers to show any special concern for their interests, above and beyond that required by law and contract. To do so would be to expect some of the benefits of ownership without assuming any of the costs.

I would like to suggest that this conclusion does not follow, and that the "Hansmann argument" lends itself to a less minimalist view of business ethics, what I refer to as a "market failures" approach. The issue turns on what, precisely, the "least-cost assignment of ownership" amounts to. Hansmann argues that the patron group that assumes ownership will be the one with the lowest (net) costs of ownership, not just for itself, but for all patron groups. This is equivalent to saying that the outcome of the contest for ownership among patron groups will be Kaldor–Hicks efficient.[3] It is this feature of Hansmann's

[1] The problem for stakeholder theory is sufficiently evident that Hansmann dedicates only two paragraphs of *The Ownership of Enterprise* to it (2000: 44). There is a more ample discussion in Hansmann and Kraakman (2003).

[2] For my own contribution, see Heath 2006a, 2007.

[3] Hansmann writes: "I use the term 'cost-minimizing' here to mean 'efficient' in the economist's very broad sense of that word—that is, to refer to a situation in which there is no alternative arrangement

argument that encourages Boatright to argue that workers themselves are better off handing the firm over to shareholders to run, and thus that they have (implicitly) chosen law and contract over managerial obligation as a way of protecting their interests.[4] Shareholder primacy, Boatright says, "best serves the interest of all stakeholder groups" (2006: 115).

One could quibble about whether Kaldor–Hicks efficiency is sufficient to motivate this conclusion (it seems to require Pareto efficiency). The point is moot, however, because Hansmann's own argument in favor of the view that free contracting will generate Kaldor–Hicks efficient ownership arrangements is problematic. As I will attempt to show, his claim would be correct if the contest for ownership were a cooperative game, with the final outcome falling within the core (or to put it in plainer terms, if the patron group that eventually assumed ownership had to outbid, not just each of the other groups taken singly, but every possible coalition of those groups as well, with their shared resources). His own argument against stakeholder theory, however, suggests that cooperation between patron groups will have its own costs (and that these will typically be quite high). Thus one cannot realistically posit coalitions. This suggests that the contest for ownership should instead be modeled as a competitive game between patron groups, with the outcome being an ordinary Nash equilibrium. If this is the case, then there is no reason to think that the outcome will be Kaldor–Hicks efficient. As a result, it is incorrect to look at prevailing ownership arrangements with the presumption that they are advantageous for all parties, or that they minimize total costs of contracting. This means that even if there is a strong case for shareholder primacy in the legal structure of the corporation, it would be a mistake to assume that this resolves all of the moral concerns that have traditionally animated stakeholder theory. It is possible for patron groups to get stuck in arrangements where law and contract alone put them in an extremely disadvantaged relationship with the firm, but where for largely fortuitous reasons they are also unable to assume ownership, or to participate effectively in shared ownership. In these situations, one can make the case, from a social welfare perspective, for managerial restraint (e.g., forbearance from sharp practices), even though the affected party has not contracted for any fiduciary duty.

---

that could make any class of patrons better off, by their own subjective valuation, without making some other class worse off to a greater degree" (2000: 23).

[4] Boatright writes, "Whether a corporation is owned by investors, employees, customers, suppliers, or some other constituency is determined by the costs and benefits of ownership as reflected in the market choices made by each group. The standard argument holds that whatever the assignment of ownership, the resulting system of corporate governance is optimal not only for the constituency with control and a claim on the residual but also for all other constituencies" (2002: 48).

5.1. Normative Foundations

The convergence thesis articulated by Hansmann and Kraakman is based on an analysis of developments in the legal structure of the corporation (Kraakman et al. 2004). There is, however, a deeper level at which convergence has occurred. During the high water mark of debates between stakeholder theory and shareholder primacy, proponents of the different positions disagreed, not just about the ideal legal form of the corporation but also about the more fundamental normative standards that should be used to evaluate the law. In particular, shareholder primacy theorists often appealed to various rights doctrines (such as libertarianism or contract theory narrowly construed), while stakeholder theorists appealed to broader moral views, which typically incorporated a general concern for social welfare. What is striking about Hansmann's defense of shareholder primacy is that it takes as its point of departure the same normative presuppositions that have traditionally motivated stakeholder theorists. The argument does not rest on any controversial doctrines about property or freedom of contract, but instead, on a general concern to do what is "best for society." Thus, it constitutes a version of what Boatright calls the "public policy" argument for shareholder primacy (2000: 57; Moore 1999: 121). Because of this, it would be mistaken to assume that the "end of history" has come about simply through a capitulation of stakeholder theory to the doctrine of shareholder primacy. There has also been a significant change in the way that the latter doctrine is understood and defended. In order to understand the force of Hansmann's argument, it is necessary to situate it first with respect to this evolving tradition.

The question of whose interests management should assign priority to acquired significance with the emergence of the large, publicly traded corporation in the early twentieth century, along with the "separation of ownership and control" that went along with it. As shareholders became more anonymous and further removed from the day-to-day operations of the firm, it became increasingly unclear why managers should assign special priority (or "primacy") to their interests (Berle and Means 1932: 309–312). Customers, suppliers, and employees also make important contributions to the success of the firm. Why should managers focus on maximizing shareholder value, when they could instead provide employees a wage premium, or offer customers below-market prices, or guarantee a stable market to suppliers? Furthermore, the experience of the Great Depression, followed by the effective national mobilization of industry that occurred during the Second World War, suggested to many that corporations had responsibilities to "society" that extended beyond merely the interests of their owners.

The naïve defense against these sorts of suggestions involves an appeal to the *property rights* of shareholders. According to this view, just as people own houses and cars, and are entitled to the benefits that arise from them, shareholders are the owners of the firm, and as a result, are entitled to the benefits (or the income that arises from it) (Hart 1995). The problem with this claim, as first-year law students are invariably told, is that "property" is not a unitary right, but is typically analyzed as a bundle of rights, which can easily be disaggregated (Honoré 1961). In fact, ownership of the corporation provides a clear-cut example of how such rights can be disaggregated (or "unbundled"), simply because shareholders in publicly traded firms lack most of the rights typically associated with property ownership. For example, they lack the right of possession or use of the firm's assets. They cannot simply walk onto the premises whenever they like, they cannot issue direct orders to managers, they cannot withdraw their investment, and they enjoy limited liability. Indeed, there are many features of the modern corporation that are intended to create something of a buffer between the day-to-day assets and operations of the firm and its shareholders. The "business judgment rule," for instance, gives management extraordinary discretion when it comes to many aspects of the firm's operations (Easterbrook and Fischel 1991: 93). Thus, one cannot simply point to the property rights of shareholders as the basis for the claim that managers should serve their interests—this may be one of the rights that was unbundled and removed when the separation of ownership and control was established.[5]

The central feature of the property argument is that it looks for some characteristic that shareholders possess, which other constituency groups do not possess, that could serve as the basis of an entitlement to the loyalty of management. If the property relation fails to license such an entitlement, one might be inclined to look around for some other characteristic that shareholders have that could serve as a more plausible basis. Thus, theorists have pointed to a variety of distinguishing features of shareholders, such as being residual claimants, or having a greater exposure to risk, or being in a position of special vulnerability vis-à-vis management, as the basis for their special status within the firm. I refer to these as "moralizing arguments," because they claim that shareholder primacy is merely an acknowledgment and recognition of an independently given moral fact. Marcoux (2003), for instance, argues that shareholders are distinguished by the position of extreme vulnerability that they stand in with respect to management. The fiduciary relation between managers and owners, according to this view, is strictly analogous to that between a doctor and a patient (Marcoux 2003: 6–8). In both cases, the information asymmetries are so great that the principal is unable to tell if his agent has conducted herself in a proper fashion. Since it is not reasonable to expect

[5] Kent Greenfield summarizes the argument as follows: "Shareholders do not have a complete 'bundle of rights' to make them 'owners' in the traditional sense, nor are they owners in any other way that would distinguish their contribution to the firm from the contributions of other stakeholders" (2006: 126). The force of this argument, he argues, is the reason that "no prominent contemporary corporate law scholar uses property rights as the primary rationale for shareholder dominance" (2006: 47).

the information asymmetry to disappear, the courts impose a set of fiduciary responsibilities, including a "duty of care" and a "duty of loyalty." The reason that shareholders are the beneficiaries of these duties is that they, unlike other constituency groups, have no explicit contract to protect their interests. They are residual claimants—they get what is left over after the firm's other contractual obligations have been met (i.e., after wages have been paid, lenders have been repaid, etc.) and are therefore the most vulnerable.

The problem with these moralizing arguments is that they fail to explain the rather sharp distinction that is drawn between shareholders and other constituency groups. If one wants to decide questions of managerial obligation by looking at which groups are most vulnerable in their dealings with the firm, the answer is going to be rather variegated and likely to change over time. Employees, for instance, have often dedicated years of their lives to the firm, and in many cases have made what can best be described as an asset-specific human capital investment (e.g., acquiring knowledge of procedures and operations that make them more productive employees, but that have no market value to any other firm) (Blair 1999). They have thereby made their employer a monopsonist for their particular labor skills, giving the corporation a degree of market power over them. Furthermore, the employee's entire livelihood often depends upon the relationship with the firm. Shareholders, on the other hand, often have very little invested in the firm, and usually have diversified holdings, so that in many cases the overall impact upon them of managerial decisions is rather slight. Furthermore, it is extremely easy for investors to sell their shares if they do not like the way management is behaving—which in turn depresses the share price, making the firm vulnerable to hostile takeover (Freeman and Evan 1990: 340–342; Boatright 1994: 396). Thus, it is far from obvious that shareholders are uniquely vulnerable in their relations with the firm, such that they should be the exclusive beneficiaries of managerial loyalty.

Frustration over the inconclusiveness of these arguments has led many theorists in recent years toward a more contractual view (Maitland 1994). While shareholders may not be automatically entitled to the loyalty of management by virtue of any intrinsic characteristic they possess, it is nevertheless they who have contracted for managerial services. For example, it is often maintained that shareholders are the ones who have hired managers to do a particular job, namely, to maximize shareholder value, and so managers are obliged, as professionals, to carry out this task with dedication and loyalty. There are two slightly different versions of this contractual account of managerial obligation. The first maintains that managers are hired by shareholders, and so are effectively their agents. The second points rather to the terms under which shareholders were persuaded to invest in the firm. There is nothing to stop a firm from adopting a corporate charter that specifies something other than profit as the central objective of the organization—indeed, firms such as the New York Times Company explicitly do so—then attempting to attract investors (Easterbrook and Fischel 1991: 36). Investors who purchase such shares cannot then complain that management is serving other constituencies, having entered into the transaction with full knowledge that they intended to do so. In the standard case, however, corporate charters do not place any objectives above profit-maximization, and so investors enter into the transaction with the expectation that their interests will be assigned priority.

There are problems, however, with both versions of this argument. The central issue is that the "contracts" in these cases are not explicit and have been compromised by courts in myriad ways. The typical corporate charter, for instance, does not mention profit-maximization, or even returns to investors, as an explicit objective (Stout 2008: 169). Furthermore, court decisions such as *Dodge v. Ford*, which interpret the obligations toward shareholders quite strictly, are distinct outliers and have largely been rendered obsolete through the application of the business judgment rule (Stout 2008: 166; Lee 2006: 10). The overwhelming trend has been to grant managers considerable discretion in deciding how—and even to what extent—they go about advancing the interests of shareholders (Marens and Wicks 1999). As far as the agency model is concerned, managers do not sign any agreement with shareholders. Their contract is with the corporation, and courts treat them as having a fiduciary obligation toward the firm (Robé 2011). Thus, "the relationship between shareholders and directors/managers," as Eric Orts puts it, "is not a direct agency, but rather a special form of quasi-agency governed by an overlay of state corporate law" (1998: 311).

Proponents of the contractual view treat these sorts of legal points as mere technicalities, masking the substance of the underlying relations (thus it is common to speak of managers as having a fiduciary relationship to shareholders, even though technically they do not). Yet the argument that purports to render explicit the underlying promise or contract typically presupposes the desirability, on independent grounds, of some form of shareholder primacy. Those who are opposed to shareholder primacy argue that the technicalities in question are not mere technicalities, but are barriers designed with the explicit objective of impairing the ability of shareholders to impose their interests (Blair and Stout 1999). They point to the fact that the parties are prohibited from contracting around the various provisions of corporate law (regarding, for example, the governance structure of the firm), as evidence that these are not merely contractual provisions, which the parties themselves might have arrived at, but represent the imposition of an independent set of considerations, reflecting a broader social interest (Clark 1985).

Thus the contractual view is problematic, because there are almost never any explicit contracts that commit managers to maximizing shareholder value. While something like shareholder primacy is clearly implicit in all these transactions, and is certainly the expectation that most parties bring to the table, the extent to which the firm can or should permissibly deviate from it is an open question (Boatright 1994: 397–398). Furthermore, for those willing to countenance legal reform, it is always an option to argue that existing constraints on shareholder primacy should be tightened, in order to strengthen the hand of "other constituencies." After all, the law is very far from being libertarian, even in areas as private as corporate law. Decades of labor, civil rights, consumer protection, and even contract law have effectively eliminated the presumption that contracts will be respected and enforced merely because they have been entered into by competent adults. As Michael Jensen and William Meckling point out, "new laws as well as court decisions often can and do change the rights of contracting parties ex post" (1976: 311). So even if it were true that shareholders and managers had some kind of an agreement, which involved assigning priority to the interests of shareholders over other constituency groups, this does not really settle the question—from either a legal or a moral point of view—whether this agreement should retain its authority in cases where it conflicts with the claims or interests of other groups.

As a result, one can see a very noticeable slide in the arguments made by "contractualists" away from a libertarian toward an essentially welfarist normative standard (see Maitland 1994). This is apparent in Frank Easterbrook and Daniel Fischel's enormously influential book, *The Economic Structure of Corporate Law*, which makes rhetorical appeal to the voluntariness of contract as though this were the central normative criterion, but at crucial points shifts the argument over to the efficiency properties of legal regimes that respect voluntariness (e.g., 1991: 36–39). The claim is not that the contracts have intrinsic authority, but that it is best for everyone, overall and in the long run, if these contracts are respected and enforced. The latter argument can be referred to, following Boatright, as the "public policy" argument for shareholder primacy. "Put simply, the argument is that institutions in which management is accountable primarily to shareholders provide the most socially beneficial system of economic organization" (Boatright 1994: 401).[6] After more than a century of tinkering with different models of corporate governance (public ownership, cooperatives, co-determination, managerialism, etc.) and market structure, a particular constellation has emerged that seems to be better than most. It features regulated competitive markets, shareholder-oriented firms, and an active market for corporate control. There is nothing intrinsically important about shareholders, or peculiar to their relationship with managers, that speaks in favor of assigning priority to their interests. They just seem to be better at running companies—or holding managers to account for the way that they run companies—than any other constituency group, or coalition of constituency groups. For example, one of the advantages of shareholder primacy is that it

[6] It follows, he says, that "the shareholder–management relation is not 'ethically different' for any reason that is unique to that relation" (Boatright 1994: 403).

gives the firm's managers a single, reasonably coherent objective, one that is also reasonably easy to observe and quantify. (One need only consider the problems with so-called triple-bottom line accounting to see the enormous advantages that come from having a single bottom line [Norman and MacDonald 2004].) As a result, it is much more difficult for managers to make excuses for poor performance when they are being held to account by shareholders (Jensen 2002). Similarly, the orientation of most shareholders toward maximization of returns, combined with an active stock market in which they can unload their shares, combine to impose a "hard budget constraint" on firms, which in turn has salutary effects on the motivation of managers (not to mention employees throughout the firm) (Kornai 1992: 143–144). So according to proponents of this argument, while shareholder primacy may have some deleterious side-effects (as critics of the "profit motive" have amply detailed), it tends on balance to be better than the alternative.

5.2. The Hansmann Argument

Among theorists who are broadly sympathetic to the norm of shareholder primacy, the problem with the arguments outlined above is not that any of them are wrong, but simply that any one, taken singly, is insufficient to serve as an adequate defense of that norm. Clearly there is some sense in which shareholders are the owners of a standard business corporation; their right to elect the board of directors that oversees the operations of the firm means that managers work for them, at least in some attenuated sense of the term; the fact that they are residual claimants is a significant difference between them and other constituency groups; and a well-managed, profit-oriented firm, operating in a properly regulated, reasonably competitive market, creates important benefits for all of society, not just for its owners. There is an element of truth in all of this. What makes Hansmann's argument in support of the shareholder primacy doctrine particularly forceful is the way that he combines all four of these observations in a simple, compelling way.

Overall, Hansmann's argument can be considered a version of the "public policy" argument, because his primary normative criterion for the evaluation of corporate governance structures is social welfare maximization. In *The Anatomy of Corporate Law*, he and Reiner Kraakman state this quite explicitly:

> As a normative matter, the overall objective of corporate law—as of any branch of law—is presumably to serve the *interests of society as a whole*. More particularly, the appropriate goal of corporate law is to advance the *aggregate welfare* of a firm's shareholders, employees, suppliers, and customers without undue sacrifice—and, if possible, with benefit—to third parties such as local communities and the beneficiaries of the natural

> environment. This is what economists would characterize as the pursuit of *overall social efficiency*. (2004: 18)

The most original aspect of Hansmann's analysis is his observation that the "ownership of enterprise" is in fact much more dynamic than most people assume. It is conventional to distinguish four major constituencies, or "patron groups" of the firm, according to the contribution that they make: workers provide labor, suppliers provide inputs, lenders provide capital, and customers consume outputs (and thereby provide revenue). In the case of a nonprofit enterprise, all of the contributions are acquired on contractual terms. In particular, employees provide labor on fixed terms (e.g., an hourly wage or salary) and banks provide capital on fixed terms (e.g., the interest rate charged on a loan). There are no residual claimants, which means that any surplus generated is simply reinvested in the organization. There is no obstacle, in principle, to having the entire economy organized this way. In practice, however, there are advantages to having one of the constituency groups acquire the residual claim—the primary one being that it allows the firm to acquire the relevant input on more flexible terms.

Hansmann's great insight is that the standard business corporation and the cooperative are simply variations on the same blueprint—one in which a particular constituency group surrenders its contractual entitlements in return for a residual claim. He gives the example of a dairy cooperative, which is a firm that is owned by the suppliers of its primary input, milk (Hansmann 2000: 13). In a typical dairy cooperative, farmers agree to sell their milk to the cooperative at a below-market price and receive in return the right to a profit-share, which is disbursed at the end of the year. A mutual insurance company functions in a similar way, except that it is owned by its customers (the policy holders) rather than its suppliers. In a cooperative insurance scheme, policy holders typically pay an above-market rate up front for their policy, in return for a substantial rebate at the end of the year. The legal form of the standard business corporation conceals what is, in effect, a very similar structure. A public corporation is essentially a lender's cooperative. Investors provide what amounts to a loan to the company at a below-market price (i.e., an interest rate of zero), in return for a profit-share, disbursed on a periodic basis.

Naturally, accepting such a residual claim entails a degree of risk, and so constituency groups that put themselves in such a position typically do so only when they are also given "formal control" of the firm, including control of the process through which managers are hired, fired, and compensated. Part of the bargain involves the understanding that management works for the owners and will make a good-faith effort to maximize the value of the residual claim. It is these two elements, formal control and the residual claim, that form the complex that Hansmann refers to as "ownership" (2000: 11). It can be used to characterize the structure of a cooperative just as well as it can a business corporation. Indeed, it is in no way a foregone conclusion that lenders will be the ones to assume ownership. Furthermore, it is often open to one constituency group to buy some other group out, thereby effecting a transfer of ownership. Whether anyone has an interest in doing so will depend upon the relative merits of having a contractual relationship and having an ownership relationship with the firm.

Consider, for example, the conversion of a rental apartment building into a condominium. As Hansmann observes, a condominium, despite being a *sui generis* legal form, is essentially a customer cooperative—the consumers of rental housing band together to acquire ownership of the property (2000: 195–226). One can see the pros and cons of this arrangement by considering the choice that individuals face when deciding whether to rent an apartment or purchase a condominium. Renting involves entering into a contractual relationship with a corporation owned by lenders—those who have invested the capital. This has various disadvantages. Foremost among them is the fact that the building manager works for the landlord, and so may be relatively unresponsive to the needs and complaints of tenants. There is also the fact that the expenses involved in moving—not just in terms of money but also time and disruption—creates a certain amount of "lock in," which gives the landlord a degree of market power over tenants (which may be used, on some occasions, to impose extortionate rents on a relatively captive group of consumers). These can be considered costs associated with having a contractual relationship with the corporation that provides one's apartment. There are, however, certain advantages, which can be seen by comparing the situation of the tenant to that of the condominium owner. Having a contractual relation typically has less upside but also a lot less downside. The tenant has no stake in the long-term value of the property and so is spared the trouble of worrying about maintenance and upkeep. He is insulated from all the risks associated with property ownership. Condo fees can go up quite unexpectedly if major repairs need to be done. Rent stays fixed for the length of a lease, and if it goes up after that, exit is relatively easy—whereas with a condominium the only way to get out is to find someone else who wants in. Furthermore, while the employees of a condominium association are undoubtedly more responsive to the needs of residents, residents also acquire the responsibility for overseeing management of the building, and they are the ones who suffer the consequences of mismanagement. These are some of the costs of ownership.

When a person is trying to decide whether to rent or own, what she typically does is compare the costs of having a relatively fixed contract (renting) to the costs of ownership. Hansmann's observation is that the constituencies of a firm are essentially making the same decision, and it is not uncommon for ownership to shift from one group to another, as is the case with "condo conversion," whereby the customers take over from the lenders, or else demutualization in the insurance industry, whereby lenders take over from customers. In both cases, one constituency group is essentially buying the other one out, because its members calculate that they would be better off substituting an ownership relation with the firm for a contractual one in their own case, and substituting a contractual relation for an ownership relation in the case of the outgoing ownership group (e.g., substituting bank loans for equity capital).

Given this analysis, the question then becomes, why is shareholder ownership such a dominant organizational structure? After all, cooperatives enjoy rather dramatic tax advantages in most jurisdictions, since they are able to treat residual earnings that are disbursed (such as the annual "rebate" offered by cooperative insurance schemes) as a business expense, whereas shareholder-owned firms must pay tax on all residual earnings (or "profits"). Furthermore, as Hansmann observes, cooperatives cannot be subject to insuperable legal obstacles, since they are able to compete effectively with standard business corporations in certain sectors (2000: 157–158). Dairy coops and mutual insurance companies are two examples, already mentioned. Beyond that, many travel agencies are worker coops. Partnerships are also, in many cases, essentially worker coops, and they do very well in certain areas such as law firms. There are also instances of successful customer-owned firms, including coop banks, electricity generation facilities, agricultural supply stores, as well as certain retail chains. Thus it cannot be the case that the deck is stacked against cooperatives in some deep, structural sense—as is sometimes argued (Miller 1981)—because cooperatives do well in certain sectors.

Hansmann identifies a range of distinct costs of ownership, which each constituency group must compare to its own costs of market contracting in order to determine how much it is willing to bid for the firm. For example, exposure to certain types of market risk is one of the costs of ownership, since the residual claim leaves the owners more fully exposed than they would be under a contract with fixed terms: a stock is riskier than a bond, a profit-share is riskier than a salary, etc. The greater willingness of investors to bear such risks, compared to workers, is sometimes presented as *the* reason for shareholder primacy. In Hansmann's view, it is just one of the factors that goes into a larger calculation, which each constituency group must undertake in order to decide whether to vie for control of the firm. Furthermore, the mere fact that one group is exposed to greater risks than some other does not, in itself, warrant any sort of claim upon the loyalties of management. The exposure merely increases that group's incentive to acquire ownership of the firm, and thereby to secure the loyalty of management.

The factor that Hansmann draws particular attention to, which traditional analyses have tended to ignore, is the political costs of ownership, which includes both the "costs of collective decision making," as well as the agency costs associated with inadequacies in such mechanisms. The owners are in a position to give managers instructions on how to run the firm, but before doing so, they must themselves come to a decision about how the firm is to be run. If the process for doing so is costly and time-consuming, or unstable (such that directives change from month to month), subject to manipulation, or produces outcomes too detailed to permit a clean separation of ownership and control, then this will diminish the benefits of ownership. Generally speaking, the more conflict there is within the ownership group, the more divergent their interests, the greater the costs of collective decision making. Hansmann argues, therefore, that homogeneity of interest is one of the major factors determining the costs of ownership. He observes that cooperatives are most successful when the members of the ownership class are all similarly situated vis-à-vis the firm. A dairy cooperative works because the one input, milk, is an undifferentiated commodity, and as a result, there are few potential conflicts of interest between the owners. If the cooperative starts using eggs as well and brings in a group of poultry farmers as members, it immediately creates a host of potential problems. How much to pay for milk versus eggs, how to determine profit shares, how to allocate voting rights? Furthermore, all sorts of investment decisions that might once have been uncontroversial have the potential to become politicized. What if the coop wants to invest in new cheese-making equipment, which stands to benefit the dairy but not the poultry operators? Because of this, cooperatives will often strive to keep the ownership class homogeneous, by acquiring other inputs through market contracting, rather than bringing in new classes of members.

The same consideration applies, with even greater force, in the case of worker cooperatives. These are most successful when all employees are doing the same type of work, on the same salary scale. There is already ample potential for conflict between employees, given the conflicts of interest that exist between older and younger workers (over, for instance, the value of investment expenditures, the financing of pension liabilities, etc.). When one brings in different classes of worker, doing completely different types of work (e.g., manual vs. white-collar employment), the potential for conflict is greatly exacerbated. Most obviously, the determination of wages becomes an immediate source of controversy. Thus, within worker-owned firms with a heterogeneous class of employees, one can often see adversarial behavior, and hence collective action problems, arising within the ownership group. (The case of United Airlines, in which three separate employee unions persisted in maintaining an adversarial relation with the firm, despite having assumed a majority ownership stake, is a particularly well-known example [Gates 1998: 50–51].) This is why worker cooperatives also often bring in workers on contract, particularly into managerial positions, where the existence of hierarchical relations between members would be a source of obvious conflict.

From this perspective, one of the central advantages that investors have is that the input they bring to the firm is extremely (one is almost tempted to say maximally) homogeneous. While there are no doubt differences between different classes of investor, these pale in comparison to the differences that exist within the other constituency groups. It is also relatively easy to determine how much benefit each investor is deriving from his or her relationship to the firm. Compare that to the case of a worker cooperative, where some of the benefits of association with the firm will take the form of wages, but others will be provided in-kind. Unless workers are similarly situated, some may be receiving more in the way of in-kind benefits than others. This can, in turn, create real difficulties when it comes to determining compensation and deciding what constitutes fair wage differentials. The result is that investors often have the lowest costs of ownership, which means that they are well positioned to make themselves the owners, and thereby make their "interests the objective of the firm and the end of management's fiduciary duty" (Boatright 1999: 178).

In this way, Hansmann's argument brings together all four of the standard arguments for shareholder primacy, while at the same time subsuming them into a general form of the public policy argument. According to his view, the moral responsibilities of managers should track their fiduciary obligations, which are a part of the "bundle" of rights that individuals acquire when they become owners of the firm. There is nothing special about shareholders in this regard. Managers should faithfully serve the interests of shareholders in a standard business corporation, just as they should be loyal to the "tenants" in a housing cooperative or the "employees" in a worker cooperative—because these are the groups that have, in each case, contracted for the loyalties of management, acquiring it in return for acceptance of a residual claim. The idea that managers "work for" or are "agents" of the owners is part of the agreement that constitutes the ownership relation. All of the "intrinsic" features that shareholders possess, which have been appealed to over the years as arguments in favor of shareholder primacy, are not so much theoretical arguments in its favor as practical considerations that feed into the deliberations of investors, when they consider whether they want to acquire ownership of the firm. Thus, the fact that shareholders are residual claimants, or superior risk-bearers, or are particularly vulnerable in their relations with managers, has no intrinsic normative significance; these are simply factors that go into the cost–benefit calculation being undertaken when they decide whether to become owners. Similarly, the "agreement" that shareholders enter into with other constituency groups does not have normative significance because it is a contract, voluntarily entered into (as a pure libertarian would have it), but because it is presumed to be a welfare-maximizing arrangement. In Hansmann's view, it would be exceptionally difficult, if not impossible, for an outsider to determine which constituency group was in the best position to minimize the transaction costs generated by the governance structure of a firm. Allowing the parties to work out a private solution, through contracting, is the best way of getting to the "least-cost assignment of ownership"—and hence at achieving the welfare-maximizing outcome. Thus, the contracts should be respected, not as such, but because of their heuristic value. The arrangements between managers and shareholders are not sacrosanct, on the grounds that they are voluntary, it is rather that their voluntariness creates a strong presumption in favor of the claim that they are welfare-maximizing, and therefore worthy of deference.

5.3. The Force of the Argument

Before going on to criticize some aspects of this argument, I would like to dwell briefly upon its very significant merits. The redescription of the standard business corporation as a "lender's cooperative" is a transformative way of thinking, one that casts a number of rather old debates in an entirely new light. The first thing that must be acknowledged is the straightforward rhetorical force of the redescription. Thanks to more than a century of trenchant criticism of the capitalist system, the very term "profit" has come to carry a negative connotation, which in turn creates a degree of moral opprobrium associated with the concept of "profit-maximization" (e.g., Bakan 2004). Hence the prima facie implausibility, for many people, of the claim that profit-maximization could be a moral obligation for managers. And yet, in the case of cooperatives, it seems unproblematic to say that managers should faithfully serve the interests of the ownership group. Indeed, I am not aware of any business ethicist having ever challenged the notion that in a worker cooperative, managers should serve the interests of the workers. Many customer coops proudly declare their allegiance and loyalty to their members as well, without attracting much in the way of concern. No one ever says, "But what about the suppliers, or the lenders? They have interests too." Hansmann's suggestion is that the profit-oriented firm is just serving its members, in the same way that any cooperative does.

This way of looking at things sharply reverses the burden of proof, when it comes to the doctrine of shareholder primacy. Instead of leading us to ask "What is so special about shareholders, such that they are entitled to the loyalty of management?" it leads one to ask "Is there anything different about shareholders, such that they should not be entitled to the loyalty of management, the way any other constituency group would be if it assumed ownership of the firm?" The answer, it would seem, is "no," simply because there is nothing special about profit. It just happens to be the case that, with lenders, the input that they provide is money, and so the benefits provided to them by the firm will also take a monetary form. In the case of other cooperatives, some of the benefits may be provided in-kind (in the form of a less alienating work environment, for instance, in a worker's cooperative, or higher quality goods, in a customer cooperative). But this does not mean that managers are relieved of the obligation to maximize with respect to the provision of these benefits.

They still have to do the best that they can for the members of the cooperative, as well as assign priority to their interests.

This perspective also exposes some major deficits in the arguments of theorists such as Ronald Dahl, who would like to see greater worker control of firms. In his *Preface to Economic Democracy* (1985), Dahl takes issue with the fact that firms are "undemocratic." He argues that the very same considerations that speak in favor of democratic control of the state speak in favor of democratic control of the firm. "*If* democracy is justified in governing the state, then it must *also* be justified in governing economic enterprises" (1985: 111). He never stops to consider the possibility that firms might already be subject to democratic control—by their owners. He simply assumes that democratic control means control by workers, not shareholders. From Hansmann's perspective, the appropriate response to this would be to point out that if workers want to assume ownership of the firm, then they will quite naturally acquire democratic control of it as well—that is how worker cooperatives function. But if workers opt for a purely contractual relationship with a firm that is owned by some other constituency group, which is to say, if they opt not to shoulder the costs associated with ownership, then they cannot turn around and start to demand the benefits of ownership.

There is a more general point here. If one sees the firm as a "nexus of contracts" between a disparate group of individuals, then it is obvious that not everyone involved in the firm is going to have a democratic say in how that firm is run. "Economic democracy" of the sort that Dahl imagines is not really a democracy for everyone with a stake in the firm. He is not imagining that suppliers and customers will be given the right to vote, or be consulted, or elect representatives to the firm's board of directors (even though that is how things work in the relevant sort of cooperative). He is imagining that only workers will be given such rights. Thus, the appeal to democracy is a red herring, what he is really doing is promoting a particular structure of ownership.

Of course, this does raise the question of why firms are owned by just one constituency group, and whether or not they should be owned in some joint fashion by all constituency groups. An arrangement such as this could legitimately be described as "more democratic" than one in which a single constituency exercises the prerogatives of ownership. Something along these lines has been proposed by stakeholder theorists, such as R. Edward Freeman, who argues that an "agent" within the firm (such as a manager or director), "should serve the interests of all stakeholders" (1994: 417). Similar arguments have been made by Greenfield (2006). The claim is that ownership, in Hansmann's sense of the term, should be jointly exercised by everyone involved in the firm.

Against such proposals, however, Hansmann's analysis provides a simple yet powerful reply. Perhaps the single most important impediment to the success of cooperatives is the political cost arising from heterogeneity of interests amongst the ownership class. The major source of comparative advantage enjoyed by investors, as an ownership group, is that they have relatively similar interests (not identical, of course, just less heterogeneous than any other group). This is witnessed by the fact that cooperatives can be extremely successful even in capital-intensive industries, so long as the ownership group consists of individuals who are similarly situated in their relations with the firm. This is why taxi companies and travel agencies are often worker cooperatives, whereas bus companies and restaurants are not. The problem then with stakeholder ownership is obvious. If heterogeneity of interest within constituency groups is such a formidable obstacle to effective ownership, amalgamating several constituency groups into a single ownership class is guaranteed to dramatically exacerbate the problem. The central evidence for this, Hansmann suggests, is "the nearly complete absence of large firms in which ownership is shared among two or more different types of patrons, such as customers and suppliers, or investors and workers" (2000: 44).[7]

If the agency costs associated with collective ownership are so great, patron groups might all benefit by designating the "least-cost" group as the owner. It is important to remember that a major portion of the costs associated with ineffective ownership are the result of collective action problems, which reduce the overall productivity of the firm and hurt everyone, not just the owners. Thus, handing over the firm to its investors may result in costs for each of the other constituency groups, but to the extent that it produces better governance, it can also produce compensating benefits. Thus, there is nothing implausible about the suggestion that workers might prefer to have a contractual relationship with an investor-owned firm, rather than participate in a cooperative, or even that workers implicitly *pay* investors to run the firm, because the latter are able to do a better job of it. This should be no more surprising than the fact that many people prefer to live in a rental apartment, rather than buy a condominium.

5.4. Problems with the Argument

There is much in this argument that is correct, and that is important for business ethicists to acknowledge. Yet there are also problems. Hansmann claims that ownership will go to the constituency group able to minimize governance

---

[7] For a response to this argument, see Philips et al. 2003. They assert: "Managerial opportunism is a problem, but it is no more a problem for stakeholder theory than the alternative. Indeed, there may be reason to believe stakeholder theory more resistant to managerial self-dealing" (2003: 484). The argument presented in support of this claim, however, is entirely hypothetical, based on speculation about how managers and stakeholder groups might respond to the incentives they face. Hansmann, by contrast, offers empirical considerations in support of his view. For a similar argument, that looks rather to the public sector for empirical evidence regarding the effects of multi-principal agency problems, see Heath and Norman 2004.

costs for the firm as a whole: "The least-cost assignment of ownership is therefore that which minimizes the sum of all the costs of a firm's transactions. That is, it minimizes the sum of (1) the costs of market contracting for those classes of patrons that are not owners and (2) the costs of ownership for the class of patrons who own the firm" (Hansmann 2000: 22). The second component follows fairly immediately from his analysis and seems uncontroversial. The group that has the lowest costs of ownership, or more specifically, the one that has the lowest costs of ownership compared to its costs of market contracting, will be the one with the largest budget when it comes to bidding for ownership of the firm. But what reason do we have for thinking that having this group assume ownership will also result in lower costs of market contracting for the other groups (or if it does increase costs for these other groups, that the total increase in cost will be less than the decrease in cost enjoyed by the ownership group)? Hansmann does not actually provide an argument, and a moment's consideration will show that the question is a bit complicated to resolve.

In order to simplify somewhat, imagine that a firm is being auctioned off, sold to the highest bidder by an otherwise neutral third party (such as the state).[8] How much each constituency group is willing to bid is determined by a number of factors. First, it is important to keep in mind that the total value of the firm to its assembled constituency groups may vary, depending upon which group assumes ownership. The firm is, after all, an institution that exists in order to resolve collective action problems, and its productivity is almost entirely a function of its success in doing so. To the extent that one group does a better job than some other at running the firm, there may be more to go around for all. Thus, it is something of a simplification to talk about simply minimizing the costs of ownership, as Hansmann does. Because the benefits can vary as well, what matters is the net benefit to each constituency group of its transactions with the firm, that is, gross benefit minus transaction costs (whether those be costs of ownership or costs of market contracting). We can assume, for the sake of argument, that benefits are always positive—otherwise the constituency group will simply opt not to deal with the firm.

In the same way that the total value of the firm can vary, depending upon which group controls it, the cost of market contracting for each party can vary as well, depending on who the owner is. Employees, for instance, may find that they get a better bargain when dealing with a consumer coop than an investor-owned firm, for completely contingent, firm-specific reasons. Thus, one cannot simply compare "costs of ownership" to "costs of market contracting," because the costs of market contracting can change, depending upon who has ownership. Thus, in order to model the auction of a firm, one must

[8] This is for simplicity of presentation. It does not affect the results if one imagines the firm already being owned by one group, entertaining bids for a transfer of control to some other.

TABLE 5.1 } **Standard case**

Ownership by	Benefit to			
	Workers	Suppliers	Investors	Customers
Investors	1	3	5	2
Workers	3	1	2	2

consider the net benefit each constituency group would derive under multiple scenarios, depending upon which group has ownership. The amount that one group would bid against some other group is determined by the difference between the first group's cost of ownership and its cost of contracting under that second group's ownership.

Consider table 5.1, which presents a slightly simplified representation of the choice between having a firm controlled by its workers or its investors. Rows represent a given ownership scenario, columns show how much each group benefits under that scenario. Benefit is represented in utils, since there are a variety of non-monetary costs and benefits that go into the calculations of each group (although one may assume that utils are interpersonally comparable and map onto money in a linear fashion). Table 5.1 then shows a subset of the decision problem, representing the contest between workers and investors for control of the firm.

The amount that each group is willing to bid for the firm is the difference between its benefit level when it constitutes the ownership group and the benefit level it enjoys under market contracting when the other group assumes ownership. Workers, for instance, get a benefit of 3 from ownership, versus only 1 from market contracting with an investor-owner firm, and are therefore willing to bid 2 against investors in an effort to keep the firm out of their hands. Investors, by contrast, are willing to bid 3 against workers—more than any other constituency group—and so they are able to win the contest for ownership. One can see that in this case, a move from worker to investor ownership would be Kaldor–Hicks but not Pareto efficient. Workers would rather own than have a contract, but the benefit is not great enough to outbid investors. They are harmed by the transition, yet the benefit to investors is more than sufficient to compensate the workers, and so the outcome is Kaldor–Hicks efficient.

The fact that the benefit to investors would be sufficient to compensate workers in this case is entailed by the fact that investors are willing to bid more for the firm. This is probably the sort of case Hansmann had in mind when he asserted that the aggregate costs of contracting would be minimized for all groups. His thought was presumably that, if the costs of contracting were not minimized, this would give some other group the budget needed to outbid the current ownership group, and so ownership would pass to this other group, under whom the costs of contracting would be minimized.

Problems arise, however, when elevated costs of contracting are split among multiple constituency groups. For example, table 5.2 shows a situation in which both suppliers and workers would be better off under worker ownership. It is not hard to imagine such a scenario—perhaps the investor-owned firm would "play hardball" with suppliers in a way that a worker-owned firm would not. Thus one can imagine that worker ownership could enhance productivity, by reducing shirking among employees, while at the same time generating modest spin-off benefits for suppliers. Yet one can imagine also that, while the sum of these benefits (4), is greater than the benefits enjoyed by investors from ownership (3), the former is unfortunately divided up between two parties. Thus, no one constituency group—neither workers nor suppliers—is in a position to outbid investors for control of the firm. The investors become owners because their costs of ownership are lowest, even though they are not in a position to minimize the costs of market contracting for other patron groups. Investor ownership in this case is not Kaldor–Hicks efficient, while worker ownership would be.

Now what Hansmann may have been thinking was that in such cases, employees and suppliers could pool their resources in order to outbid investors. (In other words, that the auction should not be modeled as a competitive game amongst constituency groups, but rather as a cooperative game.) Yet having employees and suppliers form such a coalition would no doubt entail transaction costs, increasing the costs of ownership (in cases where they agreed to joint ownership) or increasing costs of market contracting (in cases where they agreed to some sort of single-group-ownership-plus-compensation arrangement). Indeed, if Hansmann's central argument against stakeholder theory is to be believed, these transaction costs will be quite high, precisely because they involve two constituency groups with radically different interests. Thus one cannot just add up the benefit to workers and suppliers and conclude that together they have the resources to outbid investors, since achieving this kind of collaboration is likely to erode some of the benefits that each group would receive under a regime of worker ownership.

Of course, the constituency groups, who are being treated as the "players" in this game, are themselves coalitions. This is unobjectionable, however, because the costs of forming these coalitions are already factored into the payoffs shown in the tables (since Hansmann treats these as part of the costs

TABLE 5.2 } **Complicated case**

Ownership by	Benefit to			
	Workers	Suppliers	Investors	Customers
Investors	1	2	5	2
Workers	3	4	2	2
Suppliers	1	4	2	2

of ownership). When we start talking about the formation of new coalitions, however, among constituency groups, for the purposes of bidding for ownership, it is clear that the payoffs will have to be rewritten. And with the new set of payoffs, it is not clear that the coalition will have the resources or the incentive to outbid every other group for ownership of the firm. For this reason, one cannot say that having ownership pass to the individual group that will benefit the most from it will also minimize the costs of market contracting for the other groups.

5.5. Lessons for Business Ethics

This problem with Hansmann's analysis in no way diminishes the broader significance of his work for business ethicists. What it suggests is merely that the argument needs to be handled with a degree of caution, and that it should not be taken to imply more than it does. For example, a superficial reading of Hansmann's argument could easily lead to the view that managers need not concern themselves with the interests of non-ownership groups at all—that they can adopt the same strategic, adversarial orientation toward these groups as they would any other party with whom they engage in market transactions—because the interests of these groups are fully protected by "law and contract." Thus workers may not get to enjoy the benefits of ownership, but that's fine because they do just as well with contracts. This view, which I refer to as the "let them eat contracts" position, does not follow from Hansmann's argument, and is even in tension with the transaction cost theory of the firm that is at the heart of the analysis. After all, if there were not serious imperfections in the relevant labor markets, the firm would not need to have employees at all, it could use outside contractors. Thus one should expect the firm to be in a position to exercise market power over employees. The question is how this market power should be handled.

For concreteness, consider the claim made by Boatright, which he takes to be a consequence of Hansmann's analysis:

> Each [constituency] group has the opportunity to seek the best protections or safeguards for their own interests, which is to say the return on the firm-specific assets that they provide to a firm. Usually, non-shareholder groups are better served by safeguards other than control, which is left to shareholders. This outcome is not only efficient but also morally justified because it best serves the interest of all stakeholder groups and results from voluntary agreements or contracts made by all the relevant groups. (2006: 115)

The claim that shareholder primacy "best serves the interest of all stakeholder groups" harbors an important ambiguity, since "all" must not be taken to mean "each stakeholder group, taken singly" but rather "all stakeholder groups, in the aggregate." In other words, Hansmann does not even purport to show that the least-cost assignment of ownership creates a situation in which each stakeholder group is better off than under some other assignment. This would require that the transfer of ownership to the least-cost group constitute a Pareto improvement. On some occasions it probably does, but to claim this as a general rule would be Panglossian. Hansmann claims only that it is a Kaldor–Hicks improvement. In other words, he does not claim that workers are better off under shareholder ownership than they would be under worker ownership, and so he would not say that workers have chosen contract over control. The correct thing to say would be that they were outbid for control. The consolation, such as it is, lies in his claim that the amount that shareholders benefit from shareholder ownership is greater than the amount that other constituencies would have benefited from some other form of ownership, and so the benefits to shareholders of shareholder ownership outweigh the losses suffered by the others. In table 5.1, for example, workers are better served by control than by contract. But investors are even better served by control, and so are able to wrest it away. "Society" as a whole is better off, even though workers are not.

But even this, it turns out, is not true as a general rule. Ownership by one group may impose losses upon other constituency groups that are not outweighed by their own gains, and yet ownership does not change hands because the losses are divided up among the various groups in such a way as to create transaction costs that prevent the group that could maximize social welfare from assuming ownership. Thus even if one were to examine the situation from a straightforwardly utilitarian perspective, and were willing to sacrifice the interests of one group in the name of greater benefits for some other, Hansmann's argument fails to justify any sort of complacency about the extent to which shareholder primacy will serve the interests of society at large. In other words, it fails to justify the "let them eat contracts" position.

What then does it contribute? Most importantly, it diminishes the temptation to want to solve the problems of non-ownership constituency groups by fiddling with ownership structures (e.g., by imposing additional fiduciary duties on directors, requiring them to concern themselves with stakeholder interests, or by giving workers democratic control over various decision-making processes within the firm). On the one hand, these kinds of interventions attempt to redistribute the benefits of ownership without redistributing the costs. At the very least, this is likely to disrupt the calculus that made one group willing to shoulder the burdens of ownership in the first place. More generally, Hansmann's argument suggests that if this sort of diffusion of ownership amongst the constituency groups were workable, the parties themselves would have already come to such an agreement. Groups that stood to benefit from participation in ownership would have pooled their resources and acquired ownership from investors. The fact that individuals do not spontaneously form "stakeholder-oriented" firms, with ownership shared across multiple constituencies, suggests that there may be good reason not to form them by legislative fiat either. The only circumstance in which doing so could be recommended would be if legislation dramatically reduced the transaction costs associated with this sort of pooling. This seems unlikely, if Hansmann is correct that political costs constitute a significant barrier to ownership. The history of state-owned enterprises certainly suggests that so-called multi-principal agency problems can be a source of enormous mischief in organizations (Heath and Norman 2004: 259–261).

Furthermore, in cases where voluntary contracting fails to assign ownership to the group able to minimize transaction costs for the firm as a whole, it is extremely unlikely that any legal mechanism could reliably detect this, much less improve upon the outcome. Consider the example provided by table 5.2. Here worker ownership is the arrangement that maximizes net benefit, even though voluntary contracting winds up assigning ownership to investors. Yet the example is extremely abstract—intended only to prove a conceptual point. In the real world, it would be virtually impossible to know that a firm was in such a situation, because figuring it out would require the ability to assess counterfactual claims about a complex organization in an environment of uncertainty, involving parties with an incentive to misrepresent both their circumstances and interests. Thus the only realistic alternative to a regime of voluntary contracting is a "one-size fits all" arrangement, in which some sort of standard multi-fiduciary structure is imposed upon all firms. The fact that the parties themselves never voluntarily enter into such arrangements stands as a forceful criticism of such proposals.

Thus the Hansmann argument, in a suitably modest form, suggests three things:

1. There is nothing inherently wrong with the idea of managers assigning priority to the interests of the owners of a firm or a cooperative, and thus there is nothing wrong with the norm of shareholder primacy in cases where investors form the ownership group. Furthermore, there is typically a good reason why firms are owned by investors, namely, investors are the group able to derive the greatest net benefit from ownership. The success of cooperatives in certain sectors shows that the existing regime is not one of imposed shareholder primacy, but rather one of voluntary contracting, in which shareholders most often are the ones who win the contest for ownership.
2. Legal interventions aimed at reassigning ownership, or dispersing it across multiple constituency groups, are unlikely to produce any improvements over the arrangement that the parties themselves

have contracted to. The fact that other constituency groups, or coalitions of constituency groups, have not assumed ownership of the firm should generate the presumption that their costs of ownership would be prohibitively high or that the benefits would not be that great.

3. Knowing what the relative advantages of ownership are for a particular firm, at a particular point in time, is extremely "impacted" information, which only the parties immediately involved in transactions with the firm are likely to have access to. A relatively unheralded feature of the existing regime is that it gives the parties involved a measure of discretion in how they choose to arrange ownership. Many proposals for reform involve replacing a voluntary regime with a mandatory one, which is likely to carry its own, rather significant efficiency losses.

Critics of shareholder primacy have exposed a number of very serious flaws in the existing structure of corporate law (or at least its effects as implemented) (e.g., Greenfield 2006). And yet their proposals for reform almost invariably slip into the deep, well-worn rut of multi-fiduciary stakeholder theory. This approach is subject to a number of powerful objections (Heath 2006a; Orts and Strudler 2009). Many of these objections, however, concern the motivational and incentive problems that such a governance structure would create. These problems are ones that many ethicists are inclined to dismiss, on the grounds that they would not be such serious problems if people were simply to act more ethically. The result is something of an impasse, in which each side's arguments do little to reach the other. For those who are unsatisfied with this state of affairs, Hansmann's argument offers an attractive third option. Rather than conceiving of the existing arrangement as one of imposed shareholder primacy, it is better to think of it as a voluntary contest for control that shareholders typically win. For those who think that shareholders win a bit too often, or that their victory has too many negative consequences for the losers, there are two avenues of improvement. Rather than trying to make everyone a winner, one could either improve the contest, in order to provide greater reassurance that the most deserving party has won, or else focus on improving the outcome for the loser.

One could begin by focusing on the terms of the competition for ownership, in order to ensure that there is at very least a level playing field with respect to the four primary forms of cooperative organization. One could even make the case for tilting the field further, in such a way as to encourage non-investor groups to assume ownership (beyond the existing rules governing the taxation of profits, which already offer a large subsidy to cooperatives). For example, in some jurisdictions customer cooperatives are required to do a relatively large percentage of business with their own members. This is tantamount to a restriction on their ability to engage in market contracting. These cooperatives would have lower transaction costs if these restrictions were lifted. Another major problem with cooperatives is that there is no well-established "market for control" within the ownership class, such as the stock market provides for investors. This has a number of consequences: cooperatives are not subject to takeovers, members have constrained exit options, there is nothing equivalent to the "stock price" to assess managerial performance, etc. All of these diminish the attractions of the cooperative form, and could in principle be addressed through institutional and legal innovation. This would require a certain measure of creativity, but would probably be time better spent than on further investment in the stakeholder paradigm.

The second major focus could be on remedying the defects of the contractual regime that constituency groups, most importantly workers, are subject to. The first step would be to acknowledge that the firm is often in a position to exercise significant market power over its constituency groups, and that law and contract typically offer them incomplete protection. Yet for largely fortuitous reasons, these constituency groups may be unable to assume ownership of the firm without worsening their situation. In the same way that a landlord has power over his tenants, which she is often in a position to abuse, a manager exercises considerable market power over her workers, which she is also quite often in a position to abuse. It is important to keep in mind that firms are created precisely because of imperfections in market contracting, which threaten either to undermine the possibility of cooperation entirely or drastically limit its benefits. The substitution of transactions administered by an organizational hierarchy for market transactions is a response to this underlying failure of the market (Williamson 1981: 564). Certain terms in the relationship between employees and the firm are contractual, and are subject to the adversarial norms of market contracting. Yet the core of the relationship is non-contractual, and even—with respect to many details—outside the scope of legal regulation. It is governed by the norms and expectations that govern authority relations. This is precisely why this organizational form has comparative advantage over market contracting—because it is fundamentally different. For example, it explains why employees often do a better job than outside contractors, and the two cannot be substituted for one another willy-nilly.

Seen from this perspective, one can see the obvious problem with the "let them eat contracts" view. Boatright himself cites "management's willful violation of agreements, market failures, and externalities or third-party effects" (2006: 123), as serious problems that may arise in contractual relations, and that call for the exercise of moral restraint. The guiding idea here should be that of market failure. Workers in a shareholder-owned firm are perhaps best thought of as stuck between a rock and a hard place. On the one hand, they are unable to secure the loyalty of management by assuming ownership of the firm. On the other hand, they are unable to take full advantage of the protections offered by the market—if they could, they would be self-employed contractors, not employees. The dominant moral imperative for managers should be to avoid taking advantage of this situation. One way of doing so would be to refrain from taking advantage of the market power they exercise—by not exploiting asymmetric information, or their own monopsony power with respect to that individual's firm-specific skills. For example, the issue of managers lying to employees could be taken much more seriously than it is, both legally and from the moral point of view. The argument against lying, however, would be that such misrepresentation undermines the efficiency of labor markets, not that managers have some special fiduciary relationship to workers (Greenfield 2006: 200–202).

The resulting conception of business ethics would be one in which managers continue to owe fiduciary duties only to the owners of the firm and continue to assign primacy to their interests, yet are subject to a set of deontic constraints in dealing with other constituency groups, constraints that reflect the institutional preconditions for "healthy" marketplace competition (Heath 2001, 2006b). Stakeholder primacy is preserved, in the sense that if there is a conflict between the interests of various constituency groups, management should assign priority to the interests of shareholders. If, however, the conflict is one between the interests of shareholders and the principle that managers should refrain from taking advantage of market power in dealing with other constituencies, then the principle trumps the interests. This is because the promise of healthy marketplace competition is what makes it permissible for managers to show partiality to the interests of just one constituency group in the first place (rather than being obliged to concern themselves with the interests of society as a whole). So if the question is one of how windfall revenue should be allocated, then the interests of shareholders should be accorded primacy. But if the question is whether windfall gains should be achieved by deceiving employees, or "churning" them in order to avoid paying higher salaries or benefits, or underfunding the pension scheme, the interests of employees should come first—not because the manager is obliged to concern himself with the interests of employees per se, but because he is morally prohibited from advancing the interests of shareholders by taking advantage of market failures. Both imperatives—the duty of loyalty to shareholders and the moral constraint in dealing with other constituencies—are a straightforward consequence of the way that the corporation tries to mesh administrative hierarchy with market contracting in the overall pursuit of economic efficiency.

PART II }

Cooperation and the Market

6 }

Contractualism: Micro and Macro

One of the central and most attractive features of contemporary social contract theory is the idea that principles of justice exist in order to divide up the "benefits and burdens of cooperation." There are many circumstances in which individuals are able to engage in mutually beneficial interaction, but on the condition that each exercise some restraint in the pursuit of his or her individual interest. Thus the situation calls for a measure of voluntary self-restraint, which each individual must (by and large and in general) be persuaded to undertake. The structure of the interaction, however, underdetermines the choice problem, in the sense that there are many different cooperative arrangements, each of which involves a different allocation of the burdens and benefits, but all of which are mutually beneficial. Thus a set of principles is required, in order to specify the precise modalities of cooperation—who does what, who gets what, who decides what, etc.—in a way that will be acceptable to all.

The central contractualist idea—articulated at its highest level of generality—is that the principles of justice specify the terms under which individuals would voluntarily *agree* to undertake a *cooperative* interaction. Both ideas are important: the fact that each individual must be induced to agree is what accounts for the deontological flavor of these principles (i.e., the fact that they do not always recommend maximizing aggregate welfare); while the fact that the interaction is cooperative explains why individuals might nevertheless be willing to accept something less than their maximal claim. This analysis, however, leaves open an important ambiguity with respect to the level at which these principles should be taken to apply. Perhaps the most natural way to apply them is at the level of particular interactions, such as business partners trying to set up a joint venture, siblings trying to divide up an estate, or shipwreck survivors allocating resources on a desert island. These people can benefit by cooperating with one another, yet failure to agree upon specific modalities has the potential to erode these gains. Thus a set of principles that can command convergence with regard to the specification of these modalities has obvious appeal. Furthermore, applying the principles of justice at this level is responsive to our everyday sense that these problems must each be solved in a way that is fair to the individual participants (e.g., that the partners, siblings, or survivors should each be treated fairly with respect to this particular interaction).

I refer to this way of applying the central contractualist idea as *microcontractualism*, on the grounds that it treats the principles of justice as constraining individual conduct at the action-theoretic level. It has obvious appeal, but is also subject to certain powerful objections. In particular, this way of applying contractualist principles seems to bear more than a passing resemblance to libertarianism (indeed, David Gauthier's contractarianism, which is a paradigm instance of a microcontractualist analysis, is sometimes lumped together into the broader family of libertarian theories[1]). There is an important difference, which is that libertarianism typically takes whatever the parties happen to agree to with respect to the division of benefits and burdens as authoritative, whereas contractarians—including Gauthier—instead favor the division that would be normatively prescribed, through the application of a set of principles that reflect the choices individuals would make, under more-or-less idealized conditions. There is, however, one important similarity, which is that both libertarianism and microcontractualism leave individuals without redistributive obligations toward those with whom they choose not to cooperate. Of course, while both Robert Nozick and Gauthier view this as an attraction of their respective views, many other philosophers regard it as a *reductio* (illustrating, if nothing else, the unhelpfulness of many of the moral "intuitions" that are routinely appealed to in contemporary moral philosophy). There is, however, a more firmly specifiable worry, which is that this feature of the view makes it possible for individuals to game the principles of justice, allowing them to achieve outcomes through selective association that could not be achieved within the scope of an ordinary cooperative interaction. This violates a stability property that is plausibly regarded as an important desideratum for any theory of justice.[2]

[1] Most influentially, by Will Kymlicka (2002: 128–138). Kymlicka does not actually explain why he considers the view libertarian and says nothing about its strongly egalitarian features (such as Gauthier's claim that the state may rightfully confiscate the portion of Wilt Chamberlain's earnings that constitutes economic "rent"—which is to say, almost all of it). See David Gauthier 1986: 273. The egalitarian aspect is somewhat concealed by Gauthier's unorthodox, and ultimately unsuccessful, derivation of the minimax relative concession principle. If one looks instead at the equivalent result in bargaining theory, the Kalai-Smorodinsky solution, one can see that it incorporates an egalitarian "symmetry" axiom.

[2] For the general flavor of this, see John Harsanyi 1963: 194–220; Terje Lensberg 1988: 330–341. The idea is that in an n-person solution, having one person exit with his just share should not cause all the remaining participants to want to renegotiate their shares.

The most common response to this problem has been to shift the level at which the principles are applied, so that instead of being used to resolve the modalities of particular cooperative interactions, they are instead applied to "social institutions" more generally, and in the extreme, to "society" as a whole. This is, of course, the way that the classical social contract theorists conceived of the doctrine—as providing the terms governing the "civil condition" as a whole—and it is echoed in John Rawls's famous description of society as a "cooperative venture for mutual advantage" (Rawls 1999: 4). I refer to this as *macrocontractualism*, for obvious reasons. Shifting to this level of analysis makes the social contract much more metaphorical, which has certain disadvantages, but it also helps to minimize the problem of selective association. Individuals cannot evade their obligations toward others simply by avoiding any sort of cooperative interaction with them; by virtue of belonging to the same society, they are part of a generalized system of cooperation, and so are subject to certain obligations that apply to everyone. Of course, the problem may still recur at the boundary, if one stops short of treating all of humanity as party to the contract. Rawls, for instance, treats "the" system of cooperation as secured by the basic structure of society, largely coinciding with the institutions of the nation-state. This, in turn, generates a set of reduced obligations toward foreigners, a position that has attracted a certain measure of resistance among those who feel that it represents a pinched, perhaps even ungenerous, response to the human condition.

In this paper, I would like to focus on a different, less well-known problem with macrocontractualism. By shifting the analysis up to the level of "society as a whole," it is easy to lose track of the fact that individuals also expect their particular interactions and private associations to be fair, above and beyond whatever contribution the outcome may make to the fairness of the entire society (so that, for instance, even in the context of an unjust society, particular institutions or domains of interaction might nevertheless be just). Our sense is also that a just society is one in which all component institutions and associations—families, schools, churches, contracts, wills, corporations, etc.—can also be deemed to be just, in a relatively self-standing fashion. Yet macrocontractualism seems to lack the resources needed to ensure this. Indeed, contractualists such as Rawls wind up adopting a surprisingly voluntarist standard to judge the particular cooperative projects that individuals may undertake (again, not far off from libertarian views). Hence the puzzle that emerges: if we start out at the bottom, with a micro perspective, insisting that particular cooperative interactions and small-scale institutions be internally just, then we have no assurance that this will all add up to a "just society" at the macro level. If, on the other hand, we start out at the macro level, and insist that society as a whole be just, then we lose the ability to insist that small-scale institutions and interactions be internally just.

My objective in this paper will be to map out in slightly greater detail how this puzzle comes about, and then suggest a strategy for resolving it that takes us beyond the standard flavors of contractualism. This involves adopting a cultural-evolutionary perspective, then interpreting the principles of justice in terms of a set of pragmatic-structural biases in the transmission and reproduction of social norms.

6.1. Minimally Controversial Contractualism

I would like to begin by providing an outline of what I refer to, perhaps tendentiously, as "minimally controversial contractualism," then show how the puzzle follows quite immediately from it. The central advantage of contractualism, from the standpoint of its adherents, is the leverage it provides in responding to various forms of moral skepticism. First and foremost, contractualists are inclined to treat rational egoism as a serious concern, and therefore to worry about what Christine Korsgaard calls "motivational skepticism" (1996: 311–334). Even if we had the power to discern moral facts, or had access to clear and distinct moral intuitions, contractualists worry about how individuals are to be persuaded to respect these judgments in practice, especially when the moral rules demand that we set aside our self-interest—often in rather dramatic ways—in favor of the good of others. Contractualists generally would like to have something to say to the person who is unmoved by altruistic appeals. The contractualist approach takes as its point of departure the observation that adherence to moral rules typically produces benefits for others that are greater than the losses to the individual. Thus when jointly adopted, they produce mutual benefit, which is to say, they establish a system of cooperation. Morality involves sacrifice, but it also produces reward. So for those who worry about motivational questions, the focus on these rewards provides, if not a reason to act morally, at least a rationale for the way that morality constrains individual self-interest.

I mention this because, among critics, the contractualist focus on cooperation—and thus on mutual benefit—is often portrayed in a negative light, as though it were motivated by a desire to limit the scope of our obligations (Barry 1989: 163). Invidious comparisons are drawn to the expansive, almost promiscuous extension of moral duty among utilitarians, many of whom believe that we have unbounded obligations to improve the happiness of all living things, or luck egalitarians, who believe that we are literally responsible for bringing about the Kantian *summum bonum* (a task that Kant himself believed could only be plausibly undertaken by an omnipotent God, and even then would require an eternity to achieve). And yet the problem with these sorts of high-minded ideals is that they tend to lack motivational efficacy. They overrule considerations of self-interest so completely that it becomes a serious question whether any of us could ever be justified in, say, buying a cup of coffee, much less a pastry to go with it. Because they show such wanton disregard for the interests of the individual, these views make it difficult to see how one could convince a person not already in their grip to take them seriously. They are, as it were, all stick and no carrot.

Contractualists are generally willing to sacrifice some of this loftiness in the interest of producing norms that are more likely to have motivational efficacy. Rawls articulated this ambition quite clearly when discussing what he called the "strains of commitment." Agreements that require us "to accept the greater advantages of others as a sufficient reason for lower expectations over the whole course of our life," make what he describes as an "extreme demand" on individuals, compliance with which may well "exceed the capacity of human nature" (Rawls 1999: 177–178).[3] He took this as grounds for limiting the range of fair outcomes to those belonging to what John Nash defined as the "feasible set" in cooperative interactions (Nash 1949: 158). The important point is that the focus on cooperation, and hence the willingness to limit the scope of our obligations both with respect to the persons to whom they are owed and the benefits that are subject to their claims, is not always—or even usually—motivated by a desire to minimize the claims that others can make on us. It is more often a consequence of a genuine concern about motivational skepticism. Mutual advantage seems to provide a happy *via media* between the vulgar appeal to self-interest and the question-begging reliance upon existing moral commitments.

With respect to the content of our moral judgments, contractualists are also worried about skepticism, particularly when it comes to the principle of equality. While some philosophers seem content simply to posit a commitment to equality, as some sort of ultimate value, not susceptible to further justification (see, e.g., Cohen 2008: 7), contractualism recommends itself to those who would like to have something to say to those who do not already share this commitment (or worse, who actively distance themselves from it, on the grounds that it is nothing but a rationalization of envy). It is important to recognize that equality is an extremely demanding moral ideal, one that can impose onerous obligations upon the individual. Furthermore, since any serious commitment to equality must involve a willingness, at times, to level down, equality can enter into tension with other, quite plausible, moral ideals.[4]

[3] Rawls continues, "In fact, when society is conceived as a system of cooperation designed to advance the good of its members, it seems quite incredible that some citizens should be expected, on the basis of political principles, to accept lower prospects of life for the sake of others" (1999: 178).

[4] For example, what Larry Temkin calls "the Slogan" strikes most people as being extremely plausible at first glance. See Temkin 1993: 248–255. Explaining what is wrong with it requires considerable subtlety.

All of this generates a significant burden of justification, particularly toward those who can be expected to be the losers in any egalitarian redistribution.

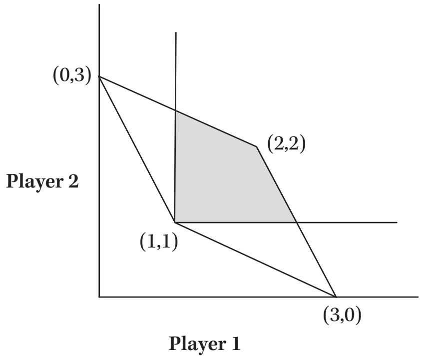
In response to this challenge, contractualists have a rather simple and very powerful claim. Starting with Hobbes, the fundamental argument for the principle of equality has been that it arises out of a symmetry condition that must be satisfied in order to secure agreement. Whether it is "splitting the difference," or creating equal shares, or flipping a coin to choose the winner, everyone is familiar with the way that equalization can be used as a technique to minimize, and often eliminate, a particular type of objection to a cooperative enterprise. Thus equality is not simply posited as an ultimate value, it is derived from the constraints that must be satisfied in order to achieve agreement. To the "losers" in any egalitarian distribution, then, who want to know why they should be the losers, one can point out that the losses are entirely hypothetical. Without an ongoing system of cooperation, there would be nothing to lose. Yet the cooperative scheme is made possible only by the willingness of all participants to play along, a willingness that is, in turn, brought about by the fact that the benefits and burdens of the system are distributed and borne equally.

The standard way of illustrating this is with a prisoner's dilemma, as shown in figure 6.1. Note that the numbers need not be taken to represent utility, but could be anything that the two players are able to produce through cooperation, whether it be increased life-expectancy, calories available for consumption, travel speed, evolutionary fitness, or, of course, preference-satisfaction. The set of possible payoffs, obtainable through either randomization or repeated interaction, is shown as the diamond-shaped region in panel B. Since each individual is able to guarantee herself a payoff of 1 through straightforward maximization, the space of mutually beneficial arrangements is the set of points to the north-east of (1,1)—the "feasible set," or what Gauthier described as the potential "cooperative surplus." The fact that there is a continuum of possible cooperative arrangements reveals the extent to which utility-maximization (or self-interest, narrowly construed) underdetermines the interaction.[5] There are literally an infinite number of possible cooperative arrangements, even within the scope of this highly simplified model of interaction. Furthermore, if one were to imagine a simple "alternating offers" bargaining model (where one player proposes a particular cooperative arrangement, with the other having a choice of either accepting it or proposing a counteroffer) absent some penalty for delay, and among players animated only by self-interest, it is easy to see that the game will go on forever. The players will simply take

[5] This is codified in the form of the folk theorem for repeated games. See Fudenberg and Maskin 1986: 533–554.

		Player 2	
		Defect	Cooperate
Player 1	Defect	(1,1)	(3,0)
	Cooperate	(0,3)	(2,2)

A. Strategic form



B. Payoffs

FIGURE 6.1 Prisoner's dilemma

turns making self-serving proposals, which the other will reject in favor of an equally self-serving counteroffer.

Thus it is no great stretch to imagine that a set of normative principles is called for, in order to specify what should count as a reasonable agreement, and that the parties should accept such an agreement, not because it coincides with their self-interest, but precisely out of a recognition that self-interest fails to provide an acceptable basis for agreement, and that given this failure, the proposed principles are reasonable. Of course, there may be a variety of "thick" cultural resources that the parties can appeal to in order to resolve such problems (such as an inherited set of social norms that specify how different sorts of interactions should be organized). Indeed, it has been observed that when "public goods" games are played in some non-Western societies, where the practice of "the psychology experiment" are unfamiliar, individuals often respond by searching for a "cultural template" for the interaction (Henrich et al. 2001:73–78). So after thinking about the structure of the game, they may say something like "this is just like in the village, when everyone contributes to repairing the road" (Henrich et al. 2001:76). They then act as they would if it were an interaction of that sort.

If one assumes, however, that thick resources of this type are unavailable—either because there is no cultural template, or because there are multiple templates and the choice of one is controversial—there are two principles that seem to be suggested by the very structure of the interaction, or that can be applied without drawing upon any particular cultural resources. First, if there is an arrangement that makes both individuals better off, then it would seem obviously superior to one in which they are both worse off. This judgment can be made without even getting into the details of the case. Articulating this idea as a principle yields the familiar Pareto-superiority criterion. Applying that principle to any potential system of cooperation results in all points in the

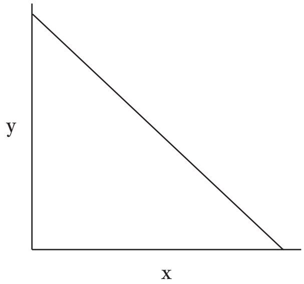


FIGURE 6.2 The set of Pareto-optimal outcomes

feasible set that are Pareto-inferior to some other point in that set being discarded as candidates for agreement. What remains are the set of Pareto-optimal points, such as one can see in figure 6.2. It is worth repeating the familiar observation that the ordering of points imposed by the Pareto-superiority criterion is incomplete, since the set of Pareto-optimal points are unranked vis-à-vis one another. This means that the set of points shown in figure 6.2 can be thought of as an indifference curve—each outcome in the set is just as good as any other, from the standpoint of efficiency.

The second principle is less self-evident, but will be familiar to anyone who has spent some time dealing with children. One obvious way of minimizing objections to a proposed distribution is to avoid giving anyone an incentive to switch places, or allocations, with anyone else. In a welfarist framework this generates a symmetry (or anonymity) principle (that no player should want to switch places with another player); in a resourcist framework it generates the envy-freeness principle (that no player should want to acquire the allocation of any other player). Either way, it suggests that the set of points in the feasible set that generate a desire to switch places on the part of any player should be discarded. Call this the egalitarian principle. What remains after its application is the set of symmetric, or envy-free points, shown in figure 6.3. Here it is worth making the less-common observation that the ordering of points imposed by this principle is also incomplete, in a way that is precisely complementary to that of the Pareto principle. Thus figure 6.3 again can be thought of as a social indifference curve—each outcome in the set is just as good as any other, from the standpoint of equality.

In the simplest of cases, under a first-best scenario, the intersection set of these two curves will be a single point (sometimes known as the efficient equal allocation).[6] Contractualism then selects that arrangement as the most

[6] It is a single point only because the problem involves a single distribuendum and is therefore one-dimensional. When extended to $n > 1$ goods (as in a typical "resourcist" framework), the set of envy-free allocations typically becomes an n-dimensional space containing multiple Pareto-optima. Thus some additional resources must be introduced in order to pick out a single solution. See Heath 2004b: 313–335.

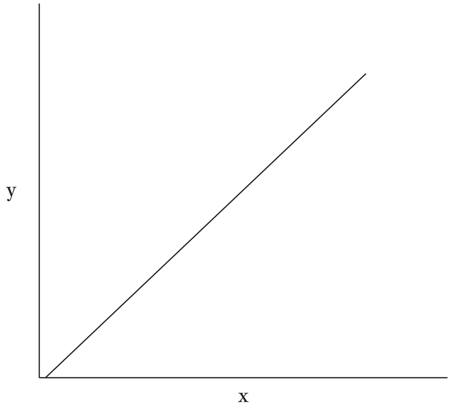


FIGURE 6.3 The set of symmetric or envy-free outcomes

reasonable, not because it has any intrinsic merit, but simply because it does not give rise to any of the obvious objections that every other point in the feasible set would give rise to. (If one thinks of this in terms of the Parfit-Scanlon "complaint" model, one can see that moving to an efficient allocation eliminates one type of complaint, while moving to an equal allocation eliminates another. Thus one can derive the two principles from the most minimal version of the complaint model, one that does not need to get into the dicey business of distinguishing "stronger" from "weaker" complaints.)

In the real world, however, situations may easily arise in which it is possible to make significant improvements with respect to one of the two principles, but only by accepting an arrangement that is worse with respect to the other. Thus the question arises how much inequality one should be willing to accept in order to achieve gains in efficiency, or how much inefficiency one should be willing to accept in order to get improvements in equality. One way of resolving this is to assume that the further the status quo is from the ideal, with respect to either principle, the more strenuously players will object to a deviation of a given magnitude from it. If one assumes that the most favored arrangement will then be the one that players object to least strenuously, the result is a prioritarian ordering, in which benefits to an individual can justify departures from equality, but where these benefits "count" for less, the further away one gets from this ideal. The most well-known formula from making this trade-off is the Nash Bargaining Solution, which favors the arrangement that maximizes the product of the benefit received by each player. This is shown in figure 6.4 as a social indifference curve *N*, which is contrasted with the utilitarian solution *U* (exhibiting complete indifference to the distribution of benefit between the two individuals), and the difference principle *D* (exhibiting complete indifference to the allocation of the better-off individual, so long as that person remains the better-off).

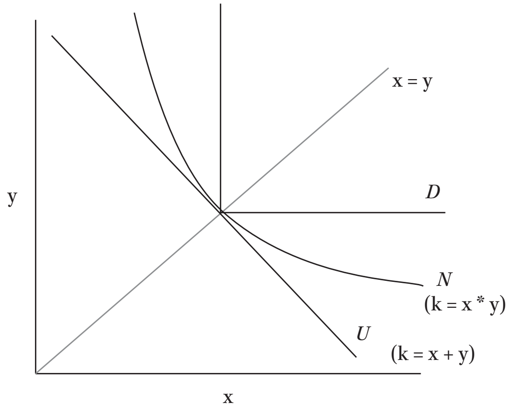


FIGURE 6.4 Social indifference curves

To see how the bargaining solution can be applied, consider a case such as Taurek's numbers problem.[7] There are five people on one island, one person on another: you have the opportunity to save the inhabitants of only one island. The maximizing solution (*U*) says you should save the five (and ex ante, with each individual having an equal probability of being on either island, this solution Pareto-dominates the alternatives). Yet the "fair" solution (*D*) would assign each individual an equal chance of being saved, which can be accomplished by flipping a coin, in order to decide whether to rescue the one or the five (this gives each person a probability of exactly ½ of being saved). Now, suppose that one is torn between these two considerations. One does not like the idea of imposing certain death upon the lone individual just because he had the bad luck of winding up on the wrong island, and yet one cannot avoid the feeling that letting five people die in order to save one is a terrible waste. What is needed is a way of balancing the two considerations against one another. The Nash Bargaining Solution does so: maximizing the product of each individual's utility gain from the rescue suggests that one construct a weighted lottery that gives the lone individual a 1 in 6 chance of being rescued.[8] (Furthermore, the solution automatically readjusts the lottery if one adds or subtracts people from either of the two islands.)

---

[7] Taurek 1977: 293–316.

[8] To see how: assign death a value of 0, life a value of 1, so that utility numbers are the same as the chances of living assigned by the lottery. One wants to construct a lottery that assigns a chance *p* of being rescued to the lone individual, and therefore a 1-p chance of rescue to the five other people, such that the product $p * (1\text{-}p)^5$ is maximized. The utilitarian solution (0, 1, 1, 1, 1 ,1) has a product of 0, the "fair" solution (.5, .5, .5, .5, .5, .5) has a product of.0156. The product is maximized at approximately: (.165, .835, .835, .835, .835, .835, .835). Note that while the numbers come out the same in this instance as John Broome's (1998) "weighted lottery," the apparatus that generates it is different. It would count as an instance of what Martin Peterson (2009) refers to as a "mixed solution."

My intention is not to suggest that this solution is uncontroversial (although, if one had a rescue team that was composed of half consequentialists and half deontologists, who could not agree on what to do, and so one wanted to split the difference between them, this would not be a bad way to do it). My reason for describing it as minimally controversial is that it is formulated at a higher level of generality than any of specific solution concepts offered by Rawls, Gauthier, and other contractualists. Indeed, all of these theories can be regarded as instantiations of the more abstract two-principle schema—specifying the equalisandum more precisely, developing a corresponding formulation of the two principles, specifying a mechanism for trading off equality against efficiency, and of course, specifying the level at which cooperation is to be conceived. What I have outlined is a more "generic" (in the narrowest etymological sense of that term) form of contractualism.

6.2. The Puzzle for Contractualism

Suppose one were to accept something like this "minimally controversial contractualism" and agree that it provides an adequate template for the specification of a set of "principles of justice." The question then becomes, how do these principles become effective in everyday life? How does the "rational" become "real"?

Perhaps the most natural approach, in answering this question, is to build the principles directly into one's model of practical rationality, by assuming that agents in some way make use of these principles in deciding what to do. This is Gauthier's approach.[9] Thus he argues that in cases where the strategic equilibrium of an interaction is Pareto-optimal, agents reason in accordance with the standard canons of rational choice theory, but when faced with a potentially suboptimal equilibrium, they switch gears and begin to cast about for a cooperative solution. Once the players have established that their interaction partner is likely to cooperate, they apply Gauthier's favored contractualist solution concept ("minimax relative concession") to the feasible set of the anticipated interaction, then carry out the actions needed to bring about that outcome. When both parties do the same, they are able to coordinate on a cooperative solution. (One can find a similar architectonic in "strong reciprocity" models of the evolution of cooperation.[10])

This "microfoundational" approach is extremely intuitive, in part because almost every theory in normative ethics has this structure (i.e., morality is

[9] This approach is shared by contractualists ranging from T. M. Scanlon to Ken Binmore. I will not discuss either of these views here, simply because both theorists downplay the importance of cooperation.

[10] See, most recently, Bowles and Gintis 2011: 20–21.

thought to have practical effect because people take moral considerations into account when deciding what to do). When combined with the contractualist emphasis on cooperation, however, it generates some perverse consequences. This is because the principles of justice, on this conception, constrain individuals' actions only within the scope of cooperative interactions. They tell you how you should treat a person with whom you are cooperating. They do not tell you, however, with whom you should be cooperating. In this respect, the use of a prisoner's dilemma, in particular a two-player prisoner's dilemma, as the central model in the development of the theory is extremely misleading. This is because in a two-player model, each player's participation is necessary to the execution of the cooperative scheme. With three players, however, it may be the case that two players can cooperate without needing to include the third, and so strategic considerations enter into the choice of interaction partner.

In order to accommodate this added layer of complexity, the solution needs to be formulated in the language of cooperative game theory. This is designed to model interactions in which players can not only act on the basis of individual strategies, but can also form coalitions that can act on the basis of joint (i.e., cooperative) strategies. In order for the outcome of a multiplayer interaction to be stable, then, it must not only be the case that no individual has an incentive to defect, but that no coalition (i.e., proper subset of the total number of players) has an incentive to defect. An outcome that possesses this property is described as being in the core of a game. Many games, however, do not have a core—which means that for every outcome achievable by the group as a whole, there will always be some proper subset of the total set of players that could do better for each of its members by leaving the "grand coalition" and acting on its own.

To take the simplest example of this, consider a group of three adventurers who discover a treasure chest deep in the jungle. It takes two to carry the chest out, but only two. It is not difficult to imagine that, for any proposed arrangement in which the three adventurers take turns carrying the chest, then split the reward into ⅓ shares, there is a more attractive arrangement in which just two of them carry it, then split the reward into ½ shares (all it takes is for the value of the reward to significantly outweigh the disutility of carrying it). Thus while the three of them have the option of cooperating with one another as a group, they need not do so. It is possible for two of them to cooperate while excluding the third; indeed, it is advantageous for the two of them—any two—to do so. Thus the game has no core. This raises two problems: first, the resulting division of the treasure (½, ½, 0) seems to be, as Gauthier puts it, "transparently unjust." And second, the question of who gets to participate in the winning coalition and who does not seems open to being resolved in a completely unprincipled manner—two players might decide to cooperate simply because they like each other, or because they were born in the same town, or because they share what is referred to in civil rights law as "prohibited grounds for discrimination."

While this example may seem fanciful, the structure of the interaction is actually quite common, simply because the expansion of any cooperative scheme beyond a certain size often generates diminishing returns, which may give members of that scheme an incentive to admit fewer than the total number of potential cooperators. This is the classic problem afflicting worker cooperatives, for instance, which is why they were often received with hostility by egalitarian socialists (see, e.g., Webb 1891). Under a profit-sharing regime, worker co-ops have an incentive to bring in new members up until the point at which the average profit is no longer increasing. Thus they will hire fewer workers than capitalist firms, which continue to hire labor so long as the marginal contribution to profit is positive (Ward 1958: 578).

Because of this, worker co-ops will stop bringing in new members at a point at which absolute profitability could still be increased by hiring more labor. This could be simply a deadweight loss associated with that organizational form. A deadweight loss, however, is nothing but an unrealized opportunity for cooperation. Thus there is an incentive for the co-op to expand production by creating a secondary class of workers, brought in on a fixed wage, on the same terms that they would be in a capitalist firm. (One can see this sort of an arrangement in the structure of a typical law firm, with its division between "partners" and "associates.") This is, of course, also a cooperative scheme; the fact that net revenue is positive at the margin means that both the co-op members and contract workers benefit. The problem is that it creates two tiers of workers within the firm (i.e., it creates a situation where "some are more equal than others"). Furthermore, because the introduction of contract labor increases profitability without expanding the number of co-op members, it increases the profit-share of each member, and thereby encourages the charge that they are exploiting those who were not lucky enough to get in "on the ground floor." Many people—including, over the years, many socialists—have had the egalitarian intuition that workers in a firm doing the same job should have the same status, and be entitled to the same rewards. And yet the interaction here has no core. The members of the cooperative would rather not hire the contract workers at all than bring them in on equal terms.

The general problem here is that of determining what Gauthier calls the "appropriate cooperative infrastructure" (1988: 397). If there is a potential cooperative interaction between several people, is it acceptable for them to form coalitions first, and then have the coalitions enter into a cooperative arrangement with one another? If so, then determining the proper modalities of cooperation will involve applying the principles of justice several times, to several distinct cooperative surpluses. The results of this will almost inevitably be different from those that would be obtained by applying the principles of justice just once to the grand coalition. Suppose that four players can generate a cooperative surplus of $12, but a particular pair of them can generate a surplus of $8. A straightforward egalitarian division among the grand coalition would produce an allocation of ($3, $3, $3, $3). If, however, the pair is allowed to form a coalition first and split the $8 between themselves, then enter into an agreement with the remaining two to realize the additional $4, a two-step egalitarian division will produce a final allocation of ($5, $5, $1, $1). There are times when this seems appropriate, even necessary, if we hope to make the application of the theory at all tractable. Consider the case of two firms entering into a contract. One is inclined to assess the fairness of the contract by examining the way that it divides up the benefits of the particular exchange that it facilitates, while ignoring the question of how each firm then engages in an internal division of the advantages it receives (and certainly without trying to level the advantages across members of the two firms). And yet in other cases, such as the worker's cooperative, the two-step application of the principles seems to create a loophole that individuals can use to achieve outcomes that are entirely contrary to the spirit of equality.

Gauthier admits, quite directly, that he has no solution to this problem.[11] The natural temptation, of course, is to say that all this strategizing about who to cooperate with is foreign to the idea of justice. People should not be able to cherry-pick their interaction partners in such a way as to minimize their redistributive obligations. Thus the principles of justice should always be applied to the grand coalition. One should be obliged to treat as a partner in cooperation anyone with whom one can cooperate, without coalitions, partial agreements, or side deals. It is easy to see, though, that this pushes the entire framework in the direction of macrocontractualism. After all, it suggests that one has obligations of justice, not only toward those with whom one actually chooses to cooperate, but toward those with whom one merely might cooperate. This will certainly be everyone within a very large radius. So without even getting into the special problem of children, the handicapped, the sick, and the elderly—those who may not be in a position to offer any cooperative benefits to anyone[12]—there is already good reason to want to break with the

[11] Gauthier 1993: 47. Anthony Simon Laden was probably correct to point out that, within this framework, the biggest issue of justice becomes what sort of game the players wind up playing, not the particular outcome they receive (since the structure of the game is going to determine opportunities)—see Laden 1993: 48–52. In the treasure-chest example, it is certainly true that the metagame, in which it is decided which of the two will be the ones to carry it out, is where the action occurs. But Gauthier was certainly right, as well, to point out that the strategic structure of our interactions is usually unchosen.

[12] Shifting to a macrocontractualist perspective allows one to solve this problem—what Peter Vanderschraaf (2011: 119–147) calls the "vulnerability objection"—rather easily. Assuming diminishing returns to consumption, there are advantages to be had from a social system that permits individuals to shift in and out of contributory roles while maintaining some level of consumption. It follows quite immediately from the folk theorem that a system in which individuals share the benefits of their labor when they are active, and receive benefits when they are inactive, can be sustained as an equilibrium of a repeated game. The fact that some people may never become active need not undermine the equilibrium—again, for obvious folk-theoremy reasons. For an intergenerational model with a

action-theoretic, microcontractualist perspective, and to apply social contract principles to "society" as a whole.

The result is the familiar macrocontractualist framework, with its stylized representation of society as a "cooperative venture for mutual advantage," and the theory of justice interpreted in terms of principles that would bring about agreement, not in the circumstances of choice that individuals find themselves in, but in some hypothetical founding "social contract." This allows one to insist that the "basic structure" of such a society—the basic framework of law, the major social institutions that determine life chances, such as the education system and the labor market—treat everyone equally, without worrying too much about the fact that not everyone will be making a positive contribution to this cooperative project, and even among those who do, the nature and quality of the contribution made will vary enormously. This puts an end to all strategizing about interaction partners, by assuming that, for the purposes of determining entitlements and responsibilities, everyone can be assumed to cooperate with everyone else, and that differences in contribution will all come out in the wash, as it were, when generalized across society as a whole.

The downside of this construct is that it solves one problem at the expense of creating another. Even while endorsing the idea that the major set of institutions in our society should be just, we also tend to judge particular cooperative arrangements, not in terms of their contribution to the justice of society as a whole, but on a relatively self-contained basis, using the same set of principles. Strictly speaking, the macrocontractualist should not be doing so. In particular, if one truly believes that the principle of equality is derived from the contract thought-experiment, then one should not be using any sort of egalitarian intuitions when judging particular interactions or institutions. Rawls, it should be noted, was consistent in this regard, in that he refrained from any attempt to assess the fairness of particular interactions (or, in the later formulation, refrained from using "political" principles of justice to assess them). His acolytes have sometimes not paid enough attention to passages such as the following, from *A Theory of Justice*:

> It is a mistake to focus attention on the varying relative positions of individuals and to require that every change, considered as a single transaction viewed in isolation, be in itself just. It is the arrangement of the basic structure which is to be judged, and judged from a general point of view. Unless we are prepared to criticize it from the standpoint of the relevant representative man in some particular position, we have no complaint against it. Thus the acceptance of the two principles [of justice] constitutes an understanding to discard as irrelevant as a matter of social justice

similar structure—agents shifting in and out of contributory and non-contributory roles—see Heath 1997: 361–76. Critics of contractualism typically err in presupposing that "reciprocity" must involve direct reciprocity, and so fail to appreciate how flexible and robust systems of indirect reciprocity can be.

> much of the information and many of the complications of everyday life. (Rawls 1999: 87–88)

It seems clear the Rawls's macrocontractualism lacks the resources to say anything critical about the "two-tiered" workers' cooperative—indeed, this sort of inequality seems to be one of the "complications of everyday life" that must be discarded as irrelevant. First of all, Rawls makes it clear that corporations (and cooperatives) are outside the basic structure of society, since they are voluntary associations (Rawls 1999: 126). This means that their internal structure (and division of advantages) cannot be directly assessed as just or unjust. Second, the impact that the inequality between the two tiers of workers would have on inequality in society at large would be difficult to assess (and, to the extent that allowing cooperatives to hire on contract is likely to mitigate their otherwise lamentable tendency to generate unemployment, it might even generate benefits for the "worst-off representative individual"), and therefore qualifies as one of the factors that would be too complicated to assess. Thus the only criterion that seems available to assess this organizational structure is that of voluntariness and conformity to law—did it come about through an exercise of the rights and liberties that individuals are accorded by the first principle of justice, and in conformity with the relevant enabling legislation (i.e., corporate or cooperative law)? If so, then the outcome is a matter of pure procedural justice.

It is therefore not obvious that Rawls is in a position to say anything about the firm that differs in any significant respect from the libertarian-contractual view that one finds in, say, Frank Easterbrook and Daniel Fischel (1996). As long as everything is clearly announced in advance, everyone freely accepts the terms, and everyone retains a right of exit, it would seem that there should be no limits on terms of employment that a firm can offer (or the structure of shares that it can issue). Needless to say, Rawlsians have often felt that they have quite a lot more to say about these issues. Setting aside those who simply apply the difference principle directly to the distribution of advantages within the firm (ignoring all of the reasons that one cannot do this[13]), Rawlsians have tried all manner of subtle strategies to derive constraints on the way that firms can treat their workers, shareholders, customers, etc. To take just one example among many, Nien-hê Hsieh has argued (in "Rawlsian Justice and Workplace Republicanism") that the standard capitalist firm is organized on the basis of authority relations that raise troubling issues of social justice. He posits a "basic right to protection against arbitrary interference" as part of the basic structure of society, then tries to show that a right to exit from employment does not provide adequate protection of this right (Hsieh 2005: 128).[14]

[13] See remarks in Rawls 1997: 789–790.

[14] Hsieh's argument is considerably more subtle than the standard "Rawlsian" approach, which simply ignores the fact that corporations are not a part of the basic structure.

His argument, however, appeals to the cost that this type of exit typically imposes upon workers, and claims that workers cannot reasonably be expected to shoulder this burden. "Reasonable" here means, of course, "unjust," but that simply begs the question, since the idea that there is a conception of justice governing these relations that is in some sense egalitarian is precisely what needs to be shown.[15]

One can see the problem crop up in other areas as well. It explains, presumably, Rawls's extremely ambivalent attitude toward the inclusion of the family in the basic structure. While claiming that families are part of the basic structure (on the grounds that they are essential to the orderly reproduction of society as "a scheme of social cooperation over time"), he then goes on to describe them as associations that arise within that structure, which are not subject to "political" principles of justice. This is because he does not want to see their "internal affairs" subject to principles of distributive justice such as the difference principle. Yet it seems obvious that there are certain family structures, particularly those involving gender inequality, that we are inclined to regard as unjust. And yet it is difficult to see how Rawls could apply any standard but respect for basic rights, voluntariness, and right of exit.[16]

The puzzle for contractualism, therefore, stated at its highest level of generality, is simply that we are inclined to apply principles of justice—particularly conceptions of equality—at both the micro and the macro level simultaneously. On the one hand, contractualists are like most people, in that they tend to worry about the overall distribution of advantages in society at a very high level of abstraction. Thus it is widely thought that the GINI coefficient, or the poverty rate, or gender inequality in various occupational spheres, reveal something important about the "justice" of the institutional arrangements of a society. At the same time, we tend to judge particular interactions and allocation rules according to primarily internal factors, without looking at total endowments, or their impact on the distribution of advantages in society at large. For instance, an enormous amount of concern has been expressed in recent years about the "socioeconomic health gradient"[17] in Western societies, and the persistence of inequality in life expectancy. And yet no attempt has been made to correct this by modifying the distribution of health care

[15] Admittedly, in later work Rawls states that the internal affairs of associations must be governed by "some conception of justice (or fairness)," just not a "political" conception (1997: 790).

[16] This is, of course, the flip side of the problem that many philosophers have had with Rawls's approach to global justice. Because Rawls picks the basic structure as the "subject" of justice, to which the contract will be applied, he treats only relations at this level as subject to a norm of equality. This leads him to claim that inequalities that arise at both the sub (micro) and supra (international) level are unproblematic from the standpoint of justice, so long as they do not undermine the capacity of the basic structure to secure rough equality. For some reason, this has been felt to be more of a problem at the supra-national than at the sub-national level.

[17] See Hertzman and Siddiqi 2009: 27–29.

resources to the benefit of low-SES individuals. This is not to say that the distribution of health care resources is not considered subject to egalitarian norms—on the contrary, people have extremely strong egalitarian intuitions in this domain; it is precisely these egalitarian intuitions that constitute the primary source of resistance to the use of cost-effectiveness measures in this area (see, e.g., Nord 1999). Yet what distinguishes these egalitarian intuitions is that they are all "local" to the domain of health. When point systems are used for the allocation of scarce resources, for instance, patients are not given extra points for being poor or for having low education levels. The criteria that are considered salient are all related, in one way or another, to the patient's health status (Daniels 1981). As Jon Elster has observed: "Those who are entrusted with the task of allocating a scarce good rarely if ever evaluate recipients in the light of their past successes or failures in receiving other goods. Local justice is largely noncompensatory. There is no mechanism of redress across allocative spheres" (Elster 1995: 133). The question for the macrocontractualist then becomes, where does this "local" conception of equality come from? It cannot be derived from the "social contract," because not only is it not contributing to equality at that level, it will often be exacerbating it. The microcontractualist, on the other hand, needs to explain where the concern about "social inequality" at the macro level comes from.

The problem is that people have a fairly standard set of ideas about fairness, which we apply both to particular interactions, to medium-sized institutions taken singly, and to "society as a whole." But it is not clear how one can consistently move from one level to another. There is a compositional fallacy in thinking that if you guarantee that the distribution of the cooperative surplus conforms to a set of principles of justice at the lowest level of individual interactions, adding up the results of these interactions will produce a distribution of the aggregate cooperative surplus (i.e., at the level of society as a whole) that conforms to these same principles. (So, for instance, just because every cooperative interaction between men and women is one that respects principles of gender equality, it does not follow that society as a whole will exhibit what is conventionally thought of as "gender equality." It depends also on the pattern of association that prevails between men and women.) The flip side of the coin is that guaranteeing that society as a whole respects certain principles of justice provides no assurances that this will percolate down successfully and produce interactions at the lower levels that respect anything like the same principles of fairness.

I describe this as a "puzzle" and not an "antinomy" because the problem can easily be resolved by anyone willing to bite the bullet and simply apply the principles of justice at a particular level, damn the consequences at the other. Microcontractualists can adhere consistently to their view by rejecting any sort of "patterned" conception of justice at the macro level. Macrocontractualists can adhere consistently to their view by embracing voluntarism as the primary standard of rightness at the interactionist level. It is only if one wants to judge things at both levels using at least similar principles that there is a problem.

6.3. A Cultural-Evolutionary Perspective

Before concluding, I would like to indicate briefly one of the directions that contractualists might go, in order to find a solution to this difficulty. I do so not because I have any proper solution worked out, but simply because I would like to offer some resistance to the almost inevitable temptation to avoid the problem by severing the link between justice and cooperation. Thus I will be stating the position somewhat baldly, without providing much in the way of argument in support of it, much less tying up all the loose ends.

My earlier presentation of "minimally controversial contractualism" followed the conventions of the genre, in that it treated the contract thought-experiment as though it were intended to provide foundations for what Annette Baier has referred to as a *normative theory*, namely, "a system of moral principles in which the less general are derived from the more general" (Baier 1985: 232). This is the standard contractualist view: the constraints that must be satisfied in order to achieve agreement are used as the basis for a derivation of one or more extremely general principles, such as efficiency or equality, which then serve as "supernorms," from which more specific norms, such as "don't lie," or "don't steal," can be derived. Normative authority flows down, as it were, from the more to the less general (the same way that it does in a Kantian or a utilitarian view).

An alternative way of interpreting these very general principles is to treat them as explicitative vocabulary (in Robert Brandom's sense[18]) that we introduce in order to talk about broader patterns in our practices of normative inference. (This is a general characteristic of moral vocabulary: a term like "ought," for instance, serves the explicitative role of transforming imperatives into assertions, allowing us to reason about them, by embedding them in conditionals.) According to this view, primary normative authority rests with the low-level moral norms, which form part of a complex artifact that is reproduced through cultural inheritance. Thus we learn, from our parents, teachers, and peers, a set of specific rules to govern our conduct in everyday life, which include a wide range of techniques for managing conflict and creating

[18] Brandom 1994: 105–107. Brandom actually uses the term "expressive," which is a somewhat unfortunate choice of terms for the purposes of reflecting on moral vocabulary, because "expressive" in Brandom's sense of the term has nothing to do with the "expressivist" tradition in metaethics. In Brandom's sense of the term, first-order, thick moral concepts are not expressive. It is only the vocabulary that gets introduced in order to talk about these first-order judgments that is expressive.

habituated patterns of prosocial behavior. This is the standard repertoire of rules that any parent is familiar with: not to hit people and grab things, how to form a queue and wait one's turn, techniques for allocating goods both divisible ("you cut I choose") and indivisible ("eeny-meeny-miny-mo"), deference to legitimate authority, suppression of our tendency to enjoy cruelty, and so on. In principle, there need not be any commonalities between the way that one type of situation is handled and the way that we approach some other. In other words, a culture might have (and many do have) a very specific way of dealing with one area of social life (e.g., marital obligation) and another quite different way of dealing with some other area (e.g., social labor), so that if one were to ask in general terms "what we owe to each other," the answer would simply be "it depends." As both formal models of cultural transmission and generations of ethnographers have shown, cultural inheritance is able to sustain almost anything as a normatively enforced pattern of behavior, and consistency across domains tends not to be an important feature (or at least not upon surface inspection) (Boyd and Richerson 2005a: 166–188).[19]

Nevertheless, since culture forms a system of "descent with modification," it exhibits evolutionary dynamics that are in several key respects comparable to those that prevail in the biological realm.[20] Each social norm must compete for adherents with other variants that inevitably crop up, both from internal deviance and dissent, as well as from contact (and often conflict) with other cultural groups. The structure of this competition, however, is not neutral with respect to all variants. Our innate psychological dispositions, for instance, although seldom determinative, certainly make some patterns statistically more likely to be reproduced than others. (For example, while compulsory incest and compulsory incest-avoidance have both been normatively enforced at different times in different societies, incest-avoidance has been far more common as a norm. Similarly, while there have been and are societies that practice polyandrous marriage, polygyny has been by far the more common norm.) The fact that we are inclined to find certain actions easier, or more gratifying, or more repulsive, shows up as a *bias* (in the non-pejorative sense of the term) in the cultural inheritance system.

While our innate psychology provides a set of *content biases*,[21] cultural evolution is also subject to a set of *pragmatic biases*, which arise out of the structure of social interaction. I would argue that the pragmatic considerations that

---

[19] This is not to deny that there are human universals: see Brown 1991. It is noteworthy, however, that most of the ideas that we take to be central to morality are not on the list, and therefore stand in need of explanation in cultural-evolutionary terms.

[20] Of course, this means that it is in many other respects non-comparable. See Richerson and Boyd 1995: 69. The most important difference involves what they refer to as "guided variation."

[21] Shaun Nichols, for instance, has argued that moral sentimentalism is, in effect, an illusion produced by the operation of such biases in the reproduction of social norms—such that norms with greater "affective resonance" have a better chance of reproducing (Nichols 2004).

speak in favor of the two principles of justice outlined above, efficiency and equality, also favor norms exhibiting those properties in the process of cultural evolution. Why? Because these norms favor arrangements that attract fewer complaints, and thus less motivated dissent, than any of their near rivals. Social norms, despite being enforced, still require—as a matter of sociological fact—high levels of voluntary compliance. In other words, they must be able to attract agreement—not necessarily consensus, but at least high levels of agreement. This is because the punishment system itself is stable only to the extent that it is normatively enforced (particularly so when it is decentralized and informal). Thus a certain willingness to play along in good faith is essential to the stability of normative systems. The more they attract objections, the less stable they will be, and the more they will tend to be replaced by systems that attract fewer objections.

This basic framework for understanding the principles of justice has been proposed by Jürgen Habermas, although unfortunately without much uptake. Partly this is because Habermas moves beyond the pragmatics of social interaction to make a series of more controversial claims about the way that structural features of linguistic practice bias cultural transmission.[22] There remain, however, important similarities between his account and the more minimal one sketched out here. First of all, both views hold that the principles of justice—in this case efficiency and equality—have no intrinsic normative authority, they merely articulate the end result of a process of cultural evolution. What is being posited is nothing more than a bias, the consequences of which are only felt in the fullness of time, and only when not overridden by other forces. But because the pragmatic features of interaction that favor efficient, egalitarian norms obtain in multiple domains of social interaction, different practices will tend to evolve in the same direction, or in such a way that they exhibit certain shared structural features. The result is that we are able to make some very robust generalizations about "what we owe to each other." But this is not because our more specific obligations are derived from the more abstract principles that we use to articulate these obligations; it is because the abstract principles were introduced as a way of talking about (in particular, generalizing about) the more specific obligations.[23] (Among other things, this framework is also able to explain why, within a culture such as our own, there can be high levels of convergence around low-level moral

---

[22] Habermas claims that "when it becomes linguistically channeled, social reproduction is subject to certain structural constraints; and . . . by reference to these we can—not causally explain, certainly, but—render reconstructively comprehensible, in their inner logic, the . . . structural transformation of worldviews, the universalization of law and morality, and the growing individuation of socialized subjects" (Habermas 1985: 86–87).

[23] For further discussion, see Heath 2008b: 272–273.

judgments, combined with deep and persistent disagreement over abstract principles.[24])

Of course, once we have developed the explicitative vocabulary (having achieved what Brandom calls "semantic self-consciousness" [1994: 384]), we are able to engage in an explicit attempt to direct the evolution of norms in the direction of increased efficiency and equality. For instance, we are able to make use of the principles of justice in a self-consciously "political" fashion, in cases where we recognize the need to minimize disagreement. And when designing and implementing new systems of cooperation, we can do so in a way that reflects an explicit concern for equality and efficiency, and can appeal to the authority of those principles as such. This amplifies the force of what evolutionary theorists refer to as "guided variation," further biasing cultural evolution in the direction of contractualist norms, and further enhancing the generality and authority of those norms. This autocatalytic process, which was triggered in our society by the Enlightenment and the emergence of liberal political orders,[25] has provided the central dynamic of moral transformation in our society, and is what accounts for the extreme instability of moral ideas in the past century.[26] Thus, general normative principles, in this view, are not extra gears, merely by virtue of being explicitative. But they are considerably less central to the mechanism of moral judgment than many philosophers have taken them to be.

---

[24] Jonsen and Toulmin (1988), explain their return to casuistic methods through a practical illustration of this phenomenon. The example arose from the participation by one author in a commission struck by the US government, in order to provide guidance on various bioethical questions. Commissioners were intentionally chosen with an eye toward diversity in several dimensions: "men and women; blacks and whites; Catholics, Protestants, Jews and atheists; medical scientists and behavioral psychologists; philosophers; lawyers; theologians; and public interest representatives" (17). Expecting a high level of disagreement, what he found instead was a fair degree of convergence on practical questions. "The locus of certitude in the commissioners' discussions did not lie in an agreed set of intrinsically convincing general rules or principles, as they shared no commitment to any such body of agreed principles. Rather, it lay in a shared perception of what was specifically at stake in particular kinds of situations. Their practical certitude about specific types of cases lent to the commissioner's collective recommendations a kind of conviction that could never have been derived from the supposed theoretical certainty of the principles to which individual commissioners appealed in their personal accounts. In theory their particular concrete value judgments should have been strengthened by being 'validly deduced' from universal ethical principles. In practice the general truth and relevance of those universal principles turned out to be less certain than the soundness of the particular judgments for which they supposedly provided a 'deductive foundation' " (1988: 19).

[25] What I take to be fundamental here is Pierre Manent's idea that liberal societies are the first societies to be organized according to an idea of how a society should be organized (see Manent 1995: pp. xv–xviii).

[26] The result has been as Michele Moody-Adams describes: "Much of the moral language that helps shape the economic, social, and political dimensions of the contemporary world is a product of distinctively philosophical efforts to articulate interpretations of the structure of moral experience" (Moody-Adams 1997: 194).

6.4. Implications

So how does this way of looking at things help to solve the puzzle? It does so by suggesting that the egalitarian intuitions deployed by the average person in our society are neither "built up" from a set of principles governing individual interactions, nor are they "inferred down" from a conception of how society as a whole should be ordered. They are instead a product of simultaneous, convergent cultural evolution in different domains of interaction, catalyzed by the reflexive use of explicitative vocabulary in practices of social criticism and reform. This is why our egalitarian intuitions across different domains only sometimes add up in a way that is consistent with egalitarianism at the level of society as a whole—it is because they are not derived from a commitment to an abstract principle of equality.

Adopting this perspective helps to explain some of the peculiar wrinkles in the way that we apply egalitarian ideas. Take, for example, what has come to be known as the "Titanic puzzle." It arises from a rather casual remark in Thomas Schelling's *Choice and Consequence*, in which he suggested that the R.M.S. Titanic had an inadequate number of lifeboats because passengers in 3rd class (or "steerage") were expected to "go down with the ship," and that this was somehow part of the conditions of carriage associated with the less expensive tickets (Schelling 1984: 115). The puzzle is then as follows: assuming that we find it unacceptable for passengers on the same boat to have differential access to lifeboats, on the grounds that some did and some did not pay for this safety feature, how then can we accept an arrangement under which passengers on *different* boats, having paid different prices for carriage, have access to different levels of safety?[27] (After all, different ships provide different levels of safety, in the same way that different automobiles do.)

The standard micro and macrocontractualist frameworks seem unable to capture what is troublesome about this case. From a micro perspective, there would seem to be nothing wrong at all with passengers on the same ship having differential access to lifeboats—indeed, insisting that there be enough lifeboats for everyone is likely to create a deadweight loss, since it will raise the price of tickets, thereby making them unaffordable to a thin slice of consumers who would have been willing to pay just slightly less for carriage, and who could have been squeezed onto the ship. If a group of passengers enters into a transaction with White Star Lines that is mutually beneficial, and the terms of that transaction are internally just, how can it be relevant that some other group of passengers enters into a different transaction with White Star

[27] See Satz 2010: 88. I hesitate to use this example, because it risks perpetrating an urban myth, since the account of conditions on the Titanic is entirely fictitious (indeed, the suggestion that there was a policy of denying 3rd-class passengers access to the lifeboats was vehemently denied by White Star Lines).

Lines, which is also internally just, but features a different set of terms? There is no presumption of equality between the passengers with respect to lifeboat access, because the passengers are not cooperating with one another in order to secure the provision of this good.

The macrocontractualist is in a better to position to criticize the "steerage goes down with the ship" arrangement, because he can say that, from the standpoint of society as a whole, it is a violation of equality for some people to be exposed to mortal dangers that others are able to protect themselves against. But then she is unable to explain why we are untroubled by the fact that different ships have different safety standards, and that passengers might choose one ship over another because these differences resulted in a lower price of transport. What is it about being on the same ship that somehow makes it troublesome for access to lifeboats to be distributed in accordance with ticket price?

Rather than searching for general principles from which this particular constraint can be derived, there is much to be gained simply by noting that shared transportation is a particular type of cooperative enterprise, which is subject to a distinct set of norms that have evolved and adapted over time. Typically these norms require greater forbearance than is expected in everyday interaction—willingness to tolerate greater encroachment of one's personal space, deference to the authority of the captain or driver, restrictions on one's ability to engage in activities that might jeopardize the safety of others, or slow down passage, and also a set of procedures for dealing with emergencies. It is the latter set of norms, I would suggest, that generates the "puzzle" in the Titanic scenario—the view that it is impermissible for passengers on the same boat to have differential access to lifeboats, even though there is no general social requirement that safety on different ships be equalized. Emergencies typically evoke a higher level of social solidarity than everyday interactions, and so the norms governing them are often more egalitarian. As a matter of historical record, the norm that actually governed the evacuation of the Titanic was "women and children first"—to the point where men were barred entirely from entering lifeboats on one side of the ship. (Indeed, Schelling's claim that passengers in steerage were expected to go down with the ship is simply false. Survival rates among women even in steerage were much higher than that of men, including those in first class. The lower survival rate of passengers in steerage can be almost entirely explained by the lower percentage of women traveling third class, along with the physical positioning of the lifeboats on the upper decks of the ship.) This is, one might note, not exactly an egalitarian norm—it discriminates on the basis of gender and age. It is just that within the different groups (men, women, children), it does not permit further discrimination (based on, say, seating class), but rather applies a queuing norm of "first come, first served" (see Elster 1995: 73–74).

The best way of thinking about this example, I would suggest, is to regard "travel by ship" as a particular type of cooperative practice, governed by a distinctive set of norms (i.e., "naval tradition"). Differential access to lifeboats, based on carriage class, violates these norms, in the same way that trying to buy your way into line at a movie theatre violates the norms governing the practice of queuing. When one is dealing with different ships, on the other hand, the same norms do not apply. (There are of course different norms that apply between ships, such as the obligation to divert course in order to effect a rescue. These are part of a system of generalized reciprocity that has a cooperative structure, but the overall objectives are different.) The same sort of structure can be found in a variety of social institutions. Andrew Stark, for example, has developed a particularly careful analysis of the role that egalitarian norms play in debates over parental fund-raising in American public schools. These practices are characterized by a tension between, on the one hand, very little commitment to a norm of *interschool* equality, combined with a strong commitment to *intraschool* equality (Stark 2010: 60–65). Thus parents are not allowed to fund-raise for resources or activities that will benefit their own child's class; they must provide a benefit to the *entire school*. And yet parents feel no obligation whatsoever to address the sometimes glaring disparities in per-pupil public spending that prevail across different municipal school districts. From either a micro or a macrocontractualist perspective, it is difficult to see how anyone would ever have hit upon "the school" as the appropriate institutional level at which to apply egalitarian norms. From a cultural-evolutionary perspective it is much less mysterious.

The idea that particular sets of norms are tied to particular systems of cooperation remains important, in part because it explains why egalitarian principles are so often "local" in their application. The cultural-evolutionary model proposed above suggests that distributive obligations are likely to be limited in scope to those who are directly involved in the cooperative scheme, simply because they are the ones whose voluntary compliance needs to be secured, and therefore, the ones whose complaints need to be addressed in order to secure reproduction of the norm. Thus it is hardly surprising to find "special-purpose" norms that are adapted specifically to regulate the behavior of passengers on the same ship, since these are precisely the people who need to cooperate with one another in order to get various things done. One might be tempted to call this "mesocontractualism," except that this generates the misleading suggestion that there might be some way to draw a crisp line to demarcate the scope of a particular cooperative enterprise. Determining the scope of cooperation, however, is not a specifically philosophical problem, because individuals' own ability to manage their daily affairs depends upon an ability to classify interactions using the appropriate "cultural template," to identify the type of situation they are in, and to apply the relevant norms. The scope of their obligations is ultimately defined by these norms and the attendent institution. The specifically philosophical project of offering a rationale for these norms, using abstract concepts such as "cooperation," and "equality," should not be seen as an attempt to unearth the reasons why people have the intuitions that they have about specific cases. On the contrary, the philosophical articulation and refinement of these norms primarily serve to enable a more "clairvoyant" continuation, or critique, of the practice. (For example, the low level of commitment to interschool equality in the American context provides a perfect illustration of the type of practice that one might want to use the language of equality to criticize. It is important, however, not to mischaracterize the way that this type of critique is carried out. It is generally ineffective to condemn the practice merely on the grounds that it conflicts with some supposedly foundation norm of equality. Effective criticism, as Stark illustrates, merely uses the language of equality to render explicit what is already implicit in the practices that secure intraschool equality, and projects outward from that.)

The reason that we seek to apply egalitarian norms at a more abstract level—and so worry about things like "the distribution of wealth" or "gender inequality" in society as a whole—is twofold. First, the development of the modern nation-state has generated systems of cooperation that genuinely do encompass all members of society. For example, once a uniform system of property rights is put into place—a system chosen from an enormous menu of options, each with different implications for the production and the distribution of goods—then the consequences that this system has for inequality in society as a whole becomes a legitimate object of concern and critique. The same sort of consequences flow from the development of universal childhood education, conscription, comprehensive health insurance, increased legal regulation of family relations, and so on. The second reason is that we have been and continue to be engaged in the project of constructing institutions that will enable us to cooperate on a larger and larger scale. This is being done self-consciously, as it were, explicitly using the principles of efficiency and equality in our processes of institutional design, precisely out of a recognition that we lack the shared cultural resources that could bring about consensus around some thicker set of norms. Thus we articulate our aspirations, in terms of collective action problems that we would like to see solved and systems of cooperation that we would like to see institutionalized, in the contractual language of efficiency and equality. This exercise is going to look like macrocontractualism, in the sense that both the principles and the system of cooperation are going to be conceived of in a very stylized way (so, e.g., we may talk about controlling global warming as a "public good," even though, strictly speaking, not everyone benefits).

As a result of these institutional developments, we wind up with a set of normative commitments that are, if not formally inconsistent, then at the very least in tension. We tend to have strong views about how particular interactions should be organized, in order to meet certain standards of justice, but we often become uncomfortable with the aggregate consequences of organizing these interactions in this way, and so cast about for ways to tweak or rearrange things, so that the large-scale outcome is one that satisfies our conceptions of justice at that level. Reconciling these tensions is not primarily a philosophical problem, but a practical problem—that of finding ways to bring our institutions into line with our ambitions and ideals, so that they form an integrated and consistent system.

6.5. Conclusion

Philosophical debates about the requirements of justice have been drifting, over the course of the past few decades, toward increasingly abstract concerns about "equality" at the level of society as a whole, or the entire human race. Those who have resisted this tendency have attracted a degree of opprobrium, based perhaps on the suspicion that, when push comes to shove, the real basis for their opposition to this expansive conception of equality is that they are rich Westerners who do not want to share their wealth. Examining the case more charitably, however, it quickly becomes apparent that there are a variety of motives. Perhaps the most common reason for concern is that it makes the theory of justice utopian in the pejorative sense of the term (and therefore useless when it comes to addressing any real-world political questions). Elizabeth Anderson has suggested, along these lines, that "in focusing on correcting a supposed cosmic injustice, recent egalitarian writing has lost sight of the distinctively political aims of egalitarianism" (Anderson 1999: 288).

There are, however, somewhat narrower, more philosophical reasons for concern. The major one, which I have focused on here, is that by applying the principle at this abstract level, one loses sight of the role that equality—or more diffuse conceptions of fairness—play in mediating interpersonal relations and institutional decision-making. University professors, for instance, care very much what salary the person down the hall is drawing and tend not to evaluate the fairness of that arrangement in terms of its contribution to global equality. The principal of a school, faced with excess demand for enrollment, is concerned to implement impartial admission procedures, but feels no need to calculate what impact this will have on inequality in society as a whole. A group of entrepreneurs, going into business together, settle on a division of ownership shares based on a conception of fairness, typically one that ties contribution to reward among the partners. The examples can easily be multiplied.

The most natural account of what people are doing when they apply these sorts of norms, and in particular, why they gravitate toward egalitarian norms, is that these norms are conflict-minimizing—as, indeed, anyone who has participated in this type of decision-making can attest. (The best way to discover the attractions of egalitarian norms is to experience the consequences of failing to apply egalitarian norms.) This is the intuition underlying the contractualist idea that the principle of equality (and, more obviously, the principle of efficiency) might be valued for its ability to bring about agreement in cooperative enterprises. Developing this plausible-sounding idea into a fully specified theory, however, has proven difficult. If one treats contractualism as having the structure of a "normative theory," with a single set of very abstract principles which then get applied in a way that generates more specific normative constraints, then one quickly winds up caught up in the puzzle described in the first two sections: there is no way to establish a consistent micro-macro link. It can be avoided, I have suggested, by instead adopting a cultural-evolutionary framework, and viewing micro-macro consistency as a social project, rather than as a logical requirement of normative theories.

7 }

Efficiency as the Implicit Morality of the Market

The idea that "business ethics" involves some sort of a contradiction in terms is a widespread misperception. The attitude, however, arises as a response to a genuine phenomenon, which is that business ethics is in certain respects deeply counterintuitive. The reason for this, I will argue, is that business ethics is structured by the norms of a market economy, and the market is governed by a set of third-best normative principles. Furthermore, within this framework, business ethics is typically undertaken as an exercise in non-ideal theory. By contrast, our everyday impulse, when making moral judgments, is to reason from first-best normative principles, under the presuppositions of ideal theory. To get from this to a third-best, non-ideal framework involves such a dramatic set of concessions in the direction of the "merely empirical," that many people find the end product so dissipated as to be bereft of normative authority. In other words, business ethics winds up being so far removed from "the best" that many people have difficulty recognizing it as being "ethical" in any worthwhile sense of the term.

My objective in this paper is to undo this impression, by showing how a set of third-best principles can nevertheless provide a basis for a robust set of constraints on the behavior of individuals in economic contexts. In order to get to this, I will begin by articulating what I mean by first, second, and third-best principles, as well as clarifying the distinction between ideal and non-ideal theory. This involves a slight regimentation of terminology, but nothing outside the scope of what has been circulating in the (increasingly voluminous) literature on the subject. In the end, what I want to argue is that the guiding idea in business ethics should be the principle of Pareto efficiency. This principle underlies what Christopher McMahon has called "the implicit morality of the market," which he explains as follows:

> The implicit morality of the market consists primarily of the hypothetical imperatives which are generated by economic theory when the achievement of economic efficiency is taken as an end. Certain conditions must

> be satisfied if a free-enterprise system is to allocate resources to producers and distribute products to consumers in a Pareto-optimal way. And from these conditions various requirements on the behavior of economic agents—they might be called "efficiency imperatives"—can be derived. (McMahon 1981: 255)

This in itself is not overly controversial. The claim that I want to make, the one that is controversial, is that these "efficiency imperatives" are pretty much all there is to business ethics, at least with respect to market transactions. Indeed, I will argue that the Pareto principle forms the normative core of what I have called a "market failures" approach to business ethics, which provides a framework for thinking about all of the issues that are traditionally classified under the heading of "corporate social responsibility."

This approach is counterintuitive for two reasons. First of all, there is the well-known fact that market interactions involve a "suspension of altruism" (McMahon 1981: 253), and so the implicit morality of the market does not include some of the obligations that one finds in everyday morality. This is largely a consequence of the fact that markets are competitively structured, and so require an adversarial orientation on the part of actors, a feature that is hardly specific to markets, but is rather a generic feature of competitive interactions (Applbaum 1999). (There is a "suspension of altruism" in competitive sports as well, at least toward one's opponent.) This is something that many people still find difficult to accept, but two centuries of "invisible hand" rhetoric have certainly made the idea less than entirely foreign. The second counterintuitive feature, which is slightly less familiar, is the fact that the norms of market interaction assign priority to efficiency considerations, and so require a suspension of our everyday concerns about fairness or equality. This is not, I will argue, the result of distributive considerations being normatively insignificant, it is simply a concession that we are forced to make when switching from the second-best to a third-best framework.

What I would like to emphasize, however, is that neither of these concessions has the effect of transforming markets into a "moral-free" zone (as some have described it [Gauthier 1982]). The idea that it does is sometimes abetted by a persistent confusion between efficiency in the engineering sense and efficiency in the Pareto sense, which one can see even in the quote from McMahon above (which describes the efficiency imperatives, misleadingly, as hypothetical imperatives).[1] The Pareto principle states that if some transformation of the status quo is able to make at least one person better off, by his or her own lights, and no one worse off, then from an impartial point of view (and *ceteris paribus*) the outcome of that transformation is normatively superior to the status

[1] For a more extended discussion of this distinction, see Heath 2001b. For an example of the stubborn refusal to grasp this distinction, see Stein 2002: 68–70.

quo. When one looks at the structure of a collective action problem, it is obvious that this principle is not simply a counsel of prudence, or a canon of instrumental rationality, because individual self-interested action will often fail to bring about such improvements. Thus a commitment to Pareto efficiency serves as a genuine constraint on the pursuit of individual self-interest; it is, as John Rawls pointed out long ago, a principle of justice (1971: 58–62).

Yet even among those who do not make the mistake of treating efficiency as a purely instrumental principle, it is still often regarded as a somewhat impoverished, or at least uninspiring, moral ideal. As a result, a conception of business ethics based entirely upon this principle seems to bear the taint of moral laxity. The proper way to defend it, I will suggest, is not to show that efficiency is adequate as a stand-alone morality, or that it is superior to other principles at the level of some abstract idealized conception.[2] The proper defense, I will suggest, focuses on the reasons for adopting a third-best framework, then tries to show that efficiency is the most attractive normative ideal within that framework.

7.1. The Ethics of Third Best

My classification of moral theories into categories of "first" and "second" best is, of course, an invocation of Richard Lipsey and Kelvin Lancaster's celebrated distinction in their article, "The General Theory of Second Best" (1957). Although the way that I intend to use the term is rather different in the details, there is an important lesson in the Lipsey and Lancaster paper, the spirit of which I would like to preserve. What economists found surprising in Lipsey and Lancaster's analysis was the way that it overturned the conventional assumption that, when a first-best outcome is unobtainable, the best course of action will be to approximate the conditions required to bring about that outcome, with the thought that this will bring us as close as possible to it. Lipsey and Lancaster's analysis is aimed specifically at the first fundamental theorem of welfare economics, which shows that a perfectly competitive market will also be Pareto-optimal. It is well known, of course, that the conditions that must be satisfied in order for perfect competition to obtain are highly idealized (Schultz 2001). The significance of this had often been underestimated, however, because economists succumbed to what I have elsewhere called the "frictionless plane fallacy" (Heath 2009).[3] They

[2] This was the error committed by the Richard Posner (1973) and the early enthusiasts of the law and economics movement. For useful discussion, see Ronald Dworkin (1980).

[3] Versions of this fallacy can be found all over. Perhaps the most prominent is Friedman (1982: 120), cited in chapter 1. Also Buchanan, "The case for the market on grounds of efficiency depends on the extent to which actual markets do approximate, or can be modified to approximate, the ideal market" (1985: 15).

assumed that if these conditions are approximately realized, then the favored outcome—Pareto efficiency—would also be approximately realized (specifically, that one will achieve the *constrained* Pareto optimum). What Lipsey and Lancaster were able to show is that, if the conditions required for perfect competition cannot be satisfied (for some "merely empirical" reason), then satisfying them as much as possible will not (except *per accidens*) produce an outcome that is as close as possible to the Pareto-optimum. On the contrary, it will almost always be worse, not better. As a result, the type of policy recommendations that one would be inclined to make within the first-best framework, about how the economy should be organized, have no authority once it is recognized that the first-best outcome cannot be realized. As soon as a single, recalcitrant fact makes it impossible to achieve the first-best, one must switch to the second-best framework. And at that point, any presumption about what the best course of action is must be suspended. Second-best reasoning is therefore not just a shadow, or an approximation, of first-best reasoning; it is a very different exercise.

What I would like to propose is a generalization of this idea (one that makes slightly more metaphorical use of the vocabulary than the extension proposed by Robert Goodin [1995]). Moral theories may be constrained by a variety of "merely empirical" circumstances, many of which are facts of human psychology (understood broadly, to include the scope of developmental plasticity and social learning, the limits of computational power, memory, and attention, and the types of biases identified by behavioral economics). A first-best framework is one that ignores all of this *in the formulation of its principles*. The idea is that one must first identify the demands of morality and that only once this has been settled can one move on to the question of how it ought to be implemented, or brought about in the world. It is only at this second stage that all of the familiar frailties of human nature may be taken into consideration.

A familiar example of a first-best normative theory would be act-utilitarianism, which claims that morality requires perfect altruism on the part of all agents—we must be indifferent between our own happiness and that of any other person, and we should be prepared to make arbitrarily large personal sacrifices whenever doing so can be expected to produce compensating benefits for others (calculated on a 1:1 ratio). This is a perfect description of the evolved psychology that we would no doubt possess if we were all biological clones, but of course it fails to reflect very closely the psychological dispositions of our species such as it exists. Furthermore, very few people believe that this kind of self-transcendence could be achieved even through the most complete and intensive process of socialization. Yet for theorists of the "first-best," this does not impugn the ideal. All it requires is that the theory be extended, from the context of "ideal theory," in which full compliance can be assumed, to that of "non-ideal theory," in which it is recognized that individuals may fail, for various reasons, to do what is morally required of them. Most obviously, this means that they may require some non-moral incentive in order to comply with the moral rules. Thus many theorists of the "first-best" bring in social institutions at this point, with the specification that their role is to bring about the realization of an antecedently specified moral end. Beyond this, the set of normative principles must be extended to include an account of what constitutes an appropriate response to non-compliance; thus one must articulate a theory of just punishment, of restorative justice, and so on. All of this will be oriented toward getting as close as possible to satisfaction of the first-best principle.

This might be the end of it—and according to many theorists it is the end of it. However, one might also find that the movement from ideal to non-ideal changes the framework to a degree that it no longer makes sense to pursue first-best principles. Thus one might feel inclined to reformulate the normative principles, in recognition of the human frailties introduced in the transition from ideal to non-ideal theory.[4] There could be a variety of reasons for this, depending on the normative theory in question, but three in particular seem salient. Most obviously, the implementation problems that arise at the non-ideal level might make it impossible to apply the principles. (For example, one might find act-utilitarianism attractive as a first-best ideal theory, but then decide, for a variety of reasons arising from the non-ideal conditions in which we live, that it is impossible to make meaningful interpersonal comparisons of utility, and therefore that it no longer makes sense to try to maximize the aggregate of utility.) Second, one might have reservations about the amount of force that would need to be employed in order to achieve an acceptable level of compliance with the principles. The institutionalization of norms, including moral norms, relies upon a combination of internal motivation and external sanctions (Parsons 1951). Generally speaking, deficits on one side of the ledger must be made up for by surpluses on the other. With some normative principles it will naturally be harder to elicit voluntary compliance from individuals than with others, and so institutionalization will require more extensive reliance on external sanctions. And yet one might also think that there is a limit on the amount of coercion society should be willing to apply in order to achieve its moral ideals. If a particular principle exceeded this threshold (according to some independent specification of where that threshold lies), then one might think it was time to consider a new, less motivationally demanding principle.

---

[4] David Schmidtz approaches the issue from a somewhat different direction, but makes essentially the same claim: "compliance is an endogenous variable; the extent of compliance is not externally determined but is instead a function of the principles chosen. When we choose a principle, and any particular way of trying to put it into practice, we choose a compliance problem at the same time. We cannot set aside compliance as something to address later, because our task of choosing a principle we can live with is a task of choosing a compliance problem we can live with" (2011: 778).

Finally, one might think that a first-best principle became self-defeating, once its implementation problems were taken into consideration. For example, if the motivational demandingness of a principle would make it very costly to implement, and yet the principle itself condemns costliness, then the principle itself might recommend its own replacement with another, less costly, alternative.

This final idea is what underlies the familiar argument that marks the transition from act-utilitarianism to rule-utilitarianism. There has always been some question as to whether the latter articulates a different set of normative principles, or whether the "rules" are just an implementation strategy for the former. Nevertheless, it is clear that many rule-utilitarians (or rule-consequentialists) consider their doctrine to be based upon genuinely different principles. Brad Hooker, for example, in *Ideal Code, Real World*, claims that the introduction of real-world constraints should lead consequentialists to adopt rule-consequentialism as their preferred moral theory, and that this represents, not just an application, but rather a modification of act-consequentialism (2003: 78–79).[5] Not only does it shift the focus of maximization from actions to rules (thereby reducing information problems), but in Hooker's formulation it also requires that the rules be susceptible to internalization (thereby reducing motivational problems). These are substantial differences, in Hooker's view, which he emphasizes as a way of rebutting the charge that rule-consequentialism, upon closer inspection, collapses into act-consequentialism (2003: 95). The differences are reflected in the fact that a rule-consequentialist and an act-consequentialist, facing identical empirical constraints and possessed of the same information, might genuinely disagree about what morality requires one to do in a particular situation.

There is no need to assess the success of this argument, for our purposes it is sufficient to observe that the distinction between ideal and non-ideal theory, as conventionally drawn, does not get at the difference between these two forms of consequentialism. One can easily imagine a set of rule-consequentialist principles being elaborated under the assumption of full compliance with rule-consequentialist principles (and thus not addressing questions of punishment, civil disobedience, and so forth). Thus it is important to develop some additional vocabulary, in order to distinguish different forms of "idealization" that normative theories may be subject to. This is, in fact, a fairly constant theme in the recent literature on "ideal and non-ideal" theory (Simmons 2010; Robeyns 2008; Valentini 2012). My major suggestion is that the vocabulary of "first-best," "second-best," (etc.) be used as a way of drawing this distinction

[5] Hooker also argues that there are independent reasons for favoring rule-consequentialism, the view is not just derived from act-consequentialism.

between the different levels of idealization at which normative principles can be formulated. (In this respect the Lipsey-Lancaster analogy is slightly inapt, because they continued to use the Pareto principle as a normative standard even within the second-best framework. A closer analogy can be found in the work of economists who, realizing that the Pareto-efficiency principle will never be satisfied in practice, adopt a weaker normative standard such as Kaldor–Hicks efficiency as a guide to action in second-best contexts. Although many downplay the significance of this shift, it actually represents a major change in the normative principle.)

Of course, the most celebrated contemporary example of a second-best theory is John Rawls's theory of justice, and in particular, the difference principle. It is an exercise in "ideal theory," since Rawls assumes full compliance with the principles of justice in the development of his theory (1999: 7–8, 215–216). It is second-best, however, because it has as its core a system of "mutual benefit contractualism," which is motivated in part by what he calls the "strains of commitment." One of his major complaints about utilitarianism is that the expectation that individuals be willing to give up their own life projects, in order to benefit others, is simply too "extreme" a demand, given the facts of human psychology. (For parties in the original position, "it is unwise if not irrational to choose principles which may have consequences so extreme that they could not accept them in practice" [1999: 155].) Thus our overall approach to thinking about questions of justice, according to Rawls, should be guided by some conception of what people might reasonably be motivated to conform to.

A similar pattern of reasoning underlies Rawls's derivation of the difference principle, which makes explicit appeal to incentive problems that would arise with any stricter form of egalitarianism. At a first-best level, one need not worry about trade-offs between equality and efficiency. In the real world, however, people may be inclined to work less if their earnings are taxed away at a rate of 100 percent. Such a tax rate would, in turn, result in lower levels of production, and hence losses in the efficiency dimension. The fundamental motivation for the difference principle is the idea that we should be willing to tolerate certain inequalities, if doing so will result in the worst-off being better off than they would have been under more egalitarian arrangements. Rawls suggests that these inequalities may be regarded as "concessions to human nature" (1958: 173).

Many of Rawls's readers have felt that this line of reasoning is precluded by his adoption of an "ideal theory" framework. G. A. Cohen (2008), for instance, has interpreted the "ideal theory" constraint as one that precludes any appeal to incentive considerations in the formulation of the theory of justice. "Rawls assumes full compliance," he argues, "so if justice requires that people work without compensation, then one must assume that that is what they will do, at least when it comes to formulating the principles of justice.

The question of how to get them to *actually* do it is simply an implementation problem, not an issue that should affect the formulation of the normative principles."

The superficial plausibility of this line of argument is due to a confusion, which is in many ways encouraged by Rawls's discussion, but which the regimentation of terminology proposed here is intended to remedy. Once the distinction between first-best and second-best frameworks has been drawn, the central difference between Rawls and Cohen can be stated simply: Cohen is working at the level of first-best, ideal theory, while Rawls is also working at the level of ideal theory, but within a second-best framework. He is developing a set of principles of justice that are, in part, a response to the human frailties that make certain first-best theories unworkable, or unsuitable as candidates for a theory of justice. Most importantly, Rawls is working under the assumption of limited altruism. (In later work he also proposes, as one of the attractions of the second-best theory he has developed, that it is compatible with a wide range of first-best theories, including both Kantianism and utilitarianism [Rawls 1996: 170]. In this case, the "merely empirical" detail that prevents the implementation of first-best theories is what he calls the "fact of pluralism." This is one way of interpreting what it means for a theory of justice to be "political.")

This is a familiar landscape, albeit one that I am attempting to redescribe in at least slightly different terms. The one new suggestion I would like to make, however, is that the process from which a first-best framework can be transformed into a second-best framework is iterative, and can in turn be used to generate a third-best framework (and, should one so desire, a fourth-best, fifth-best, etc.).[6] Once a second-best normative principle is formulated—such as the difference principle, which accommodates the "real-world" need for trade-offs between efficiency and equality—new implementation problems may still arise with the transition from ideal to non-ideal theory, which are different from those that motivated the initial adoption of the second-best framework. For example, Rawls's shift from first-best to second-best is based on the need to incorporate incentive considerations in the formulation of the theory of justice. One might find, however, that the incentive structure needs to be legally imposed, and any effort to do so raises significant compliance problems. If these problems are serious enough, it might warrant a further modification of the *normative* principles, in order to favor institutional arrangements that are more easily enforced.

In this way, normatively weaker frameworks can be generated through an iterative process, in which ideals are modified in response to empirical

[6] One can find a similar terminological usage in Yew-Kwang Ng (2009: 196–202), although the details of his proposal are quite different from my own.

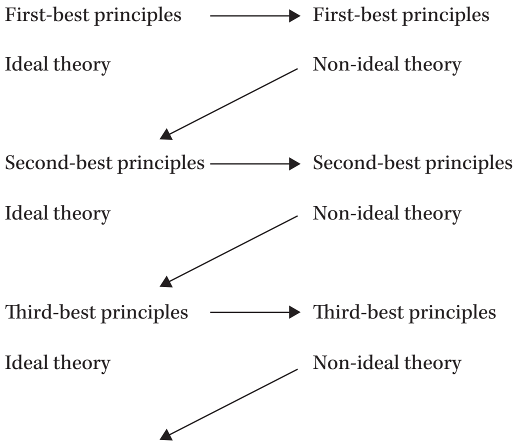


FIGURE 7.1 Iterative process for generating normative frameworks

constraints that arise at the previous level and are taken to vitiate those ideals (figure 7.1), but then in turn may encounter new constraints, which may militate in favor of further modification of the ideals. Each time the process is repeated, it generates what we might refer to as a lower level of idealization.

It should be noted that the principles need not extend in scope to "all of society," since it may be the case that the facts of human psychology only constrain the implementation of certain principles in certain contexts. Thus it may be the case that in face-to-face interactions, such as Cohen's idealized "camping trip" (2009), there is no obstacle to the realization of first-best principles of distributive justice, and so some form of strict egalitarianism is in order. With interactions between strangers, on the other hand, it is difficult to achieve the same level of social solidarity, and so one might think, with Rawls, that the state should be governed by second-best principles (which anticipate the need for trade-offs between efficiency and equality, and so have a "built in" formula for making such compromises).

The fact that the analysis can be extended to generate a third-best framework is useful, I will argue, because there are domains of interaction in our society where second-best principles, of the sort that Rawls articulates, are subject to insurmountable difficulties at the level of implementation, leading to the creation of a set of institutions governed by third-best principles. In other words, what I want to suggest is that the schema presented in figure 7.1 is not simply an abstract exercise, but is actually instantiated in our society. In particular, I will argue that a market economy is best seen as a response to implementation problems encountered when trying to institutionalize a second-best theory of justice over the allocation of goods and resources, and so constitutes a third-best framework.

7.2. The Morality of Cooperation

This mechanism for generating *n*th-best normative frameworks can be used in the development of any normative-ethical theory. In order to pursue my analysis of the market economy, I would like to start by applying it to a specific view, which I refer to as "minimally controversial contractualism." The central idea is well expressed by Rawls, in the early pages of *A Theory of Justice*, in which he points out that every system of cooperation of necessity generates both a common interest and a conflict of interest. Parties to the interaction have a common interest in maximizing the benefits of cooperation, but they experience a conflict of interest when it comes to deciding how the benefits and burdens of cooperation are to be distributed (since "they each prefer a larger to a lesser share" [1999, 4]). Thus in order to secure agreement with respect to the particular modalities of cooperation, the individuals involved must accept at least two normative principles: first, a principle of efficiency, which specifies the concept of maximization with respect to their common interest, and second, a principle of distribution, which specifies some conception of equalization as a way of resolving the conflict of interest. This analysis provides conceptual foundations for the commonly held view that the Pareto-efficiency principle constitutes an attractive conception of maximization, while some form of equality principle (such as envy-freeness) provides a plausible way of characterizing fairness. Each principle provides only a partial ordering of the set of possible cooperative arrangements, yet when taken together, they provide either a unique solution, or else they narrow it down to a very small set of focal options (Baumol 1986).

Because of this complementarity between the two principles, when a contractualist framework of this sort is conceived of as a first-best normative theory, its central characteristic is that it never requires any trade-off between efficiency and equality. Indeed, it is fairly easy to show that in an idealized scenario, such as a two-person two-good economy without production, the "contract curve" (i.e., the set of Pareto-optimal allocations) necessarily intersects the set of envy-free allocations, and so there will always be outcomes that are both efficient and equal (in that sense) (Baumol 1986; Kolm 1997; Heath 2004b).

With this in mind, it is not difficult to see why many theorists might have begun to think of the market as simply an implementation mechanism for a first-best normative theory of this sort (e.g., Dworkin 1981). Although it is common to talk about the "equality-efficiency tradeoff" (Okun 1975), one of the major accomplishments of twentieth-century welfare economics was to show that a market economy could, in principle, be used to achieve both perfect efficiency and perfect equality simultaneously (Stiglitz 1994). The first fundamental theorem showed that under the conditions that characterize "perfect competition," the outcome of market interactions will be Pareto-optimal.

More importantly, for our purposes, the second fundamental theorem of welfare economics showed that, with the proper allocation of initial assets, a competitive market could generate any one of the set of possible Pareto-optimal outcomes. This means that whatever outcome would be picked out by one's favored conception of equality, there is no obstacle in principle to the design of a market that would generate exactly that outcome. So it is possible to achieve both equality and efficiency without requiring any trade-offs.

This is the framework that informed the so-called "socialist calculation" debate of the early twentieth century (Stiglitz 1994), as well as the early discussions of optimal taxation in the economics literature. The preferred vehicle for achieving greater equality is a lump-sum head tax imposed upon individuals, combined with redistribution to an initial allocation from which individuals can trade to the preferred optimal market equilibrium. Because it is a head tax, it does not distort incentives (nothing that a person does can help her to avoid it), and so the tax-and-transfer scheme can achieve equality without imposing any inefficiency losses. And yet as Joseph Stiglitz points out,

> Governments do not engage in lump-sum redistribution—and for good reasons. They do not have the information required to implement such taxes in an equitable manner. Governments clearly believe that different individuals should pay different taxes. As a basis of taxation, they inevitably rely on observable variables, like income or wealth, variables that are alterable. Hence the taxes are distortionary. Once we recognize that redistributions are inevitably distortionary, we must also recognize that changing the distribution of endowments has an effect on the overall efficiency of the economy. (1994: 45)

If one accepts the idea that non-ideal factors will necessarily force trade-offs between efficiency and equality, then one has no choice but to formulate a new set of principles as well. This is because both the Pareto efficiency and the envy-freeness criteria are deficient when it comes to ranking imperfect states relative to one another. The Pareto principle classifies a huge number of states as Pareto-noncomparable. The envy-freeness principle is even more limited, in that it offers no way of ranking imperfect states. Without introducing some controversial procedure for quantifying envy, there is no way of saying that states are more or less envy-free (Arnsperger 1994: 167–169; Heath 2004b: 324). So once one begins to take seriously the need for trade-offs between efficiency gains and equality considerations, a different set of normative principles is required (or at least a different way of specifying the guiding ideas that led to the adoption of the Pareto efficiency and envy-freeness standards).

It is here that most of the action has occurred, over the past several decades, in the debates over equality. The most prominent proposals in the literature, from Nash's Bargaining Solution (NBS) to Rawls's difference principle, are best interpreted as formulae for trading off equality against efficiency. At a certain level of abstraction, all of these can be conceived of as forms of prioritarianism, in that they seek to maximize the benefits of cooperation, but assign different weights to the benefits going to different individuals depending upon how much inequality there is between those individuals. This is most obvious in the case of the NBS, which seeks to maximize the product of individual utility, and less obvious in the case of Rawls's difference principle, which assigns lexical priority to the interests of the worst-off individual.[7]

What is important to emphasize is that these normative principles are all tailored for a second-best framework, in which trade-offs between efficiency and equality may be required. They provide us with normative guidance in situations in which the allocation winds up off the $x = y$ axis (a state of affairs that need never arise in an ideal, first-best world).

To make this more concrete, consider how such principles would be applied in a scenario like Cohen's camping trip (2009). Suppose, for example, that some of the campers bring along fishing rods and happen to catch some fresh fish. In Cohen's view, it would be most natural for them to share everything equally with the rest of the group, even if the others contributed nothing to the production of this bounty. One can imagine, however, a situation in which the people with the fishing rods begin to grow tired of the activity and feel that they are not really getting enough out of it. Or suppose that more people without fishing rods join the group and so the people who are catching the fish find that they are doing a lot of work, yet are not able to enjoy more than a small fraction of their catch. Under these conditions, it might prove beneficial to adopt a slightly different arrangement, whereby the person who catches the fish gets to keep a certain fraction of it for his own personal consumption, with the remainder being shared out equally within the group (Martin 2012: 37). This is essentially an "incentive pay" system. It typically will have the advantage of increasing the amount of fish that get caught, and so can easily be designed in such a way that it will benefit everyone. Yet it has the disadvantage of pushing the allocation in the direction of increased inequality: first, between those who have fishing rods and those who do not, and second, among those who have fishing rods, between those who are better at fishing and those who are worse, or perhaps even just between those who get lucky and those who don't. Making a decision about how much of the fish the person who catches it is entitled to keep amounts to making a decision about how much of a compromise one is willing to accept, in the equality dimension, in order to increase benefit in the efficiency dimension, and vice versa.

The important thing to note about this new arrangement is that, although it solves a number of implementation problems—by introducing individual

[7] It becomes more apparent when represented visually, as in figure 6.4, which makes it obvious that the difference principle is really just an extreme variant of the prioritarian indifference curve.

incentives to increase production—it gives rise to a new set of implementation problems that have a slightly different character. In order to design an incentive scheme that will produce the best outcome, according to a prioritarian social welfare function, detailed information is required about how producers will respond to incentives (specifically, what their labor supply curve looks like). Note that this information is not required under the first-best scenario, where everything is divided up equally; it is an issue that arises only at the second-best level, with the implementation of second-best principles. In order to get this information under ideal theory conditions, one can imagine simply asking people "how much would you fish, if you got to keep *x* per cent of your catch?" But under non-ideal conditions, one must take into consideration the possibility that producers will misrepresent their preferences in order to increase their share. And so one might want to institute a *revelation mechanism*, changing their incentives in such a way that they are motivated to disclose the truth (either verbally or through their actions). The question then is whether this will require further modification of the normative framework—that is, whether this will require a shift to a third-best framework.

One of the characteristics of the "incentive pay" system is that it still allows the group to exercise direct control over the distribution of fish, particularly between producers and non-producers, by adjusting the fraction of the catch that the producers get to keep. If the information required to make these adjustments is unavailable, however, and so instead control is shifted over to a revelation mechanism, then the group may have to relinquish its ability to directly control the distribution. As a result, instead of being able to apply an egalitarian norm to "pattern" the distribution, the group may only be able to institute certain tolerances, and may need to restrict itself to indirect manipulation of the outcomes. For example (and here we begin to see the point of this lengthy thought-experiment), if the campers were to switch over to an exchange system, so that the people doing the fishing get compensated for their efforts by having certain services provided for them by the other campers (e.g., cleaning and cooking of the fish), and the parties are left free to negotiate the terms of these exchanges as they see fit, then it becomes difficult to anticipate in advance what the distributive consequences of the arrangement will be. The campers might nevertheless find it attractive, because of the sharper and more flexible system of incentives it provides. The efficiency gains might be so great, in other words, that the campers are willing to surrender control over the distributive consequences, only intervening in cases where the departure from equality becomes sufficiently egregious. This is, I would argue, a third-best framework, requiring a set of normative principles that constitutes a weakening of prioritarianism. Instead of trading off equality and efficiency in a continuous fashion, the way that (for example) the NBS does, it requires instead the privileging of efficiency, with the concern over equality showing up only as a set of boundary constraints on the range of acceptable Pareto improvements.

TABLE 7.1 } **Norms of cooperation**

Normative framework	Form of cooperation	Norm
First-best	Spontaneous cooperation	Egalitarianism
Second-best	Institutionalized cooperation	Prioritarianism
Third-best	Competitively induced cooperation	Constrained efficiency

One could translate this into a general claim about competitively organized interactions. There are some cases in which people cooperate in an absolutely effortless manner, guided entirely by a sense of justice, or honor, or fair play, or what have you. Under such conditions, no trade-offs between equality and efficiency are required. In the more ordinary run of cases, the internal motives of individuals must be supplemented with some external incentives. In this case, we say that the system of cooperation must be institutionalized, in Talcott Parsons's (1951) sense of the term. Institutionalization requires trade-offs, so we now require a set of second-best normative principles, which specify how the inevitable compromises between efficiency and equality are to be assessed. In some cases, however, there are significant obstacles to directly institutionalizing a system of cooperation. This is particularly true when the interactions are large-scale and anonymous, and so internalized moral constraints on free riding are weakened. Under such circumstances, we may be able to further expand the benefits of cooperation by organizing a competition, in essence harnessing the free-rider incentive and deploying it in such a way as to generate beneficial outcomes as a byproduct. The major problem with such an arrangement is that we lose the ability to pick and choose outcomes, and so our ability to ensure that the system as a whole satisfies an egalitarian constraint is further attenuated. For this reason, I believe that competitively organized interactions constitute a third-best framework and require a set of third-best normative principles for their assessment (table 7.1).

This is, I would argue, the best framework for approaching a normative assessment of the market economy. It also explains why certain features of the ordinary operations of a market economy continue to strike many people as morally problematic, despite the centrality of the market in our society.

7.3. Markets as a Third-Best Institutional Arrangement

When it comes to justifying capitalism, the major obstacle was well summarized by Friedrich Hayek, when he admitted that "the manner in which the benefits and burdens are apportioned by the market mechanism would in many instances have to be regarded as very unjust *if* it were the result of a deliberate allocation to particular people" (1976: 64). Hayek's solution was to suggest that the result was acceptable because the market was not a system of direct allocation, but instead was a spontaneous order. This is not a particularly compelling argument, however, because we obviously have a choice as to whether we want to keep the market or abolish it and replace it with something else. Hayek tacitly acknowledges this, when he goes on to write that market allocations "are the outcome of a process the effect of which on particular people was neither intended nor foreseen by anyone when the institutions first appeared—institutions which were then permitted to continue because it was found that they improve for all or most the prospects of having their needs satisfied. To demand justice from such a process is clearly absurd" (1976: 64–65). So what turns out to matter, in the end, is not that the institution was initially an unplanned order and so beyond the scope of justice, but that it is Pareto-improving, or at least, by-and-large Pareto-improving ("they improve for all or most the prospects of having their needs satisfied"), which is why we choose to keep it. But to say that it is Pareto-improving is to suggest that it is a system of cooperation, and if it is a system of cooperation, then it seems not only reasonable to expect that it be organized in a way that is just, it seems like a paradigmatic example of the case where such demands are appropriate. Far from being "clearly absurd" to demand justice from such an institution, it is actually somewhat difficult to see how such an institution could possibly evade the demands of justice.

To put the point in slightly different terms: there are a variety of mechanisms through which people can produce cooperative benefits. For instance, if there is a large object, which is beyond the strength of a single individual to move, people can combine their efforts in order to accomplish the task. We call this an *economy of scale*, and it is a well-known mechanism of cooperative benefit. Another such mechanism is a *gain from trade*. When two people are each assigned a heterogeneous basket of goods, it may be possible to generate mutual benefit by reassigning those goods, in such a way as to take advantage of differences in their preferences. If one person does not like carrots and another person does not like potatoes, then they can exchange carrots for potatoes, making themselves both better off. (Similarly, if people have different abilities, it may be possible to create mutual benefit by reassigning productive resources.) The important point is that the exchange is a perfectly ordinary cooperative interaction, and there is no reason a priori to think that the benefits should not be divided up in accordance with whatever principles of justice govern every other cooperative interaction. Furthermore, if one thinks that the primary function of the principle of equality is to pin down the modalities of cooperation, then it seems obvious that the benefits of trade should be divided up in an egalitarian fashion.

Thanks to modern welfare economics, we have an extremely elegant way of depicting (and even quantifying) these gains from trade (De Marchi 2003). The gains are, of course, in welfare, since the total supply of goods remains

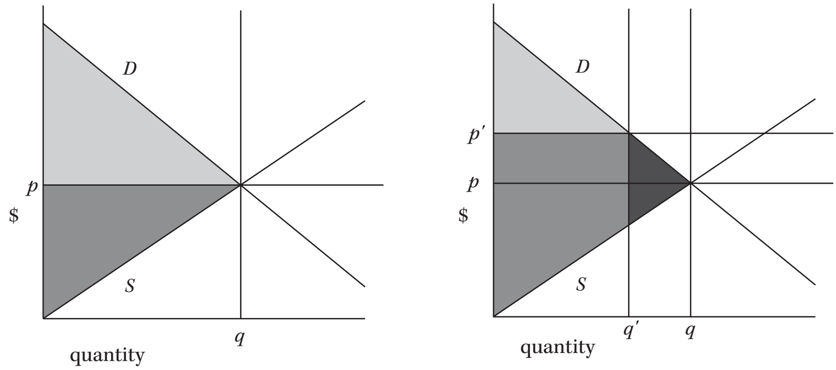


FIGURE 7.2 Gains from trade

unchanged throughout—it simply gets allocated in a way that increases everyone's utility. The left-hand side of figure 7.2 shows a standard supply-demand diagram for a particular good, where the price at which the transaction occurs determines the quantity that will be exchanged—*p* indicates the market-clearing price, which is the equilibrium in a competitive market. Although this is quite familiar, many non-economists are unaware that the triangle formed by the demand curve *D*, the price line at *p*, and the *y* axis actually represents the utility gain to consumers, when quantity *q* is exchanged at price *p*. This is known as the consumer's surplus, and it follows fairly immediately from an understanding of what the demand curve represents. (Intuitively, the fact that consumers would be willing to pay the higher price in order to purchase the first few units of the good means that when they get all *q* units for the low price of *p*, it means that they are "saving" the difference—which comes to them in the form of a utility gain. This same principle applies at every point to the left of *q*—or "inward from the margin"—so the volume of the triangle specifies precisely the difference between what the consumer would have been willing to pay and what he actually had to pay for the quantity of the good purchased.) By parity of reasoning, one can see that the triangle formed by the supply curve *S* and the price line represents the utility gain to the supplier and is accordingly known as the producer's surplus.

One of the major advantages of this way of representing the cooperative interaction is that it shows quite clearly how the gain from trade gets divided up between the two parties. It is *entirely mediated by the price*. If the price goes up, then the consumer's surplus is reduced while the producer's surplus is increased. If the price goes down, then the consumer's surplus is increased while the producer's surplus is reduced. This is why, since Aristotle at least, it was universally assumed that the question of distributive justice in exchange relationships should be answered *through some specification of the price at which goods should trade*, because this is going to determine how the benefits of the exchange are divided up. Thomas Aquinas formulated a particularly influential version of the doctrine, which identified the moral issue quite specifically with the relative size of the consumer's and supplier's surpluses ("If the one man derive a great advantage by becoming possessed of the other man's property, and the seller be not at a loss through being without that thing, the latter ought not to raise the price, because the advantage accruing to the buyer, is not due to the seller, but to a circumstance affecting the buyer"[1947: 1514 (2-2, q. 77)]).

Thus for centuries it seemed obvious—to both cloistered academics and angry peasants—that there was an issue of distributive justice that arose in market transactions. Nothing that has been said by modern welfare economics has changed that. What has changed is simply a greater awareness of the relationship between price and efficiency. This is due to the less obvious but extremely important observation that, if the price goes up or down, not only will it change the size of the consumer's and producer's surpluses, but it will also affect the quantity of the good being exchanged. (On the right-hand side of figure 7.2, if the price rises from p to p', the quantity exchanged will tend to decline from q to q'.) This obviously generates a change in the two surpluses—in this case, the size of the consumer's surplus is reduced, while the producer gains an economic *rent*, shown as the rectangle bordered by p, p', q', and the y axis. More importantly, however, there is a deadweight loss, the magnitude of which is given by the area of the triangle to the right of the line q', below the demand and above the supply curve. This triangle (known as Harberger's triangle) represents a portion of the utility potentially realizable through exchange that is, in effect, being thrown away because of the increase in price.

If two individuals are brought together to trade, they may settle on a price that is anywhere in the region between the supply and the demand curves. However, if more sellers are introduced into the market, it creates a collective action problem—a competitive dynamic—that will tend to drive down prices, while if more buyers are introduced, it creates a collective action problem that drives up prices. These two tendencies push in opposite directions, with the result that prices in a reasonably competitive market will move toward the market-clearing level. This has the useful consequence of maximizing efficiency. It not only benefits the parties to the interaction, but the emergence of a publicly observable price also provides an extremely important source of information to other market actors, regarding the relative scarcity of the good. This can in turn be used to guide both production and consumption decisions, in order to improve the allocation of goods and resources across the economy as a whole.

Unfortunately, there is no reason to think that the price level that clears the market will generate a division of the benefits of cooperation that satisfies any sort of egalitarian constraint. Indeed, from the perspective of fairness, the outcome will be more-or-less arbitrary. As Alasdair Macintyre put it, "What is necessarily absent in such markets is any justice of desert. Concepts of a just wage and just price necessarily have no application to transactions within those markets" (1995: p. *x*). Similarly, Hayek argued that the central advantage of market economies over planned economies is that the latter, being subject to "conscious direction," "would always have to aim for prices that are considered fair," and yet "an economic system in which everyone received what others felt he deserved could not help but be a highly inefficient one" (2002: 18).

For a good example of how market prices are orthogonal to questions of distributive justice, one need only consider the mixed attitudes that people have toward so-called price discrimination. If the diagram in figure 7.2 is used to represent a competitive market, then the demand curve will not represent the preferences of a single consumer; it will be an aggregate, representing the preferences of a reasonably large group. Within that group, preferences may vary enormously. Some consumers may experience rapidly diminishing returns, others less so. Or consumers may have returns that diminish at the same rate, but experience vastly different levels of satisfaction. Suppose, for instance, that the curve shows demand for televisions. It may be the case that everyone wants only one television (and so individual returns drop to zero after the first unit purchased), but some people simply care more about having a nice television and so are willing to pay more for it. Suppliers could set the price very high and still sell some televisions, but they could sell even more by setting it lower. Thus the people who would be willing to pay $1,000 for a television derive an enormous benefit from the marginal consumer who won't spend more than $400. In order to induce the marginal consumer to buy, suppliers wind up having to drop the price they charge everyone to only $400. Thus if one looks at the consumer's surplus, one can see that it is being divided up amongst the consumers in a highly inequitable fashion.

For suppliers, the difference in willingness-to-pay among consumers creates an obvious incentive to partition the market—to somehow isolate those who are willing to pay $1,000 from those willing to pay $800, from those will only pay $600, etc., and to charge them each a different price. The question is: would it be unfair for firms to do this? Straightforward price discrimination is illegal in the United States (under the Robinson-Patman Act), although mainly for antitrust reasons, not because the practice is regarded as inherently discriminatory or unjust. There are certainly cases where consumers have reacted with hostility to the practice. For example, it caused a minor scandal when it was suggested that Internet retailers, most importantly Amazon, were using a customer's past purchase history as a way of making an educated guess about willingness to pay, and so offering different prices to different consumers on the same products. (A University of Pennsylvania study, which detailed widespread consumer ignorance of price discrimination, was entitled "Open to Exploitation: America's Shoppers Online and Offline" [Turow, Feldman, and Meltzer 2005].)

On the other hand, to the extent that the targeting is successful, it is unclear what basis the inframarginal consumer has for complaint. After all, if you genuinely want something more than someone else wants it, where is the injustice in being charged more for it? Price discrimination in this case reduces inequality in the satisfaction that different consumers derive from a purchase. This may explain why de facto price discrimination, or thinly veiled price discrimination, is tolerated throughout the economy. For example, the most obvious form of price discrimination occurs when goods—such as new types of televisions—are introduced at a relatively high price, which then declines over time. This has the effect of segmenting consumers, so that those who really want it and are therefore loathe to wait, will buy it right away, while those who want it somewhat less will wait for the price to decline. This is price-discrimination-in-everything-but-name, and yet it is tolerated by consumers (and the law). The same can be said for coupons, customer loyalty programs, overpriced "options" on goods ranging from automobiles to hamburgers, exotic rules governing airplane fares, and so on, all of which are different ways of achieving price discrimination.

In my view the reason that there is no consistent response to these practices—and indeed, why the issue of price discrimination is seldom mentioned in the literature on capitalism and distributive justice—is that the price at which goods trade is basically arbitrary from the standpoint of distributive justice, and so it is easy to imagine cases where price discrimination would be intuitively just and many other cases where it would be intuitively unjust (Dietsch 2010: 230).

Here we can see the compromise that is at the heart of capitalism. We could choose to organize the economy as a system of direct cooperation (through some sort of central planning mechanism), the problem is that without prices it is incredibly difficult to decide how to allocate goods, resources, and labor in such a way as to best satisfy our needs. By allowing free exchange and competition to develop, it becomes possible to achieve significant efficiency gains. The downside is that, in so doing, we relinquish direct control over the system of allocation, and so are unable to ensure that it satisfies the norm of equality. This constitutes, I have suggested, the adoption of a third-best framework, wherein we refrain from applying the egalitarian norms that we would normally apply to assess cooperative interactions. We try to get market-clearing prices because of their efficiency-promoting qualities, while acknowledging that the distributive consequences of this will be pretty much arbitrary from the standpoint of justice. To the extent that we do bring in considerations of equality and distributive justice, these take the form of outside boundaries or constraints, such as the minimum wage, or certain restrictions on "unconscionable" contracts (Trebilcock 1993).

7.4. The Determination of Wages

A system that aims to set prices at the level at which all markets will clear is sometimes referred to as a system of "scarcity pricing." The idea is that the price of a good should reflect its relative scarcity—how much supply there is, relative to demand, how much demand there is, relative to the supply. If people suddenly decide that they prefer tea to coffee, tea will become relatively scarce, and the price will rise in response to that. If growers increase production, then scarcity will decline, along with the price. If one is trying to decide whether a particular parcel of land should be used to grow tea, coffee, or something else, then a system of scarcity pricing is incredibly important, because it provides a basis for determining how that land should be used, in order best to satisfy human need. It is largely for this reason that many socialist economists came around to the view that not only capitalist economies, but production in socialist states should also be governed by a system of scarcity pricing (Lerner 1977: 236).

As I have argued in the previous section, adopting such a system requires a certain measure of compromise, because it involves forfeiting direct control over the distributive consequences of the exchange system. One can have scarcity prices or one can have just prices—prices that attempt to achieve a fair division of the cooperative surplus—but one cannot have both. This means that the ordinary operations of the market economy have a number of mildly counterintuitive implications with respect to the purchase of everyday goods. For example, a majority of Americans have the moral intuition that it is unethical to raise the price of umbrellas when it rains or the cost of shovels when it snows (Hauser 2006: 92–93). Underlying this is the plausible thought that it is wrong to charge someone more for something just because her need for it is urgent. Everyday morality in fact tends to recommend the opposite—that one should be more accommodating toward the person whose need is greatest. But this amounts to a rejection of the basic principle used in a market economy to establish prices.

The area where scarcity pricing generates the most counterintuitive results, however, is with respect to wages—which are, in a market economy, scarcity prices like any other. The general rule is that firms must pay their workers enough to keep them from leaving, but not more than it would cost to replace them. In practice, this means that wages are affected by a variety of factors that, intuitively, seem quite unrelated to the actual work that is being done—such as how many other people are willing to do that work and what outside options existing workers have. (As Hayek put it, the central purpose of the price system is "solely that it shows individuals that what they have previously done, or can do now, has become more or less important, for reasons with which they have nothing to do." Because of this, "the compensation of the various services changes without taking into account the merits or defects of those involved" [2002: 17].)

Our everyday moral intuition is that how much a person is paid should reflect a variety of factors, including how much that person produces, how difficult the job is, and how much skill or training is required. These are all, in effect, "just price" intuitions—they are about how the cooperative surplus should be divided up. If the employee is creating a very large benefit, then we tend to think that she should get a fairly large share of it. If the job is particularly burdensome, then she should be compensated, and so on. In a competitive labor market, however, these factors affect wages only indirectly, insofar as they influence the number of other people able and willing to do a particular job. This means that the two can easily come apart, most obviously in cases of work that is quite demanding, but which for one reason or another many people are willing to do.

There is a persistent tendency, however, to want to read moral categories into the operations of the price system for labor. Consider the following, very typical example:

> [I]ncomes result from the joint operation of demand and supply in the labor market, which respond roughly to people's contributions and efforts. On the one side, the more an employer thinks an employee will promote his, the employer's, ends, the more he will pay for the employee's labour; on the other side, the more strenuous the labor is or the more training it requires, the more an employee will insist on being paid to do it. Contribution and effort are therefore rewarded on, respectively, the demand and supply sides of the labor market. (Hurka 2012: 60)

This is, unfortunately, not what economic theory tells us about the way that either the supply or the demand for labor affects its price. It is, for example, simply not true that "the more strenuous the labor is or the more training it requires, the more an employee will insist on being paid to do it." This sort of "insisting" is precisely what is precluded by the competitiveness of the labor market—if one potential employee demands a higher wage, he or she will simply be underbid by someone else, until wages are depressed to the market-clearing level. Training is a sunk cost that has no impact on the price of labor, except insofar as it affects the scarcity of a particular labor type. (For example, being able to play the piano competently takes an incredible amount of training, and yet because so many people have that training, it has very little market value.)

There is a closely related error, which has become widespread in philosophical discussions of distributive justice and markets, and so is perhaps worth commenting on. It is often maintained that markets somehow reward people for being "talented." The popularity of this idea is probably due to Robert Nozick's famous "Wilt Chamberlain" argument in defense of income inequality (1974: 161–162). Nozick's thought was that Chamberlain, being such a talented basketball player, could charge a slight premium on tickets for all those wanting to watch him play. Spectators would happily accept this surcharge, with the result that Chamberlain would receive an income that was significantly larger than that of the other players (or members of society generally). This gave rise to the idea that certain people, by virtue of a natural endowment that allows them to do something special, are able to go out into the market and command higher salaries (e.g., Cohen 1991). Yet in the Chamberlain example, it is not his "talent" as such that allows him to command the higher salary, it is that he possesses abilities that are relatively scarce.[8] If spectators could go next door and witness the performance of another basketball player indistinguishable from Chamberlain, then Chamberlain would be unable to charge his entry premium, regardless of how objectively "talented" he is. (David Gauthier gets this point exactly right, in emphasizing that the issue is not one of salary, but rather of Chamberlain's ability to command an economic *rent* [1986: 273]. Wilt Chamberlain is, in effect, a monopolist in the market for the supply of Wilt Chamberlain services.) This is counterintuitive, however, because the concept of rewarding talent makes sense, while the concept of rewarding scarcity has no intuitive moral plausibility—it only makes sense when one looks at the downstream consequence of its incentive effects.

There is also a commonsense view that compensation should be linked to productivity, so that how much you get paid should somehow be linked to how much you produce ("the more an employer thinks an employee will promote his, the employer's, ends, the more he will pay for the employee's labor"). This is, again, not true. Employers typically pay no more than they need to, regardless of how much they think an employee will contribute. The fact that, according to neoclassical economic theory, the wages of labor are determined by its marginal product has occasionally been taken as a vindication of the idea that "contribution" matters. It actually proves the opposite. It suggests that every worker's wage is determined, not by the value of what he or she actually produces, but rather by the value of what the marginal worker produces—which, given diminishing returns to scale, will be less than what the average worker produces.

There is not much to be gained from puzzling over this, however, because the set of idealizing conditions under which this model of wage-determination applies makes it of limited applicability. In the real world, in which labor markets are segmented, there is unemployment, and wages often contain an element of economic rent, the amount that a worker gets paid is typically determined by the scarcity of that particular type of labor, which is to say, how

[8] Cohen observes that people need not be talented, in the ordinary sense of the term, but "*so positioned that, happily, for them, they do command a high salary and they can vary their productivity according to exactly how high it is*" (2008, 120). What he is describing is better known as "market power."

much it would cost to replace her.[9] For instance, it is generally thought that segmentation of the labor market combined with crowding in certain areas accounts for a fraction of the wage differential between men and women.

The fact that wages are not determined by the value of what workers produce is also reflected in the fact that workers in particular sectors are unable to capture the benefits of productivity gains in that sector for more than a brief period of time. Productivity in some sectors of the economy, such as manufacturing, and more recently, data management and record-keeping, increased dramatically over the course of the twentieth century, whereas other sectors have remained almost entirely stagnant. In the long run, however, labor is highly mobile between occupations, so if wages rise in one sector as a result of productivity gains, more workers will start to look for those jobs, not only depressing wages, but also pushing up wages in other sectors, in order to induce workers to remain in those occupations.[10] This is why wages tend to reflect trends in average productivity, across the economy as a whole, and not productivity in particular sectors.

The important point is that this system, which is designed to direct labor to its most valuable employment, does not correspond to any of our prior moral ideas about what people "deserve" or what is "fair" (unless one defines either of these terms as "doing one's part to direct labor to its best employment"). In particular, it means that wages will be affected by a variety of external factors, which are in no way linked to the "internal" quality or properties of the work being done.

This basic principle of wage-determination is to some extent masked in wealthier countries through a variety of mechanisms. Most importantly, the minimum wage stops the price from dropping to the level that would clear the market. Furthermore, large organizations, both corporations and public sector employers, do an enormous amount of cross-subsidization across employee groups, and so insulate their workers to varying degrees from market determination of wages. Labor unions do much the same. Pay equity schemes push even more strongly in this direction, since they typically do not permit consideration of supply and demand conditions in the determination of wages. Many highly skilled groups also belong to professional associations, which limit entry to certain labor markets and therefore raise wages. Thus the basic principle of scarcity pricing in the determination of wages tends to operate in the medium-to-long term.

When rich countries begin to engage in international trade with poor countries, however, the basic principle is suddenly put on display in particularly

---

9 For an accessible discussion, see Harford 2006: 25–28.

10 This is the mechanism underlying what is known as the Baumol effect (see Baumol and Bowen 1966).

stark form. One can see workers in different factories, in different countries, doing what amounts to the same job, producing essentially the same products, but receiving vastly different wages. The fact that what workers are being paid bears no relationship to what they are producing, or how hard they are working, or under what conditions, becomes impossible to ignore. The only way to understand the wage differential is to see that wages are a scarcity price, and in underdeveloped countries labor is not scarce, compared to capital. This is reflected in the fact that average productivity is low. More concretely, it is that factory workers in poor countries have terrible outside options (in many parts of the world, it is estimated that the marginal productivity of labor in the agricultural sector is zero). There are, of course, a variety of factors that exacerbate this, particularly the fact that many developing nations engage in coercive tactics aimed at blocking the formation of trade unions and the introduction of collective bargaining. Yet these policies are only likely to depress the labor share of national income by a few percentage points, they do little to change the essential fact that labor is poorly rewarded for reasons that have nothing to do with its intrinsic qualities, or the moral entitlements of the workers doing it. Furthermore, the only way to create a durable increase in these wages is to increase labor productivity in the country as a whole.

7.5. Markets Not a System of Natural Justice

Much of normative political economy in the twentieth century was dominated by an extremely intense, sophisticated, and wide-ranging debate over the "morality of the market," divided between, on the one hand, those who saw it as a system of thinly veiled exploitation, and on the other hand, those who regarded it as a system of "natural justice." And while opinion is still polarized, by the end of the twentieth century it had become clear that neither of the extreme positions had stood up very well to scrutiny. Careful work led the leading analytical Marxists to abandon exploitation as a useful normative concept for the evaluation of markets (Roemer 1985; Cohen 2000). But just as importantly, better understanding of the distributive dynamics of capitalism, and in particular the determination of wages, led to the abandonment of the claim that market outcomes satisfy any plausible conception of fairness, or that they map onto our everyday understanding of desert. Thus it became clear that the mere fact that an outcome was the result of a set of market processes was insufficient to generate the presumption it was either just or unjust, in any robust (i.e., first- or second-best) sense of the term.

This has resulted in there being only two major normative models of the market left standing. The first is the family of arguments claiming that the market economy is a desirable arrangement because it maximizes social welfare. The most forceful of these arguments are the Paretian ones, which draw their inspiration from the first fundamental theorem of welfare economics, in order to suggest that a well-structured, competitive market economy produces not just utilitarian gains (where some might benefit while others lose), but Pareto improvements (where everyone benefits). Naturally, this Paretianism must be hedged and qualified in numerous ways, in recognition of the fact that any action that affects a sufficiently large number of people is bound to produce both winners and losers. Nevertheless, it remains the case that markets are systems of exchange, and people engage in trade because they all expect to be better off after the exchange than they were before. So to the extent that exchange is voluntary, and the market system encourages individuals not just to exchange, but to maximize the number of exchanges, then there is a strong presumption that the results will be Pareto-improving. Furthermore, experiences with central planning during the twentieth century lent considerable plausibility to the view that the market not only "makes possible a better life for all than any would have if each were to live solely by his own efforts" (Rawls 1971: 4), but that it makes possible a better life for all than any would have under any other set of large-scale economic institutions.

The second major view, and the only serious alternative to Paretianism, is the type of deontological libertarianism that abjures any attempt to evaluate social institutions by looking at the outcomes they produce. The standard approach, following Nozick, is to claim that particular transactions can be either just or unjust, but that any outcome that is produced through a set of just transactions, no matter how large or complex the set, is just *eo ipso*. Thus if individuals have rights to property, which they are entitled to alienate through voluntary exchange, then the outcome of any market transaction will be just by virtue of its voluntariness, and not its specific terms. From this perspective, the expectation that prices should be "just," or that wages should be "fair," reveals a deep confusion about the nature of justice and fairness. Even the fact that exchanges are Pareto-improving is just a happy byproduct, from this perspective, not at all important to the justification of the institutions that permit them.

A sustained analysis of libertarian ideas is beyond the scope of this chapter, or indeed this book. It is perhaps adequate here simply to observe that the overall approach suffered a significant setback during the civil rights era in the United States, because an enormous amount of the racial discrimination that prevailed in the American South was a consequence of people voluntarily choosing to associate or disassociate themselves from others on the basis of racial characteristics. In this and many other areas, important legal precedents were set establishing the state's prerogative to overrule private transactions in cases where their aggregate effects were sufficiently deleterious. Furthermore, these legal precedents acquired enormous moral prestige, making it difficult for libertarians to attract many adherents to the view that the voluntariness of market transactions alone is sufficient to establish their justness. As a result, if one looks carefully at the work of various theorists described as "libertarian," they often turn out to be crypto-Paretian. Typically, they will spend a lot of time talking about freedom and the voluntariness of market exchange, but when it comes time to address the normative question, they will claim that the voluntariness is important because it generates such a strong presumption that the exchange is Pareto-improving. This is typically combined with some empirical claim, to the effect that attempts to achieve Pareto improvements through anything other than market mechanisms are doomed to failure. Thus the "liberty" of the free market winds up being celebrated, but only because of the presumption that the exercise of this liberty maximizes welfare.[11]

This general analysis of the state of play is what motivates my basic contention that Paretianism is the "last principle standing" in the debates over the moral justification of the market (and what vindicates McMahon's claim that efficiency is the "implicit morality" of the market). Admittedly, what I have presented is an "argument from elimination" with respect to other views, and there is nothing to stop someone from coming up with a new, more exotic justification for the market. Nevertheless, for the moment Paretianism is the only serious game in town. Furthermore, Paretianism is clearly the guiding idea in the multifarious ways in which the modern welfare state seeks to regulate, modify, and sometime supplant the market (Heath 2011). It is also the guiding idea that informs judicial interpretation of modern business and corporate law (Hansmann and Kraakman 2004: 28).[12]

Now, to say that efficiency is the implicit morality of the market, and should provide the guiding idea in business ethics, is not to say that managers should always be asking themselves, before engaging in a particular course of action, whether it is likely to be Pareto-improving or not. On the contrary, the market is designed to promote Pareto efficiency as a byproduct of competitive behavior on the part of firms, no single instance of which will be Pareto-improving. (In the same way that an ideal of "justice" is pursued indirectly through adversarial legal institutions, where none of the parties are actually obliged to intend that outcome.) What the efficiency principle provides, instead, is an articulation

[11] John Tomasi presents an extensive survey of writers in both the classical liberal and libertarian tradition, showing how they all appeal to the idea that "institutional regimes should be evaluated in terms of the benefits they provide to all citizens subject to them" (2012: 141). He even catches many of them out appealing to the benefits provided to the poor as a justification for market institutions (127–142).

[12] Strictly speaking, Hansmann and Kraakman appeal to Kaldor–Hicks, and not Pareto efficiency. This commitment, which is very typical in the literature on regulation and corporate law, is often interpreted as an endorsement of utilitarianism. Examined more closely, however, I think it can be seen as a commitment to Pareto efficiency, modulated by a "realistic" accommodation of the fact that literal Pareto improvements are few and far between. The fact that the commitment is not genuine utilitarianism shows up in the set of problems that it is typically applied to, which in almost all cases involve redressing market failure. By contrast, one never sees the Kaldor–Hicks principle appealed to as a justification for redistributive policies, despite the fact that any pure redistribution necessarily satisfies that principle (the winners could, in principle, compensate the losers).

of the "point" of marketplace competition, which can in turn be used as a basis for distinguishing permissible from impermissible forms of competitive behavior. Thus the broader objective of promoting efficiency, when combined with a set of empirical claims about the conditions under which marketplace competition promotes that end, generates an intermediate-level deontology, which includes, *inter alia*, principles prohibiting the externalization of costs, the exploitation of market power, strategic use of information asymmetries, as well as various forms of opportunism (Heath 2004a: 84). These principles must then be further refined, in order to fit the circumstances of specific markets (with particular attention to the possibility of offsetting market imperfections that may generate conflict among the principles) in order to generate concrete rules that can directly govern managerial conduct.[13]

Although such principles are controversial among scholars in corporate law, and are widely rejected in the business press, they are not likely to generate much disagreement among ethicists. On the contrary, the standard objection to the Pareto principle among business ethicists is not that there is anything specifically wrong with it, but that it is simply too weak a constraint to count as a guiding norm of justice. In other words, what many ethicists find difficult to accept is the idea that a market could be governed by a commitment to nothing beyond efficiency. Thus the dominant impulse will be to want to make the deontology more stringent, by adding additional constraints derived from principles of fairness or various perfectionist ideas. The problem with this impulse, in my view, is not that there is anything morally objectionable about the additional principles, but simply that they are incompatible with the functional requirements of the capitalist system. So while they may be entirely appropriate at a first-best or second-best level, the creation of a competitive market requires that we set them aside.

This is an extremely non-polemical way of putting a point that is usually put more sharply: which is that these more robust normative theories are implicitly, if not explicitly, anticapitalist.[14] Of course, there is also nothing wrong with wanting to debate the merits of capitalism, but there is a lot to be said for the view that business ethics should not be anticapitalist (or that, to the extent it becomes so, it is no longer trying to answer the questions that people

---

[13] An example of off-setting imperfections: cartelization aimed at driving up the price of oil constitutes a prima facie violation of the moral constraint on seeking market power, but given the presence of significant uninternalized atmospheric externalities associated with fossil fuel consumption, monopoly pricing will tend to reduce production of the negative externality, and therefore enhance compliance with the moral imperative that prohibits cost externalization.

[14] This is, of course, the force of Milton Friedman's (1970) claim that the idea of corporate social responsibility was "pure and unadulterated socialism." Whether it is "unadulterated" is, of course, debatable. And it is worth observing that, historically, this sort of red-baiting did little to enhance the quality of debate. Nevertheless, when stripped of both hyperbole and rancor, the substance of the point that Friedman was making is correct. There is something at very least strange about the claim that

typically turn to business ethicists for an answer to). For some people the fact that these theories are anticapitalist is itself a reductio of them. My diagnosis of the problem is quite different. Everyday morality contains a variety of principles that are implicitly, if not explicitly, anticapitalist. To the extent that the more robust normative theories in business ethics are anticapitalist, what this shows is that these theories are not far enough removed from everyday morality. Our willingness to accept the market as the central organizing institution in the economy requires a willingness to accept a huge number of violations of these everyday moral principles (in order to get the compensating benefits of the proper operation of the price system). But the fact that we are willing to accept some forms of behavior that would ordinarily be classified as immoral does not mean that we are willing to accept any and all forms of antisocial behavior. The task that falls to business ethicists is the extremely tricky one of distinguishing the necessary from the unnecessary violations. (Part of what makes the task tricky is the fact that our everyday moral intuitions are an unreliable guide in this endeavor, and so the usual sort of intuition-mongering practised by moral philosophers is unhelpful.) Thus the Pareto principle is not put forward as an ideal moral standard; on the contrary, it is put forward as merely a succinct statement of what we consider it reasonable to expect within the third-best framework imposed by the market economy. So while the Paretian approach to business ethics is, in certain respects, normatively undemanding, it is also the most that a normative theory can require without becoming anticapitalist.

That having been said, the claim that Paretianism is insufficiently moral stems in part from a failure to think through its full consequences. When we turn to its specific normative implications, it is apparent that the Paretian approach does generate an extremely demanding deontology. If firms were to behave more ethically, as defined by the market failures approach, the world that we live in would be dramatically transformed. If one considers, for example, the past two waves of large-scale "ethics scandals" to afflict the American economy—the widespread accounting fraud and malpractice associated with the "Enron era," and the circumvention of financial market regulation underlying the 2008 financial crisis—none of the abuses that occurred are such that they cannot be characterized, in one way or another, as the exploitation of a market imperfection. Because of this, one need not appeal to any principle stronger or more controversial than the Pareto principle in order to condemn them.

Perhaps most importantly, the Paretian approach generates a simple and powerful moral argument in favor of respecting the law when, as is typically the

businesspeople, in their capacity as managers of corporations, are morally obliged to act in ways that are fundamentally incompatible with the operations of a market economy.

case, the purpose of regulation is to correct a market failure. Corporate crime is an enormously costly social problem (the magnitude of which is usually not fully appreciated because the losses are often dispersed over a very large number of victims). In general, the cost to society of white-collar crime—including the number of lives lost—is much greater than the cost of street crime (Barkan and Bryjak 2011). Furthermore, the culture of organized resistance to regulation in all of its forms, among businesses in various sectors, imposes huge costs upon society (in particular, the adversarialism of business–government relations generates significant deadweight losses, in the form of lobbying, litigation, and enforcement costs, as well as poorly structured regulation).

Given that white-collar crime is extremely difficult to detect and prosecute, the failure of many businesses to respect the law is one of the most significant moral challenges facing our society. But beyond this, there is also in many sectors a culture of gamesmanship, where firms and individuals avoid violating the letter of the law, but act in ways that clearly undermine its purpose. An enormous amount of "creative accounting" fits into this category (Archer 1996). There is also the phenomenon of "regulatory arbitrage," a euphemism developed in the financial sector to describe strategies intended to do an end-run around regulatory constraints. Many of the derivatives implicated in the 2008 financial crisis existed in order to help firms to game the rules (in particular, credit default swaps were popular because they helped banks circumvent capital requirements). Furthermore, a lot of sophisticated tax avoidance behavior, such as manipulation of transfer prices, involves gaming the rules. All of this is unethical from a market failures perspective.

Finally, there is the full range of "beyond compliance" obligations that a market-failures perspective imposes (Norman 2012). Lawrence Mitchell once referred to the modern business corporation (in a memorable phrase) as "the perfect externalizing machine" (2001: 49). This is something of an exaggeration, but it remains the case that most firms do not pay any attention to the distinction between creating value and displacing costs, and are happy to make money either way. Thus a sustained focus on reducing negative externalities would have a transformative impact on business culture. The same can be said for information asymmetries and market power. As Arthur Applbaum has observed, canonical textbooks on business strategy, such as the work of Michael Porter (1990), are basically guides for aspiring executives that focus almost entirely on how to make money by "creating and sustaining market failures" (1999: 194). Much of the advice is aimed at helping firms avoid what is sometimes called "commodity hell," where products are undifferentiated and so firms have no option but to compete on price. The best way to do this is by reducing the competitiveness of the market. This type of strategy is, as Applbaum suggests, difficult to justify by pointing to the good ends achieved by competitive markets. "It is one thing to think that self-interested actions lead to the greatest good by way of the invisible hand. It requires an optimism that competes with that of Dr. Pangloss to think that actions deliberately designed to undermine the mechanism of the invisible hand lead to the greatest good by way of... that same mechanism" (1999: 195).

As a matter of fact, if one looks at the narrow range of profit-seeking strategies that are permissible under the market failures perspective, it would be easy to come to the conclusion that the normative theory is too demanding. Indeed, Wayne Norman has expressed reservations about the view on precisely these grounds: "all major firms derive a certain—and often significant—percentage of their earnings by exploiting market failures: do we really want to conclude that they are all to that extent *necessarily* unethical?" (2012). The answer, I would say, is "no," but in order to get there, we need to move out of the realm of ideal theory into that of non-ideal theory. Under the assumption of full compliance, I do not think Norman or many others would hesitate to say that any deviation from the deontology prescribed by the market failures view is unethical. The world that we live in, however, contains a significant amount of noncompliance, and furthermore, because of the competitiveness of the market economy, noncompliance by one firm can put very serious pressure on all of its competitors. Just as there are some markets in which it is simply impossible to compete without acting illegally—because all of one's competitors are—there are many more markets in which it is impossible to stay in business without acting unethically. It is precisely because these circumstances are so common that business ethics, or at least the interesting, relevant part of it, takes the form of third-best, non-ideal theory.

My inclination would be to distinguish these aspects of the non-ideal framework from the ideal by classifying the considerations raised there as excusing, but not justifying, conditions. So if all one's competitors are exploiting a particular regulatory loophole, this does not make it right to do so, but it may provide one with a reasonable excuse for acting wrongly. Gregory Kavka (1983) has written, with considerable insight, about what he calls "defensive violations of moral rules," attempting to specify the conditions under which they may be acceptable in a business context. William Baumol (1974) has also developed the idea in a particularly interesting direction. He starts by criticizing the suggestion that businesses might voluntarily refrain from engaging in immoral tactics, on the grounds that competitive pressures makes this unreasonably costly. But to the extent that competition excuses the behavior, Baumol argues that it also generates a higher-order obligation to take practical steps to bring about a rule change, so that everyone is forced to stop (he refers to this as "metavoluntarism"). So, for example, firms in polluting industries may be excused for polluting, but they thereby acquire a positive obligation to lobby the government in support of regulatory changes that prohibit the particular form of pollution.

Both of these arguments show that it is possible to derive interesting and important results, working in a third-best, non-ideal framework. The fact that one is forced to move so quickly from an ideal to a non-ideal framework, when thinking about the practical implications of the market failures perspective, shows that Paretianism is neither lax nor insufficiently demanding. It is, from the standpoint of existing business practice, almost utopian. The utopia, however, is only a third-best utopia, because it governs a competitively organized domain of social interaction, which necessitates certain compromises. The need to keep these two different thoughts in mind simultaneously is what makes business ethics so counterintuitive.

7.6. Conclusion

The way that we think about the market is still very much influenced by the great contest between capitalist and communist states that dominated politics in the second half of the twentieth century. Communism as a political project was clearly pitched at a first-best normative level. It was predicated on the assumption that the abolition of private property would eliminate the major (perhaps even the sole) source of conflict in human societies. As a result, communists believed that very few "concessions to human nature" would be required in formulating the organizing principles of their society. Faced with this challenge, there was an understandable impulse on the part of defenders of capitalism to overstate its virtues, by presenting it as though it were a rival first-best arrangement. This is what motivated the long-standing tendency to characterize capitalism as a system of natural liberty and justice, rather than simply of decentralized decision-making. This was always something of an implausible claim, but the stakes were high enough, in the middle of the Cold War, that many people were willing to overlook the obvious difficulties.

In the end, the history of communism revealed precisely the dangers of trying to implement principles of social organization developed without sufficient regard for the frailties of human nature. By setting their sights too high, communist states wound up achieving worse outcomes along every dimension of normative evaluation: they were not only more unfree and inefficient, they were even more unequal than some modern capitalist economies. Capitalist societies, by setting their sights lower, in the end wound up producing more humane and just economic orders. They did so in part by instituting a division of moral labor, so that private economic institutions are accountable only to a weaker conception of social justice (namely, constrained Pareto efficiency), whereas the public sector continues to be governed by a more robust conception (namely, prioritarianism). Thus, the state engages not only in activities that are complementary to those of the market (e.g., provision of public goods, regulation of externalities, control of natural monopolies, provision of social insurance), it does so on terms that limit the scope of inequality (e.g., through progressive taxation, cross-subsidization within social insurance schemes, provision of means-tested benefits, reduced barriers to intergenerational social mobility, etc.). There is, of course, enormous variation from country to country in the level of commitment to inequality reduction, with the United States being an important outlier at the low end. Nevertheless one can still recognize the same basic template underlying all these different societies.

This complex institutional structure explains why individuals qua economic actors are not accountable to everyday, all-things-considered morality, and also why they are less accountable to considerations of distributive justice than state actors.[15] At the same time, the fact that they are subject to a weaker set of constraints should not be confused with being subject to no constraints (as proponents of the "markets as moral-free zones" have suggested). The language of ideal and non-ideal theory, unfortunately, is not sufficiently nuanced to capture these distinctions. My ambition in introducing the apparatus of first-best, second-best, *n*th-best, as a dimension of normative theorization that interacts with, but is different from, the distinction between ideal and non-ideal, is to find a more precise way of articulating the way that normative principles can be weakened, in order to render them more incentive-compatible, without being dissolved entirely.

[15] This is how I interpret Kenneth Goodpaster's influential critique of stakeholder theory, and his claim that it "represents nothing less than the conversion of the modern private corporation into a public institution and probably calls for a corresponding restructuring of corporate governance" (1991: 63).

8 }

The History of the Invisible Hand

The most persuasive argument in support of the market economy has always been that it facilitates cooperation, enabling individuals to engage in mutually beneficial interactions that otherwise might not occur. It does so, however, in a way that is not entirely transparent, either to an observer or to the parties involved. Hence the popularity of Adam Smith's "invisible hand" metaphor as a way of describing the way that markets promote cooperation. Yet how exactly this invisible hand works, or what exactly it does, has been the subject of changing views over time. Several different accounts have been provided regarding both the *nature* of the cooperative benefits in question and the *mechanism* that allows the market to enhance the provision of them. Until the early twentieth century, the dominant view was that markets were superior to other institutional forms because they allowed society *to economize on moral motivation*. I refer to this as the incentive argument for the market. Smith may be regarded as the progenitor of this line of thinking, yet the person to have given it its most sophisticated articulation is Émile Durkheim. However, the development of large-scale bureaucratic organizations in the nineteenth century, followed in the twentieth century by the emergence of centrally planned economies and large multidivisional corporations within the capitalist world, led to increasing doubts about the significance of market incentives. Thus another view came to acquire increased prominence; it focused on the claim that markets allow society *to economize on knowledge*. I refer to this as the information argument for the market. The progenitor of this line of thinking is Léon Walras, but the doctrine was given its most influential articulation by Friedrich Hayek.

The history of these arguments in favor of the market is well known to those who count themselves among its supporters. Among critics of the market, on the other hand, there is much less clarity. Indeed, one can find in many places a tendency to run together the two arguments, or else just to ignore the second strand and assume that the only considerations that speak in favor of the market are the incentives it supplies. The result is a certain unevenness in the level of sophistication exhibited by critics of capitalism. My objective in this paper is to trace the history of these different lines of thought, in order to show how the various threads came together to provide the most sophisticated contemporary case for the market. I will begin by showing how the incentive argument arose with Bernard Mandeville and Smith, and how it was refined over the course of the nineteenth century until receiving its definitive formulation with Durkheim. I will then trace the beginnings of the information argument with Walras, showing how it was refined over the course of the "socialist calculation" debate, before receiving its definitive formulation with Hayek. Beyond this, I would like to show how the information argument led to a reconceptualization of the incentive argument, and how this was used to broaden the case for the market in the work of Joseph Schumpeter and János Kornai.

My objective in all of this is not to show that the case for the market is unassailable. I do, however, want to show that it is both formidable and complex. Complexity is the most important point. What the history of reflection on the subject of the invisible hand shows is that "the market" bears a certain resemblance to "democracy," in the sense that it is a complex set of institutional practices, which for various reasons we regard as highly successful, but which are also incompletely theorized. Indeed, democratic theory is currently riven by three completely different, incompatible accounts of what provides the normative authority of democratic institutions (some people say it is the aggregation of preferences, others the process of deliberation, others the competition for political leadership). Theorists will put different emphasis upon different institutional features (e.g., elections, political parties, the free press), depending upon the account they subscribe to. I would like to show that the same is true of the market. "Capitalism" is a complex set of practices, institutionalized in very different ways in different parts of the world, and subject to different strands of normative reconstruction. Thus the importance that theorists assign to institutions such as the stock market, the tort system, regulation, or antitrust enforcement—to pick just a few examples—will depend enormously upon the judgments that they make about what benefits the market provides to society, and what sort of underlying mechanism generates these benefits. An appreciation of these nuances is essential for developing an informed assessment of the contributions that the market makes to our society.

8.1. Origins in Smith

As is well-known, the "invisible hand" was something of a throw-away line for Smith. He never intended it as a characterization of the central mechanism through which the market produced cooperative benefits. Nevertheless, it is easy to see how it became associated with the central thesis of *The Wealth of*

Nations. Smith used the phrase to describe a decision on how to invest capital and observes of every individual that "He generally, indeed, neither intends to promote the publick interest, nor knows how much he is promoting it…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" ([1776] 1976: 454 [IV.ii.9]). Smith goes on to make the more general observation that "by pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it" ([1776] 1976: 454 [IV.ii.9]). This paradoxical formulation is intended, quite consciously, as a reference to Bernard Mandeville's suggestion that, in the domain of economic exchange, private vices may amount to public virtues. The central difference between Smith and Mandeville is that Smith actually had a plausible account of how this might be so.

Indeed, critics of Mandeville were not wrong to castigate his work as mere contrarianism. While declaring, of the grumbling hive, that "every Part was full of Vice / Yet the whole Mass a Paradise" ([1705] 1997, 27), Mandeville fails rather dramatically to provide a plausible account of how this might come about. His central contention is that vice produces public benefits by generating employment. The mechanism that he appeals to is in each case an instance of what would later become known as the "broken window" fallacy, after the young boy in Frédéric Bastiat's parable, who creates work for the glazier by breaking a shopkeeper's window. Substitute "Sharpers, Parasites, Pimps" and "Players" for stone-throwing young boys and you have the substance of Mandeville's argument. Yet as Bastiat wrote,

> [I]f you come to the conclusion, as is too often the case, that it is a good thing to break windows, that it causes money to circulate, and that the encouragement of industry in general will be the result of it, you will oblige me to call out, "Stop there! Your theory is confined to that *which is seen*; it takes no account of *that which is not seen*." *It is not seen* that as our shopkeeper has spent six francs upon one thing, he cannot spend them upon another. *It is not seen* that if he had not had a window to replace, he would, perhaps, have replaced his old shoes, or added another book to his library. In short, he would have employed his six francs in some way, which this accident has prevented. (1880: 72–73)[1]

Smith, on the other hand, has a much better account. First of all, he does not celebrate vice, but rather self-interest. The key characteristic of the market is that it allows individuals, within the scope of respect for property rights

[1] Mandeville's view is sometimes defended on the grounds that it anticipates Keynes's paradox of thrift. This is far too generous by half and is in any case a misunderstanding of Keynes. Saving itself does not generate a shortfall in demand, it is only under certain conditions that it can exacerbate such a problem.

and contract, to advance their own interests without regard for the interests of others. In many areas of life this would have noxious consequences—as Smith clearly believed, whatever the assertions of some of his critics. What he claims is that in the domain of economic exchange it does not, because it *promotes the division of labor*, which is the form of cooperation centrally responsible for the increased wealth of nations. Thus the discussion of the invisible hand in *The Wealth of Nations* refers back to one of the earlier passages in the book, this time one that does contain Smith's account of the central mechanism through which the market provides benefits to society:

> But man has almost constant occasion for the help of his brethren, and it is in vain for him to expect it from their benevolence only. He will be more likely to prevail if he can interest their self-love in his favour, and show them that it is for their own advantage to do for him what he requires of them. Whoever offers to another a bargain of any kind, proposes to do this. Give me that which I want, and you shall have this which you want, is the meaning of every such offer; and it is in this manner that we obtain from one another the far greater part of those good offices which we stand in need of. 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. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages. . . . As it is by treaty, by barter, and by purchase, that we obtain from one another the greater part of those mutual good offices which we stand in need of, so it is this same trucking disposition which originally gives occasion to the division of labour. ([1776] 1976: 26–27 [I.ii.2])

The central insight here was further sharpened by Edward Gibbon Wakefield, who in his annotated edition of *The Wealth of Nations*, distinguished two different forms of cooperation: "first, such cooperation as takes place when several persons help each other in the same employment; secondly, such co-operation as takes place when several persons help each other in different employments. These may be termed Simple Co-operation and Complex Co-operation" (Smith 1835: 26).[2] The benefits associated with the former are what we would now be inclined to refer to as economies of scale, while the latter are gains from trade (where these are taken to include, not just increased preference-satisfaction but also increased production through specialization and more efficient use of resources). One of the central differences between these two forms of cooperation is that "of the former, one is always conscious

[2] This famous quotation occurs in a set of lengthy notes written by Wakefield, which are interspersed with the text of *The Wealth of Nations* in the 1835 edition. It was popularized by John Stuart Mill, who cites it in his *Principles of Political Economy* ([1848] 2004: 134).

at the time of practicing it: it is obvious to the most ignorant and vulgar eye. Of the latter, but a very few of the vast numbers who practise it are in any degree conscious" (Mill 1848: 135). Hence the metaphor of the "invisible hand." The movement from autarky to specialization is only possible if one can count upon others to supply one in the areas of production that one is contemplating abandoning. The existence of a market, or a system of commercial exchange, makes this possible. Or as Smith put it, "it is the power of exchanging that gives occasion to the division of labour" ([1776] 1976: 31 [I.iii.1]). Yet the person who supplies these goods "to the market," as it were, need not be aware of what forms of specialization he is facilitating. The baker need not know what others will do with their time, now that they are no longer obliged to bake their own bread. When he spends the money he has earned, he is clearly engaged in a system of reciprocity that encompasses both the person to whom he has sold goods, and the person from whom he is purchasing goods. But the connection between his production and his consumption is unknown to him, and so the system of cooperation is invisible.

John Stuart Mill provided the following example of this:

> In the present state of society the breeding and feeding of sheep is the occupation of one set of people, dressing the wool to prepare it for the spinner is that of another, spinning it into the thread of a third, weaving the thread into a broadcloth of a fourth, dyeing the cloth of a fifth, making it into a coat of a sixth, without counting the multitude of carriers, merchants, factors, and retailers put in requisition at the successive stages of this progress. All these persons, without knowledge of one another or previous understanding, co-operate in the production of the ultimate result, a coat. But these are far from being all who co-operate in it; for each of these persons requires food, and many other articles of consumption, and unless he could have relied that other people would provide these for him, he could not have devoted his whole time to one step in the succession of operations which produces one single commodity, a coat. ([1848] 2004: 136)

This passage obviously harkens back to Smith's celebrated example of the pin factory, with which he begins his discussion of the division of labor in *The Wealth of Nations*. Yet Mill rather subtly modifies the example, in order to remove a major awkwardness in Smith's discussion. The operations that Mill cites, in the case of a coat, are all integrated through market transactions. One is invited to imagine the farmer selling the wool to the spinner, who sells it to the weaver, and so on. The pin factory example, by contrast, involves workers in a factory, all under one roof, engaged in a division of labor *without any intervention of the market* (through what we would now call "administered transactions"). Smith claims that "this division of labour, from which so many advantages are derived, is not originally the effect of any human wisdom"

([1776] 1976: 25 [I.i.1]). Yet the pin factory seems to belie this claim. Indeed, one can imagine that the organization of the pin factory was a direct consequence of human wisdom. Someone sat down, looked at how pins were made, and decided—quite consciously—to break the task down into its component parts. So while the example succeeds in illustrating the gains in productivity that can be accomplished through a division of labor, it manifestly fails to show that the market plays any sort of essential role in this. Smith is left claiming only that the division of labor is not "originally" the effect of human wisdom, but rather is "the necessary, though very slow and gradual, consequence of a certain propensity in human nature which has in view no such extensive utility; the propensity to truck, barter, and exchange one thing for another" ([1776] 1976: 25 [I.ii.1]). Perhaps this is so, and the market is to be credited with having "invented" the division of labor, or having initially alerted people to the benefits that could be achieved through this form of cooperation. But now that we know about this, what further purpose does the market serve?

At the beginning of the twentieth century, this question began to be posed with much greater force. In Smith's time, however, it was seldom raised. In part this was because it seemed obvious that the market permitted a far more extensive division of labor than could be created within a single factory or organization. This was particularly apparent in areas that were involved in or relied upon international trade. But there was another factor as well. One of the features of market interactions is that, because the parties are motivated by self-interest, and no one expects anything else, certain types of disputes, which characteristically occur within organizations, have no opportunity to arise. Workers, for instance, tend to be rather attentive to how much the person next to them is working, how much he or she is being paid, etc. Working together requires a degree of cooperativeness, which can easily be undermined by jealousy, shirking, one-upmanship, or even just personality conflicts. It is not just that "benevolence" is in short supply; there is an entire gamut of prosocial dispositions that are required in order to achieve face-to-face cooperation, all of which are in short supply. This is just one of the reasons that organizations become unwieldy as their size increases. There is, by contrast, something seamless about the way that markets coordinate interaction—certain social frictions simply do not have the opportunity to arise. In the early nineteenth century, however, economists were in no position to articulate this intuition. The ability to do so depended upon the development of sociological theory.

8.2. Durkheim's Refinement

When it comes to the limits of benevolence, Smith had what might be regarded as a common-sense view of the matter. The average person exhibits a certain degree of altruism, particularly toward family and friends, and can occasionally be motivated to help a stranger. But this disposition is rather limited and subject to exhaustion. When the limit is reached, the individual can be expected to act in a self-interested fashion. Cooperation, however, to the extent that it is vulnerable to free-rider problems, requires that individuals refrain from pursuing their self-interest. It therefore imposes a *motivational burden* upon individuals, which in turn generates a prima facie limit on the extent of cooperation, and therefore an upward bound on both the scale of a society and the level of social complexity. Thus an institutional arrangement, such as the market, which allows individuals to cooperate without tapping into these motives, is an important discovery. It expands the scope of cooperation by allowing society to economize on moral motivation, that is, to get more out of the level of moral motivation that it can plausibly expect (or non-tyrannically demand) from its members.

David Hume had his eye on a similar idea with his distinction between "natural" and "artificial" virtue ([1739] 1978: 477). Artificial virtues are cultivated when the motive of self-interest is tapped, in order to extend the individual's "natural" prosocial dispositions, thereby allowing for the development of more extensive forms of cooperation. His central examples were property rights and contract. Over the course of the nineteenth century, however, this common-sense analysis receded from view, as economists became increasingly attracted to models of rational action that downplayed the importance of altruistic action. Thus the *Harmonielehre* that Smith is sometimes accused of holding—positing a spontaneous harmony of interests among individuals—became an increasingly common view. Along with this, however, came increased difficulty explaining what was special about the invisible hand of the market. For example, if government also arises, spontaneously, through self-interested action, out of the state of nature, then what makes the market a superior instrument for coordinating social interaction?

One salutary consequence of this drift toward egoism—or theories of rational action that privileged instrumental rationality—is that it forced defenders of the common-sense position to provide a more sophisticated articulation of their view. Perhaps the most important attempt to clarify the logic of Smith's original position was carried out by Émile Durkheim, in his *De la division du travail sociale* ([1893] 1930). The extent of the agreement between Durkheim and Smith was somewhat obscured by the fact that Durkheim was sharply critical of the "economic" approach of the day, and chose Herbert Spencer as his chief target of attack. His central interest was in showing how, for every system of order established on the basis of self-interest, there was a background system of order that was not—hence his interest in what he called the "non-contractual elements of contract" ([1893] 1930: 184–197).

Durkheim's central theoretical achievement, when it comes to our understanding of social order, was to have escaped from the traditional oscillation between organicism and atomism, and to have proposed instead what is now referred to as the "cultural theory of social integration" (Lukes 1975). The standard organicist view compared human society to a beehive, with individuals simply being born to perform certain functions for the benefit of the hive as a whole. The atomist view started instead with a Hobbesian state of nature, treating individuals as unwilling to perform any action that was not in their own interest. The inadequacy of both theories can be articulated most clearly in terms of Kant's characterization of the "unsocial sociability" of man. The organicists had difficulty explaining the unsocial part, while the atomists had trouble explaining the sociability.

Durkheim agreed with the atomists that social order—the predominance of regular, cooperative interaction, in accordance with stable, reciprocally shared expectations—needed to be understood in terms of the specific motives that individuals have for acting in an "orderly" fashion. Orderliness was achieved, in his view, because individuals acted, not just on their own private desires and interests, but on the basis of what he called the *conscience collective*—most importantly, a set of action-guiding beliefs that were socially transmitted, both horizontally—from person to person—and vertically—from one generation to the next. Following Talcott Parsons's (1951) influential regimentation of Durkheim's terminology, we now refer to such socially transmitted symbolic content as *culture*, and the most abstract set of action-guiding beliefs as *shared values*. The truth of organicism lay in the observation that the content of these values tended to reflect the "needs" of society as a whole, and so to the extent that individual action was subject to the control of the *conscience collective*, it would tend toward satisfaction of these needs. For example, a society under constant threat of attack from its neighbors would come to celebrate martial virtues, which would in turn provide individuals with the motivation to engage in concerted collective action in defense of its borders. The extent to which individuals are under the control of the conscience collective, however, is subject to variation, and individual interests, to the extent that they find expression, have the capacity to subvert shared values. Durkheim introduced the term "anomie" to describe precisely these failures of social integration (and to account for the "unsocial" component of human "sociability").

Again, Parsons provided an extremely influential regimentation of Durkheim's terminology. In order to achieve social integration, and in order to be effectively reproduced, shared values must be institutionalized in society and internalized in personality (1970: 297). Internalization in personality means that children, over the course of primary socialization, must incorporate these values into their sense of personal identity in such a way that they become motivationally effective. (For example, boys must come to want to be proud warriors, or honest brokers, or good providers.) Institutionalization means that the values become the basis for a set of sanctioned regularities of conduct, or social norms, so that those who fail to live up to the appropriate standards are punished (either physically, or through various modalities that presuppose partial internalization, such as shaming, ridicule, or stigmatization). Effective social control requires both: not just a solid body of individuals who are sufficiently motivated to follow the rules of their own accord, but a set of sanctions in place, both to discourage those who do not follow the rules and to symbolically reaffirm society's commitment to them, in order to prevent those who do follow them from becoming overly disgruntled.[3]

Durkheim describes societies that rely on nothing more than shared values as a medium of social integration as exhibiting *mechanical solidarity.* This is a peculiar term, which Durkheim chose in order to convey the idea that the society coheres only to the extent that individual action is governed by the conscience collective, or the system of shared values. Individualism, in such a society, is a force that only serves to undermine the social order—and so the society is orderly only to the extent that it can succeed in subordinating the individual to the collectivity ([1893] 1930: 100). There is, in Durkheim's view, a natural limit on the scale of a society that can be achieved in this fashion. His account of the reasons for this is complicated, and somewhat odd, but the general claim is plausible. Mechanical solidarity relies very much upon face-to-face interactions, among individuals who know each other, can track each others' behavior, and can directly enforce social norms. As the volume of a society (the number of individuals) and the density (the number of transactions) increases, the system of shared values begins to lose its grip upon the individual. This can generate anomic conditions, characterized above all by an increase in violence—both self and other-directed.

Thus the archetype of mechanical integration is the small-scale, egalitarian hunter-gatherer society (governed by a moral code roughly similar to what G. A. Cohen would later describe as the norms of the "camping trip" [2009]), in which everyone is engaged in the same suite of activities and has roughly the same set of skills and needs. These societies are, as Durkheim noted, extremely susceptible to fission. There is some irony in this fact—since the social bond between individuals is so strong, small disagreements can easily lead to rupture. If people cannot get along, one party will simply leave and form a new group (a pattern that one can see quite easily in hunter-gatherer groups that are not subject to territorial pressure [Boehm 1999]). When this option is not present—as with agricultural societies, where population growth has generated resource constraints—one begins to see the emergence of segmentary societies. The idea is that the initial group, facing internal constraints on its expansion, essentially clones itself, creating a new, small-scale society.

---

[3] Hence Durkheim's somewhat counterintuitive view that the crime rate in a society is determined by the society's need to punish criminals, and not by the number of people who actually do things that are wrong. After all, people violate social norms all the time. Society criminalizes such violations only to that extent that is necessary to sustain its system of shared values, which involves both deterring potential deviance and symbolically reaffirming these values, for the edification of some and the appeasement of others.

("We say that these societies are segmentary in order to indicate that they are formed through repetition of aggregates that are similar to one another" [Durkheim 1930: 150].) So instead of seeing a large, loosely integrated society, one finds instead a loose federation of small, tightly integrated social groups. This is the basic form of tribal social organization. Durkheim's most important observation is that the segmentary form of these societies is a reflection of the limitations of human sociality, or more specifically, the upper bound that exists on the scale of a society that can be achieved through shared-value integration (i.e., mechanical solidarity).

The most common form of such a society is the clan-based tribal society, where integration in each segment is bolstered by having it organized around an ascriptive source of solidarity, such as family membership or lineage. But there are other forms. One could see similar structures in many ancient and medieval cities. In the middle ages, for instance, a European city might give the appearance of being a large-scale society with diffuse social ties. Yet closer attention would reveal that it was in fact a federation of *parishes*, each of which presided over a particular neighborhood and exercised extremely tight control over the day-to-day lives of individuals (and each parish was often described as being like a small village) (Kaplan 2007: 62–63).

The division of labor was important, for Durkheim, because it allowed a different form of social solidarity to emerge, what he referred to as organic solidarity. Organic solidarity emerges when individuals engage in activities that are complementary to one another, paradigmatically, through a division of labor. The importance of the division of labor, from the standpoint of social integration, is that it creates a society where there is less antagonism between the interests of the individual and those of the collectivity. As a result, it is less important that there be shared values, or that individuals be socialized in such a way that their interests are completely subordinate to those of the community. Thus it is possible to create a social order on the basis of constrained instrumental action—instrumental, because individuals are seeking to advance their interests in a self-serving way, but also constrained, because the system of shared values fixes the rules of the game, determining how far individuals can go in seeking to advance their interests. Within such a system, assuming that the constraints hold, individual self-interested action can occur without necessarily generating anomic conditions. As a result, it is possible to organize a much larger-scale society, because it is not necessary to exercise as much social control over the individual. The system of socialization and social control is required only to guarantee the integrity of the constraints, it is not needed to motivate the specific actions that individuals perform. Instrumental action, combined with natural complementarities of interest, take over the role that was once served by socialization into a set of shared values. Thus the market can be thought of as a type of institutional kluge, used to extend the "natural" boundaries of human sociality.

8.3. Problems with the Incentive Argument

The extent to which Durkheim is the true disciple of Adam Smith can be seen by comparing his analysis to the account of "spontaneous order" that emerged at roughly the same time in the Austrian school of economics. Carl Menger, for instance, sought to generalize the "invisible hand" model, arguing that "law, language, the state, money, markets... [the] prices of goods, interest rates, ground rents, wages, and a thousand other phenomena" are to "no small extent the unintended result of social development" (Menger 1985: 147). The key task of the social sciences, he argued, was to answer the question "how can it be that institutions which serve the common welfare and are extremely significant for its development come into being without a common will directed toward establishing them?" (Menger 1985: 146). The central distinction is between planned (those based on a "common will") and unplanned orders. But the category of unplanned or "spontaneous" orders includes a number of rather distinct phenomena, which after Durkheim one would be inclined to distinguish. In particular, Menger treats institutions that are integrated informally through social norms as being of the same genus as those that are integrated through systems of constrained instrumental action. From Durkheim's perspective, this analysis confuses institutions that are subject to mechanical integration with those that exhibit the type of organic solidarity brought about by the market. The former type may, of course, be "unplanned" in the sense that the institutions in question are subject to cultural evolution—and thus various "functional" constraints—that no particular individual fully understands. Individuals who grow up in a culture that celebrates martial virtues may have no idea why their culture is this way. But that does not mean that a citizen army in such a society, composed of willing volunteers, can usefully be described as an unplanned order. The difference between this and the market is that in the former case the behavior is fully scripted by the culture, and the outcome is directly enforced by others. In the case of the market, by contrast, there is no such direct enforcement of the outcome, only the general constraints. Often no one even knows what the outcome will be.[4]

Thus Menger, with his excessive enthusiasm for "invisible hand" explanations, wound up obscuring what was distinctive about the market. If the law, the state, and the market are all just institutions that arose in a spontaneous, unplanned fashion, then there is no way to explain how the market facilitated the development of a new form of social integration, and hence the breakdown of segmentary patterns of cooperation and the emergence of large-scale "horizontal" cooperation. As a result, the analysis tends to underplay the qualitative difference between market and non-market societies. Durkheim, by contrast,

[4] A point that was actually made quite well, with respect to the market, by Hayek (2002: 10).

is able to explain why the market permits the emergence of entirely new social formations.

So what is the problem with the incentive argument for the market? What eventually came to undermine Durkheim's analysis—as an account of what the invisible hand of the market contributes to society—was precisely the same ambiguity that afflicted Smith's example of the pin factory. While the market may be good at promoting the division of labor, it is not the only mechanism capable of doing so. And since it is the division of labor that permits the emergence of large-scale cooperation, and not the market per se, if it were possible to find other ways of instituting a division of labor, it would be possible to have organic solidarity in the absence of the market.

One of the other institutional kluges used to extend human sociality—one that predates the market—is that of hierarchy. Shared-value integration presupposes a system in which socialization meets social control halfway—where individuals are motivated to respect the system of norms, but where they are also punished for any instances of deviance. If one uses the system of shared values, however, not to institute a particular system of, say, joint economic production, but rather just to institute an enforcement agency, then one may be able to rely upon social control alone, that is, the system of sanctions, in order to enforce a particular economic system. For example, one way to organize a common field system is through shared values: everyone works the field together, hard work is praised, laziness is stigmatized, and so on. An alternative is to create a system of forced labor, where peasants work the fields in order to avoid punishment from, say, an aristocratic class that monopolizes possession of weapons and the use of force. Note that the latter system does not allow one to dispense with the system of shared values entirely. While the peasants are motivated primarily by instrumental reasons, the integrity of the punishment mechanism still depends upon genuine commitment on the part of those who are called upon to do the enforcing. Thus one will typically find societies that are organized this way possessed of an "ideology," which legitimates the actions of the ruling class, first and foremost in its own eyes. Shared values are being leveraged here, by focusing their bonding energies on regulating the use of force, which in turn permits the construction of a social order based primarily on the threat of sanctions.

Hierarchical societies therefore represent another break with the segmentary structure—one that Durkheim paid too little attention to. Partly this is because in Europe the emergence of hierarchy and of markets tended to be blended. In other parts of the world, where there was a stronger, more centralized state, it was more obvious that the hierarchical mode could subordinate segmentation.[5] (This is what motivated Marx to distinguish the "Asiatic

[5] Christopher Boehm's discussion of clan societies, and how they resist the emergence of hierarchy, is particularly interesting on this point (1999: 106–124).

mode of production"—where redistribution of the surplus occurs through the state—from "feudalism," which has a much more decentralized, local character.) Hierarchies have long been used to institute a division of labor. The institution of *corvée* labor under feudalism, for instance, gave the aristocracy access to a workforce that was often deployed in the service of public works projects, such as roads and bridges. It is significant to note the very few instances of egalitarian agricultural societies—organized on a segmentary pattern—ever succeeding in constructing a large-scale irrigation system. The collective action problems are too great. Since large segments of the population must be freed from agricultural labor for long periods of time, a significant amount of reciprocity, including a division of labor, is required. The great irrigation systems of antiquity were typically constructed after the emergence of a hierarchical state.[6]

The emergence of bureaucracy in the late nineteenth century involved a significant rationalization of hierarchical modes of integration. In particular, it overcame one of the major weaknesses of hierarchical forms, which is the ability of those who occupy positions of authority to use the hierarchical apparatus in order to resist any attempt to evict them from their positions. Bureaucracies impose formal procedures both for the ascension to power and for orderly exit. Thus, as hierarchical modes of organization were rationalized, they became a competing source of cooperative benefit. Not only did the state begin to rival the market, when it came to institutionalizing certain transactions associated with the division of labor, but more obviously, large hierarchical firms began to emerge within the market itself. The internal division of labor within corporations began to seem as important a source of wealth as that achieved by the market.

The process accelerated in the early years of the twentieth century, in several different dimensions: the emergence of multidivisional firms in the 1920s (Chandler 1962), the "separation of ownership and control" associated with the transition to professional management (Berle and Means 1932), and the emergence of central planning in communist countries after the Russian revolution. Consider the case of multidivisional firms. By the end of the nineteenth century, it had been common for firms to have internal differentiation along functional lines—classically, between different "departments" such as manufacturing, marketing, finance, and sales. The firm as a whole, however, was still organized around a single line of business: railroads, oil, manufacturing, etc. A multidivisional (or M-form) firm, however, was involved in several distinct sectors. Thus it would replicate the market internally, organizing each

[6] This observation is what gave rise to the so-called "hydraulic" theory of state formation, which probably reverses the correct order of causation, but otherwise contains an important observation. For discussion, see Carneiro (1970).

"division" around a particular product line, each of which would have its own, functionally differentiated substructure (i.e., finance, marketing, etc.), with a head office engaged in overall coordination and strategy. In the early twentieth century, it was not obvious to many observers that there were any limits to this organizational form—it seemed as though anything the market could do externally, the firm could do internally. Hence the widespread acceptance of John Kenneth Galbraith's suggestion, in *The New Industrial State*, that there would be convergence between capitalism and communist nations, as both moved toward more and more planned systems of production.

One way of describing this overall trend is to call it simply "the emergence of management." The challenge that this process posed to Smith's account of the invisible hand was twofold: first, it showed that the central advantage of the market—its ability to economize on moral motivation by providing an incentive system that induced individuals to engage in prosocial action—could be achieved by bureaucracies as well, and second, it suggested that bureaucracies might be somewhat better at this than the market, as witnessed by the growth of large corporations within capitalist economies. As long as the central argument for the market was motivational, there was this weakness. After all, the manager of a modern corporation is not motivated by self-interest in the way that a traditional entrepreneur is. Customers no longer deal with the brewer, the butcher, or the baker, but rather with an employee, who works for a supervisor, who works for a manager, none of whom typically have any direct financial stake in the transaction. The firm may be interested in making a profit, but this is not to be confused with the pursuit of "self-interest" on the part of its representatives, since the profit goes almost entirely to investors (it is, as they say, OPM—other people's money). Employees, up to and including senior management, are motivated by the non-market incentives provided internally by the firm (such as "climbing the corporate ladder," or the threat of dismissal).

Thus the view became widespread that the rise of "managerial capitalism" had rendered the invisible hand otiose. Here is how Alfred Chandler put it, in *The Visible Hand*:

> In many sectors of the economy the visible hand of management replaced what Adam Smith referred to as the invisible hand of market forces. The market remained the generator of demand for goods and services, but modern business enterprise took over the functions of coordinating flows of goods through existing processes of production and distribution, and of allocating funds and personnel for future production and distribution. As modern business enterprise acquired functions hitherto carried out by the market, it became the most powerful institution in the American economy and its managers the most influential group of economic decision makers. The rise of modern business enterprise in the

> United States, therefore, brought with it managerial capitalism. (Chandler 1977: 1)

As proponents of this managerial revolution became more confident in their ability to design an effective system of social control, they became less convinced of the need for the market to achieve social integration. From their perspective, proponents of the market underestimated the human capacity to design effective non-market incentive systems.

On the flip side of the coin, many began to suspect as well that proponents of the market had underestimated our capacity to evoke prosocial behavior through more effective forms of socialization. There were two components to this. First, as the *homo economicus* model of action became increasingly well-specified, and its general assumptions became more pervasive in economic modeling, critics began to argue that proponents of the incentive argument for the market underestimated the extent to which individuals could be motivated by values and norms. The market, according to this perspective, is an institution designed for a "race of devils." Many people reasoned, therefore, that to the extent that we are not quite as devilish as economists believe, we have less need for the services of the invisible hand. Second, many people began to think that the market did not just extend the scope of cooperation beyond the limits of our capacity to be motivated by shared values, but that it actually undermined the system of shared values, thereby limiting the capacity of individuals born and raised in a market society to exhibit much social solidarity. Thus the invisible hand, according to this view, creates the demand for its own services, by eroding the moral motivation required for people to engage in direct cooperation. The market only appears to extend the scope of human cooperation because socialization in a capitalist society dramatically limits people's capacity for prosocial behavior. Eliminate the market and you remove this constraint. This was the hope underlying the vision of a "new socialist man," who would exhibit all of the virtues required for large-scale cooperation without the need for any appeal to self-interest. In the same way that improvements in our systems of social control might eliminate the need for the market, many came to believe that improvements in our techniques of socialization (or to make that sound less sinister, the creation of a less possessive, less individualistic culture) might obviate the need for the market.

8.4. The Information Argument for the Market

The emergence of central planning in the communist bloc posed the challenge associated with the rise of management in its sharpest form. Here one could see an entire economy in which the mechanism of administrative hierarchy was in the ascendant. (There were, of course, still some market transactions—workers were paid wages, which they used to purchase consumer goods in shops. The point is that all the major decisions about production and the use of resources were made by non-market institutions.) Perhaps even more awkwardly for proponents of the market, the rise of managerial capitalism led to a whole new set of new proposals for a socialist reorganization of the economy—the set of blueprints that we now lump together under the title of "market socialism." Since the managers of major corporations are now just bureaucrats, taking their orders from a group of anonymous investors scattered about from one end of the country to another, what would it matter if the state were to assume ownership, and order the managers to engage in marginal cost pricing?

It is this proposal that generated the now-famous "socialist calculation" debate, with the principal parties being Abba Lerner, Oskar Lange, and Maurice Dobb on the side of market socialism, and Ludwig von Mises and Friedrich von Hayek defending capitalism. The most important feature of this debate is that it led the defenders of the market to shift the argument away from the incentive argument toward what I have been calling the information argument for the market. Underlying this was a growing awareness of the importance of *the price system* in coordinating production in a capitalist economy. To put it in modern terms, von Mises and Hayek suggested that the incentives supplied by the market were more important as a *revelation mechanism*—one that induced individuals to reveal private information, then compiled it in a useful form—than as a motivational system. The invisible hand was important because it produced a system of publicly observable prices, each reflecting the relative scarcity of a particular good or service at a particular point in time in the economy. This led Hayek to criticize the habit of talking about incentives "as though their primary purpose were to induce individuals to exert themselves sufficiently" (2002: 17). In his view, the most important function of prices "is that they tell us *what* we should accomplish" (2002: 17).

Thus Hayek described competition as a "discovery procedure" (2002), rather than as a system designed to enhance individual motivation. Most importantly, it was a discovery procedure whose results could not be replicated by a bureaucracy. The central weakness of organizational hierarchies, from this perspective, is that they involve what Ronald Coase called the "supersession of the price mechanism" (1937). This may be an advantage when it comes to organizing particular transactions, but only when there is an enormous background of prices that are made available by the market.

The reason that the socialist calculation debate is seen as a turning point in the history of the invisible hand discussion is twofold: first, that von Mises and Hayek abandoned Smith's argument for the market, essentially granting that the market is not required in order to achieve a division of labor, or to motivate individuals. In their view, what makes the market superior as a form of economic organization is that it can achieve a more efficient allocation of goods and of productive resources, because it is able to get the prices right (von Mises 1922). The second important point is that Lange and Lerner accepted this shift in the terms of the debate. In other words, they accepted the general principle of scarcity pricing and granted that any socialist society should be just as concerned about achieving efficiency as a capitalist one (Lerner 1977: 236). This was an important shift, since in the early twentieth century many socialists denied that there would need to be prices in a socialist economy, or else believed that prices should be determined on some completely different basis than scarcity (e.g., that they should reflect the amount of labor "embodied" in each good) (Nove 1983: 95). Thus the challenge for socialists, after Lange and Lerner, was to try to show that one could get scarcity prices without having some of the characteristic features of capitalism, such as competition between firms or private ownership of the means of production.

The person who did more than anyone to accentuate the importance of prices was Léon Walras, the father of general equilibrium theory. Traditional arguments for the market had been confined to what we now call a partial equilibrium framework—looking at a particular market and showing how the forces of competition would affect the volume of goods traded. It had been well understood since Smith that the movement of prices was an important mechanism for bringing supply and demand for any particular good into equilibrium (or "clearing" the market, as we would now say) (Smith [1776] 1976: 73–74 [I.vii.9–11]). But the difficulty of moving from this sort of analysis to a general equilibrium framework—one that looked at the economy as a whole—was not fully appreciated until Walras. How is one to know whether the process that brings one market into equilibrium will not put some other one out? In order to show that the individual pursuit of self-interest, in the context of a well-regulated market economy, will lead to an outcome that is best for all, one must show that competition can clear all markets simultaneously. If it cannot, then the efficiency gains achieved in one sector may be at the expense of losses in some other.

Walras's approach to the problem ([1874] 1954), and his major formal insight, was to picture the economy as something like a giant system of linear equations, with one variable and one equation for each particular market. One can find a solution to each equation representing each particular market, yet it is easy to see that there may be no solution that can simultaneously satisfy every equation (i.e., the set of equations as a whole may have no consistent solution) (Blaug 1985: 571). Furthermore, there is an obvious compositional fallacy in moving from the claim that for each equation there is a solution to the claim that for all equations there is a solution. Yet the latter is precisely what would need to be shown in order to vindicate Smith's claims about the invisible hand. If a set of prices can be found that will clear every market simultaneously, then it could be argued that individual self-interest was producing a public benefit, in the sense that it was making at least some people better off without harming anyone. Thus, proof of what Paul Samuelson would later call "the Invisible Hand Theorem" (that a perfectly competitive market would also be Pareto-optimal) became very closely tied to the proof for the existence of a general equilibrium.

Walras's analysis revealed how deep the chasm was between partial and general equilibrium conditions, and therefore exposed the insufficiency of the evidence marshalled by early economists in support of the invisible hand. Samuelson, for instance, in a particularly sharp rebuke, argued that "Adam Smith, in his famous passage, had no right to assert that an Invisible Hand channels individuals selfishly seeking their own interests into promoting the 'public interest'. . . Smith has proved nothing of this kind, nor has any economist since 1776" (Samuelson 1967: 610). One can see here that Samuelson is thinking within a Walrasian framework—it is only if one is particularly attuned to the compositional fallacy involved in moving from partial to general equilibrium that one would think that the various arguments advanced by Smith gave him "no right" at all to the claims he made about the public benefits of markets.

When one begins to think of the economy as a giant mathematics problem, with the solution being the instantiation of a set of variables, it is natural to start thinking about the central output of the market as being a set of numbers, namely, prices. Thus von Mises and Hayek argued that the central advantage of the market is not that it supplies individuals with incentives, but that it supplies individuals with the *correct* incentives, because it is able to produce the right information about the relative scarcity of goods. When the supply of a good is tight, and consumers begin to bid up the price, this sends a signal to producers, telling them not just that more of the good is required, but precisely how much more of it is required. This signal is extremely difficult to replicate bureaucratically, because individuals seldom have an incentive to reveal the truth about their own preferences. The market, however, forces them to "put their money where their mouth is." Since the only way to raise the price of something is for someone actually to pay more for it, market signals have a level of credibility that other economic signals may lack.

The general importance of prices and the difficulties associated with replicating them bureaucratically were made perfectly clear over the years by the experience of communist planners, particularly in the former Soviet Union. Despite the general stigma associated with prices among communists (prices were considered a symptom of the persistence of the "commodity form"), it quickly became apparent that working without them was completely impossible. Alec Nove described the conundrum faced by central planners: "How should one generate electricity? Should power stations be large or small? Is it prohibitively costly to invest in coal mines in north-east Siberia? What type of insulating material is cheaper? Is it worth investing in a new process for producing sulphuric acid? One cannot answer such questions without using prices of some kind, whether real or shadow, and the prices so used must reflect costs, which in turn reflect relative scarcity of means" (Nove 1983: 106). Thus, Soviet planners found themselves having to generate their own set of prices, without the use of the market. This turned out to be enormously difficult:

> There is one other important aspect of Soviet experience with regard to pricing. This is the sheer volume of work involved. Estimates of the number of prices, fully disaggregated, range from 10 to 12 million. It is essential to appreciate that actual prices must be fully disaggregated. There is no such thing as 'the price' of footwear or of agricultural machinery, only prices of specific kinds of shoes or ploughs from which an overall price index can be derived ex post. If millions of prices are to be fixed, whoever determines or approves them must collect information (on costs and on demand), information which needs cross-checking, in view of the possible interest of the information-providers in higher prices. This is a hugely labour-intensive, time-consuming task. A general review of prices is therefore undertaken at long intervals (e.g., 1955, 1967, 1981), which means that, even if they do bear some relation to cost in the first year of their operation, they soon cease to have such a relation as costs after. Furthermore, even if it were decided that prices should be flexibly adjusted to changes in supply-demand relationships, it would be quite impracticable to do this administratively, that is, through price control by the government and/or the planners. There would simply be far too many prices to control. (Nove 1983: 101–102)

These difficulties had already been clearly identified by Hayek, who pointed out two things: first, that the information required in order to "solve" the optimization problem was extremely local, and therefore not easily observable by a central authority, and second, that relying on agents with local knowledge to communicate this information up the chain of command would be difficult within an administrative system. Thus even if the price calculation problem was solvable in principle, it could never be solved in practice.

One can see the force of Hayek's contention even in the present day, particularly in cases where new markets have been created as part of privatization of deregulatory schemes. One particularly good example involves the creation of a market for emission permits for sulfur dioxide in the United States in the 1990s. As Tim Harford reports, even after exercising due skepticism about industry claims, the US Environmental Protection Agency estimated that it would cost between $250 and $700 per ton to reduce sulfur-dioxide emissions. But bids in the initial auction, held in 1993, generally fell far short of this. By 1996, the price of permits had dropped to a mere $70 per ton. As it turned out, not only was it a lot easier to reduce sulfur-dioxide emissions than industry had claimed, it was even much easier than skeptical regulators had suspected. As Harford puts it:

> The regulators discovered that getting rid of sulfur dioxide was so cheap that few people were willing to pay much for the right to keep producing it.... The clever thing about the auction was not that the sulfur emissions were reduced—that could have been required by law—but that legislators all over the world found out how much sulfur scrubbers *really* cost. It created a basis for further legislation: not making rules in the dark but in full knowledge of the (modest) cost. (Harford 2005: 99)

It is probably worth emphasizing that none of the corporate actors that constituted the market for these permits wanted to reveal this information to the government. On the contrary, when asked to provide it in advance, they systematically lied. But then they were forced to put their money where their mouths were and actually choose between reducing emissions and buying permits. Many of them—more than expected—chose to reduce emissions, and as a result the price of permits began to fall. Each intended only his own gain, but wound up thereby revealing extremely important information to the government, and to society at large. Each was "in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention."

8.5. Secondary Incentive Arguments

I have been discussing the history of arguments in support of the market, arguments that point to the existence of some mechanism at work that channels the self-interest of individuals in such a way as to produce cooperative benefits. I have tried to show how Adam Smith's emphasis on the incentive structure provided by the market has gradually been replaced by an appreciation for the information that it provides. The former argument was always controversial, in part because it tied the analysis of the invisible hand very closely to the fortunes of the *homo economicus* model of action. This led naturally to the accusation that economists overvalued the invisible hand, precisely because they overestimated the centrality of self-interest in individual motivation. This led, in turn, to the seductive conclusion that if one could only make people more moral—or better yet, create a less individualistic, possessive culture, in which our naturally prosocial impulses could find expression—one would have no more need for the market.

The information argument for the market completely changes the terms of this debate, by showing that, even if one could produce individuals who were always keen to do their part in any cooperative scheme, there would still be a problem figuring out what the right thing to do was, because it would be extremely difficult to know exactly what should be produced, how it should be produced, who should be consuming it, and so on. Furthermore, certain forms of cooperative behavior, far from ameliorating the situation, might actually worsen it. For example, if consumers decided to respond to a shortage of some good by sharing it, rather than by trying to outbid one another for it, then the price would not rise, the market would fail to generate the right signal, and no additional resources would be shifted into production of this good. This makes it difficult to see how any improvements in either socialization or social control could eliminate the need for the market, when it comes to determining the set of prices that serve as the basis for economic decision-making.

Thus the central case for the market shifted over time, away from the incentive argument toward the information argument. Yet the incentive arguments have not entirely disappeared. One of the characteristics of the general equilibrium models, which emphasize the importance of prices, is that they are static. Some of the most striking features of real-world markets, however, are the effects that they have over time. This is particularly obvious when it comes to economic growth and technical change. Both of these seem to be unintended yet beneficial consequences of market interactions. Yet we have nothing like an adequate model of either process, and so analysis of the mechanism that generates these outcomes remains primarily qualitative.

One obvious metaphor that is used to describe the dynamic properties of markets is that of an evolutionary process (Schumpeter 1943: 82). The core of an evolutionary system is a mechanism of selection. Proponents of the market have always put a lot of emphasis on the way that markets reward success. Over the course of the twentieth century, however, it became apparent that an equally important function of markets is the way that they penalize failure. And this is, of course, the core of any selection mechanism. Nature produces the illusion of teleology and directional change only because it is particularly ruthless at eliminating unsuccessful variants. Well over 99 percent of all species that have existed on this planet are now extinct. Markets may also succeed, not so much because they promote success, but because they are uncompromising in dealing with failure, so that the only things left over are successful.

There are two well-known phrases that have emerged from this school of thinking: the first is Joseph Schumpeter's concept of "creative destruction" and the second is János Kornai's concept of the "hard budget constraint." Both of these can be seen as invisible hand mechanisms, where it is precisely the fact that individuals act in their self-interest that generates the collective benefit.

Schumpeter's analysis of creative destruction is motivated by the observation that capitalist economies are particularly good at generating innovation, both in terms of production processes and product qualities. Indeed, even critics of the market, who are willing to countenance an "end to growth," may be given pause by the prospect of an end to technological innovation. Consider, for example, being given a choice between returning to a nineteenth-century level of material well-being, yet preserving all of the benefits of modern medical technology, versus retaining one's present level of material well-being, yet returning to a nineteenth-century level of medical technology. Many people would choose the former without hesitation, suggesting that technical innovation is a non-negligible component of the public benefits generated by the market economy. Yet it is clearly generated through an "invisible hand" mechanism, namely, the pressures of competition. Furthermore, it is obvious that the incentives are doing all the work in this domain. As Schumpeter argued, economists have put far too much emphasis on price competition:

> In capitalist reality as distinguished from its textbook picture, it is not that kind of competition which counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organization (the largest-scale unit of control for instance)—competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives.... It is hardly necessary to point out that competition of the kind we now have in mind acts not only when in being but also when it is merely an ever-present threat. It disciplines before it attacks. (Schumpeter 1943: 84–85)

One of the major weaknesses of bureaucracies is that, by contrast, not only do they have difficulty promoting innovation, but they have a tendency to actively resist it. (Consider, for instance, the battle between Minitel in France—a classic product of bureaucratic planning, centralized thinking, and government monopoly—and the Internet.) Indeed, this is true not only of government agencies, but even of large corporations. One need only look at the history of large high-technology firms over the past forty years to see the extraordinary effectiveness of markets at destroying firms that at one time seemed impregnable—and thereby preventing stagnation in the field. Thus part of the account of how markets promote innovation must involve their extraordinary effectiveness at destroying the institutions capable of resisting it. From IBM to Microsoft to Nokia, it is quite surprising how little ability large firms have to protect themselves against competition from new entrants or products.

The second major way in which markets penalize failure has to do with the role of the stock market and of outside lenders in the firm. One of the central problems with the public sector—both in the former communist bloc and in democratic welfare states—is that officials sometimes have difficulty controlling their own senior managers. Part of this has to do with what Kornai called the "soft budget constraint" (1992: 140–145). While managers are told to achieve some objective within a certain budget, governments often lack the means to enforce this, simply because they are unable to threaten any sort of credible punishment. Partly, this is because they are constrained to act in the public interest, and so cannot simply close down a department, or fire the staff, without considering what the impact will be upon the broader public. Private investors, on the other hand, have no such concerns. If a company loses money, or fails to meet its quarterly targets, investors will sell their shares. If it fails to make payments on its loans, lenders may foreclose, or force it into bankruptcy. The fact that they are concerned only about the return on their investment in this case has beneficial social consequences, insofar as it allows them to exercise more effective discipline over managers.[7] It is precisely because investors lack any sense of social responsibility that they are able to impose a hard budget constraint.

These ideas have received sharper formulation with the development of "principal–agent theory," which tries to model explicitly the difficulties of achieving proper incentive-alignment among agents. John Roemer has argued that the rise of this theory has generated another shift within the economics profession, away from the information argument and back to the incentive argument for the market (2010: 6–7). Indeed, some have argued that Hayek, in his contribution to the socialist calculation debate, conceded far too much to his opponents by accepting their assumption that managers were "loyal and capable" (Hayek 1940), and so would simply do what they were told to by the planner. In this respect, he took a large number of incentive problems off the table, by assuming that socialist ownership would be no more likely to generate agency problems than investor ownership. The experiences of the twentieth century, however, suggested that managers do not simply do what they are told, and in complex organizations with significant information symmetries, can in fact be quite difficult to control. Under these conditions, the structure of ownership can have a significant impact on the kinds of agency problems that an organization experiences.

Of course, the principal–agent literature tends to paint a very pessimistic picture, when it comes to estimating the chances of cooperative behavior in any sort of bureaucratic organization, public or private, under capitalism or socialism. Indeed, some of the early work done by agency theorists on the theory of the firm took the form of a critique of transaction cost theory, pointing to the fact that many of the mergers and acquisitions undertaken by large corporations are unsuccessful, and upon closer examination, turn out to have never had any reasonable prospect of lowering transaction costs (Eisenhardt 1989). The best explanation for these changes in firm structure, it was suggested, was simply self-aggrandizement on the part of senior managers, undertaken at the expense of shareholders—in other words, that they were the product of agency problems.

---

[7] Looking at the history of social-democratic countries in their dealings with "dying industries" shows that there is a real concern about the consequences of eliminating the sort of discipline imposed by investors. Of course, there are also concerns about the abruptness of the way that markets may deal with such transitions. But most people will agree about what the long-term outcome should be and have legitimate concerns about the ability of the state to achieve those outcomes. Therefore, the emphasis tends to be on strategies for easing the transition—social safety net, active labor-market policies, etc.—rather than on a suspension of the underlying mechanism.

Now it seems obvious that there is no room for the "invisible hand" to intervene and restore efficiency in these cases, since they involve problems that arise within the organizational structure of the firm. There is, however, an indirect route. This involves the role that the stock market plays in helping to discipline managers in a capitalist economy. The basic thought is that if managerial misbehavior becomes too egregious, it will depress the stock price, and therefore make the firm a more attractive target for takeover or buyout. For example, if a "vanity" merger produces an inefficient conglomerate, then this makes it a tempting target for corporate raiders, who can buy the company, fire the managers, break it up and sell off the pieces (a scenario that was immortalized by Bryan Burrough and John Helyar in their book *Barbarians at the Gate* [1990]). In this way, the *market for control* came to be seen as an important efficiency-promoting device, arising not from the market generally, but rather the stock market specifically.

Once again though, Hayek's essential point about incentives stands. Incentives are not just useless but can be positively counterproductive when they are not properly targeted. In order to target them correctly—to reward the behavior that one wants to encourage and to punish the behavior that one wants to discourage—one must have good information. This applies just as much to the stock market as it does to ordinary commodity markets. Thus it was not long before people began to argue that the importance of the stock market was not that it disciplined management, but that it provided a publicly available source of information about the firm and its prospects, in the form of the stock price. The so-called "efficient markets hypothesis," which in Paul Samuelson's formulation was a general claim about prices, acquired more specific interest as a thesis about financial markets. Eugene Fama's claim that "security prices at any time 'fully reflect' all available information" (1970: 383) was taken as the canonical formulation. And so once again there was a shift away from emphasis on the market as a source of incentives ("the hard budget constraint") toward the market as a "revelation mechanism," or a source of information.

In this case, however, the triumph was less complete than it was in the case of ordinary commodity markets. The efficient markets hypothesis immediately attracted a number of very forceful criticisms (see Malkiel 2003), which came to a head after the 2008 financial crisis. In particular, it became apparent that much of the (rather surprising) inattention shown to the quality of the underlying assets in various derivative markets was in fact due to widespread belief that all the relevant information would already be reflected in the price (and so old-fashioned forms of due diligence were not required). As a result, the efficient markets hypothesis may well have proven to be self-undermining, insofar as it contributed to the rather spectacular mispricing of credit risk that prompted the crisis. This has a number of different implications, but at a certain level of generality what it suggests is that "invisible hand" arguments in favor of market efficiency are probably best focused on what is sometimes called the "real economy," rather than the stock market.

8.6. Conclusion

By way of conclusion, I would just like to reiterate my earlier assertion that the market economy is a very complex set of institutions. Not only that, but society is constantly in a process of shifting the boundaries of markets, through nationalization and privatization, as well as changing the rules through regulation. Thus it is very difficult to say what "the" major benefit of markets is, or how "the" invisible hand works. The central idea is clearly to have economic transactions organized on a decentralized, competitive basis. But how we structure this competition, which competitive strategies we promote, what sort of actors we allow to enter the competition and what constraints we put on them, are all highly variable over time. Furthermore, the way that we adjust these structures is strongly influenced by our sense of what the major benefits of the competition are, and how they arise as a byproduct of this competition. Both the information and the incentive arguments point to important aspects of this and in a sense the two are obviously intertwined. Without the right incentives, it is difficult to get correct information, while the value of information is often that it allows us to target incentives in the right way. If anything is clear, it is that we cannot just pick out one strand, the case for markets necessarily involves some complex appeal to the two.

9 }

The Benefits of Cooperation

There is an idea, extremely common among social contract theorists, that the primary function of social institutions is to secure some form of cooperative benefit. If individuals simply seek to satisfy their own preferences in a narrowly instrumental fashion, they will find themselves embroiled in collective action problems: interactions with an outcome that is worse for everyone involved than some other possible outcome. Thus they have reason to accept some form of constraint over their conduct, in order to achieve this superior, but out-of-equilibrium outcome. A social institution can be defined as a set of norms that codify these constraints.[1] Simplifying somewhat, one can then say that social institutions exist in order to secure gains in Pareto efficiency.

This theory is one that I take to be in large measure correct.[2] My concern, however, is that it tends to be formulated at too high a level of abstraction. By focusing on the structure of the interaction—a structure that is often specified simply in terms of the utility functions of participants—the theory tends to abstract away completely the mechanism through which social benefits are produced. Thus major philosophical writers working in the social contract tradition, such as David Gauthier and John Rawls, make no attempt at all to specify how cooperation improves the human condition. Rawls, for example, states simply that "social cooperation makes possible a better life for all than any would have if each were to live solely by his own efforts," without saying how (Rawls 1999: 4).[3] Gauthier focuses entirely upon the role of institutional

[1] One can see this idea most clearly in the work of David Gauthier (1986) or Andrew Schotter (1980).

[2] The dominant opposition to this view stems from those who consider "conflict" to be a more important determinant of the structure of social institutions than the "consensus" that Pareto efficiency supposedly creates. As we shall see, the version of the "consensus" approach used here predicts much higher levels of social conflict than the more traditional versions, and thus goes a long way toward undermining the force of the "conflict" objection.

[3] When he does go on to discuss more concrete institutional proposals, his analysis suffers from the catallactic bias diagnosed below (especially his proposal for an "exchange branch" of the state to handle efficiency issues, 249–250).

constraints in resolving "prisoner's dilemmas," but with no systematic analysis of what people are typically trying to accomplish when they get into these dilemmas.

Social scientists interested in the subject of cooperation have not achieved much greater clarity. One often hears talk of "social capital," for example, as a generic resource, a fund of trust and solidarity that individuals can draw upon in order to overcome collective action problems without having to institute a formal system of sanctions (Putnam 2000).[4] This is often accompanied by some specific empirical examples of cooperative arrangements that rely upon this sort of trust, but very seldom is there any discussion of the type of cooperative projects that social capital gets used for, or what sort of benefits cooperation can produce. When a more abstract mechanism does get mentioned, the discussion tends to focus upon the best-known instance of such, which is the *gain from trade* achieved through market exchange (or the division of labor). Thus social capital is often characterized as a resource that is used to reduce transaction costs, in order to facilitate exchange and reduce deadweight losses (Putnam 2000; Fukuyama 1995). Although this is no doubt true, it represents only one of the ways that social capital can be used.

In this article, I would like to correct this deficit, by specifying five different mechanisms of cooperative benefit. I refer to them as mechanisms, and not simply as "social goods," because they represent different ways in which individuals can help each other to achieve each others' objectives, whatever those objectives may be.[5] Although each of these mechanisms is, on its own, well known, there has been no systematic attempt to classify them, or to draw out the implications that such a classification has for social contract theory, or for political philosophy more generally.

What are some of these implications? The first thing one notices when laying out such a typology is that much of contemporary social contract theory has been marked by what might be referred to as a *catallactic bias*, which results from a tacit conceptual privileging of gains from trade as the primary mechanism of cooperative benefit.[6] One can find the primary source of this

---

[4] James S. Coleman, who introduced the concept, writes, "Social capital is defined by its function. It is not a single entity but a variety of different entities, with two elements in common: they all consist of some aspect of social structures, and they facilitate certain actions of actors— whether persons or corporate actors—within the structure. Like other forms of capital, social capital is productive, making possible the achievements of certain ends that in its absence would not be possible" (Coleman 1988: 98).

[5] For an example of an analysis based upon a list of goods, rather than mechanisms, see Knight 1992. When he does specify mechanisms (25), he presents only one: gains from trade, along with two abstract categories—gains from cooperation and gains from coordination.

[6] The term "catallaxy" was introduced by Friedrich A. Hayek as a way of referring to a system of order established through exchange. See Hayek 1976. The highest expression of this catallactic perspective can be found in the work of James M. Buchanan, who essentially redefines efficiency in such a way that only gains from trade count as efficiency-promoting. See Buchanan 1986.

catallactic bias in the influence exercised by the so-called invisible hand theorem of welfare economics, which shows that the equilibrium of a perfectly competitive market will be Pareto-optimal (Blaug 1985: 585–597).[7] It is well understood that these results obtain only under a set of highly idealized conditions, which are never satisfied in the real world.[8] What is more seldom recognized is that this theorem focuses only upon one mechanism of cooperative benefit to the exclusion of all others.[9] Thus the reason that real-world markets fail to achieve Pareto-optimal outcomes is not just because of imperfections in the system of competitive exchange (such as transaction costs, or externalities), but because there are entire other categories of efficiency gain that these institutions do not even presume to promote. Indeed, I will attempt to show that not only does market exchange fail to promote certain types of cooperative benefit but also that the relevant set of institutions often interferes with our ability to produce such benefits. Thus we often face hard choices when it comes to determining the structure of our social institutions, along with significant social conflict over the type of cooperative benefits that should be accorded priority.

The major impact of catallactic thinking in political philosophy has been to promote a highly misleading account of the role that the welfare state plays in a capitalist economy. More specifically, it has encouraged the widespread perception that, when it comes to efficiency, the central functions of the welfare state are all residual. According to this view, the central purpose of the market is to maximize the number of efficiency-promoting exchanges concluded. The central economic role of the state, when it comes to efficiency, is to assist in this function: ensuring that the system of property rights is respected; preventing collusion and other anticompetitive practices; regulating the price level in cases of natural monopoly; internalizing externalities, either through taxation or regulation, to compensate for incompleteness in the system of property rights; and finally, forcing certain exchanges to occur, through taxation and public goods provision, in cases where it proves impossible to organize a private market. Above and beyond this range

---

[7] Also known as the "first theorem" or "first fundamental theorem" of welfare economics, e.g., see Myles 1995: 37–40.

[8] The extent to which these limitations prevent one from drawing normative conclusions is often not fully appreciated, however. In particular, there is a widespread failure to recognize the significance of the "second-best theorem." See Lipsey and Lancaster 1956. This is, however, not my focus in this article.

[9] Of course, the exclusion of economies of scale is explicit. However, the significance of the assumptions made about information, uncertainty, and preferences is less obvious. Sophisticated economists have long understood these limitations (e.g., Arrow 1974: 7), but these reservations have not fully percolated through the broader intellectual community. Consider, for example, Gauthier (1982). He acknowledges that the presence of externalities limits the "invisible hand" reasoning that he seeks to deploy, but he ignores entirely the impact of asymmetric information. For an accessible survey of these and other issues, see Stiglitz 1994.

of efficiency-promoting functions, all other activities are then classified as "redistributive" in character.[10]

The result, I will argue, has been a serious misunderstanding of the contemporary welfare state, one that dramatically overstates its redistributive character. This has, in turn, led philosophers to dramatically overestimate the extent to which its programs require an egalitarian justification. Elements of the "social safety net," for instance, such as state pensions, unemployment insurance, and socialized medicine, are routinely described as redistributive (and thus, tacitly, as egalitarian). This is, I will argue, a misclassification. The social safety net is first and foremost a set of risk-pooling arrangements, which are organized in the public sector primarily in the interest of promoting efficiency gains. Furthermore, the mechanism of cooperative benefit that is exploited in this case is fundamentally different from the gains from trade that are achieved in market exchange. Thus there is no reason to think of the state's role in this domain as residual. On the contrary, the state often takes a lead role when it comes to promoting certain forms of cooperative benefit. Distinguishing the various mechanisms of cooperative benefit therefore provides not only a more satisfactory articulation of the different forms of cooperation that are enabled by the welfare state but also a more sophisticated approach to defending its normative foundations.

9.1. Forms of Cooperation

I would like to start out where social contract theorists typically do start out, with the state of nature. It is useful when trying to conceptualize the benefits of cooperation to imagine a state of society in which there is a complete absence of cooperation, and then to consider what would be missing in such a scenario. So imagine two Robinson Crusoes, living on closely adjoining islands. They have the choice of either ignoring one another or interacting with one another. Why might they choose to interact?

In answering this question, we should begin by distinguishing between benefits that can be achieved through purely instrumental action and those that require some form of cooperation and constraint. One Robinson may derive enjoyment simply from watching the other go about his business, but this is of no particular interest to the contractualist, because it can be achieved quite easily through straightforward instrumental action (in the standard run of cases). It requires no constraint, and thus no agreement.

[10] Perhaps the most important example of this in contemporary political philosophy is Rawls, *A Theory of Justice* (1999: 242–251). For a more explicit version, see Barr 1998. Adam Przeworski (2003: 44), describes this as the "markets whenever possible, the state whenever necessary" view.

Within the set of benefits that do require cooperation to produce, we can also distinguish between two very important types: those produced when individuals agree to engage in actions that have positive externalities, and those produced when individuals agree to refrain from actions that have negative externalities.[11] For example, purely instrumental action might lead the two Robinsons to raid each other's food stocks, which would in turn force each to take costly defensive action. This is a classic prisoner's dilemma. Thus they might both benefit from an arrangement under which they refrain from raiding each other's food stocks. Since the benefit here comes from the elimination of a negative externality that arises from social interaction, however, it is not really an inducement to cooperation. After all, the two Robinsons could achieve the same result by finding a couple of even more isolated islands and avoiding each other completely. Thus the agreement does not capture the "upside" of social interaction; it simply allows them to eliminate one of the many "downsides." The focus here will be quite specifically on ways in which individuals are able to generate positive externalities, and where they depend upon some structure of reciprocity in order to motivate them to do so.

The other major constraint in the analysis that follows is that I will be ignoring all of the benefits that arise from what we might typically think of as primary socialization. Individuals obviously derive significant benefits from language acquisition, the development of norm-conformative competencies, and various other aspects of sociocognitive development that occur during childhood, via social interaction. There is some question as to how much "cooperation" is required in order to produce these benefits to the individual, given the extent to which parental investment in the young, for reasons of inclusive fitness, is independent of any system of reciprocity (Hamilton 1963: 354–356; Smith 1964: 1145–1147).[12] Without getting into these debates, however, simplicity alone may serve as an adequate motive for confining the analysis to the type of cooperative benefits that fully developed adults are able to provide for one another.

With this in mind, five primary mechanisms are immediately apparent: (1) economies of scale; (2) gains from trade; (3) risk pooling; (4) self-binding; and (5) information transmission.

> *A. Economies of scale:* The most obvious form of cooperation arises because of the simple fact that not all jobs can be done by one person. Nature sometimes organizes things in such a way that people can do

[11] This broader way of speaking of externalities, not as an effect on a third party, but merely as an effect upon another, is one that I find very helpful. It is used in this way by Gauthier (1982: 96) and passim.

[12] To see the significance of the distinction between kin (or inclusive fitness) altruism and reciprocal altruism, see Hammerstein 2003 and also Boyd and Richerson 2005b.

better working together than they can working individually. One Robinson may have a large rock, blocking his view of the sea, which he would like to have moved. Unfortunately, it is too heavy for him to move on his own. With the assistance of the other Robinson, he would be able to move it. If the other Robinson has a similar job over on his island, then the two are in a position to engage in mutually beneficial cooperation.

This example is actually just a "lumpy" version of what are commonly known as economies of scale. Some jobs are such that adding another person generates a disproportionate increase in output. If one individual is able to produce an output of x per unit of labor, an economy of scale is present when adding a comparable unit of labor from another individual increases output by more than x. This increase can be either continuous (as with a harvest, where each new worker's contribution speeds up the process, thus reducing the chances that the crop will be rained on), or it can be lumpy (as with a barn-raising, where certain parts of the structure can only be assembled with a minimum number of hands).

This mechanism is so familiar that we often refer to it simply as "cooperation." However, despite the obviousness of the benefits, self-interest alone is often not enough to secure cooperation of this type. If the cooperative system requires reciprocity over time, each individual may have an incentive to defect after the work that benefits him has been performed (and anticipating this, others may refuse cooperation from the outset). If the benefits are produced and divided up in a single shot, then each individual has an incentive to "shirk" (i.e., contribute as little effort to the collective project as possible). The extent to which benefits can be secured under such conditions without constraint will then be determined by the observability of effort (Holmström 1979: 74–91).

B. Gains from trade: This second mechanism is slightly more subtle than the first and has therefore attracted greater theoretical interest. Plato's account of the ideal city in the Republic starts from the observation that people are naturally suited for social life, because they have different needs and abilities. In cases where these differences are complementary, they can serve as a source of cooperative benefit. In particular, individuals may be able to achieve benefits by rearranging the distribution of goods, or of tasks, among themselves. This is what motivates various forms of exchange. These gains are often analyzed under two headings:

Consumption: Gains from trade can be achieved because individuals have different needs and tastes. One person may not like carrots and another may not like potatoes. If they trade vegetables, they are able to improve both of their consumption bundles, each from his or her own perspective. It is important to realize that if everyone had identical tastes and needs, no such gains could be made. There are, of course, a variety of reasons why people have different needs, but one of the most important is that people are of different ages, and that their preferences change as they age. Gender is also an important source of complementarities.

Production: Gains from trade can also be realized because individuals have different abilities. Even if everyone had identical tastes, and therefore sought identical consumption bundles, it would still be advantageous to have a division of labor. Some people are better at certain types of work, and so it makes sense to have them carry out these tasks "for everyone," while instituting a structure of reciprocity such that their other needs will then be taken care of by others.

The important point about this class of gains is that they can be achieved without actually increasing the stock of goods, in the former case, or the stock of productive resources, in the latter. When individuals cooperate in order to achieve economies of scale, the results of this cooperation are quite tangible. When they change the allocation of goods, however, in order to better satisfy the preferences of all, or the allocation of resources, in order to better take advantage of productive abilities, the gains are less obvious. It is because of the more subtle character of these gains that they were felt to be very much a discovery among early economists. Edward G. Wakefield, for example, in his heavily annotated 1835 edition of Adam Smith's *The Wealth of Nations*, drew particular attention to the difference between "simple cooperation" ("when several persons help each other in the same employment") and "complex cooperation" ("when several persons help each other in different employments"). "Of the former," he wrote, "one is always conscious at the time of practicing it: it is obvious to the most ignorant and vulgar eye. Of the latter, but a very few of the vast number who practice it are in any degree conscious" (Smith 1835: 26).[13]

Cooperation is essential to achieving gains from trade, simply because all of the individual advantages come from the reciprocation of the other. Of course, there is a long-standing project in political philosophy that seeks to show that these sorts of gains would also arise "spontaneously" in a state of nature, or that market exchanges constitute, in David Gauthier's term, "a morality free zone."[14] These arguments studiously ignore the free-rider strategies that arise in contexts of economic exchange (notably theft and fraud) (Schultz 2001: 67–70).

C. Risk pooling: Thanks to the work of the classical economists, both of the above sources of collective benefit are well known. Both mechanisms

[13] John Stuart Mill saw fit to reproduce lengthy segments of this discussion in his *Principles of Political Economy* ([1848] 2004: 134).

[14] With regard to the former, see Hayek 1979 and on the latter, see Gauthier 1986: 84.

deal with the reproduction of what Marx called the "material basis" of society. It is an unfortunate feature of the human condition, however, that we are afflicted not only by scarcity but also by uncertainty. Our ability to plan for the future is severely compromised by our inability to determine precisely what the future has in store. One of the primary benefits of collective action is that it enables individuals to reduce the subjective dimension of this uncertainty. This is due to a phenomenon referred to loosely as "the law of large numbers."

The world is full of risk. Knowing the probability of various events is extremely useful when it comes to engaging in practical deliberation. Unfortunately, what matters to most of us when we make our plans is not just the background probability of an event, or the long-run average that we can expect, but also the degree to which actual outcomes can be expected to diverge from the average. When confronted with stochastic variability, people worry not just about the mean but also the variance. In this context, the "law of large numbers" becomes important, because it offers us a way of reducing this variance. For example, we know that a fair coin has a 50 percent probability of landing heads, but we also know that flipping it ten times is quite unlikely to produce exactly five heads and five tails. However, it is also well known that as the number of tosses increases, the frequency will tend to converge with the probability. In other words, increasing the number of trials induces *statistical stability* (Hacking 2002: 190–192). This is the basic connection between "large numbers" and the reduction of variance (or standard deviation).

To see how groups of individuals can benefit from this, it is important to remember that individuals are often risk-averse. Consider a subsistence farmer who under normal conditions is able to produce ten tons of grain, which is enough to feed his entire family well throughout the winter. However, his land is also subject to a highly localized blight, which sometimes wipes out the entire crop. Suppose that the chances of this blight striking his field in a given year are 20 percent. Although the expected annual output of his field is therefore eight tons, he would gladly swap a guaranteed revenue of eight tons for the gamble that he faces between ten tons or nothing. That way, his family would have a bit less to eat, but they would never risk starvation.

On his own, this is something that he cannot achieve. Suppose, however, that there are one hundred small farmers who find themselves in identical circumstances, all facing the danger of this highly localized blight. They might agree to a "risk-pooling" arrangement, under which farmers who lose their crop in a given year are compensated by those who do not. Under this arrangement, the objective risk of blight does not diminish: twenty of the hundred farmers can, on average, expect to lose their crops. However, with the risk-pooling arrangement, each farmer can expect a revenue that will be, with 95 percent probability, between 7.2 and 8.8 tons.[15] Adding in more farmers narrows that band, until it comes very close to eight tons precisely.

Thus insurance is a mechanism that generates a convergence between the subjective utility associated with a gamble and its mathematical value. What is important to recognize is that risk pooling is, like gains from trade and economies of scale, a sui generis source of collective benefit.[16] It shares with (allocative) gains from trade the characteristic that its benefits are invisible: it simply increases everyone's utility. However, unlike these sorts of gains from trade, the benefits of risk pooling are available even to individuals with identical preferences. Many theorists have encouraged confusion on this point, by conflating the benefits that come from pooling risks with those that come from trading risks.[17] Insurance, for example, is often analyzed as an exchange between a risk-averse and a risk-neutral individual (and thus the welfare benefits are classified as merely a gain from trade). A typical "mutual society" arrangement, however, involves risk pooling among individuals who may all be equally risk-averse.[18] The benefits come not from the fact that risks are transferred but from a reduction in the variance achieved through the "law of large numbers" effect, and the utility gain this provides to risk-averse individuals.

Thus risk pooling is a more subtle mechanism of cooperative benefit than either gains from trade or economies of scale. Furthermore, because an understanding of the key probability concepts was a relatively late development, and has largely been confined to specialists, it was only in the mid-nineteenth century that a more general appreciation of the social significance of risk pooling began to develop. Thus it was not until 1855, with the publication of Émile de Girardin's *La politique universelle*, that the first serious attempt was made to integrate an understanding of insurance into social contract theory (de Girardin argued that the primary function of the social contract was to found a system of "universal insurance") (Girardin 1855).

D. Self-binding: Individuals are subject to dynamic preference inconsistency. Everyone discounts future satisfaction to some degree. However,

---

[15] The standard deviation is 0.4 tons. See Moss 2002: 28–31.

[16] The law of large numbers is also at the root of the "wisdom of crowds" phenomenon: the fact that two heads are generally better than one, even in the absence of complementarities. This connection is stated most explicitly in the Condorcet Jury Theorem, but see also Surowiecki 2005.

[17] For an example of a discussion that unhelpfully blurs the distinction between the two mechanisms, see Barr 1998: 111–112, or Easterbrook and Fischel 1991: 53.

[18] Ian Hacking (2003) refers to the former as the "Lloyd's of London" model of insurance. The mutual society model, on the other hand, which involves symmetry in the relation between all policyholders, is the arrangement that predominates in the standard categories of health, life, home, and automobile insurance.

theorists who use something like interest rates as their model of how much compensation individuals require in order to defer satisfaction are likely to be misled. This is because the sort of smooth exponential function suggested by the analogy to interest rates produces preference orderings that are temporally invariant: if *a* is preferred to *b* at time *t*, then it will also be preferred at any other time. In reality, individuals seem to discount satisfaction quite sharply in the very near term, with the curve then smoothing out as it heads off into the future (Ainslie 1992; Rachlin 2000). In other words, the difference between today and tomorrow often seems like a very long time compared to the difference between November 7, 2017, and November 8, 2017, when seen from the present.

This feature of our discount rates has the capacity to generate temporary preference inversions. George Ainslie has provided a number of striking illustrations of this phenomenon (Ainslie 2001: 33). For example, given a choice between a check for $100 that can be cashed right away and a check for $200 that can be cashed in three years, many people will choose the former. Many of these same people, however, when given a choice between a $100 check that can be cashed in six years and a $200 check that can be cashed in nine years will take the $200 check. This preference structure will generate dynamic inconsistency: the individual who would take the $100 check right away and chooses the $200 check that can only be cashed in nine years will want to change the latter decision in six years' time. Ainslie refers to this as "hyperbolic" discounting, since the rate of discount is so highly exaggerated in the short term.

For a long time, this sort of inconsistency was castigated as a form of irrationality (i.e., as akrasia, or weakness of the will).[19] There is no reason to go so far. The fact that a set of preferences interact with one another in such a way as to generate instability over time is not necessarily irrational. What matters for our purposes is that the forms of instability we see are eminently predictable. Something that is judged to be the "lesser good" from afar may start to look more attractive in the very near term. This is usually a temporary phenomenon; our evaluation reverts back to the original ranking after the event has passed. Because we are all familiar with this pattern, we often take preventative steps in order to ensure that we will not act upon such temporary preference reversals. Thus we have a range of self-binding mechanisms that we use to constrain our own future choices.

[19] Jon Elster provides an extreme formulation of this thesis. Following Henry Sidgwick, he regards zero discounting as the only "rational" preference, and thus treats all time preference as a symptom of irrationality (Elster 1979: 66).

In some cases, we are able to carry out these self-binding strategies alone. One of the great advantages that we achieve from social interaction, however, is the ability to enlist others in our aid. In fact, it has been widely noted that compulsive and addictive behavior is strongly associated with social isolation (Rachlin 2000: 89; Elster 2000: 76–77). In the company of other people, we may entrust goods to our neighbors for safekeeping, authorize family members to make decisions on our behalf, or instruct our friends to ignore our demands. In some cases, what we anticipate is the impairment of our own decision-making abilities (e.g., the individual who gives away his car keys before he starts drinking). In other cases, it is the tendency toward heightened impatience in the near term that we seek to control (e.g., the person who opts for an increase in employer contributions to her retirement plan, in lieu of increased salary). We often rely upon the cooperation of others to carry out these self-binding strategies (for the simple reason that, with other people, we can give instructions in advance that specify what we want done, which they can then follow). Jon Elster refers to this as the "giving away the key" approach to self-control (Elster 2000: 66).

Our ability to enlist others to help us maintain self-control has sometimes been cited as one of the most important advantages of cooperation. David Hume, for example, was quite clear that the problem of heightened impatience in the near term offered the primary explanation for "the origin of civil government and allegiance" (Hume [1739] 1978: 537). By authorizing others ("civil magistrates, kings and their ministers, our governors and rulers") to punish me, should I fail to choose the greater good, I am able to guard against my own preference reversals. Of course, at the time that I am forced to refrain from choosing the lesser good, this will seem like a hardship. I am willing to commit to it in advance only because I also discount this hardship to the same degree that I do the lesser good. Thus "the provision we make against our negligence about remote objects, proceeds merely from our natural inclination to that negligence" (Hume [1739] 1978: 536).[20]

Contractualists have often shied away from assigning too great a role to this mechanism because of the traditional association between

[20] Incidentally, it is sometimes thought that Hume's argument relies merely upon our tendency to privilege our "short-term" over our "long-term" interests. Yet this alone is not enough to explain the phenomenon of self-binding (and thus the reasons we have for enlisting the assistance of others). A person who merely privileges his short-term over his long-term interests has no reason to take preemptive action to constrain his own future choices, since the lesser, nearer good will always appear superior to him. What Hume's argument relies upon is the preference reversal that occurs as the lesser good becomes nearer. The anticipation of such reversals is what motivates precommitment, and such reversals occur not merely through impatience, but rather through heightened impatience in the near term. It is only this specific sort of infirmity that is capable of becoming "a remedy to itself."

dynamic preference inconsistency and irrationality. Hume himself was part of the problem, insofar as he characterized the conflict as one between "passion" and "judgment."[21] Viewed in terms of discounting, however, there is no reason to regard self-binding strategies as irrational. Insofar as other individuals help us to carry out these strategies, it represents another sui generis mechanism of cooperative benefit.

E. Information transmission: Another major advantage of social interaction is that it allows individuals to economize on learning costs. Sometimes this does not require cooperation. One person may simply observe another's behavior and imitate it (Tomasello 1999). However, there are a variety of ways in which we can encode and transmit information to one another, and this type of interaction normally requires cooperation. This is actually not specific to humans. Some animals (such as vervet monkeys or ground squirrels) emit alarm calls when they detect a predator, which communicates important information to their fellows. This behavior is altruistic, since the individual who emits the alarm call generally attracts attention to himself and is therefore more likely to be killed than the others. Thus communication systems of this sort are vulnerable to free-rider problems (Dugatkin 1999: 17–20).

Humans have, of course, vastly more sophisticated means at their disposal, yet the same free-rider problem persists. Most of what we know about the world is not a product of direct experience, or trial-and-error learning: it is acquired through linguistic communication with others. It is by tapping into this system of communication that we gain access to the vast cultural inheritance of humanity. It gives us knowledge of events far removed from us, both in time and in space, which in turn allows us to plan our own activities with much greater success. Clearly we all benefit from an arrangement in which individuals volunteer such information when it seems relevant and refrain from crude falsification of their reports when it suits their interests to do so, but the preservation of such a system will require cooperation.

Although information often seems to spread in a quasi-natural fashion—consider the way a rumor moves through a crowd—it is important not to lose sight of the fact that communication is underpinned by a generalized system of trust (Heath 2001a: 86–107; Brandom 1994: 3–66).[22] The norm of truth-telling is the most apparent example. As the story of "the boy who cried wolf" reminds us, individuals whose behavior is

[21] Hume writes, "Tho' we may be fully convinc'd, that the latter object excels the former, we are not able to regulate our actions by this judgment; but yield to the solicitations of our passions, which always plead in favour of whatever is near and contiguous" (Hume 1978 [1739]: 535).

[22] The role of cooperation in transmitting information is typically underemphasized in the so-called memetics literature, with the central role assigned to imitation. See Dennett 1995: 342–369.

> determined by their own interest, rather than by the norms governing the relevant language game, undermine the integrity of the system. A certain amount of social learning is possible simply through observation and imitation, but the more powerful mechanisms of cultural transmission that we have available all involve cooperation. This mechanism of cooperative benefit is one that contractualist theorists such as Jürgen Habermas have placed considerable emphasis upon, particularly in his account of the role of language in cultural evolution (Habermas 1985: 77–92).[23]

Although I am not aware of any omissions, this list need not be exhaustive. The primary goal is to show that human beings derive a wide range of different benefits from cooperation with their fellows. Each of the mechanisms outlined above has, at one time or another, been held up as the key to understanding the motivations underlying the social order. Hume, for instance, thought that the self-binding function of social institutions explained the "basis of civil association." De Girardin thought the purpose of the social contract was to create "universal insurance." John Locke thought the goal was to protect property, and thus to promote and secure the gains from trade. Of course, however, there is no reason why any one of these mechanisms of cooperative benefit must predominate. Instead, "civil association" may be conceived of as a set of social institutions whose role is to secure as many of the benefits along each of these different dimensions as possible.

Critics may see all sorts of troubling omissions from this list. Some of these are intentional. For example, I make no mention whatsoever of reproduction, which would seem to be a social activity par excellence. The reason for leaving it off the list is that, strictly speaking, reproduction is not something that requires cooperation, and is not necessarily an activity that generates mutual benefits. The systems of cooperation that exist in the domain of reproduction tend to be secured by institutions such as the family. It is my conviction that these institutions are best analyzed, not in terms of their role in facilitating reproduction (for which they are not strictly necessary), but rather in terms of how they secure cooperative benefits using one of the mechanisms outlined above. We learn more about the family, I would argue, by analyzing the sexual division of labor and the gains from trade it facilitates than we do by focusing on reproduction. In other words, there is no sui generis form of benefit associated with the institution of the family life: it is simply one way of securing benefits that arise from other sources.

Similarly, it may seem that security is one of the major reasons that humans have always banded together into groups (security being understood as protection against danger, rather than pooling of risk). However, it seems to me

---

[23] Habermas himself is loath to accept the "contractualist" label, for reasons that are somewhat beside the point here. Reasons for including him in this camp can be found in Freeman 1990: 122–157.

that there is no sui generis mechanism of cooperative benefit here either. Individuals are able to provide some security for themselves all alone, but in a group they are able to take advantage of economies of scale, complementarities of ability, and information transmission. Thus security is best analyzed as one of the social goods produced through cooperation, not as a separate mechanism of cooperative benefit.

Finally, I do not mention anything pertaining to social status, the "social bases of self-respect," self-esteem, or mutual recognition. Although it might be argued that healthy self-esteem is one of the positive benefits of social interaction, I am inclined to analyze it in terms of its opposite, and regard low self-esteem as a potential negative consequence of social interaction. Social interaction, according to this view, elicits dominance behavior, which in turn generates status hierarchy. Self-esteem is generated by the perception of one's own position in this status hierarchy.[24] Refraining from dominance behavior is therefore a form of cooperation, insofar as it eliminates one of the many "downsides" of social interaction. Self-esteem, from this perspective, is not a social good, but merely the absence of a bad. Since the analysis in this article is confined to the "upside" of social interaction, no mention is therefore made of these issues.

9.2. Institutional Solutions

The analytic perspective that is being proposed here takes as its point of departure the suggestion that the primary function of social institutions is to secure cooperation. This in turn suggests that various institutions can be usefully analyzed by considering the role that the different mechanisms outlined above play in securing these cooperative benefits. One way of gaining purchase upon this question is to consider the type of collective action problems that these social institutions resolve, in terms of the free-rider strategies that the institutional constraints prohibit. The social contract tradition has focused considerable attention upon the practice of promising, along with its legal correlates in contract law. This is, of course, a perfectly generic mechanism that can be used to overcome a wide range of collective action problems. However, we also have a range of more specific institutional arrangements, designed to address the free-rider strategies that arise with specific mechanisms of cooperative benefit. For example, when it comes to economies of scale, cooperation is particularly vulnerable to shirking. If effort is unobservable, then individuals have an incentive to invest less than their full effort in the task (Holmström 1982: 324–349). One of the basic institutional building blocks that we use to

[24] For an analysis of the "politics of recognition" in terms of status, see Fraser 2003.

overcome these problems is the formation of teams or groups (March and Simon 1958: 59–70). Individuals are sorted into relatively small groups, then encouraged to build strong trust relations through exercises in reciprocity. Group bonding or "identification" may give individuals a strong incentive to cooperate, in order not to let down "the side."

These "internal" incentives are integrated in various complex ways with the superior external incentive schemes that team organization permits (Galbraith 1971: 138–139). Effort is more directly observable within small groups, and among people who know each other well. Nevertheless, a team might still opt to impose supervision upon itself. One need only consider the role that a coach plays on a typical sports team to see these structures in effect. The military also provides a clear-cut example. Similar team structures are reproduced in economic contexts, through the formation of corporations, or work groups within larger corporations. Thus the catallactic theory of the firm, which regards the corporation as simply a "privately owned market," obscures the primary rationale for this institution, by trying to shoehorn it into a model of efficiency gains based purely on exchange (Alcian and Demsetz 1972: 795). Even within a transaction cost framework, the primary function of the corporation can be understood much more naturally as an attempt to create an environment that is insulated from certain patterns of market behavior, in order to foster reciprocity, trust, and therefore more effective teamwork (Heath 2001b: 151–159).

In the case of complementarities of needs and abilities, cooperation generally takes the form of exchange, which then generates a series of free-rider strategies that exploit this practice. The fundamental problems are theft and fraud. Each exchange is mediated by a promise to pay, which the parties then usually have an incentive to violate.[25] The institution of exclusive property rights represents one of the most sweeping mechanisms for the elimination of such strategies. Property rights can be thought of as an all-purpose mechanism for "internalizing" externalities. They allow individuals to "capture" some part of what would otherwise be positive externalities generated by their effort, while also allowing them to "deflect" negative externalities generated by the activities of others (Nozick 1974: 280). This facilitates the development of exchange, by giving individuals the confidence that the complementarity needed to sustain specialization will be actualized.

In the case of risk pooling, the two most common free-rider problems are moral hazard and adverse selection (Rasmusen 1989: 165–169). Moral hazard arises because individuals who are indemnified against a given risk have

[25] The availability of free-rider strategies in this context is consistently downplayed, both by Humeans who emphasize the "conventional" nature of property arrangements, and by proponents of "spontaneous order," who for ideological reasons want the market to be conceptually prior to any formal systems of constraint, such as the law. See Schultz 2001: 60–70.

a reduced incentive to take precautions aimed at avoiding the loss. In cases where the risk in question is influenced by the agent's behavior, this means that the decrease in subjective uncertainty produced through risk pooling will be accompanied by an increase in the objective probability of the unfortunate event. The other free-rider strategy, adverse selection, occurs when some individuals face a higher probability or magnitude of loss, and yet are able to conceal this fact. By entering into risk-pooling arrangements, they are effectively able to transfer costs associated with this elevated risk level onto others. Both of these free-rider strategies can create very straightforward collective action problems (Rasmussen 1989: 224–235).

Many traditional social institutions have had important risk-pooling functions. Indeed, in an economic environment characterized by high variability in returns, with a mean return only slightly above the subsistence level, the benefits to be achieved through risk pooling tend to far outweigh those that are achievable through trade. Generations of ethnographers, for instance, have been impressed by food-sharing practices in hunter-gatherer societies. Indeed, the presence of extensive networks of sharing and gift giving is largely responsible for creating the widespread impression that people in such societies are less "acquisitive" or more "collectivist" in their social attitudes (Polyani 1957: 49). Yet sharing only looks altruistic when compared to market exchange. Its primary economic function is to reduce the variability of returns. In other words, it merely serves to secure a different type of cooperative benefit, one that is of much greater importance than trade in a near-subsistence economy.

Evidence for this hypothesis is reflected in the fact that food-sharing arrangements are typically more dominant in the "hunting" than in the "gathering" segment of foraging economies. Among the Aché of Paraguay, for instance, the mean standard deviation in returns to hunting (and honey collecting) across days is several times greater than it is for gathering, and a far greater percentage of meat (and honey) captured is shared outside the nuclear family (Hill, Hurtado, and Kaplan 1990). This system of sharing breaks the "feast and famine" cycle of irregular returns produced by the reliance on these two food sources (which provide around 75 percent of the caloric intake of the Aché) (Hill, Hurtado, and Kaplan 1990: 129). It is, in other words, an insurance mechanism. A gift economy offers a more flexible version of the same mechanism: in times of plenty, an individual can give away whatever is in surplus, in order to "call in the favor" when returns are scarcer.

The development of agriculture and animal husbandry created much more stable returns, and therefore diminished the need for risk-pooling arrangements in these central areas of economic life. But risk pooling remained an important function of various other institutions. In the European Middle Ages, for instance, the church provided various types of "insurance," some more obvious than others. Monasteries, for example, often sold "corrodies," which functioned very much like life annuities. In return for a lump-sum payment up front, the corrodian was entitled to a standardized room and food ration for the rest of his or her life (Lewin 2003: 37–48). Early pension homes, usually funded by craft guilds, were organized along the same lines. Other guilds ran the equivalent of pay-as-you-go defined benefit pension schemes (often means-tested) among their members. Furthermore, farmers often organized their inheritances through the "sale" of a pension to their children. They would transfer ownership of the land in return for (contractually specified) fixed periodic payments of cash and crops, or often just room and board in what had previously been their own home (Lewin 2003: 49–51).

Many guilds also provided the equivalent of disability insurance. More expansive schemes, like the Chatham Chest in England, provided disability insurance to seamen from the late sixteenth century until the early nineteenth (Lewin 2003: 216). The development of industrial capitalism naturally disrupted many of these craft- and guild-based arrangements. Thus the late eighteenth and early nineteenth century saw the widespread emergence of more "associational" insurance arrangements. During the late eighteenth century, for example, so-called friendly societies enjoyed enormous popularity in England (and "*sociétés de secours mutuels*" in France). Working men all paid a monthly fee to belong to these organizations, with the understanding that the group would pay for medical emergencies affecting any member. These societies also supported the widows and orphaned children of group members. In the early nineteenth century, approximately one million people belonged to such organizations in England (Hacking 1990: 48).[26] This says a lot about the importance of the benefits generated through risk pooling. Since there was absolutely no actuarial basis for calculating the anticipated liabilities of these groups, there was very little guarantee of their ongoing solvency. Nevertheless, the advantages of the limited protection offered were sufficient to make these organizations extremely popular, and "level premium" friendly societies persisted well into the 1920s.[27]

Institutions that enforce self-binding strategies are slightly more difficult to find. This is because, of all the mechanisms of cooperative benefit outlined above, "giving away the key" is the strategy most open to abuse. The free-rider strategy in this case is straightforward betrayal of confidence. Once Ulysses has been tied to the mast, and has instructed his crewmen to disregard all his future orders, there is very little that he can do to prevent himself from being taken advantage of (Elster 1979). As a result, self-binding institutions tend to arise only where there are very high-trust relationships. The family

[26] For further discussion, see Esping-Anderson 1990: 88–92.

[27] The former observation was apparent to nineteenth-century commentators as well, e.g., see Ansell 1835: 4. On the persistence of "level premium" societies into the twentieth century, see Emery 1996: 195–226.

obviously represents the primary locus of such relations. When legally recognized, generally they take the form of explicitly fiduciary relations. Guardianship represents the most formal mechanism through which one person can be legally empowered to act in the interests of another. There are also the so-called helping professions, which include counseling, social work, human resources development, addiction control and rehabilitation, therapy, and the like. All are largely focused upon helping individuals overcome the effects of hyperbolic discounting.

Even though institutions with an exclusively self-binding focus are rare, many social institutions have a self-binding dimension, or are structured in such a way that individuals can use them to control short-term temptations. For example, Ainslie has argued that money, because it is not consumed directly, serves as a stand-in for a temporally extended sequence of satisfactions, and therefore can be used as a psychological self-control mechanism (Ainslie 2001: 100–101). This is why people have such a strong tendency to focus on prices when trying to exercise self-control (rather than thinking in terms of opportunity costs). To take a somewhat different example, Elster has argued that constitutions serve an important self-binding function (Elster 2000: 88–89).[28] Legislatures know that on certain issues they may be inclined to act impetuously, and so seek to preempt this by entrenching their present, presumably more well-considered, judgments.

When it comes to information transmission, lying and misrepresentations are the primary forms of free-rider strategy. As a result, institutions whose primary function it is to generate the pool of shared knowledge that individuals in a society rely upon often incorporate procedures that enforce the norm of truth telling. In specialized research, in particular, peer review lends official sanction to this norm. Much the same applies to the code of journalistic ethics that governs mainstream media outlets.

Schematizing very roughly, one can show how the different types of social institutions that make up the macrostructure of a modern economy facilitate cooperation across these five dimensions: (1) corporations institute teamwork, and thus help achieve economies of scale; (2) the system of property rights helps individuals to achieve gains from trade; (3) the insurance industry provides a basic set of risk-pooling arrangements; (4) the helping professions, along with certain types of fiduciaries, provide a wide range of expertise in developing and assisting in the implementation of self-binding mechanisms; and finally, (5) the media and publication industry provide a general mechanism for the transmission of information. Of course, each of these institutions

[28] Elster notes certain disanalogies, however, since constitutions bind collectivities (i.e., legislatures) and individuals only indirectly.

typically promotes cooperation in several different dimensions. Corporations, for instance, also produce significant benefits for their employees through internal diversification (e.g., making it less risky for workers to develop highly specialized skills) (Heath 2001b: 160–161).[29] The suggestion is simply that the corporate organizational form permits individuals to achieve economies of scale that would be impossible (or prohibitively expensive) to organize through market contracting, and that this provides the primary explanation for its success (Williamson 1985: 90–98). Instead of subsuming these various institutions under the market, and treating their "outputs" as just different classes of services being exchanged in order to produce gains from trade, I think it is helpful to see them as exploiting fundamentally different mechanisms of cooperative benefit.

9.3. Conflicts

Analyzing the benefits of cooperation in terms of five distinct mechanisms suggests a multidimensional understanding of social institutions. Institutions promote efficiency, but they do so in very different ways, and they need to control very different free-rider strategies, depending upon the primary type of cooperative benefit that they are focused upon delivering. Yet already, just surveying the basic institutional building blocks used to create cooperative benefits in these various dimensions, it is easy to see the potential for conflict that they create. In an ideal world, in which individuals always voluntarily refrained from free riding, there is no reason that cooperative benefits could not be secured along all the relevant dimensions simultaneously. In the real world, however, in which institutional constraints and external incentives must be supplied in order to motivate cooperative behavior, conflicts often arise. Arrangements designed to facilitate the production of one form of cooperative benefit may simultaneously undermine the arrangements needed to secure some other.

Consider the case of the competitive market, an arrangement that has as its primary goal the maximization of gains from trade. This example is particularly illustrative, because we have available a formal statement of the conditions that must be satisfied in order for the relevant set of institutions to achieve efficiency in this one dimension. The "invisible hand theorem" (or "first fundamental theorem of welfare economics") shows that the competitive equilibrium of a market economy will be Pareto-optimal as long as certain "standard" conditions obtain (Barr 1998: 78; Blaug 1985; Myles 1995). The

[29] John Kenneth Galbraith's discussion (1990: 85–97), pays due regard to the diversity of functions the modern corporation serves. For reasons of his own, however, Galbraith chooses to subordinate all of these various mechanisms of cooperative benefit under the general rubric of "planning."

list is quite long, but includes, *inter alia*, constant returns to scale, individuals with well-behaved utility functions, symmetric information, a complete set of futures markets, and in cases of uncertainty, a complete set of insurance markets. In other words, the theorem shows that markets achieve perfect efficiency, so long as every other mechanism of cooperative benefit is excluded from consideration, either by assuming that no such benefits are possible, or that all such benefits are freely available. To see how uninformative this is, consider how we would respond to someone who proposed a model for the "optimal" production of scientific knowledge based upon the assumption that both material resources and labor were available in unlimited supply and at zero cost. Whatever its technical merits, such a model would give us very little assistance with real-life policy questions.

Thus the interesting questions arise only when one looks at the market as a real-world institution and considers how the constraints that it imposes interact with the institutional arrangements needed to supply other types of cooperative benefits. As one might expect, conflicts develop along all four dimensions.

> *A. Economies of scale:* The conflict between economies of scale and gains from trade is well known (it forms the central theme, for instance, of John Kenneth Galbraith's critique of "the competitive model" in economics) (Galbraith 1952: 110). Natural monopolies represent the most extreme manifestation. Within a system of private property, exchange alone does not have an optimizing structure. Between two individuals exchanging goods, any price level that gives both parties some portion of the "cooperative surplus" is a potential equilibrium. Thus there may be a deadweight loss in the exchange, because not all of the mutually beneficial exchanges that might occur will occur at the prevailing price level. In a sense, "not enough" exchanges take place. The standard solution is therefore to introduce more buyers and sellers, in order to provoke competition among them. This drives prices toward the level at which the market clears, reducing the size of the deadweight loss.
>
> So in order to secure gains from trade, it is important to have not only exchange but also competition between multiple sellers and buyers. Unfortunately, the tendency for economic activities to exhibit economies of scale works at cross-purposes with this need for competition. In the case of electricity, for instance, once a distributor has hooked up one person's house to a generator or substation, it costs only a tiny bit more to hook up that person's neighbor, the neighbor's neighbor, and so on. Thus it is impossible for any other distributor to compete in that neighborhood. Water supply, sewage, landline telephony, and cable television all have the same structure, giving the typical distributor a natural monopoly within a given territory.

At this point, we need to decide which is more important: the gains that come from the economies of scale or those that come from maximizing the number of exchanges. If we decide that the former are the more significant, we will allow one dominant firm to emerge, then try to minimize the deadweight losses by regulating the price level at which it does business. If we decide that pricing is more important, we might artificially break up the firm, or take antitrust measures designed to prevent a single dominant supplier from emerging. Either way, in trying to hit one target we are very likely to miss the other. Thus each of these institutional compromises generates potential conflict among parties who stand to benefit more from one or the other type of cooperative gain. The fact that, over the course of decades, the "pendulum" tends to swing back and forth between these two strategies is a symptom of the fundamental nature of the underlying conflict.

B. Risk pooling: Similar conflicts arise between the imperative of trade maximization and that of effective risk pooling, although they are less often noted. The classic argument for private property is that the pooling of economic effort encourages free riding, and thus generates a "tragedy of the commons." Consider the standard story. When all peasants are allowed to graze their animals in the common pasture, everyone has an incentive to overgraze, simply because the costs of doing so are largely externalized (i.e., borne by the other peasants who graze their animals on the pasture). Similarly, the benefits of keeping one's animals off the land, in order not to overgraze it, would largely be enjoyed by those who disregard these concerns and continue to overgraze. The solution is to divide up the commons into a set of individual plots, then restrict each peasant to grazing his animals on his own land. This effectively internalizes all the externalities, and thereby eliminates the free-rider strategy that generated the tragedy.[30] Everyone lives happily ever after.

Looked at from the perspective of risk management, however, this happy ending is not assured. The system of property rights, which resolved the tragedy of the commons, also has the effect of "unpooling" a certain type of risk. If the pasture is afflicted by a localized blight, then a system of common landholding will automatically divide the loss equally among all the peasants. When a misfortune befalls the commons, "there followeth not the undoing of any man, but the loss lighteth rather easily upon many

[30] This example, of course, draws its force from the elimination of a collective action problem that arises through negative externalities, whereas in this article I am only considering arrangements that resolve collective action problems involving the failure to produce positive externalities. However, it would be quite easy to vary the example to provide the correct illustration. The Jamestown bowling story, for example (in which no one is willing to work the fields, because the crops are divided up equally among all, regardless of work effort), would suffice. See Ellickson 1993: 1315–1400.

than heavily upon few" (to quote the English Insurance Act of 1601). Thus the commons offers a form of insurance to the peasants.[31] When it is broken up into private holdings, this may result in better land management practices, but it also exposes everyone involved too much greater risk. It is precisely the mechanism of "internalizing" the losses that, on the one hand, encourages more responsible land management, but on the other hand, generates much greater volatility in returns to individuals.

The general problem is that often the only way to "pool" a risk associated with a particular activity is to pool the costs and benefits that result from it. This sort of risk pooling may generate moral hazard, but it may also generate completely unrelated free-rider strategies, such as shirking in the production of the benefits.[32] The important point is that the institutional arrangements needed in order to take advantage of the law of large numbers are often the exact opposite of the arrangements needed to deliver gains from trade or specialization. The development of specialization itself may exacerbate risk. Individuals who invest in the development of specialized skills are faced with the prospect that demand for these skills may disappear. Instead of having a hundred underemployed general construction workers, there may be ten totally unemployed stonemasons or plasterers. As a result, certain forms of specialization may never develop if nothing can be done to shield individuals from exposure to the risks associated with the acquisition of such competencies.

It is interesting to reconsider the history of nineteenth- and twentieth-century "class struggle" in this light. What the early industrial revolution achieved, from this perspective, was a massive "unpooling" of risk. This may have generated increased production, better use of material resources, and better distribution of the ultimate product, but it also exposed members of the working classes to unprecedented levels of risk (unemployment, disability, penury in old age, and so forth), against which the institutions of traditional rural society had once offered them some protection. (For example, instead of having generalized underemployment spread across the countryside, industrialization produced

---

[31] This fact has been noted by anthropologists (e.g., see Winterhalder 1990: 68–70), but the impact of the observation has not been widely felt in normative economics and political philosophy. The "tragedy of the commons" argument for property rights is still typically presented as a straightforward efficiency gain.

[32] Rasmusen (1989) presents a simple model of an insurance arrangement in which "there is a trade-off between efficient effort and efficient risk allocation" (187). The fact that this trade-off arises from a conflict between two different mechanisms of cooperative benefit is obscured, however, by the fact that Rasmusen analyzes insurance contracts as a trade between a risk-neutral insurance company ("perhaps because it is owned by diversified shareholders") and a risk-averse individual (183). This treats the central mechanism as a gain from trade, obscuring the role played by the law of large numbers.

highly localized bouts of complete unemployment concentrated in urban areas.) It is unclear to what degree the appalling conditions of working-class life, documented by so many contemporary observers, were due to the distributive consequences of the market mechanism, and to what degree they were caused by inefficiencies imposed by the market, through its undoing of traditional risk-pooling arrangements. The Marxian reading of the history of working-class struggle, which regards it as essentially a conflict over distributive shares, has cast a long shadow over this discussion. As a result, little attention has been paid to the suggestion that this struggle might also have involved substantial conflict over the type of efficiency gain that should be assigned priority. (Although he does not put it in these terms, this is the plausible thesis at the core of François Éwald's analysis of the welfare state (Éwald 1986). He treats its development as a process through which private insurers, and ultimately the state, came to take on many of the risk-pooling functions that had been discharged by precapitalist rural institutions, filling the vacuum that had been created through the emergence of the market. His account is valuable because of its emphasis upon the incongruity between these insurance arrangements and the logic of market relations.)

C. Self-binding: One can see a similar tension between many of the institutional innovations needed to establish a market economy and the self-binding strategies that people adopt. Savings is perhaps the area where the tension between the gains from trade and self-binding strategies are the most apparent. In a pure market economy, there would be no mandatory savings, in deference to the fact that people's preferences over present versus future consumption differ. A regime of private voluntary savings generates advantages through the complementarities between those who have a greater interest in consuming now, and those with a greater interest in consuming later. However, this generates a persistent problem with dynamic preference inconsistency. Most people have trouble saving, and a significant percentage of the population will not save anything under any circumstances. Thus financial innovations aimed at reducing the transaction costs associated with market exchanges (and thus reducing deadweight losses) may have had the unwanted side effect of undermining savings. Economist David Laibson has argued that financial innovation, by increasing the overall liquidity of assets, has made it increasingly difficult for individuals to create "golden eggs" (Laibson 1997: 443–477). The difference between savings and checking accounts has become purely nominal; the introduction of ATMs has meant that everyone has access to their money at all hours (and so withdrawing a fixed amount at the beginning of the week can no longer be used as a self-control mechanism); reverse mortgages allow people to drain the asset value of their homes; and, of course, consumer credit has rendered the practice of "saving up" for a major purchase almost obsolete. This sort of "easy money" is a mixed blessing to consumers, in the same way that a twenty-four-hour beer store is a mixed blessing to the alcoholic.

In this context, one of the major attractions of state pension systems around the world is precisely their mandatory character. Of course, the insurance-like character of pension plans means that often there is no conflict between the compulsion needed to secure self-binding and the overall efficiency goals. Nevertheless, we should not harbor any illusions about where the principal benefits of these programs come from. The mere fact that people are forced to save is a source of huge long-term welfare gains. At the same time, these pension arrangements clearly restrict the extent of private contracting, and hence gains from trade, in the domain of savings and investment.

To take a more concrete example, many people in modern industrial societies have difficulty controlling their diet. Obesity, along with certain of its consequences such as Type II diabetes, are rapidly becoming major health issues in the United States. There is, however, enormous resistance to any proposal that seeks to use institutional resources to change diet (other than simply providing individuals with more information). Thus taxes on junk food or soda, along the lines of the taxes that exist on alcohol and tobacco, are rejected on the grounds that they are paternalistic, or that they interfere with "consumer sovereignty." (It is sometimes argued that tobacco and alcohol are addictive, whereas junk food is not. This is a specious distinction, when looked at through the lens of hyperbolic discounting. It also ignores the fact that cigarette sales exhibit considerable price elasticity [World Bank 2000].) This deference to market choice leaves many individuals forced to devise their own self-binding strategies. They may subscribe to diet services that provide them with meals, so that they no longer do their own food shopping. Of course, because of the voluntariness of these remedies, they are only successful with individuals who are very close to being able to exercise effective self-control. Those who require more irreversible precommitment strategies are increasingly opting for "stomach stapling" and other forms of gastric surgery as a way of controlling food cravings. The obvious inefficiency of these arrangements should clearly be marked down as one of the "costs" imposed by our commitment to maximizing the gains from trade in the market for food. It is precisely the fact that our social institutions offer individuals so little assistance in developing self-binding strategies that they are forced to turn to drastic alternatives whose sole advantage is that they can be carried out individualistically.

D. Information: Finally, the patent and copyright system offers a straightforward example of a conflict between the institutional arrangements needed to promote the production and dissemination of ideas and those

> required to achieve gains from trade. Both offer incentives to producers by granting them a short-term monopoly over the sale of a particular expression of an idea, or else its practical application. Furthermore, cross-licensing or "patent-pooling" arrangements between competitors create the possibility of legal cartelization of an industry (and hence price fixing).
>
> Theorists who try to square the circle through talk of "intellectual property"—treating ideas and information as though they were commodities to be bought and sold like any other—are willfully disregarding fundamental differences in the nature of the cooperative benefit in the two domains. Ideas are nonrival in consumption, and thus the problem of scarcity, which is fundamental to the structure of property rights and exchange, simply does not arise. In the case of intellectual production, efficiency involves creating the incentives required for an optimal rate of innovation.[33] In practice, this has often meant imposing restrictions on trade that create significant deadweight losses.
>
> To take just one example, after Wilbur and Orville Wright discovered the rudiments of a workable aircraft, they promptly patented various aspects of their design. They subsequently refused to grant any licenses to other designers who had been experimenting with aircraft, forcing many to abandon their efforts. They also engaged in protracted litigation against Glenn Curtiss, who developed a superior solution to the problem of lateral control (the aileron, which is now used in all modern aircraft). These disputes carried on well into the First World War, impeding the production of aircraft for military purposes. It was only at the request of the Secretary of War and the Navy in 1917 that the US government intervened to impose a resolution (Bittlingmayer 1988: 227–248).
>
> The lessons to be learned from examples like this are fairly clear. As Adam Jaffe and Josh Lerner observe, "Patents are blunt instruments. Because of the complexity of the evolution of technology, the monopoly that they create will sometimes retard rather than encourage competition. This means that, in the best of worlds, a patent system is a compromise among competing objectives" (Jaffe and Lerner 2004: 51). This is reflected in the fact that public policy with respect to patents also exhibits a pendulum-like swing, moving back and forth over the course of decades between a system that favors patentees and one that favors more open use.

There is a clear pattern that emerges from this discussion. The market economy does an extremely good job at promoting gains from trade but is often inefficient or obstructive when it comes to providing cooperative benefits of

[33] I am grateful to Patrick Turmel (personal communication) for this formulation of the distinction.

the other four types. Thus these other forms of cooperative benefit are often organized through "administrative" rather than "market" transactions.[34] Economies of scale have been exploited successfully in the private sector, but not so much through market exchange as through the internal structure of the corporation (Williamson 1985: 90–96). Markets greatly exacerbate the risk that individuals are exposed to. The production of information also occurs largely in the interstices of markets relations; very seldom is information bought or sold. Some of these problems are corrected within the corporation. However, when it comes to correcting the failures of the market, the most important actor is, and has always been, the state. Unfortunately, the full range of state action in this regard has seldom been adequately appreciated, because the mainstream understanding of the state has also been marked by an overemphasis on gains from trade, along with a failure to appreciate the extent to which the pursuit of these gains can conflict with the objective of achieving other types of cooperative benefit.

9.4. Rethinking the Welfare State

The catallactic perspective encourages theorists to regard the efficiency problem as essentially solvable (i.e., simply a matter of getting the incentives right). As Galbraith put it, the competitive economy is often interpreted as offering a solution to the problem of efficiency (Galbraith 1952: 24). This in turn has encouraged the view that the welfare state plays nothing but a residual role. Because markets already achieve the overwhelming majority of efficiency gains, state action must either be aimed at patching up gaps in the system of property rights, or it must be motivated by some other consideration entirely, such as a concern over distribution. The influence of this idea is widespread. For example, it is the fundamental presupposition underlying both the "right-wing" view that political action is dominated by "rent-seeking" behavior, and the "left-wing" view that the primary function of the welfare state is to secure distributive justice.[35]

There are, however, an enormous number of state activities that enhance overall quality of life and yet are difficult to fit into this taxonomy. Various aspects of the "social safety net," for example, are difficult to classify in the traditional model, and so are described as redistributive welfare programs. For example, the money paid to individuals from programs such as unemployment insurance is often treated as a redistributive income transfer rather than as an insurance payment (Rosen et al. 2003: 88). Similarly, state pensions are said to generate a transfer from the young to the old (especially when funded

[34] For this distinction, see Shipman 1999: 297.
[35] For the former, see Tullock 1970 and Downs 1957; for the latter, see Rawls 1999: 240–250.

on a pay-as-you-go basis), motivated by a set of egalitarian commitments (such as "reduction of poverty among the elderly"). This analysis ignores the fact that defined-benefit pension plans are essentially insurance systems, which offer protection against the risk of outliving one's savings. Risk-averse individuals will normally want to "oversave," in order to increase their confidence that they will have adequate retirement income. By pooling their retirement funds, they are able to achieve the same level of confidence with a much lower rate of savings. Of course, this means that the pension system generates a transfer among individuals, but that does not mean that it is governed by an egalitarian-redistributive logic. All insurance generates transfers ex post: car insurance transfers wealth from those who do not have accidents to those who do, just as pensions generate a transfer from those who die young to those who live longer. In neither case, however, does the transfer follow an egalitarian logic: it could well be regressive from the standpoint of overall income. Thus the conflict over privatization of pension systems, which is spearheaded by those who point to the much higher rates of return obtainable through private savings, is not really a conflict between efficiency and equality, but between two types of efficiency gains. Those who recommend individual retirement savings accounts are, in effect, interested in securing cooperative gains from complementarities. Those in favor of defined-benefit pension systems assign higher priority to the gains obtainable through risk pooling. From the perspective of the latter, concern over the rate of return reveals a fundamental conceptual error (Ghilarducci 2000: 69–92).

One can see a more abstract version of this tendency to misclassify insurance arrangements, and thus to overstate the egalitarian character of the welfare state, in Ronald Dworkin's resource egalitarianism. By relying upon a Walrasian auctioneer to achieve an envy-free and efficient distribution of resources, Dworkin is presupposing the basic framework of the invisible hand theorem (Heath 2004: 313–335). The attendant catallactic bias shows up in the fact that he introduces insurance only after the auction has been conducted. At that point it is brought in, not as a source of sui generis cooperative benefit, but merely as a mechanism used to buffer the egalitarian distribution from the effects of brute luck (i.e., as a way of maintaining some approximation of dynamic envy-freeness) (Dworkin 2000: 73). He fails to observe that the distribution of the benefits produced through the insurance system itself raises an independent question of distributive justice. In particular, he does not ask whether risk aversion should be regarded as an expensive taste or as "bad preference luck."[36] It also encourages him to interpret the payment of

[36] On the latter, see Dworkin 2004: 348. Dworkin does state that risk aversion is the basis for the utility gain associated with insurance (Dworkin 2000: 95), but he goes on to elaborate this point in a way that suggests otherwise. For example, he writes, "I buy insurance on my house because the marginal utility loss of an uncompensated fire is so much greater than the utility cost of the premium"

TABLE 9.1 } **Private and public institutional forms**

Mechanism of cooperative benefit	Private sector	Public sector
Economies of scale	Corporations	Natural monopolies
Gains from trade	Markets	Public goods
Risk pooling	Insurance industry	Social safety net
Self-binding	Helping professions	Social work
Information transmission	Media	National statistics, certification

taxes under the welfare state ("tax as premium") as essentially governed by an egalitarian-redistributive logic, rather than an efficiency-promoting one (Dworkin 2000: 99–109).

The way to avoid such problems is to stop thinking of the state purely in terms of residual activities. Markets were designed to facilitate exchanges, and thus gains from trade, not to create mutual-security societies, or to encourage laboratory research. Thus it would be no surprise to discover that when it comes to providing certain classes of cooperative benefit, the state has often taken the lead role. In the same way that the private sector contains institutions generating benefits in all five dimensions, one can very easily find ways in which the state discharges each of these five functions (see table 9.1). The state is not only the insurer of last resort but also provides a number of fundamental insurance services, without which the market economy itself could barely function, such as limited liability and deposit insurance (Moss 2002: 314–315). The state enforces a wide range of seemingly "paternalistic" policies, such as mandatory retirement savings, that are in effect self-binding strategies. More importantly, the state provides extensive "social work" services, dealing with problems of substance abuse, delinquency, child neglect, domestic strife, and so forth, all of which are related in important ways to problems of hyperbolic discounting. Finally, beyond its role as a straightforward producer of information (ranging from the census and weather forecasts to scientific research undertaken with state subsidy) the state plays an essential role certifying information produced in the private sector and enforcing standards of veracity (e.g., drug testing, food labeling, truth in advertising, financial auditing, and so on). It is able to take a lead role in this domain because various state agencies enjoy a level of credibility that few private institutions can match.

Rather than trying to shoehorn these various functions into the model of gains from trade, it is better to regard each as exploiting a sui generis mechanism of cooperative benefit. This is helpful in a variety of ways. The most important

(Dworkin 2000: 97). It goes without saying that the loss of the house is worse than the payment of the premium, but that fails to explain why anyone buys insurance. People buy insurance because—or more to the point, to the extent that—the expected utility associated with the gamble between losing and not losing one's house is less than the loss of utility associated with paying the premium.

is that it forces us to keep in mind just how partial the picture of society is that standard welfare economics models supply. When doing cost-benefit analyses for the development of economic policy, for instance, it is not adequate to accommodate stochasticity simply by taking certain key variables in the model and declaring them to be averages. One must also consider the variance, in order to weigh the risks that particular arrangements impose upon individuals. In the calculations concerning privatization of state electricity supply, for instance, cost-benefit analyses were often made on the basis of average prices, with surprisingly little attention paid to the disutility for consumers generated by price volatility. The way that state monopolies level the peaks of these price shocks is essentially an insurance function, yet its value was not calculated as one of the efficiency gains associated with that mode of supply. (As a result, many governments that privatized electricity were forced to backpedal and impose price caps, in response to consumer outrage over price volatility.)[37] The catallactic perspective, which encourages us to equate efficiency with exchange, prevents us from articulating the logic of the conflicts that erupt in these and other domains. The issue (lower mean versus lower variance) is not distributive, in the narrow sense of the term. It is over which type of collective action problems we should resolve, at the expense of which others.

9.5. Conclusion

It is common among contractualists to think that social institutions must be governed by two fundamental norms: first, a principle of efficiency that specifies "how much" cooperative benefit should be produced, and second, a principle of equality that specifies "who gets what" in the distribution of these benefits. It is then often assumed that, because there is a common interest in maximizing the cooperative benefit, yet a conflict of interest over who will get what share, that efficiency is somehow easier to achieve, or less controversial, than equality. Increased attention to the prisoner's dilemma has gone some way toward showing that efficiency gains cannot be taken for granted, since individual self-interested action will not, in many cases, lead to Pareto-efficient outcomes. However, there has still been a tendency to think, on this basis, that the real problem with efficiency gains is just an incentive problem. According to this view, the choice of institutional arrangement should be uncontroversial: the problem is simply the technical one of bringing the free riders under control. I have tried in this article to show that even this assumption is too sanguine. The institutional arrangements needed to bring one set of free riders

---

[37] See Trebilcock and Hrap (2003: 6). For a similar experience in Alberta, see National Energy Board (2001: 24).

"under control," in order to resolve a particular collective action problem, often preclude the institutional arrangements needed to resolve some other collective action problem. This naturally generates a conflict of interest among parties who are not indifferent to the benefits produced from these two classes of interactions. The most clear-cut example of this is the privatization strategy in a tragedy of the commons, which is needed in order to resolve externality problems and to kick-start the market economy, yet at the same time eliminates a certain sort of risk-pooling arrangement, and thus leaves individuals exposed to greater variability in returns.

As a result, efficiency gains are often just as controversial as "pure" distributive issues.[38] Unfortunately, very little thought has been put into the question of what principles should govern our choice when we are forced to trade one form of cooperative benefit off against another. The suggestion that society should be committed to promoting economic growth, for example, still seems self-evident to many, despite the fact that this maximizes only one form of cooperative benefit. The need to break free from such patterns of thought should become more urgent with the recognition that each mechanism of cooperative benefit is also subject to diminishing returns. For example, the growth of the market leads to a steady decline in the utility gains to be achieved through further trade (since the most significant complementarities of need and ability will be the first to be exploited). If expansion of the market also has the effect of increasing the level of risk exposure of the average worker, then one can easily imagine a scenario in which the welfare losses associated with the latter begin to outweigh the increasingly marginal gains from trade. The only way to appreciate the problem this creates, however, is to break free from a "monistic" conception of the mechanism that generates the benefits of cooperation.

---

[38] This observation has been made before, in different ways, see Knight 1992: 34–37.

PART III }

Extending the Framework

10 }

The Uses and Abuses of Agency Theory

The spectacular corporate scandals and bankruptcies of the past decade have served as a powerful reminder of the risks that are involved in the ownership of enterprise. Unlike other patrons of the firm, owners are residual claimants on its earnings (Hansmann 1992: 11). As a result, they have no explicit contract to protect their interests, but rely instead upon formal control of the decision-making apparatus of the firm in order to ensure that their interests are properly respected by managers. In a standard business corporation, it is the shareholders who stand in this relationship to the firm. Yet as the early twenty-first-century wave of corporate scandals demonstrated once again, it can be extraordinarily difficult for shareholders to exercise effective control of management, or more generally, for the firm to achieve the appropriate alignment of interests between managers and owners. After all, it is shareholders who were the ones most hurt by the scandals at Enron, Tyco, Worldcom, Parmalat, Hollinger, and elsewhere. Indeed, one of the reasons that Enron's collapse was particularly damaging to its employees was that so many of them were also shareholders, through the company ESOP and their 401k plans.

The type of managerial attitude toward investors that proved so damaging in these scandals was best illustrated by an internal memo written by Hollinger International CEO Conrad Black, who described a self-dealing transaction conducted by Richard Perle, a member of the firm's Board of Directors, as containing a "good deal of nest-feathering," yet complained only about the exclusion of management from the benefits. "They should treat us as insiders with our hands cupped as the money flows down, and not as outsiders pouring in the money," he wrote (Labaton 2004). It was the prevalence of such attitudes toward shareholders ("outsiders pouring in the money") that led an investigative committee struck by the Board to later describe Perle as a "faithless fiduciary," and Black as having run a "corporate kleptocracy" (Norris 2004). Black was subsequently convicted on three counts of fraud relating to the most egregious transactions. It seems plausible to suppose that this illegal conduct was surrounded by a fairly broad penumbral region of unethical conduct.

One of the central tasks of theoretical business ethics is to provide a conceptual framework that will allow us to articulate more clearly the intuitive sense we all have that "nest-feathering" and similar forms of conduct are unethical, so that we can state with greater precision the nature of the moral obligations that have been violated. In approaching this task, the first place that business ethicists might reasonably be expected to look is to *agency theory*.[1] After all, the relationship between owners and managers is a textbook example of a principal–agent relationship (e.g., Campbell 1995: 79–86; Milgrom and Roberts 1992: 170). Furthermore, deception and misappropriation of funds by the agent represent perfect examples of the type of moral hazard problems that are an endemic feature of principal–agent relations. Thus one might expect business ethicists to embrace agency vocabulary as a way of stating with greater precision the exact nature of the moral obligations that were violated at Hollinger, Enron, and elsewhere.[2] One might also expect business ethicists to insist that greater attention be paid to agency relations, and to the potential moral hazard problems that they harbor, as a way of avoiding such scandals in the future. Indeed, many have done so.

However, the reaction to the scandals among business ethicists has been far more mixed than one might expect. Part of the reason is that many business ethicists have spent considerable time and energy downplaying the importance of shareholders in the organizational structure of the firm, and trying to show that managers have important moral obligations to other "stakeholder" groups (see Blair 1995; Clarkson 1998; Kelly 2001). Many deny that managers should be regarded as "agents" of the shareholders in any significant sense of the term. Thus they do not regard the recent spate of corporate scandals as grounds for renewed attention to the agency risks that exist in the manager–shareholder relation. On the contrary, some have gone so far as to blame agency theory—and the teaching of agency theory in business schools—for creating the corporate culture that led directly to the scandals (Ghoshal 2005: 75–76). Rakesh Khurana, Nitin Nohria, and Daniel Penrice of the Harvard Business School have suggested that the "doctrine of shareholder primacy" combined with agency theory "led directly to many of the worst profit-maximizing abuses unmasked in the recent wave of corporate scandals" (2005).[3] Along similar lines, Brian Kulik (2005) has argued that "agency reasoning" on the part of Enron executives led to the creation of an "agency culture" and an organizational structure within the firm that encouraged corrupt behavior.

[1] For a complete technical overview of this theory, see Laffont and Martimort 2002.

[2] Allen Buchanan (1996) has provided what is perhaps the most sophisticated development of this approach.

[3] Some theorists, such as Quinn and Jones (1995), simply equate agency theory with the doctrine of shareholder primacy, which leads to the suggestion that anyone committed to stakeholder theory must reject agency theory.

So which is it? Is agency theory a part of the problem, or a part of the solution? In order to get clear on this question, it is important first to get clear on the sort of theoretical commitments that are essential to agency theory (in order to distinguish between agency theory itself and certain incorrect interpretations that have become widely promulgated). It is also important to be more specific about the ways that agency theory can be used to analyze relations within the firm, in order to determine whether it is the use or the abuse of agency theory that has become a source of mischief. Finally, it is important to be more specific about the circumstances in which moral obligations can arise out of agency relations. Only then is it possible to develop a more balanced appreciation of the contribution that agency theory can make to the study of business ethics. Thus I will begin with an outline of three major objections that have been raised against the use of agency theory by business ethicists. In the second section, I show how some—but not all—of these objections can be met, before going on, in the final section, to present what I consider to be the most fruitful use of agency theory. Taking agency theory seriously, I will argue, provides the closest thing one can get to a proof that, without certain forms of moral constraint, it would be impossible to organize a successful business firm, much less have a productive market economy. Thus, agency theory should be embraced by business ethicists, not because it promotes an empirically accurate understanding of the firm—it does not—but because it shows how unworkable modern capitalism would be in the absence of any sort of business ethics.

10.1. Objections to Agency Theory

Agency theory, in the sense that the term is used here, is an approach that involves the application of game theory to the analysis of a particular class of interactions, namely, "situations in which one individual (the agent) acts on behalf of another (the principal) and is supposed to advance the principal's goals" (Milgrom and Roberts 1992: 170). This is already a potential source of confusion, since the term "agent" is used differently here than in certain other contexts, such as commercial law, where the "law of agency" assigns a much narrower meaning to the term (see Clark 1985: 56). In the legal sense, an agent is one who is entitled to negotiate on behalf of a principal, or bring the principal into a contractual relation with some third party. It is in this sense of the term that we talk about "real estate agents" or "literary agents." The game-theoretic sense is much broader, dealing (at least in principle) with any sort of interaction between two individuals where one is trying to influence the actions of the other. Indeed, perhaps because of the potential for confusion on this score, some agency theorists have taken to redescribing their work as simply "the theory of incentives" (e.g., Campbell 1995; Laffont and Martimort 2002).

Yet while disputes over the use of the term "agent" have given rise to considerable misunderstanding, it is the use of game theory (or "rational choice theory") that makes agency theory genuinely controversial. This is because game theory comes freighted with a number of substantive theoretical assumptions, including most prominently, a commitment to an instrumental (or "economic") model of rational action. Thus individuals are represented as expected utility-maximizers (who, when faced with a problem of interdependent choice, select actions that represent an individually best response to the anticipated actions of the other individuals). This immediately raises the dander of many ethicists, since economic models of rationality are famous for either classifying all moral action as irrational, or else rationalizing it through the "discovery" and ascription of some underlying non-moral incentive. Unsurprisingly, this forms the basis for the most widespread and immediate objection to agency theory.

10.1.1. SELF-INTEREST

Ethicists often complain that agency theorists, by adopting an economic model of action, thereby assume that rational individuals are self-interested, or that they act only from egoistic and not altruistic motives. This is, from their point of view, equivalent to endorsing moral skepticism and is therefore not a helpful point of departure for the development of a system of applied ethics. Of course, the standard response to this criticism is to say that the economic model of rationality implies no such thing. Utility is defined with respect to the preferences of individuals, and preferences reflect whatever desires individuals happen to have, egoistic or altruistic.[4] David Gauthier made the point most succinctly, when he observed that, according to the economic model of rationality, "it is not the interests *in* the self, that take oneself as object, but interests *of* the self, held by oneself as subject, that provide the basis for rational choice and action" (1986: 7; see also Hausman and McPherson 1996: 52–53). Thus what creates the need for incentives in principal–agent relations, strictly speaking, is not the fact that the principal and the agent have egoistic preferences, but merely the fact that they have different preferences. Principal–agent theory is about how individuals manage situations involving "goal incongruity" between two or more persons (Dees 1992: 37–38). It does not matter whether they are selfish or not; what matters is that each acts in pursuit of his or her own goals, and that the goals of the other show up only insofar as they affect that agent's goals, or ability to satisfy these goals.

---

[4] For an example of a typical—and typically acrimonious—exchange between a business ethicist and a game theorist on this point, see Solomon 1999 and Binmore 1999.

On these grounds, some business ethicists have concluded that agency theory is perfectly harmless. Allan Buchanan articulates this view well when he writes that

> [i]f, in applying principal/agent theory, it were necessary to assume that motivation is exclusively or primarily self-interested, this would greatly reduce if not vitiate the enterprise. However, we need not do so. Instead, we can proceed on the assumption that the conflicts of interest that give rise to agency-risks may result from a variety of motivations, on the part of agents and principals. All that is necessary is that there be conflicts of interest. (1996: 421)

Of course, in fairness to those business ethicists who have complained about the self-interest assumption, it should be noted that one can search the economic "theory of the firm" literature for a very long time before finding an actual example of an agency analysis that ascribes altruistic motives to any of the parties involved. Even if the theoretical framework does not force them to do so, agency theorists often do make unflattering empirical assumptions about individual preferences, by stipulating in their models that, for example, work effort has negative utility, money rewards have positive utility, and that individuals have no other relevant motives (Dees 1992: 29). Strictly speaking, however, such assumptions are not essential to the economic model of rationality, and so theorists like Buchanan are quite correct to point out that agency theory per se entails no commitment to such claims.

It would be premature, however, to conclude on this basis that the economic conception of rationality is neutral from the standpoint of ethics. There are a number of other substantive theoretical commitments associated with the instrumental model, which are hostile from the perspective of the ethicist, and which cannot be purged from the model so easily.

The first of the two outstanding problems stems directly from the tendency among game theorists to "black box" all questions of motivation. While this theoretical strategy does allow them to sidestep disputes over altruism and egoism, it also leaves them without a developed theory of preference-formation, and thus without any ability to model the way that preference changes arise out of social interactions (Knight 1992: 18). Preferences are taken as given and are also taken to be independent of strategies. Thus players in a standard game-theoretic model cannot change each other's preferences (i.e., utility functions) through their actions. This is closely related to the fact that in standard game-theoretic models players are explicitly precluded from communicating with one another (using any sort of independent semantic resources, such as language; they are still able to draw inferences from observing each other's actions, and so are able to "communicate" in this sense [Nash 1951]). Furthermore, insofar as they *are* able to communicate with one another, standard game-theoretic solution concepts, like Nash equilibrium, do not apply

(Heath 2001a: 73–78). This non-trivial restriction on game-theoretic models is often conveniently forgotten by those who are eager to apply them to the analysis of empirical interactions.

In any case, the fact that there is no generally accepted or robust theory of endogenous preference-change in games means that agency theorists have devoted almost all of their time and attention to studying the way that external incentives can be used to bring about greater alignment of goals in cases of incongruity. This often turns into a classic case of economists searching where the light is best. For instance, in their widely used management textbook on organizational theory, game theorists Paul Milgrom and John Roberts (1992) dedicate an entire chapter to the subject of moral hazard and agency relations within the firm. They canvas an exhaustive range of strategies for controlling employee shirking, including monitoring, incentive contracts, performance pay, ownership stakes, employee bonding, and promotional systems. At the same time, they fail to mention such absolutely elementary factors as whether or not employees enjoy their jobs, and whether they love or hate the firm that they work for (1992: 179–192).[5] Similarly, in their chapter on human resources policy, Milgrom and Roberts have a lengthy discussion of employee retention strategies, which does not once mention the fact that employees sometimes feel a sense of loyalty toward the firm (and that managers have it within their power to cultivate such loyalties). On occasion, this theory-induced aphasia borders on the comical, as when they develop a "case study" of human resources policies in Japan that manages to avoid mentioning the issue of employee loyalty altogether. "The control structure of Japanese firms, which gives considerable power to the employees as a group" is explained, not as a way of promoting loyalty and building *esprit de corps*, but rather as a way of enabling employees "to protect their valuable employment rights" in the face of labor-market rigidity (Milgrom and Roberts 1992: 350). It is *greater fear of losing their jobs*, we are led to believe, that makes Japanese workers more willing than Americans to accept sacrifices on behalf of their employer (cf. Fukuyama 1995: 185–193, 255–266).

Once again though, this emphasis on external incentives is not a necessary consequence of the commitment to the economic conception of rational action. Nothing intrinsic to agency theory prevents theorists from taking an interest in the way that "internal" incentives—e.g., preference change—can be used to overcome agency problems, it is just that game theorists have no idea how to model such processes, and so have largely chosen to ignore them (in very much the same way that, prior to the advent of game theory, economists had no good way to model information states, and so largely chose to

[5] In this respect, their discussion falls significantly below the level of sophistication exhibited in older classics such as March and Simon (1958: 65–81).

ignore the impact of asymmetric information on market exchanges). Thus the emphasis on external incentives is simply a case of methodologically induced bias, which could be corrected through the development of more sophisticated modeling techniques—or even just frank acknowledgment of the need for qualitative analysis in this domain. So again, there is no reason in principle for the ethicist to object to the use of agency theory.

The second outstanding problem, however, has no quick fix. It involves the commitment, on the part of the agency theorist, to the view that individuals will behave opportunistically whenever given the chance to do so. There are two components of opportunism in the standard (i.e., dictionary) sense of the term: first, that of taking advantage of circumstances as they arise, and second, that of acting without regard for principle. Entering into a cooperative agreement, then reneging once the other party has performed, is the paradigm example. Agency theorists routinely assume that regardless of what people say they are going to do, they will always update their plans as the situation unfolds, and renege on any prior commitments whenever it is in their interest to do so. Thus a farmer may hire workers who promise to harvest his crop, but find himself facing a strike threat at a critical time during the season, when it is too late to bring in replacement workers (Milgrom and Roberts 1992: 128). An insurance company may agree to indemnify any policy-holder who suffers a particular sort of loss, but then drag its feet when the time comes to pay the claim (e.g., by proposing unusual legal interpretations of certain exclusion clauses). Employees may agree to give some particular job their full attention, but then shirk in various ways in situations where their effort level is unobservable, and so on.

Along with this characterization of opportunistic behavior comes the assumption that individuals are unable to credibly commit themselves to refraining from opportunism, unless they are able to create some external incentive structure that changes their own future incentives (such as posting a bond to guarantee performance). Promises to perform are basically cheap talk, and the rational principal will disregard them when it comes to managing agency relations.

Ethicists are unlikely to regard this as a satisfactory framework for analysis, since it suggests that rationality encourages individuals to exhibit a variety of vices, including fickleness (in Machiavelli's sense of the term), dissimulation, treachery, and guile. It also follows very closely upon this that rational individuals will treat each other with distrust and suspicion. Thus agency theory seems to take some of the worst assumptions about human nature and build them into its central definition of rationality. Furthermore, in this case, the standard evasive response is not available to the agency theorist. Unlike the egoism postulate, which is in fact peripheral to the instrumental conception of rationality, the assumption of opportunistic behavior is absolutely central to the model. The fact that agents are unable to make commitments is one of the defining postulates of non-cooperative game theory (Nash 1951) (and again, all of the standard solution concepts do not apply in cases where that assumption is relaxed [Heath 2001a: 86–92]). What we typically refer to as "opportunistic" behavior is a direct consequence of agents acting in accordance with the general game-theoretic principle known as *sequential rationality*. This is simply the view that, in a multi-stage game, a rational strategy must not only be utility-maximizing at the point at which it is chosen, but each of its component actions must also be utility-maximizing at the point at which it is to be performed. The sequential rationality postulate is what licenses, among other things, the use of backward induction as a method for solving multi-stage or repeated games (Fudenberg and Tirole 1991: 72–74), as well as the "subgame perfection" solution concept, which is the most uncontroversial refinement of Nash equilibrium (Selten 1975). It is so deeply entrenched that, in most cases, game theorists don't even bother to mention it. Eric Rasmusen, for example, in his widely used textbook on game theory, discusses the principle only once, in order to explain why he will not be mentioning it again:

> The term sequential rationality is used to denote the idea that a player should maximize his payoffs at each point in the game, re-optimizing his decisions at each point and taking into account the fact that he will re-optimize in the future. This is a blend of the economic idea of ignoring sunk costs and rational expectations. Sequential rationality is so standard a criterion for equilibrium now that often I will speak of "equilibrium" without the qualifier when I wish to refer to an equilibrium that satisfies sequential rationality. (Rasmusen 1989: 95)

"Opportunism," from this perspective, is just a somewhat moralizing way of describing the phenomenon of re-optimization, and as such, is not easy to get rid of as a game-theoretic assumption. On the contrary, it comes very close to capturing the essence of the strategic conception of rationality. Central to this conception is the consequentialism postulate, which states simply that the value of an action is a function of its anticipated consequences, and nothing else (the commitment to re-optimization follows almost immediately from this consequentialism). Yet consequentialism precludes the possibility that a rational agent might incorporate *deontic constraints*—principles associated directly with actions, independent of their consequences—into his or her deliberations (or what Nozick [1974: 28–32] refers to as "side constraints"). Since genuine loyalty, commitment, conformity to social norms, and respect for moral rules are all forms of deontic constraint, this is a very significant restriction. It is what leads game theorists, for instance, to dismiss all "non-payoff relevant" communication as cheap talk. Since individuals will simply say whatever it is in their interest to say (regardless of what honesty might dictate), everyone else, knowing this, will be inclined to ignore them. This generates the well-known game-theoretic result, established by Vincent Crawford and Joel Sobel, that "once interests diverge by a given, 'finite' amount, only no communication is consistent with rational behavior" (1982: 1450).

Thus when a critic like Eric Noreen claims that "at the heart of agency theory, as expounded in accounting, finance and economics, is the assumption that people act unreservedly in their own narrowly defined self-interest with, if necessary, guile and deceit" (1988: 359), he is only partly mistaken. While it is incorrect to say that self-interest, narrowly defined, is at the heart of agency theory, it is correct to associate agency theory with the view that people act unreservedly, using guile and deceit—not even when necessary, but whenever it is advantageous for them to do so. Thus the image of employees loafing around whenever the boss isn't looking, faking disabilities, calling in sick during hunting or fishing season, exaggerating the difficulty of their assignments in order to make their performance appear more impressive, and so on, is a non-accidental consequence of the agency perspective.[6] It is a case of what Lex Donaldson refers to as "guilt by axiom" (1990: 373).

In a previous iteration of these debates, Oliver Williamson's transaction cost approach to the theory of the firm was taken to task for assuming the individuals sometimes behave opportunistically (Ghoshal and Moran 1996: 19). Agency theory goes much further, claiming that agents, insofar as they are rational, always act opportunistically. Indeed, some of the early victories of agency theory came from its ability to account for cases, such as conglomerate mergers, where managers fail to adopt the firm size that would be optimal from the standpoint of transaction cost minimization (Eisenhardt 1989: 68). The implicit suggestion was that Williamson was overly optimistic about the possibility of controlling opportunism through the mere substitution of hierarchies for markets. Opportunism was, as far as agency theorists were concerned, more pervasive than even Williamson had realized.

Thus business ethicists do have some legitimate concerns about the agency theory framework, insofar as it incorporates a controversial conception of rationality, one that presupposes the correctness of a certain form of skepticism about moral rules. Yet even then, it is unclear that these concerns need ripen into full-blown complaints. After all, most agency theorists are not in the business of doing normative theory. In other words, they are not telling people how they should behave (and thus are not directly recommending opportunism, guile, and deceit as laudable forms of behavior). Their goal typically has been to develop a positive theory of the firm, to offer merely empirical explanations of why organizations take on particular forms, structured by particular sets of incentives. If, in doing so, they make certain unflattering assumptions about human nature, why should that be any cause for alarm? So despite whatever reasonable reservations ethicists may have about the conception of

[6] Most of these examples are drawn from Milgrom and Roberts 1992: 170.

practical rationality underlying agency theory, it remains to be seen how much of a problem that theory can be expected to create for those trying to understand (or promote) business ethics.

10.1.2. SHAREHOLDER PRIMACY

A second issue with agency theory that has been a source of concern among ethicists is the close connection many see (and many others assert) between agency theory and the doctrine of shareholder primacy. Margaret Blair, for instance, in her influential work on "team production" theory, starts out by defining "the principal–agent" model of the firm as the view that "public corporations are little more than bundles of assets collectively owned by shareholders (principals) who hire directors and officers (agents) to manage those assets on their behalf" (Blair and Stout 1999: 248). Similarly, Milgrom and Roberts, after defining the principal–agent relationship, go on to assert that "senior executives of corporations are charged with advancing the interests of the stockholders, who are the owners of the corporation," and that these executives are therefore "agents of the stockholders" (1992: 181). Michael Jensen and William Meckling, after offering a brief introduction to agency theory, argue that "the relationship between the stockholders and the managers of a corporation fits the definition of a pure agency relationship" (Jensen and Meckling 1976: 309), and proceed to analyze the firm on that basis.

Jensen and Meckling present this as though it were purely an empirical observation—a "positive" claim about the structure of the firm, not a "normative" claim about how the firm should be organized. Yet it is not clear that describing a particular relationship as a "principal–agent" relationship can ever be normatively neutral. This is because, in any sort of social interaction, both parties influence each other to varying degrees. Thus any purely positive definition of the agency relationship is bound to create ambiguity concerning who is the agent and who is the principal. Donald Campbell, for instance, in his textbook on incentive theory, states that "the principal is the individual whose welfare is to be served and this welfare is affected by an agent who makes decisions on behalf of the principal" (1995: 8). He then illustrates this with the standard example of a person taking a taxi from the airport, with the passenger as principal and the driver as agent. Yet who is to say that the passenger is the principal, and not the agent? Both individuals make decisions that affect the welfare of the other (e.g., the passenger decides whether to pay, how much to tip, and so on). Recall the old saying that a chicken is nothing but an egg's way of making another egg. In this case, the passenger may be nothing but the cab driver's way of earning a fare. The only way to infer the "correct" agency relationship, using Campbell's definition, is to understand the phrases "whose welfare is *to be served*" and "makes decisions *on behalf of*" in normative terms. The principal is the one whose welfare ought to be served; and the agent is the one who is under an obligation to serve the principal faithfully (and, typically, is in a position to abuse an information asymmetry).

With Milgrom and Roberts's definition, this normative structure is much more apparent. As we saw earlier, they define the agent as the one who is *supposed* to advance the principal's goals. One can see, however, that with this sort of definition, it is not uncontroversial to say that the relationship between managers and shareholders is that of agent to principal (Newton 1992: 100–101; Blair and Stout 1999: 252). Indeed, proponents of normative stakeholder theory would regard it as straightforwardly question-begging to say that the manager is supposed to advance the interests of shareholders, to the exclusion of other constituency groups. Similarly, it is not obvious that employees are agents of their superiors. Workers also depend upon managers to make decisions that will protect their jobs and preserve the value of the firm-specific human capital that they have accumulated (Blair 1999: 67). Or to take a less controversial example, with respect to the management of defined-benefit pension schemes, it is quite clear that employees are the principals, with senior managers of the firm serving as their agents.

Jensen and Meckling define an agency relationship as "a contract under which one or more persons—the principal(s)—engage another person—the agent—to perform some service on their behalf that involves delegating some decision-making authority to the agent" (1976: 308). Yet stakeholder theorists are fond of pointing out that managers have no explicit contract with shareholders, nor do they stand in a fiduciary relationship to them. They have contracts with the firm, and are fiduciaries for the firm (Blair and Stout 1999: 292). The relationship between the firm and its shareholders is in turn very complicated, making it difficult to say that shareholders have "hired" managers, or engaged them "to perform some service on their behalf." The standard response is to say that there is an "implicit" contract in this case (Easterbrook and Fischel 1991: 90–93), but again, that will be disputed by anyone who does not accept the general thrust of the shareholder primacy doctrine. Typically, the sort of implicit contracts that are posited are simply a consequence of the theory of the firm that the person who is doing the inferring happens to subscribe to.

But despite these controversies, none of it adds up to a criticism of agency theory per se. Anyone who tries to map the principal–agent framework onto the relationship between shareholders and management is clearly presupposing the doctrine of shareholder primacy (i.e., the managers ought to serve the interest of shareholders). Thus it would be question-begging to argue that managers should serve the interests of shareholders *because* they are agents of the shareholders. But a theorist could quite easily employ agency theory as a framework for understanding various relationships within the firm without presupposing the doctrine of shareholder primacy. Indeed, one of the central motivations for Blair's team production theory is to bring into sharper focus the problem of managerial opportunism with respect to firm-specific investments made by non-shareholder groups, such as workers. As the reference to "opportunism" suggests, this analysis has a strong agency-theoretic flavor. Agency theory provides a very good characterization of many of the problems that have arisen with defined-benefit pension funds (such as moral hazard problems associated with information asymmetries), all based upon the assumption that employees are the relevant principals. Agency theory has also been employed quite usefully in the analysis of cooperatives, in order to understand some of the "costs of ownership" that are incurred when workers, customers or suppliers take over ownership of the firm (Hansmann 1992: 35–38). Thus agency theory in no way presupposes shareholder primacy. Indeed, it is worth recalling that R. Edward Freeman makes liberal use of agency vocabulary in his work on stakeholder theory. He even introduces an "agency principle" in his Doctrine of Fair Contracts, specifying that "any agent must serve the interests of all stakeholders" (1994: 417; 1998: 134). In his view, the best way to think of stakeholder management is in terms of a set of agency relationships between members of the board of directors and the various constituency groups that have a "stake" in the success of the firm. Thus the connection between agency theory and the doctrine of shareholder primacy is not especially close. Many agency *theorists* are committed to the doctrine of shareholder primacy, but agency *theory* is not.

10.1.3. MISPLACED LOYALTY

Another prominent line of objection to agency theory, this time one that condemns both positive and normative uses of the theory, is based upon the claim that agency relationships, even fiduciary relationships, cannot serve as a genuine source of moral obligation. Under the best of circumstances, they serve only to transmit moral obligations from principals to agents. More often, however, agency relationships are used as an excuse for unethical conduct, as agents seek to avoid responsibility by claiming that they are merely "following orders" or "serving the client." From this perspective, agency theory is nothing but a giant distraction, a way of "passing the buck" when it comes to confronting the problem of unethical behavior in business. Either the agent's action is ethical, in which case the agency relationship has nothing to do with it and the source must be traced back to some obligation imposed upon the principal, or it is unethical, and the agency relationship serves only to obscure that fact, by suggesting that it was done out of "loyalty" or "obligation" to the principal. In both cases, the agency relationship has nothing to do with the moral obligations that individuals are subject to, and so business ethicists gain nothing by focusing upon it.

Kenneth Goodpaster has tried to provide a principled basis for this critique, by introducing what he calls the *nemo dat* principle (1991: 68). The reference is to the Latin term (and legal rule), *nemo dat quod non habet*, or "nobody gives what he doesn't have." Goodpaster uses this to draw attention to the fact that agency relationships are unable to create moral permissions where previously none existed. Principals cannot (ethically) hire someone to do on their behalf what they could not (ethically) do themselves (1991: 68). In a similar vein, Richard DeGeorge takes pains to emphasize that, "acting for another does not give one ethical license," and that "all persons are ethically responsible for their actions, whether performed under command or performed on behalf of another" (1992: 65–66). Yet since the agency relationship cannot be a source of moral permissions, it is then claimed, whether or not managers act as agents of shareholders, or of anyone else for that matter, is a question that is simply lacking in moral significance.

This view does have some prima facie plausibility. It is a well-known feature of conventional morality that, say, promising to help a friend commit a crime does not generate a moral obligation on one's part to commit that crime. To allow this would be to permit the unlimited "laundering" of unethical acts into ethical ones. Yet many people seem to believe that professional roles do permit laundering of this sort. Thus, for example, what might ordinarily be regarded as lying is sometimes presented, not just as permissible, but as morally obligatory, when done by a lawyer who is seeking to advance the interests of a client. Arthur Applbaum draws out the absurd consequences of such a view of role obligations by developing a profile of Sanson, the "executioner of Paris," who carried out his duties with consummate professionalism throughout the final years of the *ancien régime*, the French Revolution, the Terror, and the Thermidor (Applbaum 1999: 16–27). Sanson remained above the fray throughout, insisting that he was merely a loyal agent, carrying out legal executions, and was thus not to be held responsible for any of the excesses committed by one or another of the various principals he had served.

More generally, Applbaum develops a thought-experiment involving two societies, Badland and Roland. Badland is essentially a Hobbesian state of nature, in which each individual pursues his or her self-interest in a purely instrumental fashion, and thus "no one avoids harming another unless there are penalties discouraging such harm, and all craftily engage in manipulation and deception if doing so will advance their ends" (Applbaum, 1999: 7). In Roland, by contrast, "people have the same motivations, but do not pursue their own interests. Rather, each appoints a trustee who pledges to advance the trustor's interests through a blind trust, and each trustor is also a trustee." As a result, in Roland "exactly the same conflicts are fought, the same manipulations occur, the same harms inflicted, but each actor is acting as a faithful professional in fulfillment of obligations to a client" (Applbaum 1999: 8). In what sense, Applbaum then asks, is Roland any better than Badland?

Many business ethicists have seen the relationship between managers and shareholders as essentially equivalent to the relationship between trustees and trustors in Roland. Rather than denying that managers are agents of shareholders (as stakeholder theorists are inclined to do), they simply deny that any such relationship can be a source of moral obligation. Alex Michalos, for instance, in his critique of "the loyal agent's argument," attributes the following view to theorists (like Milton Friedman) who view the obligation of managers toward shareholders as paramount: "As a loyal agent of some principal, I ought to serve his interests as he would serve them himself.... He would serve his own interests in a thoroughly egoistic way. Therefore, as a loyal agent of this principal, I ought to operate in a thoroughly egoistic way on his behalf" (1995: 45 [format altered]). Michalos goes on to criticize this argument, claiming that it adds up to little more than an attempt to "launder" egoism into altruism. More polemically, Lisa Newton has argued that, from the perspective of agency theory, "the entirety of corporate enterprise seems... to be dedicated to the enrichment of the rich, to satisfy the greed of the truly greedy" (1992: 100). Furthermore, "it follows for agency theory that there can be no such thing as corporate responsibility for community welfare, for the community figures nowhere in the principal–agent relationships" (Newton 1992: 101). Thus she refers to the moral framework encouraged by agency analysis as "theory-compelled irresponsibility."

There is, however, some danger of equivocation in the way that this argument is formulated. With respect to agents, it is important to distinguish the deontic modality of permission from that of obligation. Critics of the agency perspective are perfectly correct in noting that agency relations cannot create permissions. This is in fact why theorists who are heavily influenced by the agency perspective, such as Buchanan, are at pains to specify that the moral obligation of managers is to advance the legitimate interests of shareholders (not just any old interests) (Buchanan 1996: 422–423; see also Quinn and Jones 1995: 35–36). Even Friedman qualifies his defense of profit-maximization with the stipulation that shareholders will "generally" want "to make as much money as possible while conforming to the basic rules of the society, both those embodied in law and those embodied in ethical custom" (1970). Thus no one is committing the elementary error of believing that agency relations can turn impermissible conduct into permissible conduct (or wrong into right).

What critics of the agency perspective generally fail to note is that agency relations can serve as a genuine source of moral obligation in one important sense—agency relations can transform actions that are merely permissible for the principal into ones that are obligatory for the agent. This is in fact Applbaum's final observation in *Ethics for Adversaries*. In response to the (rhetorical) question, "Why take professional roles seriously, from the moral point of view?" he replies: "Though roles ordinarily cannot permit what is forbidden, they can require what is permitted" (Applbaum 1999: 259). For example, a person who is accused of a crime, even though he may have done it, is not obliged to plead guilty, but rather is permitted to mount a defense (legally, of course, but perhaps also morally, in cases where the prosecution is seeking an unreasonably harsh sentence). Yet mounting a defense is, for the accused, merely the exercise of a permission (as witnessed by the fact that he is entitled, at any point, to change his mind and enter a guilty plea). For any attorney that he employs, on the other hand, the exercise of this permission generates an obligation to mount that defense.

Thus, from the standpoint of business ethics, if it can be shown that shareholders are merely permitted to claim the residual earnings of the firm, and that managers are their agents, it then follows that managers are obliged to serve them loyally in this regard. This is morally salient, because the relationship creates that moral obligation, by transforming a permission into an obligation. It also means, *inter alia*, that agents will be forbidden to do certain things that are permissible for principals. The *nemo dat* principle is misleading in this regard. When it comes to obligations, principals do in fact "give" that which they do not have. Thus Goodpaster's observation, with respect to the impossibility of creating permissions, does not undermine the significance of agency analysis. Relationships between individuals—particularly fiduciary relationships—are not merely a distraction; they represent a genuine source of moral obligation.

It is important to note that the shareholder's claim on residual earnings need not be "good" in order to generate an obligation on the part of the manager to maximize it. Critics of the "loyalty" argument often appeal to the intuition that the desire on the part of shareholders to "make as much money as possible" (as Friedman put it) is somehow morally dubious. But one need not show that there is anything laudable about the desire for profit in order to demonstrate the importance of agency analysis—one need only demonstrate that profit is morally permissible. This is a much lighter burden of proof, a fact that is sometimes obscured by theorists like Newton, who use abstract terms of condemnation such as "greed" to describe the motives of shareholders. Certain actions may not be morally praiseworthy, but they are not, by virtue of that fact, morally impermissible, and as long as they are permissible they may in turn become obligatory for others.

Thus the normative critique of the agency perspective is based upon a set of conceptual confusions. There does seem to be something wrong with the idea that managers might be morally obliged to maximize the profits of shareholders (or act as "agents for the greedy"). It sounds wrong when one first hears it. But upon closer examination, it turns out to be perfectly defensible, so long as one can show that it is permissible for individuals to seek a return on their savings, and that managers owe some sort of loyalty to the shareholders of the firm.

10.2. What is the Problem?

The preceding discussion has surveyed three potential problems with agency theory, from the perspective of the business ethicist: first, that it treats all motivation as self-interested; second, that it presupposes shareholder primacy; and third, that it encourages violation of the *nemo dat* principle (and thus, evasion of moral responsibility). We have seen, however, that the identification of rational choice theory with self-interest is something of an oversimplification (agency theory leads us to expect opportunism on the part of individuals, but not necessarily self-interest); that agency theory is not committed to the doctrine of shareholder primacy; and finally, that the *nemo dat* principle, correctly understood, does not diminish the moral significance of agency relationships.

Thus the only really important issue outstanding is the first one, having to do with opportunism. How important is it that agency theory downplays the significance of social norms, moral principles, and "intrinsic" motives in explaining human conduct? The standard defense of the agency theorist will be to say that this is all just positive theory, no one is recommending universal opportunism. As an empirical tool for understanding the way organizations function and for explaining various aspects of organizational structure, agency theory has proven its value. Why should that be of concern? Thus the central question for the ethicist becomes: If agency theory is merely a tool used to develop a positive theory of the firm, how much mischief could it really cause in a corporate environment? The answer is: Quite a lot.

The first step to understanding this answer lies in an appreciation of the fact that, because it is based upon a flawed conception of human rationality, agency theory generates predictions that are wildly at variance with what one can actually observe in the behavior of individuals and in the structure of organizations. In other words, it generates a positive theory that, insofar as it is falsifiable, is demonstrably false. Of course, many of the potential problems identified by agency theory are no doubt genuine—this is why the theory resonates with so many people. There is, for example, a notable tendency toward moral hazard. Similarly, individuals have a tendency to act non-cooperatively in collective action problems. Usually, however, these show up only as tendencies, even when game-theoretic analysis predicts universal defection. In particular, while moral hazard in the firm can be a serious problem, empirically it is much less of a problem than any straightforward application of game-theoretic analysis to principal–agent relations would lead one to predict. For example, while employees do sometimes shirk—everyone knows that—most of the time they shirk a lot less than they could, as a matter of fact, get away with.

The empirical limitations of game-theoretic models have, of course, been exhaustively studied and documented by experimental game theorists. It is well-known, for instance, that large numbers of individuals cooperate in one-shot prisoners' dilemmas, knowing full well that there is no possibility of reciprocation. This fairly large-scale deviation from the equilibrium strategy is not a "blip" or an artifact of some particular experimental procedure—cooperation remains stable under a wide variety of conditions: across a wide range of different cultures, among subjects playing for the first time and among those with previous experience, in large and small groups, and with a variety of different monetary rewards (Dawes and Thaler 1988; Isaac, McCue, and Plott 1985; Kim and Walker 1984; Schneider and Pommerehne 1981).

Apart from the prisoner's dilemma, the other game that has been widely studied in experimental settings is the "ultimatum game." Here, one player is given a fixed sum of money and told to propose some division of the money between himself and one other person. The second player can then either accept this proposal, in which case the money is divided up as per the offer, or reject the proposal, in which case both players receive nothing. Of course, the second player never has any positive incentive to reject any offer, since no proposed division is worse than receiving nothing. Thus rejecting the offer is a punitive action—and the threat to do so is precisely the sort of commitment that sequential rationality rules out. As a result, standard game theory suggests that the proposer should select a division that gives the second player as little as possible (in a sense, behaving opportunistically), and that this proposal should always be accepted. In reality, not only do players tend to offer much more than the instrumental analysis predicts, but proposals also tend to be rejected if they fall too low. In industrialized societies, mean offers tend to be around 44 percent, while offers below 20 percent are rejected 40 to 60 percent of the time. Experimental evidence from non-industrialized societies reflects greater variability—including examples of mean offer rates above 50 percent, combined with frequent rejection of such offers. Yet in spite of these variations, no experiment has ever come close to conforming to the expectations of standard game theory (see Henrich et al. 2001).

Given these experimental findings, it would not be surprising to find that agency theory consistently overstates the agency costs that may arise within organizations, simply because real human beings often behave cooperatively, exhibit loyalty, and refrain from acting opportunistically, even in the absence of external incentives. This fact is, of course, well-understood by sophisticated management theorists, even those deeply wedded to the agency perspective. The general upshot of a lot of agency analysis of the firm is that many organizations, especially those that exhibit what Williamson calls "information impactedness" (1973: 318), simply would not function if the only tools that managers had at their disposal were external punishments and rewards. Bengt Holmström (1982) showed very early on how imperfect observability could make it impossible to devise efficient incentive schemes for individuals working in teams. George Baker (1992) and others drew attention to the fact that, when effort or output was not fully observable, a system of sharp incentives focused upon one aspect of the task could produce results that were much worse than a system of dull incentives applied to the task as a whole. Much of the agency literature wound up sounding a very skeptical note on the subject of performance pay and provided unexpected support for the old-fashioned practice of paying employees a flat salary (Gibbons 1998). Results such as these suggested that, insofar as real-world corporations do actually succeed in extracting reasonable levels of cooperative effort from their employees, there must be more than just external incentives at work.

Given these results, one might wonder where the harm could be in business schools teaching agency theory, or in managers using it as an analytic tool. And perhaps there would be no problem, except for the fact that the limitations of the theory are often overlooked or understated. This can lead to mischief in several different ways.

10.2.1. IMPUTED INCENTIVES

People who are overly impressed by economic methodology often subscribe to the instrumental conception of rationality in a form that makes the model essentially unfalsifiable. As a result, when particular agency problems do not show up where agency theory predicts that they should, rather than concluding that there must be some relevant internal motive of deontic constraint at work, these theorists assume that the external incentive must be there, but that it simply has not been discovered yet. Economists have in fact invested extraordinary ingenuity and effort in the task of devising baroque external incentive schemes as a way of explaining phenomena that in fact admit of far more straightforward "internal" explanations. To take just one example, there are two prominent interpretations of the so-called "efficiency wage" phenomenon. Henry Ford set the relevant precedent, by voluntarily increasing the pay of his workers to $5 a day at a time when average wages in the automobile industry were less than half that. He was rewarded with a significant increase in worker productivity (so much so that he later described it as "one of the finest cost-cutting moves we ever made" [Ford 1922: 147]). The common sense explanation would be to suppose that Ford tapped into an underlying norm of reciprocity (see Akerloff 1982; Fehr, Gächter, and Kirchsteiger 1996). According to this perspective, the notion of a "fair day's work for a fair day's pay" plays a powerful role in determining employee effort levels (Hausman and McPherson 1996: 55–56). So when the "boss" agrees to pay you a rate that is, by common admission, far in excess of what he is obliged to pay, he has in essence done you a favor. And since "one good turn deserves another," you then owe it to him to put more effort into your work (or at very least, to refrain from shirking). One might also expect this obligation to be enforced informally in the relations between workers on the shop floor, thus removing an important barrier to observability and leading to a dramatic reduction in moral hazard problems.

It should also be noted that, apart from its common sense appeal, significant empirical evidence supports this "norm of reciprocity" explanation of efficiency wages (Gneezy 2003). Nevertheless, many economists have felt the need to resist this explanation. The more popular suggestion has been that, by paying workers an above-market wage rate, Ford essentially created an economic rent associated with employment at his firm. This made workers more averse to losing their jobs, by making it unlikely that they would find work at comparable wages elsewhere. This, combined with the queues of workers that began to assemble outside Ford's factory looking for work, created enough fear of dismissal to motivate the existing workers to shirk less (Fraser and Waschik 2002: 291). According to this view, the efficiency effects of the wage increase can be explained entirely through reference to traditional monetary incentives, and without appeal to any obscure "internal" motivational factors, such as a sense of fairness or a commitment to reciprocity. (Of course, few people would doubt that the "external" explanation represents a *part* of the story, perhaps even an important part. The question is whether it represents the *entire* story.)

John Boatright has argued that this methodologically induced bias toward explanations in terms of external incentives can have a psychological "framing effect" that, when translated into practical managerial decision-making, "might result in mistaken solutions to problems or even incorrect assessments of the problems to be solved" (1999: 48; see also Dees 1992: 35). For example, the agency perspective "is apt to lead to a distrust of agents and a reliance on mechanisms of control. Such an approach is warranted in certain situations, but when applied in a business setting it may result in an overinvestment in monitoring and other contractual solutions and a corresponding underinvestment in building trust in an organization, and in fostering traits like loyalty and professionalism" (Boatright 1999: 49; see also Frey and Osterloh 2002).

The more important problem, however, arises as a consequence of the assumption that, whenever a particular sort of agency cost fails to arise, there must always be an explanation in terms of external incentives. This can encourage individuals in such "agent" positions to act in a purely instrumental fashion, by leading them to assume that there must already be a system of checks and balances in place to mitigate the negative impact of any opportunistic actions that they take, even if they cannot see it. If they believed, on the other hand, that the situation called for moral restraint on their part, as the only way of avoiding an agency cost or a collective action problem, then they might be less willing to act opportunistically or non-cooperatively. They would certainly be deprived of one powerful rationalization for unethical conduct.

For example, many agency theorists downplay the significance of the "fiduciary" relationship that exists between managers and shareholders (Jensen and Meckling 1976). A fiduciary relationship implies both a "duty of care" and a "duty of loyalty," both concepts that are unintelligible as such within a game-theoretic framework. Thus agency theorists tend to resist taking these obligations at face value, instead choosing to regard them as just legal "shorthand" for a certain set of implicit contracts, ones that are ultimately structured by external incentives.[7] But such an analysis can easily lead those who are in a fiduciary role to take these obligations less seriously, and to act in a more opportunistic fashion, on the grounds that these implicit contracts already anticipate such forms of behavior. When combined with the so-called "efficient markets" hypothesis, which dramatically underplays the information asymmetry between shareholders and managers, the result can be a very straightforward rationalization of unethical conduct.

Robert Clark describes the basis of this rationalization as a "facile optimism about the optimality of existing institutions" (1985: 65). For example, it is common among those who share the "implicit contracts" perspective to regard management "nest-feathering," not as a breach of fiduciary duty, but merely as implicit compensation. Managerial misrepresentation of company accounts (i.e., "loose" accounting standards) is sometimes defended, and opportunistic behavior is excused, on the grounds that it must have already been "implicitly" accounted for. Consider the following argument, made by Lawrence Revsine:

> It is reasonable to presume that those who negotiate managers' employment contracts anticipate such opportunistic behavior and reduce the compensation package accordingly. Notice that the demand for "loose" standards is further increased insofar as managers bear some or all of the agency costs. Since they have already been "charged" for the anticipated opportunistic actions, they must now engage in them in order to achieve the benefits they "paid" for. Since loose standards facilitate opportunistic actions, the demand for such standards increases. (Revsine 1991: 18)

Here the "facile optimism" about efficient contracting is presented, not just in a way that excuses breaches of fiduciary duty, but in a way that actually puts pressure on managers to violate these duties. After all, if a manager has already been "charged" for padding an expense account, in the form of reduced compensation, then he or she would be a fool to refrain from padding it. More generally, any manager who does not take advantage of any and all "opportunities for opportunism" is essentially being suckered.

From an ethical perspective, the impact that such reasoning can have should not be underestimated. The idea that ill-gotten gains are merely implicit compensation is one of the most important "techniques of

---

[7] Frank Easterbrook and Daniel Fischel write that "the fiduciary principle is an alternative to elaborate promises and extra monitoring. It replaces prior supervision with deterrence, much as criminal law uses penalties for bank robbery rather than pat-down searches of everyone entering banks" (1991: 92). Notice how this redescription downplays the element of internal constraint in a fiduciary relation, arguing instead that it involves simply a reshuffling of external incentives.

neutralization" used by white-collar criminals to rationalize—and hence to grant themselves permission to engage in—illegal conduct (J. W. Coleman 1987: 414). Thus one can see in Revsine's argument a clear example of how a false understanding of agency theory and its implications can serve as a powerful impetus toward both immoral and illegal behavior. Of course, there is a sense in which agency theory itself is not to be blamed. Nevertheless, this false understanding is extremely widespread, so the potential for mischief that it creates merits emphasis.

10.2.2. CROWDING OUT OF MORAL INCENTIVES

As we have seen, the methodological biases of agency theory generate an overemphasis on external incentives as a way of addressing agency risks, along with a comparative neglect of internal incentives. Thus an enormous amount of time and energy has been frittered away designing increasingly clever incentive schemes, to the neglect of more obvious strategies for securing employee loyalty and dedication. Yet while this may be a waste of time, one might also be inclined to think that it also can do no harm. Even if an organization depends heavily upon voluntary deontic constraint on the part of its employees in order to avoid certain potential agency problems, surely it cannot hurt to layer on some additional external incentives, in order to create a greater alignment of interests?

Of course, the agency literature itself is full of cautionary examples of how incentive schemes can distort incentives, and thus of how poorly designed incentive schemes can exacerbate agency problems. Yet a more general problem has been almost entirely ignored, namely, that even a well-designed system of external incentives has the potential to undermine moral motivation, and thus to create agent costs where previously none existed (Fehr and Gächter 2002; Tenbrunsel and Messick 1999). This is something that was well-known to previous generations of organizational theorists (e.g., McGregor 1960), but has become so thoroughly sidelined by the rise of agency theory that serious experimental research has been required to reestablish the importance of the basic phenomenon (what is now referred to as the "crowding out" of moral incentives).

Research by Bruno Frey and Felix Oberholzer-Gee (1997) has highlighted some of the ways in which pecuniary incentives can have the effect of undermining moral motivation. In one study, they examined the willingness of citizens to accept NIMBY ("Not In My Backyard") projects, such as nuclear waste disposal sites, in Switzerland. Nuclear power plants produce benefits that are enjoyed quite widely, but impose highly localized costs (such as the dangers associated with waste storage and disposal). This gives local communities an incentive to "free ride"—to use the electrical power, but then refuse to accept either generation or disposal facilities in their region. Frey and Oberholzer-Gee found that in one Swiss village that had been identified by experts as the best disposal site, a slender majority of citizens (50.8 percent) were willing to accept the creation of such a facility in their community (and thus to act "cooperatively"). Yet surprisingly, when officials decided to sweeten the deal by offering an additional monetary payment as compensation (ranging from US$2,175 to $6,525), support for the project plummeted to 24.6 percent.

It is not difficult to imagine what went on. Based upon simple cost-benefit calculation, it is very unlikely that any community would find it in their interest to accept a nuclear waste disposal facility. The value of the power that they (along with everyone else) receive is simply not worth it, especially when there is a reasonable chance that concerted resistance to the project will result in its being located in some other community (i.e., that NIMBY free-riding is a feasible option). Thus monetary compensation is not likely to tip the balance for many people. The only way to get citizens to accept such a facility is through a moral appeal, which might lead them to overlook their self-interest in favor of the "greater good." When considering the project "from the moral point of view," citizens simply do not engage in the relevant cost-benefit calculations. They approach the question from the standpoint of what John Rawls calls "the reasonable," rather than "the rational" (1993: 48–54). Furthermore, their consent may be based upon the fact that they do not enter into these calculations.[8] Offering people external incentives has the effect of changing their perspective, so that they no longer consider the question from the moral point of view, but rather examine it from the standpoint of their self-interest (Tenbrunsel and Messick 1999). If the external incentives are inadequate from this standpoint, then the incentive scheme may easily have the effect of undermining cooperation, thereby creating real collective action problems where previously there were only potential ones.

This phenomenon has been reproduced experimentally in various ways. James Heyman and Dan Ariely (2004) provide a particularly clear illustration. They asked students to perform a somewhat boring task (dragging around circles on a computer screen). One group was paid a flat fee of $5 to participate in the experiment, another was paid a "piece rate" of 10 or 50 cents per circle dragged, and the final group was simply asked to do it as a "favor." Those who were paid 50 cents per circle dragged more than those paid only 10 cents, as an economist would be inclined to predict. However, those who were paid the

[8] This interpretation is suggested by the fact that willingness to accept the nuclear-waste facility was highly correlated with abstract support for nuclear power as a means of electricity generation. Unless individuals are basing their decisions upon a principle, this correlation is difficult to explain, since proponents of nuclear power have the same free-rider incentive as opponents. Furthermore, the impact that the money offer had upon citizens' perceptions of the risk associated with the project was studied and found to be negligible.

flat rate of $5 dragged far more circles than those who were paid a piece rate, while those who were simply asked to do it as a favor dragged the most circles of all (Ariely 2008: 68–69).

"Money," Ariely concludes, "is very often the most expensive way to motivate people" (2008: 84). More importantly, what the experiment suggests is that internal and external incentives are not necessarily complementary or cumulative, even when in theory they are correctly "aligned" to promote the same outcome. In practice they may be mutually antagonistic (e.g., if one were to take the students who were dragging the circles as a favor and start offering them money, one might easily see a decline in performance). Furthermore, there is good reason to think that the type of incentive schemes often promoted by agency theorists for use within corporations have considerable potential to undermine moral motivation. Far from intensifying work effort, the external incentive scheme may simply communicate the message that management does not "trust" workers. One need only recall the way that workers have historically responded to sharp incentives such as piece rates, along with the monitoring systems that are required in order to implement them, to see the consequences this may have.

The general problem is that agency theory has a completely "top-down" focus when it comes to analyzing relationships within the firm. Sanctions flow from the principal, who occupies a higher rank in the organizational hierarchy, down toward the agent, who occupies a subordinate role. It is a purely unilateral and one-sided relationship (Blair 2000: 71–72). Thus the "framing effect" of agency theory tends to encourage essentially Taylorian management practices. Nothing in the agency perspective, for instance, discourages the principal from acting opportunistically with respect to the agent, or even speaks to this problem (Dees 1992: 49).

Moral relations, on the other hand, are based upon trust, and are therefore typically secured through some form of reciprocity. Managers cannot dictate that employees exhibit trust, they must work to cultivate it. The standard way of doing this is to exhibit loyalty and trustworthiness in one's own conduct (Ariely 2008: 79–80). Thus moral incentives usually develop within relations that are mutual and two-sided. These can be extraordinarily difficult to cultivate in an environment in which one party also has unilateral and arbitrary control over the power to punish and reward the other. Thus an organization that seeks to cultivate trust and loyalty will often go out of its way to downplay its hierarchical structure, along with the potential for unilateral action that this creates. The type of incentive schemes that tend to flow from an agency analysis, on the other hand, often create an "ethos" that is highly antagonistic to the development of strong bonds of solidarity. (One can see here the substance of Kulik's [2005] complaint that an overemphasis on performance pay, bonuses, and other "sharp incentives" at Enron created an "agency culture," that in turn eroded the basis for ethical conduct.)

10.2.3. CRYPTONORMATIVISM

No matter how strenuously agency theorists may insist that theirs is only a "positive" theory of the firm, and thus entails no "value judgments," the fact remains that the basic approach has as its foundation a normative theory of practical rationality, one that categorizes certain forms of action as "rational" and certain other forms as "irrational." The fact that morality (or cooperation) gets consistently categorized within such models as irrational, and opportunism (or defection) as rational, might easily lead more impressionable minds to the conclusion that they should learn to ignore moral constraints (Miller 1999). This can have two pernicious consequences. First, in the interests of acting more "rationally," individuals may begin to plan their own behavior in accordance with the dictates of the instrumental model, and thus begin to act more opportunistically. Second, even if they do not change their own deliberative processes, they may begin to expect higher levels of opportunistic behavior from others, and therefore feel justified in engaging in "preemptive" defection in order to protect themselves from the anticipated defection of others. Thus Ronald Duska observes that the instrumental conception of rationality has the potential to become a "self-fulfilling prophecy." "If I think humans are always going to be selfish, and cannot help but be so, it becomes the height of foolishness to sacrifice myself, or to predict their behavior on any other than selfish grounds" (Duska 1992: 149, see also Argyris 1973: 264–266). Yet the type of "I did it to him to prevent him from doing it to me" reasoning that this generates provides another one of the classic techniques of neutralization used to excuse antisocial behavior (see Sykes and Matza 1975: 668).

There is some evidence to support this concern about instrumental rationality becoming a self-fulfilling prophecy (Ferraro, Pfeffer, and Sutton 2005). It was widely reported, for instance, that one of the only significant anomalies discovered in experimental trials of the "public goods" game in North America occurred when the game was played among economics graduate students. There the rate of cooperation fell to only 20 percent, whereas it remained over 40 percent when played by students in other disciplines (Marwell and Ames 1981; also Frank, Gilovich, and Regan 1993). In a series of follow-up questions, students were asked whether a concern over "fairness" played a role in their decisions. Whereas virtually all noneconomists answered yes, "more than one-third of the economists either refused to answer the question regarding what is fair, or gave very complex, uncodable responses.... Those who did respond were much more likely to say that little or no contribution was 'fair'. In addition, the economics graduate students were about half as likely as other subjects to indicate that they were 'concerned with fairness' in making their decisions" (Marwell and Ames 1981: 309).

This is important because, contrary to the widespread conviction that the willingness to act morally is primarily dependent upon ethical character, which in turn is instilled through childhood socialization, empirical studies have generated strong support for the contention that the willingness to act morally is in fact highly situational, and that individuals rely to an exceptional degree upon social cues in their immediate environment in order to determine what to do (Doris 2002). Thus it would be no surprise to discover that a social environment in which the dominant assumption is that "it's every man for himself" is one that would not only encourage unethical behavior, but could become positively criminogenic.

10.3. Agency Theory as Critical Theory

The discussion so far has focused upon the mischief that can be caused by an overly literal use of agency theory as a tool for understanding the relations between individuals within a firm. The problems stem from the model of rational action underlying agency theory, which is not normatively neutral, but results rather in a selective emphasis upon the consequentialist dimension of practical rationality, while ignoring the role of deontic constraint. Thus the use of agency theory as the methodological foundation of a positive theory of the firm tends to produce a highly distorted image of how these organizations function, which can in turn have undesirable effects upon behavior if naively adopted as an accurate account of reality.

This is, however, not the only way to use agency theory. There is a long-standing tradition in political philosophy, dating back most obviously to Thomas Hobbes's *Leviathan*, that uses an instrumental model of rationality as the basis for the development of a normative theory (Heath 1996). Theorists working in this tradition, rather than asserting that individuals always act in a self-interested manner, instead merely pose the question, what if individuals always acted in a purely self-interested manner? The instrumental model is then used as a foundation for a dystopian "state of nature" thought-experiment, which characterizes the condition that society would be in if individuals failed to respect any "internal" or moral constraints in the way that they pursue their objectives. It is not difficult to show that, under such conditions, individuals would become embroiled in insuperable collective action problems. With a little more work (and *pace* Hobbes), it is possible to show that no system of purely external incentives can be created that will resolve these problems (Braybrooke 1976). Thus a general case can be made for the claim that individuals should adopt some form of internal constraint, as the best way of avoiding a life that is "solitary, poor, nasty, brutish and short."

In this tradition of thought, the instrumental conception of rationality is used to construct a cautionary tale. It allows one to state with a great degree of precision what would happen in the absence of morality and other systems of deontic constraint. Agency theory can be (and has been) used in exactly the same way. Thus many business ethicists have drawn upon its results—especially the limitative results, which show how ubiquitous moral hazard problems would be, and how difficult the design of effective incentive schemes would be, in the absence of moral constraint—in order to show that corporations could not even begin to function in the absence of significant moral constraint on the part of employees or managers. Noreen (1988) has developed this insight into a powerful rebuttal of the standard "invisible hand" critique of business ethics, which claims that marketplace competition renders the constraints of morality otiose (Gauthier 1982). As Noreen puts it, "agency theory can be used to provide a series of instructive parables that illustrate the adverse consequences on social and economic systems of unconstrained opportunistic behavior" (1988: 360) and can therefore be used as a way of building the case for ethical conduct in business relations.

According to this perspective, individuals are capable of acting opportunistically, but are also capable of exhibiting restraint. The extent to which they do either is very much dependent upon circumstance, institutional context, and background culture. Agency theory offers a characterization of the dystopian extreme, in which opportunistic conduct is rampant. This provides not only a good reason for wanting to ensure that greater moral restraint is exercised (namely, to achieve a reduction in agency costs), it also provides a good explanation for the competitive advantage certain firms are able to derive from an organizational culture that promotes such restraint. Francis Fukuyama, for example, has developed this analysis as a way of explaining the competitive advantage that family-owned firms often enjoy in the incubation stage of corporate development (1995: 74–80). The fact that family members are able to draw upon preexisting trust relations allows them to avoid all sorts of contracting and agency costs that rival firms must incur. This explains why "social capital"—"the degree to which communities share norms and values and are able to subordinate individual interests to those of larger groups" (Fukuyama 1995: 10; also J. S. Coleman 1988)—is a form of *capital*. It is precisely because it can be drawn upon by individuals in order to avoid agency costs in their organizations, both by reducing agency losses directly and by reducing the need for costly monitoring.

A firm is not just a "privately owned market," as Armen Alchian and Harold Demsetz misleadingly suggested (1972: 795). It constitutes an institutional environment that is internally insulated from the competitive norms of the market and is thereby made more conducive to the emergence of cooperative or high-trust norms. Alchian and Demsetz claim that "telling an employee to type this letter is like telling a grocer to sell me this brand of tuna rather than that brand of bread" (1972: 777). Of course, at a certain level of abstraction anything resembles anything else. This particular claim, however, is not only misleading, but perversely so, because it denies precisely the most important characteristic that distinguishes hierarchies from markets, namely, the fact that firms are able to cultivate internally a set of cooperative norms (including norms establishing obedience to authority) that are specifically suspended in market transactions.

A critical agency perspective is able to explain quite clearly why, as production becomes more knowledge-intensive, successful firms typically move away from external incentives and develop an interest in organizational culture, team-building, and "shared values." It is because agency problems are caused, fundamentally, by information asymmetries. As production becomes more knowledge-intensive, the potential for such problems increases, the difficulty of creating effective external incentives schemes is compounded, and the probability of such schemes "backfiring" increases. Thus firms come to rely more and more upon internal incentives to secure the voluntary cooperation of their workers. This in turn requires treating them less like cogs, and more like partners in the production process. (In this respect, critical agency analysis vindicates several of the fundamental intuitions underlying Peter Drucker's analysis of management as a "liberal art" [2001: 3–13]).

Allen Buchanan has taken this insight one step further, arguing that agency theory not only provides a good argument for business ethics in general, but that the analysis of agency risks provides the key to understanding many of the real-world moral codes that already (implicitly or explicitly) structure activities within bureaucratic organizations. His analysis "derives important features of the ethics of bureaucratic organizations from an understanding of what bureaucratic organizations are like, in particular, from an understanding of what kinds of agency-risks arise within them" (1996: 422). Agency theory tells us where the major stress lines lie within these organizations, where cracks are most likely to appear. The implicit ethical code of the organization is then analyzed as the glue that (to a greater or lesser degree of success) holds things together. Thus, in Buchanan's view, agency analysis provides greater theoretical purchase upon these codes, helping business ethicists to gain a greater appreciation of their deep structure.

Buchanan proposes an ingenious analysis, in which he distinguishes between "first-order" and "second-order" agency risks. The former reflect the possibility of actions that impose costs upon the principal and benefit the agent directly. The latter involve actions that impose costs upon the principal, yet benefit the agent only indirectly, insofar as they make it more difficult for the principal to eliminate first-order agency risks, and thus allow the agent to continue some course of action that is a source of direct benefit. For example, while shirking would be a first-order agency problem, employees may also take actions aimed at frustrating a monitoring system that has been instituted in order to control shirking. In effect, they act to preserve the information asymmetry that creates the first-order moral hazard problem. Insofar as this is costly to the organization, it is a second-order agency problem. In this way, Buchanan is able to explain why individuals in bureaucratic organizations develop a moral allegiance, not just to meritocratic or "work ethic" principles but also to procedures that ensure accountability, proper procedure, and preservation of the "chain of command" (1996: 431).

The agency perspective is similarly useful when it comes to analyzing the obligations of senior managers. While agency theory itself does not presuppose any commitment to the doctrine of shareholder primacy, agency analysis can be used to motivate the suggestion that, in a standard business corporation, managers should bear special fiduciary obligations towards shareholders (Marcoux 2003). Unlike other patron groups, whose interests are protected by contract, shareholders are residual claimants, with only formal control of the firm's board of directors as a mechanism for ensuring that their interests are respected. The potential agency costs are simply much greater in the relationship between management and the firm's owners than they are between the firm and its other constituency groups, whose interests are protected by explicit contracts (Boatright 2006: 113). Thus the agency perspective is able to explain why courts essentially impose a fiduciary obligation upon senior managers to advance the interests of the owners of the firm (and why they limit the ability of the parties to "contract around" this obligation [Clark 1985: 64]). Agency analysis is also able to demonstrate quite clearly why conventional stock options proved so ineffective as a way of creating an external alignment of managerial and shareholder interests (Bebchuk and Fried 2004: 137–140), and it is able to show how the movement of the stock price, combined with the takeover threat, is an extremely blunt instrument for disciplining management (Bebchuk and Fried 2003; Bebchuk, Fried, and Walker 2002; Miller 1992: 171–176). This in turn helps to make the case for the claim that moral restraint on the part of managers—a genuine commitment to serving the shareholder—is an essential element in the proper functioning of the private enterprise system.

When used in this way, far from being a contributing factor to the recent spate of corporate scandals, agency theory proves to be an invaluable tool in understanding what went wrong at these firms. After all, the frauds in question occurred at precisely the points that agency theory identifies as central fault lines (Bebchuk and Fried 2003). To draw an analogy, consider what an agency analysis of the professional role of the doctor would look like. There is no question that doctors should exercise moral restraint in their dealings with colleagues, other medical professionals (nurses, technicians, etc.), patients, and their families. At the same time, an agency analysis is able to identify patients as the class of individuals who are uniquely vulnerable to exploitation in their relationship with the doctor (first and foremost because of the information asymmetries that exist between the two). Thus the case can be made for a special fiduciary obligation on the part of the doctor toward the patient (in the same way that a case can be made for a fiduciary obligation between the manager and the owners of a firm). From this perspective, it would be unsurprising to discover that most of the internal disciplinary proceedings that occur within physician associations involve abuse of patients (and not, for example, colleagues). In the same way, it would be unsurprising to discover that the most common "ethics scandals" in the corporate world involved an abuse of shareholders by management.

Of course, it is important not to think that the moral codes of bureaucratic organizations serve no purpose other than the reduction of agency costs. Steen Thomsen is perhaps being overly optimistic when he describes the moral norms that arise within firms as simply another "governance mechanism," which can be appealed to "when alternative governance mechanisms (pure markets, hierarchies, government, the prevailing social ethic) fail to achieve a social optimum" (2001: 156). This would imply a system of moral constraint containing purely "bottom-up" obligations, a structure that precisely tracked the organizational hierarchy of the firm. (The principles that Buchanan outlines, for instance, focus exclusively on what subordinates owe to their superiors. In part for this reason, Buchanan is at pains to emphasize that his is not a complete conception of business ethics.) An entirely "bottom up" moral code would be in tension with the usual system of reciprocity upon which moral obligations depend. As a result, while agency theory may serve a useful purpose in telling us where ethics is most needed within organizations, morality has its own logic, and so we may not be able to develop an ethics code that is tailored to resolve a precise set of agency problems. There will often be a *quid pro quo*, such that ethical conduct can only be elicited from the agent in one domain if the principal is willing to accept moral constraint in some other, where he or she might have preferred to exercise the freedom to act strategically.

In this context, it is worth emphasizing that the ability of a completely "intrinsic" set of motives to resolve agency problems can also be quite limited. In particular, while the presence of external sanctions can have the effect of undermining moral motivation, the absence of external sanctions can also have the effect of unraveling cooperation (as those who have acted cooperatively in the past become less willing to do so, when they see others defecting with impunity). Thus it is important, when applying the critical agency perspective, to keep in mind the limits and the instability of voluntary cooperative action. The sort of ethics codes typically recommended from the critical agency perspective will usually not be "incentive-compatible," yet that does not give the critical agency theorist license to ignore the incentives that agents face altogether, or to imagine that "ethics" is a magic bullet for resolving agency problems. The fact that a particular institutional arrangement generates an agency problem in theory may not be a problem; but if it has been shown that the arrangement generates the problem in practice, and the parties seem resistant to moral suasion, then it is time to start thinking about legal and institutional remedies. Thus when doing applied ethics, it is important to keep in mind what Rawls called "the strains of commitment" (1999: 154–155).

A lot of problems would go away if people only behaved more ethically, but the fact is, people often do not behave all that ethically. Thus merely urging more ethical behavior upon them, beyond a certain point, no longer counts as offering a solution.

10.4. Conclusion

The preceding discussion has examined two very different strategies for employing agency theory—the positive and the critical—and two very different sorts of objections that have been raised by business ethicists. The use of agency theory brings to the fore two sets of ideas that ethicists have traditionally been very uncomfortable with, first, the economic model of rational action, and second, the doctrine of shareholder primacy and the obligation to maximize profit. With regard to the first, I have suggested that business ethicists have been at least partially justified in their reservations. The economic model is based upon an inadequate conception of rational action, precisely because it classifies an important category of moral action as irrational. Indeed, it classifies all genuine rule-following as irrational and is therefore unsuitable for use as a general theory of rational action. Sophisticated practitioners of agency theory are familiar with these limitations, but a large number of enthusiasts are not. Thus agency theory can serve as a source of considerable inadvertent mischief when treated as an accurate representation of reality. I have therefore encouraged a critical use of agency theory, in which principal–agent analysis is used to provide, not a model of how firms actually work, but rather a set of "instructive parables," allowing us to see more clearly what the world of business would be like in the absence of business ethics.

In this respect, the most important use of agency theory lies in its role in combating the widespread perception that business operates outside the sphere of moral evaluation and constraint. By operationalizing a certain form of moral skepticism—one that denies that there are any genuine moral rules or deontic constraints—game theory in general, and agency theory in particular, shows what the consequences of generalized immorality would be. From this, we can extract both a normative and an empirical lesson: first, that the consequences would be unappealing, insofar as it would lead to the collapse of many mutually beneficial forms of cooperation; and second, that people are a lot more "moral"—even in the world of business—than we are sometimes inclined to believe. Agency theory allows us to see that in many cases, the alternative to ethical business enterprises is not the presence of unethical business enterprises, but rather the absence of any enterprise at all.

With regard to the doctrine of shareholder primacy, and the extent to which agency theory encourages this perspective, I have tried to emphasize that no simple connection exists between the two sets of ideas. Nevertheless, when employed cautiously, with due attention to the institutional context in which the firm operates, it is possible to use agency theory as the basis for a plausible shareholder-focused conception of business ethics. Agency theory can be used to show how the owners of a firm are in a uniquely vulnerable position with respect to the manager, and therefore why a fiduciary relation is justifiable in this case. So while a commitment to agency analysis neither presupposes nor entails a commitment to the doctrine of shareholder primacy, the gain in conceptual clarity afforded by the agency perspective does provide a powerful source of arguments in favor of that doctrine. This does not mean, however, that one could not use the agency framework to make the opposite claim, and it is certainly the case that many stakeholder theorists have sought to articulate and clarify their moral ideas using this framework.

11 }

Business Ethics and Moral Motivation

A CRIMINOLOGICAL PERSPECTIVE

One of the peculiar features of business ethics, as compared to other domains of applied ethics, is that it deals with a domain of human affairs that is afflicted by serious criminality, and an institutional environment that is in many cases demonstrably criminogenic (Braithwaite 1989: 128–129; J. W. Coleman 1989: 6–8; Leonard and Weber 1970; Sutherland 1968: 59). The oddity of this state of affairs is sometimes lost on practitioners in the field. It is common, for instance, at business ethics conferences for the majority of presentations to be concerned, not with ethical issues in the narrow sense of the term (where there is often some question as to where the correct course of action lies), but with straightforward criminality. In this respect, all the talk of "ethics scandals" in the early years of the twenty-first century has been very misleading, since what really took place at corporations like Enron, Worldcom, Parmalat, and elsewhere was, first and foremost, an outbreak of high-level, large-scale white-collar crime. Each illegal act was no doubt surrounded by a broad penumbral region of unethical conduct, yet in each case the core actions all involved a failure to respect the law.

The high incidence of crime in the corporate environment is, in itself, something of a mysterious phenomenon. Most well-adjusted adults would never consider shoplifting from their local grocery store, or stealing from their neighbor's backyard, despite having ample opportunity to do so. Yet according to a US Chamber of Commerce Study, 75% of individuals steal from their employer at some time or other (McGurn 1988). Studies of supermarket and restaurant employees found that 42 and 60 percent (respectively) admitted to stealing from their employer in the past six months (Boye and Jones 1997; Hollinger et al. 1992). The losses suffered as a result of this sort of "occupational crime"—crime committed by individuals against the corporation—greatly exceed the total economic losses suffered from all street crime combined (Snyder and Blair 1989). Yet this does not even begin to take into consideration the losses suffered from "corporate crime"—crimes committed by individuals on behalf of the corporation. During the 1990s the list of firms that were convicted of serious criminal offenses in the United States included (either the parent, a division, or a subsidiary of) BASF, Exxon, Pfizer, Banker's Trust, Teledyne, IBM, Hyundai, Sears, Eastman Kodak, Royal Caribbean Cruises, Litton, General Electric, Chevron, Unisys, ALCOA, Tyson Foods, Bristol-Meyers Squibb, and Mitsubishi (Mokhiber 2006).

The phenomenon of white-collar crime clearly casts a long shadow over discussions in business ethics. One of the most important effects has been the development of a strong emphasis upon questions of moral motivation within the field. In many domains of applied ethics, such as bioethics, it is often not clear what the right thing to do is. In business ethics, on the other hand, there is often no real dispute about the content of our moral obligations (i.e., what we should be doing), the question is rather how to motivate people to do it. The moral rules, in other words, are often quite platitudinous (e.g., don't lie, don't cheat, don't steal) and, within a given culture or society, typically coincide with legal rules. The tough questions arise at the level of compliance: what to do when a rival firm gains competitive advantage through deception, or when a supervisor orders sensitive documents to be destroyed, or even when ethical behavior simply conflicts with the bottom line (Stark 1993). As a result, business ethicists have exhibited considerable concern over the relationship between moral obligation and self-interest, whether it be in discussions of agency theory (Bowie and Freeman 1992), the question of whether "ethics pays" (Vogel 2005; Webley and More 2003), or even debates over how (or whether) business ethics should be taught (Williams and Dewett 2005).

Criminologists also have a long-standing preoccupation with motivational questions, in part because crime prevention is such a major component of their professional mandate. Considerable resources have been dedicated to the task of studying the causes of crime, and a sophisticated body of research has emerged. Given that business ethicists have cognate interests, one might expect that this research would serve as an important source of information and inspiration. Unfortunately, this resource has barely begun to be tapped. For example, instead of speculating about the motives of those who steal from their employers, business ethicists could consult Cressey's (1953) classic study *Other People's Money*, which featured extensive interviews with incarcerated embezzlers. Yet Cressey's study, a staple of the criminology literature, has been cited exactly once in the twenty-five-year history of the *Journal of Business Ethics* (less often than the 1991 Danny DeVito film of the same name).[1] This is unfortunate, since criminologists are practically unanimous in rejecting several of the more popular "folk" theories about what motivates people to

[1] The one article that cites it is Chan 2003, although Cressey's name is misspelled.

commit crimes. Yet many of these same theories continue to thrive in the business ethics literature as explanations for unethical behavior.

In this article, I will attempt to lead by example, by showing how a criminological perspective can help to illuminate some of the questions about moral motivation that have often troubled business ethicists. I will begin by explaining why criminologists almost unanimously reject three of the folk theories often proposed as explanations for white-collar crime: first, that criminals suffer some defect of character; second, that they suffer from an excess of greed; or third, that they "don't know right from wrong." I will then go on to discuss a theory that is widely accepted among criminologists, involving what are referred to as "techniques of neutralization." One of the most noteworthy features of this theory is that it is far more cognitivist than any of the folk theories—it suggests that the way people think about their actions and the situation has an enormous amount to do with their propensity to commit various crimes. I conclude by considering some of the positive conclusions that business ethicists can draw from this (including some important implications for the way that business ethics is taught).

11.1. Folk Theories of Motivation

I have spoken so far as though there were a single, unified, "criminological perspective" on the subject of white-collar crime. This is, of course, an exaggeration. Criminologists disagree with one another just as heartily as specialists in any other academic discipline, and the field of study is divided into a number of rival schools of thought (e.g., see Jones 2005). Nevertheless, there are also a number of very broad presuppositions that are widely shared within the discipline, but which may be counterintuitive to outsiders. They constitute a set of very general ideas and approaches that are mastered during early education in the field and are subsequently taken for granted. It is these general ideas that are largely uncontroversial among criminologists and make up what I am referring to as the "criminological perspective."

The first feature of the criminological perspective is that it takes as its point of departure an inversion of the everyday question that people tend to ask about crime. Picking up the morning newspaper, reading about some egregious offense, we naturally ask ourselves, "Why do people do such things?" Yet what the criminologist regards as mysterious is not the fact that some people commit crimes, but rather the fact that more people do not commit more crimes more often. This is because, when looked at from the standpoint of individual incentives, only a tiny percentage of those who could advance their interests through criminal activity actually choose to do so. Even though illegal activity is punished, the legal system typically fails to supply adequate external incentives for compliance—the chances of apprehension are remote, and the threat of punishment is highly attenuated. Thus, what the criminologist needs to ask first is "Why do people *not* commit crimes?" Only once this question has been answered can one go on to deal with the exceptions.

The standard solution to this problem is to point out some type of socialization process that individuals undergo, in the passage from childhood to membership in adult society, which aligns individual preferences with social expectations in such a way that individuals acquire a desire to comply with institutional norms. According to Talcott Parsons, this coincidence of self-interest and role expectations is "the hallmark of institutionalization" (Parsons et al. 1961: 76). Parsons used the term *deviance* in a technical sense to refer to "a process of motivated action, on the part of the actor who has unquestionably had a full opportunity to learn the requisite orientations, tending to deviate from the complementary expectations of conformity with common standards so far as these are relevant to the definition of his role" (Parsons 1951: 206). Deviance in turn evokes various "mechanisms of social control" aimed at "motivating actors to abandon their deviance and resume conformity" (i.e., restoring full institutionalization). The most significant mechanism is the imposition of external sanctions. These work to bring about a greater alignment of self-interest and social expectations, not only by realigning external incentives in such a way as to encourage conformity but also, when "internalized" by the subject, by socializing the individual in such a way that his preferences become less antisocial.

This analysis, which was enormously influential in early American sociology (and by extension, criminology), has a number of noteworthy consequences. The first is that it defines crime as a type of deviance (Parsons et al. 1961: 869–871), rather than as a simple failure of mechanism design. Thus the attempt to understand the sources of crime focuses upon failures of socialization and failures of social control—failures that are, of course, interdependent, since the primary mechanism of social control (external sanctions) also has a socializing function. This perspective also suggests that "moral" and "legal" norms within a particular society be viewed on a continuum, with the primary difference being merely that the former are enforced through what are, to varying degrees, informal social sanctions, whereas the latter are enforced using the power of the state.

This is the very general theoretical framework presupposed by the overwhelming majority of criminologists. Even so-called "rational choice" approaches to criminology are based upon variants of this view (Akers 1990). Beyond this, however, things get complicated. Applying this framework to the explanation of crime turns out to be more difficult than initially imagined, and a lot of early speculation about the causes of crime turned out to be false. Crime is widely understood to represent some form of deviance, but it is not entirely clear in many cases where the deviance lies. Naturally, before inquiring into the causes of crime, the first step must be to determine what precise form of deviance is involved. Here, it turns out that many of the traditional folk theories of criminal motivation are unsupported by the evidence. Three in particular have been debunked.

11.1.1. CHARACTER

It is widely believed among members of the public that criminal deviance is due to some failure of primary socialization. According to this folk view, criminals "lack conscience," are "sociopathic," or else possess some other character flaw that leaves them lacking the disposition to "do the right thing." Thus criminal conduct is explained as a consequence of some defect in the individual criminal's personality structure.

The problem with this theory is that it overgeneralizes in a way that is unsupported by the evidence (J. W. Coleman 1989: 202–204). Failures of socialization do, of course, occur, and sociopathy is a genuine phenomenon. However, the overwhelming majority of criminals suffer from neither. Indeed, it is precisely the ordinariness of white-collar criminals that led to a serious rethinking among criminologists in the first half of the twentieth century of the Victorian view of criminality, which regarded offenders as either genetically or psychologically inferior. As Edwin Sutherland noted, "businessmen are generally not poor, are not feebleminded, do not lack organized recreational facilities, and do not suffer from the other social and personal pathologies" (1968: 58). A certain percentage of white-collar criminals may be more egocentric and reckless than the norm, but almost all fall within the range of what is considered psychologically normal. Furthermore, an equally large number are simply "muddled" or "incompetent" (Spencer 1965: 261). There is no particular psychological trait that they all share, nor is there any trait or set of traits that set them apart in any significant way from the general population.

Indeed, the tendency to overestimate the effect of "character" upon action is an extremely pervasive error, which afflicts many of our folk theories of social interaction (Ross and Nisbett 1991; Wilson 2002: 207). The evidence of this is quite powerful. Consider, for example, the "Panalba" case, involving the pharmaceutical company Upjohn. After strong medical evidence emerged that the drug was causing a number of serious side-effects (including unnecessary deaths) and that it offered no medical benefits beyond those that could be obtained from other products on the market, the board of directors of the firm decided not only to continue marketing and selling the drug but also arranged to have a judge issue an injunction to stop the FDA from taking regulatory action (Mintz 1969). When the FDA finally succeeded in having the drug banned in the United States, the firm continued to sell it in foreign markets.

When this story is presented as a case history, respondents are almost unanimous in their conviction that the actions of the Upjohn board were "socially irresponsible" (Armstrong 1977). Attitude surveys also show that respondents in the United States regard executives who allow their firm to sell a drug with undisclosed harmful side-effects as having committed a serious criminal offense, second only to murder and rape in severity (Scott and Al-Thakeb 1997). However, when management and executive training students were put in a role-playing scenario (as members of a corporate board, faced with the same decision that confronted Upjohn), 79 percent chose the "highly irresponsible" option, of not only continuing with sales of the drug, but also taking action to prevent government regulation. The other 21 percent chose to continue selling the drug for as long as possible, only without trying to interfere with the regulatory process. Thus the range of behavior extended from "highly" to "moderately" irresponsible. Not one group chose the "socially responsible" action of voluntarily withdrawing the drug from the market (Armstrong 1977: 200). These results were obtained from ninety-one different trials of the experiment in ten different countries (Armstrong 1977: 197).

It is worth noting that Scott Armstrong, the investigator who conducted these studies, initiated them because he was puzzled by the Upjohn case and believed that his own students at the Wharton School of Management could not possibly do such a thing (Hilts 2003). Unfortunately, it was his own students who became the first group to disprove this hypothesis. Anyone familiar with Stanley Milgram's (1974) experiments would be unlikely to find this surprising. What Milgram had shown, and what subsequent studies have shown again and again, is that perfectly ordinary people are able to commit very serious crimes or moral offenses when put in the right situation. The celebrated Stanford prison camp experiment (Haney et al. 1973) taught very much the same lesson.

This is not a finding that is specific to criminology. Social psychologists have accumulated considerable evidence to show that our folk theories of character have little or no predictive value when it comes to determining the probability of "moral" versus "immoral" conduct, whereas situational factors are extremely important. In one particularly noteworthy experiment, students at the Princeton Theological Seminary were told that they needed to report to a building across campus in order to do a presentation. Some were told that they were running late, others that they were just on time, and some that they were a bit early. The experiment was designed, so that, on the way, they would pass a stranger in need of assistance. Of those who were told that they were late, only 10% stopped to help, versus 45% of those who were on time, and 63% of those who were early (Darley and Batson 1973: 105). Other studies in a similar vein have shown quite clearly that situational factors far outweigh the effects of character when it comes to determining behavior (Doris 2002: 30–60).

Yet despite the absence of evidence, the belief that criminals possess a deviant psychology or personality structure is remarkably persistent. Some have suggested that this is because the belief serves as a source of reassurance to the non-criminal segment of the population. As James William Coleman writes:

> The public tends to see criminals as a breed apart from "normal" men and women. The deviants among us are commonly branded as insane, inadequate, immoral, impulsive, egocentric, or with any one of a hundred other epithets. In seeing the deviant as a wholly different kind of person from ourselves, we bolster our self-esteem and help repress the fear that under the right circumstances we, too, might violate the same taboos. But this system of facile psychological determinism collapses when applied to white collar criminals. The embezzling accountant or the corporate functionary serving in an employer's illegal schemes conforms too closely to the middle-class ideals of American culture to be so easily dismissed. (1989: 200–201)

The idea that criminals suffer from some sort of character defect also serves the important function of absolving many institutions of any responsibility for the conduct of their members. According to the popular view, respect for social expectations, whether legal or moral, is something that is taught primarily in the home, cultivated through appropriate child-rearing techniques. As philosopher Michael Levin put it: "Moral behavior is the product of training, not reflection. As Aristotle stressed thousands of years ago, you get a good adult by habituating a good child to do the right thing" (Levin 1989: A23). He goes on to conclude that ethics courses in law schools, medical schools, business schools, and even high schools, are an "utterly pointless exercise," simply because students are fully socialized by the time they get to these institutions, and so it is too late for educators to do anything about their character.

It follows from this analysis that institutions of higher learning cannot be blamed for the conduct of their students. While Dean of the Sloan School of Management, Lester Thurow argued that business schools should be absolved of any responsibility for the unethical or illegal actions of their graduates. His argument was based upon a variation of the "garbage-in garbage-out" principle. "Business students come to us from society. If they haven't been taught ethics by their families, their clergymen, and their elementary and secondary schools . . . there is very little we can do. Injunctions to 'be good' don't sway young men and women in their mid- to late 20's. In the final analysis, what we produce is no worse than what we get" (Thurow 1987: 25). The assumption is that the way people think about their decisions is unimportant, and thus students have nothing to be taught about the moral or legal challenges that may arise in a business context. Students are programmed during early childhood to be either "good boys and girls" or bad ones. What they are subsequently taught about the ways of the world, over the course of their education, is taken to be irrelevant. Yet, this moral psychology is false (as thoroughly discredited as Aristotle's views on physics and biology). The fact that such ideas continue to circulate in the public sphere—the fact that they exercise influence in a various public policy debates—should be a source of considerable consternation.

11.1.2. GREED

There is no doubt that the vast majority of white-collar crime is motivated by what might broadly be referred to as pecuniary incentives. Typically, individuals who commit occupational crimes are seeking to enrich themselves personally, just as firms engaged in corporate crime aspire to improve their financial performance. In addition, of course, since most people prefer more money to less, there is a temptation to assume that this basic incentive is what underpins criminal conduct. Naturally, the mere presence of a pecuniary incentive is not sufficient to explain criminal conduct, since the vast majority of individuals confront such incentives on a regular basis and yet do not avail themselves of the opportunity to commit crimes. This is where greed comes in. While everyone likes money, some people seem to like it more intensely than others. Thus it may be tempting to conclude that, in the case of white-collar criminals, the intensity of their passion for money simply outweighs the various incentives that encourage respect for the law.[2]

There are many problems with this explanation. First of all, it should be noted that it does very little to explain corporate crime. Employees often break the law in ways that enhance the profits of the firm, but which generate very little personal benefit for themselves. There is an important difference, for instance, between the crimes committed at Enron by Andrew Fastow, who secretly enriched himself at the expense of the firm, and those committed by Kenneth Lay and Jeffrey Skilling, who for the most part acted in ways that enriched the firm, and themselves only indirectly (via the high stock price). Loose talk about "greed" in the corporate setting often obscures the crucial distinction between enhancing one's own compensation and enhancing the earnings of the firm. In the latter case, most of the money goes to other people, not to the lawbreaker, and thus greed—at least of the conventional sort—cannot be the primary explanation.

Greed offers a more plausible explanation for occupational crime, but even here the picture is quite complicated. Often it is not the desire for gain that motivates white-collar criminals, but rather a strong aversion to losses (there is a well-documented asymmetry in behavioral psychology between the way that individuals treat losses and gains [Tversky and Kahneman 1991]). This is

[2] This might be thought of as a defect of character, and thus merely a special case of the previous folk theory. Yet there are ways of construing the underlying moral psychology that are not committed to a "virtue ethics" framework. This, combined with the frequency of appeal to this motive, justifies giving it a separate treatment.

reflected in the fact that crime seems to be more prevalent in firms that are doing poorly than in firms that are doing well (J. W. Coleman 1989: 230–231; Lane 1953). Many white-collar criminals are certainly individuals who find themselves financially "squeezed" in some way (Cressey 1950: 742–743). In such cases, it appears to be fear or anxiety rather than greed that is the dominant motive. Yet another fair proportion of crime appears to be related to "rising expectations," when actual gains fall somewhat short of anticipated ones. In this case again, it is not exactly greed that is doing the work, but rather a sense of entitlement that develops and is subsequently disappointed.

These incentives are all very commonplace—indeed, they are too commonplace to serve as a useful explanation for criminal behavior. As Sutherland and Cressey argue, "though criminal behavior is an expression of general needs and values, it is not explained by those general needs and values, since non-criminal behavior is an expression of the same needs and values" (1978: 82). In other words, if greed combined with opportunity really caused crime in any significant sense, then there would be a lot more crime, simply because greed is ubiquitous as a human motive and the world is rife with opportunity.

Finally, it is worth noting that the "bigger" occupational crimes tend to be committed by individuals who are further up the chain of command in the firm (Weisburd et al. 1991). In part this is due to the structure of opportunities—low-level employees tend to commit less serious crimes, simply because they are not trusted with large sums of money, their work is more closely supervised, etc. Yet, if money is subject to diminishing returns, as economists typically suppose, then it is often unclear what motivates managers, many of whom are already quite wealthy, to risk everything just to gain a relatively marginal increase in income. As Coleman has observed, "criminal activities are surprisingly common among elite groups that might be thought to have little to gain from such behavior" (J. W. Coleman 1989: 243). It is also unclear why greed motivates them to commit crimes in this one particular domain of life, but does not impel them toward crime in other areas (e.g., ordinary street crime).

Indeed, one of the reasons that we ascribe an excess of greed to white-collar criminals is that we often find their motives to be inscrutable. Large numbers of offenses are clearly committed by individuals who are wealthy beyond the dreams of avarice. To the average person, the reasons these people have for stealing seem as obscure as, say, the motive that Hugh Grant had for marital infidelity. The ascription of "greed," in such cases, far from constituting an explanation for their conduct, signals rather the absence of any plausible explanatory hypothesis.

11.1.3. VALUES

One of the characteristics shared by the previous two folk theories of criminality is that they focus entirely upon the propensity of individuals, acting as individuals, to commit crimes. Yet, white-collar crime, just like street crime, has an important social dimension. If the individualistic approach were correct, then one would expect to find a fairly random distribution of white collar crime throughout various sectors of the economy, depending upon where individuals suffering from poor character or an excess of greed wound up working. Yet, what one finds instead are very high concentrations of criminal activity in particular sectors of the economy. Furthermore, these pockets of crime often persist quite stubbornly over time, despite a complete changeover in the personnel involved. For example, the petrochemical, automobile, and pharmaceutical industries have been plagued by corporate crime for years, in a way that, for example, the farm equipment or the beverage industries have not (Clinard and Yeager 1980: 340–341). Of course, some of this can be explained by the structure of opportunities in certain occupations (as with theft by dockworkers or corruption among police officers), but much of it also has to do with the formation of deviant or criminal subcultures, often with their own internal rules and normative expectations, which in turn get reproduced over time (Mars 1982).

It is precisely this observation that led Sutherland (who coined the term "white-collar crime" and did the pioneering research on the subject) to posit his "associational" theory of white-collar crime (1949). He basically treated crime as a form of learned behavior, acquired through contact and observation of the activities of other criminals. This theory has a number of defects, including the fact that, stated baldly, the explanation is regressive (who did those other criminals learn from?), but what matters for our purposes are not the merits of the theory but rather the motive that Sutherland had for proposing it. His goal was to account for the contagion-like pattern exhibited by these criminal offenses. It is precisely this pattern that overly individualistic explanations fail to account for.

One popular strategy for attempting to explain the social dimension of criminal activity is to imagine that these deviant subcultures have essentially the same internal structure as the dominant society, but that their members adhere to a different set of values, one that is not shared by those outside the group (Braithwaite 1989: 21–24; Cohen 1955). According to this view, the mechanism that produces "criminal" conduct within the subculture is the same as the mechanism that produces "law-abiding" conduct in the broader culture, namely, conformity to some set of shared expectations. The reason that the former is "criminal" while the latter is not is simply that the two groups have different values—what one calls "good" the other calls "bad," and vice versa. (So-called "labeling theory," which argues that crime is essentially an artifact of the power that dominant groups have to define certain forms of conduct as deviant, is a variation on this view.)

This sort of thinking is quite widespread. For example, after the Haditha massacre in Iraq, the US Marine Corps ordered new "core values" training for all soldiers. The senior officer in Iraq explained that although most soldiers "perform their jobs magnificently every day... there are a few individuals who sometimes choose the wrong path." In order to correct the problem, he said, "It is important that we take time to reflect on the values that separate us from our enemies" (Stout 2006).

The problem of soldiers "choosing the wrong path," by attacking unarmed civilians is a good example of criminal deviance. The way that the Marine Corps chose to render this choice intelligible was by interpreting it as the adoption, on the part of these soldiers, of a deviant set of values, namely, those of the "enemy." Thus the way to solve the problem, in their view, was to reaffirm amongst all a commitment to the official "values" of the organization. Yet, one need only think about this analysis for a moment to see that it constitutes a highly dubious explanation for the conduct in question. How plausible is it to suppose that a group of American soldiers got together and decided that there was in fact nothing wrong with terrorism (i.e., the intentional targeting of civilians), and that this change in value-commitment caused their subsequent conduct?

Criminologists give very little credence to such explanations. Research on juvenile delinquents, in particular, has shown that young offenders typically do not reject the values of mainstream society, nor do they endorse any rival system of group-specific values. "Even serious repeat delinquents mostly place higher value on conventional accomplishments than on success at breaking the law" (Braithwaite 1989: 23). They tend to partake of the same normative consensus as every other member of mainstream society: they share the same role models (e.g., "a humble, pious mother or a forgiving, upright priest" [Sykes and Matza 1957: 665]), they approve of the same standards of behavior, and so on. In other words, there is no fundamental disagreement about what is right and wrong between the majority of those who do and those who do not commit crimes. It is precisely because delinquents recognize the "wrongness" of their behavior, at some level, that they usually draw a distinction between those who are legitimate targets of crime ("fair game") and those who are not (Sykes and Matza 1957: 665).

11.2. Techniques of Neutralization

There is no question that crime involves some form of social deviance. The question that has preoccupied criminologists is "What sort of deviance?"—or more specifically, "Where exactly does the breakdown in social order occur?" While there is still considerable controversy over the correct answer to these questions, several incorrect answers have been rejected with near-unanimity. As we have seen (and contrary to popular wisdom), crime does not primarily involve a defect of character, it is not simply a matter of incentive or opportunity, and it does not reflect a rejection of society's basic moral principles. Indeed, the central question that has preoccupied criminologists for the past century, especially with regard to white-collar crime, has been "Why do psychologically normal individuals, who share the conventional value-consensus of the society in which they live, sometimes take advantage of opportunities to engage in criminal conduct?"

One way to find out why people commit crimes is to ask them. Of course, criminals can hardly be expected to have the last word on the subject, but it does seem reasonable to give them at least the first word. When criminologists did begin talking to criminals about their crimes, some interesting things turned up. One of the most noteworthy was the extent to which criminals rationalize their actions. Cressey (1953), for instance, was struck by the number of convicted embezzlers who claimed to be merely "borrowing" the money, with every intention of repaying it. Sutherland noted that one of the things criminals pick up through "differential association" are "definitions favorable to the violation of law" (Sutherland and Cressey 1978: 81), in other words, ways of describing their actions that made them seem less wrong. Gilbert Geis, studying the major antitrust case brought against heavy electrical equipment manufacturers in 1961, drew particular attention to the number of defendants who "took the line that their behavior, while technically criminal, had really served a worthwhile purpose by 'stabilizing prices' " (1968: 108).

Cressey referred to such euphemisms as "vocabularies of adjustment," which allowed the criminal to minimize the apparent conflict between his or her behavior and the prevailing normative consensus. Criminologists had traditionally described these as rationalizations, used after the fact to protect the individual from blame. Sykes and Matza (1957), however, suggested that this sort of reasoning often preceded the action as well, constituting a mechanism through which the criminal, in effect, gave himself permission to violate the law. Thus, they claimed that much of delinquency involved, not deviancy with respect to primary values, but rather a deviant use of what were, in principle, legitimate excuses for crime.[3] Through these excuses, "social controls that serve to check or inhibit deviant motivational patterns are rendered inoperative, and the individual is freed to engage in delinquency without serious damage to his self image" (Sykes and Matza 1957). Thus they referred to them as "techniques of neutralization." According to Sykes and Matza, "much delinquency is based

[3] I am tacitly introducing the distinction between excuses and justifications into this discussion (see Baron 2005). To justify an action is to show that it is, in some sense, the "right" thing to do. To excuse an action, on the other hand, is to grant that it is, in some sense, the "wrong" thing to do, but to claim that the individual cannot be blamed for performing it under the circumstances (Ripstein 1998). Sykes and Matza use only the vocabulary of "justification," but most of the patterns of reasoning they discuss are better understood as excuses.

on what is essentially an unrecognized extension of defense to crimes, in the form of justifications for deviance that are seen as valid by the delinquent but not by the legal system or society at large" (1957: 666).

Sykes and Matza draw attention to five categories of neutralization techniques, used by offenders to deny the criminality of their actions. It is important to note that each appeals to a consideration that, in some cases, provides the basis for a legitimate excuse. What distinguishes the criminal is the tendency to make overly generous or self-serving use of them.[4]

> *Denial of responsibility:* The offender here claims that one or more of the conditions of responsible agency were not met: that the action or its consequences were unintentional; that he was drunk, insane, provoked, or otherwise unable to think clearly while performing it; that he had "no choice" but to do it, and thus acted out of necessity; that it was all an accident, etc.
>
> *Denial of injury:* The offender seeks to minimize or deny the harm done, e.g., by claiming that an assault was merely intended to frighten, that stolen money was merely borrowed (or the victim too rich to notice it missing). Overly generous applications of the *volenti non fit iniuria* principle also fall into this category (the claim that the victim's consent negates the injury).
>
> *Denial of the victim:* The offender acknowledges the injury, but claims that the victim is unworthy of concern because, in some sense, he deserved it. Thus the crime is portrayed as retaliation for some offense committed by the victim (or a preemptive strike, to stave off an attack), for example, vandalism is portrayed as "revenge on an unfair teacher," thefts are excused on the grounds that the storekeeper is "crooked" (Sykes and Matza 1957: 668). Attacks on stigmatized minorities are also often justified in this way.
>
> *Condemnation of the condemners:* The offender attempts to "turn back" the charges by impugning the motives of those who condemn his actions. Thus the police are criticized for being corrupt, singling him out unfairly, prosecuting him out of malice, racism, stupidity, etc. It is sometimes suggested that it is morally unacceptable for one individual to be punished for an offense, when not everyone who has committed the same offense is punished.
>
> *Appeal to higher loyalties:* The offender denies that the act was motivated by self-interest, claiming that it was instead done out of obedience to

[4] I use masculine pronouns throughout, in reflection of the fact that the overwhelming majority of criminals—both white collar and blue collar—are men.

some moral obligation (that conflicted with the law). These obligations often have a highly particularistic character, such as loyalty to friends, family, or fellow gang-members. Offenders might also claim to have been acting for political motives, and thus characterize their behavior as a form of dissent or civil disobedience.

I have interpreted the above categories quite broadly, in order to subsume some subsequent proposals for addition to the list (e.g., Minor 1981). However, two additional techniques proposed by other authors are sufficiently different that they deserve categories of their own.

Everyone else is doing it: This is to be distinguished from cases in which the offender uses the fact that others violate the law, and yet escape prosecution, in order to condemn the condemners, or uses the fact that others break the law to show that he had "no choice" but to follow suit, and thus was acting out of necessity. In some cases, the mere fact that others are breaking the law is used to suggest that it is unreasonable for society to expect compliance. An appeal to the fact of widespread violation may also be used to remove the moral stigma associated with an offense. In either case, the goal is to show that the law is out of touch with social expectations, and therefore that enforcement is illegitimate.

Claim to entitlement: The offender may claim an entitlement to act as he did, either because he was subject to a moral obligation, or because of some misdeed perpetrated by the victim. He may, however, grant that his motive was self-interested, and yet still claim an entitlement to the act, simply by denying the authority of the law (J. W. Coleman 1989: 213). An offender may argue, for instance, that he was acting "within his rights" and that the legal prohibition of his conduct constituted unjust or unnecessary interference. Certain offenders also appeal to a more "karmic" version of this argument, claiming that their good behavior on past occasions gives them an entitlement to act badly in this one respect (Klockars 1974).

The important thing about the use of excuses is that they allow the delinquent to "have his cake and eat it too," by retaining allegiance to the dominant system of norms and values, while at the same time exempting his own actions from its imperatives, thereby freeing him to pursue his self-interest in a relatively unconstrained fashion (Sykes and Matza 1957: 667). In many cases, a cognitive norm will be violated (e.g., "stealing" is described as "borrowing"), in such a way as to allow the offender to claim that he was in compliance with a more heavily weighted moral or legal norm (e.g., "don't steal").

Consider, for example, the following letter, which was sent to two researchers investigating the use of neutralization techniques by hunters cited for illegal possession of game in the state of Colorado. In a cover letter accompanying the survey, the researchers used the term "poaching" to describe the offense. Although this is in fact the correct term, the description was vehemently resisted by many of those who responded to the survey. One of them wrote:

> I almost didn't answer this, I had to leave it lay for several days in order to calm down some. I am very proud of my almost 40 years of hunting and fishing in Colorado. For someone to put me in the same category with poachers, as far as I am concerned that puts them in the same category with antihunting groups. If that's an injustice it can't be a bigger injustice than what you did [to] me. I made a mistake once, and a young hothead game warden tried to take advantage of it to boost his arrest record point system. I misread some very complicated regulations. They write them more complicated every year to try to boost their "fine" income. (Eliason and Dodder 1999: 239)

Apart from the writer's success in squeezing perhaps four different categories of neutralizing excuse into one short paragraph, what is noteworthy about the letter is the writer's strong endorsement of the dominant social attitudes toward "poaching." Indeed, it is precisely because he abhors poachers that he is driven to adopt the rather untenable position that while he may (by his own admission) have illegally hunted game, he is nevertheless not a poacher. One can find similar attempts to defeat analyticity in the claim, often made by those convicted of white-collar offenses, that though they may have broken the law, they are not really criminals (Geis 1968: 104).

As one can see from this example, there is an element of genuine self-deception in the use that offenders make of these neutralizing excuses. Furthermore, it is still in many respects a mystery why certain people, in certain situations, seem to be more vulnerable to these sorts of self-deceptions. Thus the discussion of techniques of neutralization does not solve the problem of explaining criminal motivation. The significance of the theory lies in the way that it redirects our attention, away from the issue of compliance with primary moral norms, toward compliance with the secondary norms that govern excusing conditions. It suggests that what many criminals are doing, when they break the law, is not violating shared moral principles, but rather circumventing them—violating non-moral rules in such a way as to persuade themselves that their criminal actions remain compliant with the prevailing set of moral rules.

Despite the fact that this theory puts considerable emphasis upon the way that individuals think about their actions, it is not a fully cognitivist account of criminal motivation. There is still a core element of deviance in the criminal will that remains somewhat mysterious—although not entirely so. It is here that the social dimension of criminal behavior is clearly important. The offender will find it much easier to regard his own excuses as plausible (and thus to maintain the self-deception) if he is in a social environment in which such claims tend to be given credence, or where he is unlikely to encounter critical or dismissive voices. Thus "differential association" and the formation of deviant "subcultures" remain an important part of the story about crime. Neutralization theory, however, regards the function of these subcultures differently. Rather than sustaining an independent system of values and moral principles, different from those of the mainstream, the function of the subculture is to create a social context in which certain types of excuses are given a sympathetic hearing, or perhaps even encouraged.[5] In this way, the offender finds it easier to live with the (otherwise glaring) contradiction between his own commitment to the moral standards of society and the criminality of his actions.

There is some debate about how much this theory explains, since the use of such techniques of neutralization is not universal (e.g., Kraut 1976: 363–364). It is also not clear to what extent these techniques are used merely to provide excuses, or whether they in fact supply full-blown justifications (Hindelang 1970, 1974). It seems clear, for instance, that an appeal to higher loyalties suggests that the action was not merely excusable, but actually the right thing to do under the circumstances. In that case, the extent to which the criminal shares in the broader normative consensus of the society becomes subject to dispute. Nevertheless, the basic empirical phenomenon of neutralization is clearly an important one (see Agnew 1994; Agnew and Peters 1986; Akers et al. 1979; Buffalo and Rodgers 1971; Landsheer et al. 1994). In contemporary criminological research, it is typically embedded within a multifactorial theory of deviance, as one of several "social" factors that generate a propensity toward crime (Akers 1998: 77–87). It is worth singling out for special attention in this context, however, because it is a factor that should be of particular interest to business ethicists.

11.3. Neutralizations in Business

When crime is analyzed from the perspective of techniques of neutralization—rather than, say, faulty socialization or deviant values—it immediately

[5] In this context, one might read with interest the lyrics of Ice Cube's "Why We Thugs."

becomes apparent why bureaucratic organizations such as large corporations, as well as the "market" more generally, might constitute peculiarly criminogenic environments. These are institutional contexts that generate a very steady stream of rather plausible (or plausible-sounding) excuses for misconduct. This is the result of a confluence of factors: first, corporations are typically large, impersonal bureaucracies; second, the market allows individuals to act only on the basis of local information (Hayek 1945), leaving them in many cases unaware of the full consequences of their actions; third, widespread ideological hostility to government, and to regulation of the market in particular, results in diminished respect for the law; and finally, the fact that firms are engaged in adversarial (or competitive) interactions gives them broader license to adopt what would otherwise be regarded as antisocial strategies (Heath 2007). The other major feature of the corporation, and of the business world more generally, is that it constitutes a subculture that in many cases isolates individuals from the broader community, and thus may serve to insulate deviant ideas and arguments from critical scrutiny.

It may be helpful to consider these factors from the perspective of the seven different categories of neutralization technique. Sykes and Matza's original work was done in the context of juvenile delinquency and street gangs. However, it is easy to see that there are very familiar "business" versions of each pattern of excuse that was encountered there.

> *Denial of responsibility:* Hannah Arendt once described bureaucracy as "rule by nobody" (1969: 81). With corporate crime in particular, it is seldom the case that any one individual is clearly responsible for a particular action. Thus when a crime is committed, everyone can, with some degree of plausibility, point the finger at someone else. The person who carried out the action can blame the person who made the decision, the person who made the decision can blame the person who vetted the decision, etc. (e.g., see Vandivier 1996: 128). Due to the organizational hierarchy of the firm, individuals can always try to pass the blame up to their superiors. These superiors can, in turn, try to pass the blame back down, by insisting that their subordinates acted independently (Clinard and Yeager 1980: 45). (In this context, it is worth noting that the "ethics codes" adopted by some firms clearly facilitate the latter. By imposing upon each employee the obligation to resist any "unethical" orders, they in turn make it more difficult for these employees to shift the blame up.)
>
> The competitive structure of the marketplace, not to mention the "hard budget constraint" (Kornai 1992: 143–144) imposed by investors, also generate the perception, among many people, that they have "no choice" but to violate the law. This is, of course, predicated upon the

assumption that the bankruptcy of the firm (or personal bankruptcy, or even just losing one's job) is an evil to be avoided at all cost. For example, Geis quotes one defendant in the heavy electrical antitrust case excusing his actions in the following terms: "I thought that we were more or less working on a survival basis in order to try to make enough to keep our plant and our employees" (1968: 108). Here one can see the vocabulary of "survival" being used to blend the "necessity" defense into an appeal to higher loyalties (in this case, an altruistic concern for the plant's employees).

The competitiveness of the marketplace, and the workplace, also means that if one individual refuses to perform an illegal act, he may simply be replaced by someone else who is (or if one firm refuses to pay a bribe, the business will simply go to some other firm that is, etc.).[6] This suggests that the illegal act is going to occur regardless of what any one individual chooses and is thus subject to some sort of metaphysical "necessity." As a result, the particular individual who happens to perform the act cannot be said to have "caused" the harm that results, since one of the central counterfactuals associated with causal relations is false (it is not the case that, had he not performed the act, the harm would not have occurred).

Denial of injury: One of the most important features of white-collar crime is its often "faceless" character. In general, people have more permissive attitudes toward crime when the victim is unknown, or else an institution (Landsheer et al. 1994: 51). Most white-collar criminals never meet or interact with those who are harmed by their actions (and in many cases they would not even know how to find their victims should they choose to). This makes it more plausible to claim that no injury has occurred. In antitrust cases, in particular, many offenders simply refuse to believe that they have caused any harm. Geis quotes a Westinghouse executive, for instance, acknowledging that price-fixing arrangements were illegal, but denying that they were criminal: "I assumed that criminal action meant damaging someone, and we did not do that" (1968: 108). One can find the same steadfast refusal to acknowledge any harm by Microsoft executives, despite having been found in violation of the law in both the United States and the European Union. The problem stems from an ignorance of, or perhaps an unwillingness to grasp, a rather subtle point of economic theory, namely, that the social cost of monopoly is borne,

[6] My father, while serving in the Royal Canadian Air Force, once threatened to resign if a particular practice, which he considered unethical, was not stopped. His commanding officer stuck his fist into a pail of water than happened to be on his desk, pulled it out, and said "You see that Heath? That's the hole you'll leave in this organization when you're gone."

not by those who purchase the firm's products, but rather by those who do not purchase them due to monopolistic pricing. Typically, however, monopolists point to the satisfaction of the firm's own customers as evidence that their conduct caused no harm. This defense is based upon an economic fallacy, but it is hardly one that they have an incentive to sort their way out of.

In these cases, there is potential confusion as to the identity of the individuals who are harmed by the criminal's actions. In other cases, the mere fact that there is diffusion of the harm over a very large number of persons is appealed to as grounds for denial that anyone was injured by the person's actions. This is presumably what underlies the widespread conviction that crimes committed against large corporations are more acceptable than those committed against small ones. It may also be a major factor in the extraordinarily permissive public attitudes toward tax evasion, insurance fraud, or crimes resulting in losses that are covered by insurance. Finally, because shareholders are not entitled to any fixed rate of profit, actions that merely produce a lower rate of profit are sometimes excused on the grounds that they did not result in actual losses.

One of the most general grounds for denying injury stems from overly generous use of the *volenti non fit iniuria* principle. This is often tied to a form of market utopianism, which suggests market outcomes are to be presumed efficient until proven otherwise. Since market transactions typically involve consent, it is relatively easy for people to convince themselves that shareholders who are exploited by management could have invested their money elsewhere, consumers who purchase inferior goods ignored the "buyer beware" rule, workers who are injured "knew the risks when they took the job," and so on. One can find highly sophisticated variants of these arguments. Certain proponents of the so-called "efficient markets" hypothesis, for example, claim that the stock market fully anticipates managerial graft when determining the price at which shares trade. Since the shares of firms where managers abuse their perquisites will trade at a discount, this sort of 'abuse' does not actually harm shareholders—indeed some theorists claim that it is merely "implicit compensation" for the managers. Many "economically" minded theorists defend insider trading using more-or-less the same rationale (Easterbrook and Fischel 1991: 257–258).

Denial of the victim: The essence of this neutralization technique is the claim that, rather than merely acting opportunistically toward the victim, the offender is in fact playing tit-for-tat, and thus responding in kind to past opportunistic conduct on the part of the supposed victim. The least sophisticated version of this argument involves simply pointing at the other and saying "he started it." The more sophisticated version involves presenting the offender as exacting righteous vengeance, perhaps even sacrificing his own interests in order to ensure that the crimes of others do not go unpunished.

This category of neutralization technique is especially important when it comes to occupational crime. It is very difficult to find an employee who believes that an enhancement of the overall level of distributive justice in society would require a reduction of his or her current compensation package. Such perceptions of "underpayment inequity" can be an important source of occupational crime (Greenberg 1990). Among less skilled workers, people often confuse the fact that their *role* is invaluable to the organization with the belief that *they* are essential to the organization. Thus they feel undercompensated, ignoring the fact that it is the ease with which they can be replaced that determines their wage rate, not the value that they contribute to the firm on a day-to-day basis.

The basic structural problem comes from the difference between the adversarial orientation associated with the competitive labor market and the more cooperative orientation required for work within the firm. Labor is, as Karl Polanyi wrote, a "fictitious commodity" (1957: 72–73). When a firm hires an employee on salary, what they are doing is essentially paying to secure that person's cooperation. Yet when it comes to negotiating compensation, it is the adversarial norms of the marketplace that prevail (see Heath 2007). It can be very difficult for employees to "switch hats" so quickly, to put what are often very bitter wage negotiations behind them, and return to being "team players," devoting themselves selflessly to the interests of the firm.

All of this creates an environment in which it is relatively easy for people to convince themselves that, rather than stealing, what they are really doing is taking what they are owed, or perhaps punishing their employer for treating employees poorly (Green 1990: 81–83; Greenberg 1990). In one large-scale survey, Richard Hollinger and John Clark found that "when employees felt exploited by the company . . . these workers were more involved in acts against the organization as a mechanism to correct perceptions of inequity or injustice" (1983: 142). Furthermore, if the corporation is engaged in unethical or illegal practices, employees may regard their own theft as nothing but the seizure of "ill-gotten gains." More generally, few people in the public at large regard corporations as absolutely innocent (in the way that a person walking down the street, singled out at random and mugged, is absolutely innocent). This contributes to a general propensity to regard occupational white-collar crime as merely "just deserts" (and hence as victimless).

Condemnation of the condemners: One of the most prominent features of corporate crime is the frequency with which business executives dispute the legitimacy of the law under which they are charged, or impugn the motives of the prosecutors who enforce them. Consider, for instance, the abuse that was heaped upon New York State Governor Eliot Spitzer during his tenure as Attorney General (particularly in the *Wall Street Journal*) for exposing a wide range of dubious practices in the insurance, mutual fund, and securities industry. His major prosecutorial work was almost never discussed, in the popular press, without some mention of his "political ambitions."

More generally, corporate criminals will often contest the very legitimacy of regulation, by suggesting that the government, when it imposes constraints upon the marketplace, is actually beholden to "special interests," while the corporation represents the broader interests of the public. Since the latter is taken to be a larger constituency than the former, the suggestion is that the corporation enjoys stronger democratic legitimacy than the government. Another common strategy is to pick out one overzealous or odd regulation and use it as grounds for dismissing the need for all regulation (Clinard and Yeager 1980: 70–71), or to impugn the competence of government in general. Raymond de Sousa, for instance, argued for jury-nullification in the Hollinger International case on this basis: "I have very little confidence that the same vast bureaucratic apparatus that manages our health care, our post office or our roads somehow becomes more competent and fair when it comes to criminal justice" (De Sousa 2007: A20).

The other major strategy is to suggest that the government is motivated by some type of ideological agenda (as opposed to the corporation, which for structural reasons can have no interest other than to "give the people what they want"). Thus prosecution of white-collar offenses is seen as stemming, not from considerations of justice, but rather from some sectarian political ideology.[7] The very concept of "white-collar crime" is often dismissed as a socialist plot, despite the fact that the primary beneficiaries of such prosecutions are usually capitalists (i.e., investors). For example, when Robert Lane interviewed a group of business executives in the early 1950s, asking them how to reduce the level of corporate crime, the most common recommendation was to "stop the drift to socialism and the restriction of freedom" (Lane 1953: 164). All of the other proposals made by these executives focused upon either increasing

[7] Writing for the Heritage Foundation, Baker (2004) argues, "The origin of the 'white-collar crime' concept derives from a socialist, anti-business viewpoint that defines the term by the class of those it stigmatizes."

the quality or integrity of government, or else decriminalizing the relevant activities. Not one made any suggestion that would have enhanced compliance with the existing body of law.

Appeal to higher loyalties: "I did it for my family" remains one of the most popular excuses for occupational crime, especially among female offenders (Daly 1989). These sorts of excuses are no different in kind from the ones employed by street criminals. What is different in the business context, and what outsiders sometimes have difficulty comprehending, is the extent to which the corporation itself can serve as an object of higher loyalty. This is especially the case in more knowledge-intensive industries, which are subject to greater "information impactedness," and so rely much more heavily upon the loyalty of their employees in order to overcome internal agency problems. Considerable effort on the part of management is aimed toward cultivation of these loyalties, from dramatic initiation rituals for new employees, on-site recreational and sports facilities, personal counseling services, to the ubiquitous "team building" seminars and weekend retreats (Arnott 2000).

An unintended consequence of the intense loyalties that are developed through such techniques is that employees may sometimes feel that they are excused from any accusation of criminality, so long as their actions were undertaken for the sake of the firm rather than for reasons of self-interest. (For example, it is quite plausible to suppose that neither Kenneth Lay nor Jeffrey Skilling were motivated by any personal pecuniary incentive when they misled investors about Enron's financial condition. They did it for the sake of Enron—an organization that they both continued to insist was a "great company" even after its collapse [McLean and Elkind 2004: 419].) One study of retired Fortune 500 company managers by Marshall Clinard (1983) showed a widespread condemnation of whistleblowing, on the grounds that it conflicted with the "loyalty" owed by employees to the firm. Many believed that (with certain exceptions, such as safety violations) individuals who were unwilling to participate in illegal activities should simply quit their jobs and keep quiet, rather than "go to the government" (1983: 116).

It should also be noted that managers will sometimes appeal to the fiduciary relationship that they hold toward shareholders as an excuse for misconduct (Clinard and Yeager 1980: 72). (Depending upon the audience being appealed to, offenders will also sometimes appeal to stakeholder interests as well. Corporate crime, for instance, can be excused as an action taken to stave off bankruptcy, in order to protect workers from losing their jobs, etc.) The "we did it for the shareholders" excuse had a ring of plausibility to it, because agents are obliged to advance the interests of their principal as best they can, and this sometimes does require violations of conventional morality. Lawyers, for instance, are generally thought to be under a professional obligation to conceal information on behalf of their clients in many circumstances. Yet the loyalty argument is spurious as a defense against crime, of course, because agency relationships cannot be used to "launder" impermissible actions in this way.

Everyone else is doing it: This is an excuse for all kinds of crime, but it should be noted that it has greater plausibility in a business context than in many other cases. This is because the competitiveness of the marketplace creates certain pressures that are absent in other domains. If one doctor is performing unnecessary procedures, this does not necessarily create any pressure on other doctors to do the same, simply because it does not affect them in any material way. In business, however, illegal conduct can give a firm an unfair competitive advantage that threatens rival firms with significant losses. For example, a minor safety infraction may save a firm only a small amount of money, but if it gives them an advantage over their competitors, which allows them to land several contracts that might otherwise not have gone to them, then these slight gains will be significantly amplified. This will, in turn, create pressures on their rivals to follow suit. (It may also make the violation seem trivial, relative to what is at stake.) The best analogy here is to the dilemma that many athletes face when confronted with the problem of doping in sport (Heath 2007). In some cases, the individual faces a situation in which the consequence of acting ethically is certain defeat. Similarly, corporations are sometimes put in situations where they must offer a bribe, or arrange a kickback scheme, if they want to do business with a particular client. Thus there are clearly cases in which "everyone else is doing it" can serve as a reasonable excuse (although never, it should be noted, as a justification). This having been said, however, one must be on guard against the tendency toward overuse of this excuse. In particular, one must be suspicious of the version that treats it as a general result of microeconomic theory that the misbehavior of one firm "forces" all others to follow suit. In Clinard's study of middle managers, for instance, most ranked the "unethical competitive practices" on the part of rival firms quite low in their assessment of the causes of unethical or criminal conduct (1983: 62–63), while only one in nine felt that it was a significant factor. Primarily, this is because they felt ethical firms had a variety of different ways of protecting themselves from these sorts of tactics—including, most significantly, bringing adverse publicity or regulatory attention to bear upon the firm that was acting unethically or illegally.

Entitlement: One of the major differences between corporate crime and street crime is the frequency with which white-collar criminals simply deny the authority of the laws that they have broken. Often this is based on some variant of laissez-faire ideology (e.g., Clinard and Yeager 1980: 69), which either contests the legitimacy, or denies the efficacy, of any government interference in the market. More sophisticated apologists appeal to the "business judgment" rule, in order to condemn government interference in mere "governance" issues. Both arguments suggest that the state simply does not have the right to regulate certain forms of private transactions. Thus individual businesspeople need not appeal to any "higher good" in defense of their actions, they need only insist upon their rights. Civil rights legislation and various aspects of labor law were for a long time very publicly resisted on these grounds—shouldn't employers be free to choose who they want to employ, or which customers they want to serve? What business is it of the government's?

These sorts of ideological challenges can have very powerful effects. In the United States, for instance, where these ideas enjoy much greater public acceptance, "the problem of business resisting law enforcement by forming oppositional and criminogenic business subcultures would seem to be more widespread" (Braithwaite 1989: 129). Braithwaite draws particular attention to the Occupational Safety and Health Administration in the United States, which has encountered what he calls "an organized subculture of resistance that advocates contesting all enforcement actions, that is consistently challenging and litigating the legitimacy of the government to enforce the law" (1989: 129). It is worth pausing for a moment to emphasize how extraordinarily uncommon it is in advanced Western democracies to encounter such large-scale, organized attempts to undermine the authority of the law. The rather uncompromising tradition of individual rights in the United States, combined with the fact that the US Supreme Court for many years (during the so-called "Lochner" era) interpreted these rights in such a way as to prohibit many of the forms of government intervention in the marketplace that we see today, presumably accounts for much of this phenomenon.

It is also quite easy to find "karmic" versions of the entitlement argument, where people point to how much "good" a company does (e.g., the number of satisfied customers, happy employees, etc.) as an excusing condition for violations of law.

The power of these techniques of neutralization is amplified by the social environment created within many corporations. As Gerald Mars has emphasized, illegal conduct creates considerable cognitive dissonance for the typical perpetrator. Membership in a deviant subgroup plays an important role in "normalizing" this otherwise proscribed conduct. Without the supportive group, "the 'sinning' self threatens to overwhelm the working self." (1982: 170).

For most people, work is the center of their lives. Not only do they spend more waking hours at work than anywhere else, but they do most of their socializing there as well. Their entire circle of social interaction is often limited to family and coworkers. This is encouraged by many modern management techniques, which take a lot of the interactions that would traditionally have occurred outside the workplace and transfer them to inside the organization—creating what Dave Arnott (2000) refers to as "all-consuming organizations." One can see this trend at work in the creation of company "campuses" or "compounds," which include banking services, medical clinics, dry cleaners, daycares, and convenience stores (Arnott 2000: 72–73). A (largely) unintended consequences of this trend is that it leaves employees increasingly cut off from any contact with the broader community, and in many cases, even from their own families. Such arrangements are troublesome, from the standpoint of white-collar crime, simply because they also leave individuals quite isolated from any contact with those who might challenge the "company line" on illegal practices, or reject the excuses that are conventionally offered within the firm.

11.4. Implications for Business Ethics

There is an enormous benefit to be derived for business ethicists from this sort of foray into the criminology literature. I would like to draw attention to some of the implications that the focus on techniques of neutralization has for the way that business ethics is taught. This is an issue that is close to the heart of many in the field, since most people who do research in business ethics also teach it. Of those who teach business ethics, very few do so out of purely "academic" interest, most are also hoping, in one way or another, to improve the chances that their students will act ethically, when and if they continue on to careers in business. There is nothing wrong with such aspirations. Suppose though that we change the focus slightly, in order to bring the criminological perspective to bear. Instead of asking how an ethics course should be taught, in order to reduce the chances that students will behave unethically, let us ask how a course should be taught, in order to reduce the chances that students will go on to commit major felonies. We can then ask what advice a criminologist would have to offer. By paying careful attention to this advice, we can perhaps learn some more effective strategies for the design of ethics courses as well.

The first thing that one notices, when turning to the issue of ethics education, is that the debate over the efficacy of business ethics programs is almost entirely dominated by the folk theories of moral motivation that have been so thoroughly discredited in the field of criminology. Critics of business ethics typically argue that morality is a matter of character, or of values, and that "by the time students enroll in college-level business courses their values have already been formed, rendering ethics education a waste of time" (Williams and Dewett 2005). Defenders of business ethics education, unfortunately, have been far too willing to accept the theory of moral motivation that is implicit in this critique. Thus they have responded by trying to show that it is still possible to improve the character (Hartman 1998), or influence the values (Williams and Dewett 2005: 112– 113), of students. A more appropriate response would be to dismiss the entire frame of reference.

It is worth recalling, in this context, that the motivation most people have for obeying the law is often the same as the motivation that they have for acting ethically. This is especially true with regard to white-collar crime, where enforcement is exceedingly difficult, and the threat of legal penalties in many cases slim to non-existent (J. W. Coleman 1989: 177–180). Insofar as most people respect the law, they do so because they feel morally bound to do so. What the criminology literature tells us about this moral motivation is that it is not about character, and it is not about values. On the contrary, it is various aspects of the situation that individuals find themselves in, what they think about this situation, and what they expect others to think about the situation, that plays the major role in determining how they conduct themselves.

Too many business ethicists, unfortunately, have maintained a stubborn adherence to a discredited folk theory of character traits (e.g., Hartman 1998; Solomon 1992: 3–4). The fact that institutional context is far more important than character should be a source of encouragement for business ethicists. After all, thinking in a disciplined manner about the sort of institutional arrangements that employees find themselves working in is one of the central functions of management. One of the interesting results turned up by Armstrong, in his study of how management students would behave when confronted with the Panalba case, is that the outcome was highly sensitive to the way that he described the role that students would be playing. When told that "a resolution was passed in 1950 which stated that the Board's duty was to represent the stockholders," 79 percent of groups chose the "highly irresponsible" course of action. However, when told that a resolution was passed stating that "the Board's duty was to represent the interests of each and every one of its 'interest groups' or 'stakeholders'" the level of highly irresponsible conduct dropped to 22 percent (Armstrong 1979). Setting aside the more complicated question of whether this type of "stakeholder" orientation represents either a feasible or desirable way of achieving more ethical conduct in business (see Heath 2006a), what this result does show quite clearly is that the way individuals conceive of their obligations—and the neutralizations that are made available to them by aspects of their situation—is an enormously important factor in the decisions that they ultimately make.

This has important implications for business ethicists. On the one hand, it means that business schools—and business managers more generally—cannot simply throw up their hands and claim that it is "too late" to do anything about ethics. The best way to get people to behave ethically is to put them in a situation in which ethical conduct is expected of them and self-serving excuses are not tolerated. This is a matter of effective institutional design. Thus business ethics courses need not do anything particularly profound, such as forcing students to rethink their fundamental values, or promoting their moral development (Williams and Dewett 2005: 112). They need only teach managers how to create institutional environments that will promote ethical conduct. One way of doing this, suggested by the criminology literature, is to create an environment in which the standard techniques of neutralization used to excuse criminal and unethical behavior are not accepted.

If one takes this perspective seriously, then there is no particular reason for business ethics courses to focus on moral dilemmas, or to teach fundamental meta-ethical perspectives (Kantian, utilitarian, etc.). Students do not commit crimes because they lack expertise in the application of the categorical imperative or the felicific calculus. They are more likely to commit crimes because they have talked themselves into believing some type of excuse for their actions, and they have found a social environment is which this sort of excuse is accepted or encouraged. Thus a more useful intervention, in an ethics course, would be to attack the techniques of neutralization that students are likely to encounter, and may be tempted to employ, when they go on to their future careers. As we have seen, white-collar criminals are typically conflicted about their own actions. They know what morality and the law require of them. The problem is that they have convinced themselves that no one is really injured by their actions, or that they had no choice in the matter, or that it is permissible because everyone else is doing it, etc. Typically, the arguments they have used to convince themselves are sufficiently fragile that they can only be sustained in a supportive environment, among peers who are also inclined to view these claims as legitimate. One way to tackle this problem, "preemptively" so to speak, is to demonstrate the inadequacy of these rationalizations, for example, by tracing out the harm caused by embezzlement or expense account abuse; by articulating the logic of government regulation and the basis for its legitimacy; by explaining the concept of market failure and why unconstrained competition sometimes produces inferior results; and by exploring the tendency toward dissipation of responsibility in bureaucracies. One can imagine an ethics curriculum structured around these themes. The goal would be to bring to conscious awareness certain patterns of self-exculpatory reasoning, and to flag them as suspicious, so that students will be less likely to accept them at face value when they encounter them later in life. The goal, in other words, would be to neutralize the neutralizations.

12 }

Business Ethics after Virtue

In the provocative introduction to his book *After Virtue* (1984), Alasdair MacIntyre suggested that modern societies are living in the aftermath of a catastrophic disruption of moral thinking. By way of analogy, he asked his readers to imagine a scenario in which the practice of science had been completely lost—laboratories were destroyed, journals no longer published, and the scientific method as a whole was forgotten—and yet people retained a dim awareness of the content of scientific theories. They could recite from memory some of the major findings of modern physics, chemistry, or biology, but had no idea how to prove any of it, or even what sort of evidence or procedure led to its derivation. This is, MacIntyre suggested, more-or-less the situation that we find ourselves in with respect to morality. There was a time, he claimed, when our moral world made sense, when we could explain why we made the judgments that we made. But this world has collapsed, leaving us with nothing but "fragments of a conceptual scheme . . . simulacra of morality" (1984: 2).

The catastrophe that destroyed the old world, in MacIntyre's view, was the displacement of the concept of virtue from the center of our moral thinking. The solution, he argued, lay in a restoration of this concept to its former position of glory.

It has been thirty years since the publication of MacIntyre's book. And yet, far from seeing a restoration of the concept of virtue, what we have witnessed instead is close to the opposite. While virtue theory has undergone a modest resurgence in philosophical ethics, in other areas of the human sciences it has been subjected to withering critique. Some would even say that it has been debunked. The major lines of criticism have come from three primary sources: social psychology, sociology, and political theory. Most damning is the fact that psychologists have been unable to find any evidence that the type of character traits or dispositions traditionally posited by virtue theory have any predictive or explanatory value, when it comes to distinguishing various individual propensities to act morally. Meanwhile, several other factors, ignored by traditional virtue theory, have been shown to play a surprisingly important role. Sociologists have also found that the concept of vice is of limited use when it comes to understanding the formation of criminal subcultures. In particular, the emphasis that virtue theorists put on habituation and early socialization as ways of discouraging vicious behavior appears not to be supported by the evidence. And finally, political theorists have become increasingly impressed by the argument that major social institutions should be organized in a way that is neutral with respect to plurality of values in the background society. Perfectionism—the political expression of virtue theory—has come to be seen as a major source of parochialism and disorder in the world.

In this chapter, I will present an overview of these three different lines of critique, with particular emphasis on the lessons that business ethicists can draw from them. While there are still many apologists for virtue theory in the philosophical literature, business ethicists must be responsive to a broader constituency. To the extent that business ethicists persist in using the language of virtue in formulating their claims, they put themselves increasingly out of step with broader currents of thought in the human sciences (Brady and Logsdon 1988). There are times when resistance to intellectual trends may be called for, particularly when dealing with popular fads that sweep over the academy. Yet when confronting the results of decades of careful psychological research, along with converging evidence from several different academic disciplines, it is not intellectual heroism, but rather just obstinacy, to persist in one's views. The phrase "after virtue" must acquire a new, more positive connotation. The challenge that business ethicists face today is to construct a conception of business ethics that builds upon the results of modern moral psychology—a science that is finally, after two thousand years, emerging from the shadow of Aristotle—in order to put virtue theory behind us once and for all.

12.1. Virtue Theory

I would like to begin with a note of clarification about how I will be using the term "virtue ethics." In part because of the critical literature that has developed over the past few decades, philosophers have introduced a number of different variations on the theme of "virtue," many of which have little in common with what might be thought of as traditional virtue theory (e.g., Hurka 2001). In this paper I will be focusing on the traditional conception, which received its most influential expression in Aristotle's *Nicomachean Ethics*, was picked up and developed in the Christian tradition, and received a particularly important formulation in the work of St. Thomas Aquinas. I want to focus on this version because it is the one most often used by business ethicists. Furthermore, because of its enormous influence on Christian—particularly Catholic—moral thinking, it makes up a very large component of the "folk theory" of morality in Western societies.

There are three main components to this traditional conception:

1. *Virtues are character traits*. Although some theorists have been backing away from this characterization of virtue, it remains the dominant idea, and it is ubiquitous in the literature. Julia Annas, for example, describes virtues as "persisting, reliable, and characteristic dispositions" to act and feel certain ways (2011: 12). Geoff Moore describes virtues as "enduring character traits" (2005). Robert Solomon describes a virtue as "a pervasive trait of character" (1992: 107).
2. *Virtues are cultivated through socialization, or habituation*—what Macintyre refers to as "habitual exercise" (1984: 154). "People acquire virtues much as they do skills such as carpentry, playing a musical instrument, or cooking. They become just by performing just actions and become temperate by performing temperate actions" (Beauchamp, Bowie, and Arnold 2009: 34). And since we are most impressionable when we are young, "it is one of the tasks of parental authority to make children grow up so as to be virtuous adults" (MacIntyre 1984: 195).
3. *Virtues are oriented toward the achievement of some good*. As Aquinas put it, "the object of every virtue is a good considered as in that virtue's proper matter" (1947: 856 [1-2, q. 63]). This is often articulated as the view that the good constitutes the *telos* of virtuous action (MacIntyre 1984: 148). Furthermore, although this is not essential to the position, this good is typically substantive, rather than formal (Taylor 1985: 233).

The overall picture therefore is of virtue as a cultivated personality trait, which disposes us to act in ways that promote the good.[1]

It should be noted that there is considerable divergence of opinion over what this "disposition" involves, and therefore what role deliberation plays in virtuous conduct. Some theorists have a view that is almost behavioral, and therefore assign only a small role to deliberation. Solomon, for instance, claims that "if the practice of the virtues is often spontaneous and requires little thought, it certainly does not as such require deliberation. Indeed, too much deliberation… is evidence that one does not have the virtue in question" (1992: 195).

By contrast, most virtue theorists conceive of the virtues as "multi-track" dispositions, which admit of different expressions in different circumstances, and where there is a cognitive dimension to the agent's assessment of the circumstances (Spieker 1999: 220–222). Nevertheless, it is clear that with respect

[1] There are more or less egocentric versions of the view. The good under one construal may be our own flourishing, under some other it may be an objective good. See George Sher (1992).

to what Aristotle called the "virtues of character" (such as courage, honesty, loyalty) possession of the virtue is supposed to have more than just cognitive import; it is also supposed to have motivational force. This is why, with respect to these virtues, it is not enough just to be told what is good, we also require habituation or training, because this is what allows us to develop the right motivation.

Macintyre articulates what is probably the mainstream view, with respect to the role that deliberation plays:

> We become just or courageous by performing just or courageous acts; we become theoretically or practically wise as a result of systematic instruction. Nonetheless these two kinds of moral education are intimately related. As we transform our initial naturally given dispositions into virtues of character, we do so by gradually coming to exercise those dispositions *kata ton orthon logon* [guided by right reason]. The exercise of intelligence is what makes the crucial difference between a natural disposition of a certain kind and the corresponding virtue. (1984: 154)

Despite the fact that a multi-track disposition might lead an agent to act differently in different circumstances, it remains clear that the possession of a virtue is supposed to generate an identifiable pattern of virtuous behavior. Indeed, part of the attraction of the theory is that it is supposed to help us understand how people get from *knowing what is right* to *doing the right thing* (and therefore claims to solve what Michael Smith [1994: 6–12] called "the moral problem," of reconciling moral cognitivism with motivational internalism).

Virtue ethics, understood in roughly this way, has a number of strong advocates in contemporary business ethics: Geoff Moore, Robert Solomon, John Dobson, and Alejo José Sison. Nevertheless, it is worth noting that the relationship between business ethics and virtue theory has historically been somewhat fraught. In particular, many virtue ethicists dismiss business ethics, for one of two reasons.

The first avenue of critique is well represented by Macintyre (1979), who argues that the competitiveness of the market economy and the alienation of labor simply make business incompatible with virtue. Most obviously, the imperative of profit-maximization that forms the overarching objective of the standard business corporation bears a strong resemblance to avarice or greed, traits that are conventionally classified as vices (Macintyre 1995: p. xiii). Furthermore, both Aristotle and Aquinas were committed to some version of the doctrine of the just price, and considered it the responsibility of individual economic agents to charge no more and no less. A market economy relies instead on competition between buyers and sellers to achieve the desired price level, and so makes demands on individuals that are incompatible with the traditional virtues.[2]

[2] For further discussion, see chapter 7.

More fundamentally, modern capitalism instrumentalizes work in ways that theorists like Macintyre regard as incompatible with virtuous activity. Many workers spend their days producing and selling products to which they assign no intrinsic value (i.e., "widgets"); they do so only as a way to make money. In Macintyre's view, this means that their actions are not oriented toward the good in the right way to qualify as virtuous (1984: 196; 1979: 126–127). On top of this, the practice of business often requires people to act in ways that seem to erode virtue (Macintyre 1995: p. xiv). People who work in retail sales, for example, are often put in a position of having to promote goods that they themselves have reservations about. While many find a way to avoid outright lying, it would be hard to describe what they are doing as "honest."

The response to this criticism typically involves a certain amount of revisionism in virtue theory (e.g., Dobson 2008). David MacPherson, for instance, tries to reformulate Protestant ideas about a vocational "calling" in virtue-ethical terms, in order to argue that corporations can become communities focused on the cultivation of virtue (2013). Solomon tries to build some wiggle room into the traditional virtues. For example, he defines "fairness in business" as "a certain kind of 'atunement', a sense of value and a willingness to exchange value for value in a market that provides no ultimately objective guideposts" (1992: 210). "Honesty" turns out not to require perfect truthfulness, since "there are certain aspects of every transaction that are expected to be unknown and undisclosed" (1992: 210), and so on.

The second reason that many virtue theorists dismiss business ethics is more problematic. Here the argument focuses less on the nature of the virtues and more on the account of character and socialization. The claim is that managers have no control over the ethical conduct of their employees because by the time people enter the workforce they are already fully socialized, and so it is too late to do anything that will really influence their behavior. Similar arguments are used to show that schools have no ability to improve the ethics of their students. As Clifford Orwin puts it, "by the time a student arrives at university, and *a fortiori* several years later when he ambles on to his MBA, his ethical character is already firmly set. Whether virtue can ever be taught was already a thorny question for Plato. Whether it can be taught to adults, in a classroom, shouldn't be a thorny question for anyone" (2009: A17).[3]

From this perspective, the best way for businesses to promote ethical behavior is to screen potential employees, in order to weed out the bad apples. For those who have already been hired, there is not much that can be done to change their behavior or to prevent them from acting unethically, other than aggressive supervision. The die is already cast, apparently sometime during childhood or early adolescence. After that their "ethical character is firmly set,"

[3] For discussion of near-identical versions of this argument, see Heath (2008a).

as Orwin puts it. Thus he confesses that "the spectacle of MBA students taking oaths to be ethical fills me not with reverence but with giggles" (2009: A17).

Some business ethicists have tried to respond to this challenge on its own terms, by arguing that business schools can have a socializing effect on their students and employees (McCabe, Trevino, and Butterfield 1996). It is certainly the case that many corporations create very totalizing environments, which gives them a considerable amount of control over employees' social lives. The more important question, however, is whether the underlying premise of these arguments is correct. Is it actually the case that the primary determinant of moral behavior is ethical character?

In the following two sections, I will argue that there is no evidence for this contention. Most importantly, there is no evidence that the traditional conception of virtue has any psychological reality. Furthermore, there is considerable evidence to show that other psychological factors, not easily characterizable as virtues, do play an important role in determining whether people act morally. To the extent that it distracts managers from the latter, virtue theory winds up causing considerable mischief. While it is always possible to invent some new, more complex conception of virtue, which avoids the defects of the traditional conception, there is a question of whether it should still be described as virtue theory. Given the long-standing history and the enormous cultural influence of the traditional conception, there is good reason to believe that virtue theory provides a vocabulary for thinking about moral action that is inherently misleading.

12.2. The Psychological Critique

The suggestion that there is something wrong with the folk theory of character traits underlying traditional Aristotelian virtue theory has been around for a long time. As early as the 1920s, psychologists found that traits like "honesty" or "compassion" seemed to have weak predictive value and could easily be overwhelmed by situational factors (Hartshorne and May 1929; Ross and Nisbett 1991). (Since psychologists have such easy access to large numbers of students, some of the earliest studies were on cheating and plagiarism. This resulted in "honesty" being the most exhaustively studied trait [McCabe, Treviño and Butterfield 2001].) Stanley Milgram's (1974) experiments on obedience to authority shocked almost everyone by showing that, whatever resources of moral "character" individuals might have, these could be consistently overwhelmed by the imposition of what would appear to be fairly mild social pressure.[4] This led to a fairly broad proliferation of experiments, aimed

[4] The literature on this is voluminous, but for a balanced survey, see Thomas Blass (1991).

at exploring the different factors that govern individual behavior (summarized in Doris 2002).

As a result of these experiments, psychologists quickly learned to be extremely cautious about relying on folk theories of character to predict or explain behavior. As Lee Ross and Richard Nisbett put it, in their widely read synopsis of the literature, "some factors that we expect to be very important prove to be trivial in their impact; and some factors that we expect to be weak prove, at least in some contexts, to exert a very large influence indeed" (1991: 6). Thus psychologists became increasingly reluctant to rely on whatever "common sense" intuitions we may have about individual motivation.

This message, however, took a long time to reach philosophers. Owen Flanagan (1991) and Gilbert Harman (1999) are generally credited with being the first to note the significance of this research, and Peter Railton discussed it (1995), but the majority of moral philosophers continued to ignore it through to the end of the twentieth century. The turning point came with the publication of John Doris's book, *Lack of Character* (2002), which presented the data in a systematic and accessible format, but more importantly, framed it in a way that made the underlying challenge to virtue theory impossible to dismiss through mere hand-waving.

There are two major aspects of the psychological research on character that are significant. First is the finding that it is difficult to isolate character traits at the medium level of generality posited by virtue theory. Virtues have traditionally been taken to have cross-situational predictive validity (e.g., according to one proponent, "virtues are not simply practice-specific, but span and are applicable to all practices and situations in which an individual is involved" [Moore 2005]). Thus the expectation was that the presence of a character trait like "honesty" or "bravery" should be detectable across a range of different behavioral contexts. This does not mean that the person must always act in the same way, but it does mean that a trait should result in detectable differences in the behavior of individuals who are thought to possess it and those who do not. So, if person *a* is more likely to behave "honestly" in situation *x* than person *b*, then person *a* should also be more likely to behave "honestly" in at least some other situations than person *b* is. As David C. Funder put it: "The relevance of personality for behavior is revealed not by a lack of change across situations but by the maintenance of individual differences. Such maintenance is indexed by the correlation coefficient, which reflects the degree to which subjects who perform a given behavior more often than others do in one situation also do so relatively more in a second situation" (1999: 42).

There is no question that individuals have habits, or specific scripts that they follow in particular situations.[5] For example, at a store when the cashier

[5] In this respect, some situationists overplay their hand, by denying that individuals possess any sort of stable personality. The critique of virtue theory does not require the denial that individuals have

making change mistakenly hands over too much money, many people, when they notice the error, will automatically give it back. This is typically a settled disposition, often developed at an early age. The problem is that it does not correlate in any significant way with behavior in other types of situations, even ones that are only slightly different. Many people, for instance, when they notice that an item has scanned wrong, and that they have been charged less than the marked price, will let the error stand. Again, this may be a very settled disposition. The important point is that knowing how a person behaves in one sort of situation tells you nothing at all about how the person will behave in the other sort of situation. It also tells you nothing about whether they pay their taxes, or whether they will cheat on a test. And so if you ask the question, "Is this person honest?" the answer will be "It depends." Even the microcategory of "willingness to take advantage of employee error in order to pay less than full price for goods" is still too general to capture the morally significant dimension of the individual's behavioral dispositions.

There is, however, some evidence of the existence of character traits, as traditionally conceived, at a much higher level of generality than those posited by virtue theory. Psychologists often talk about the "big five" personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism (Goldberg 1990; McCrae and Costa 1996). There are two things that are striking about these traits: first, they have no particular moral valence and are therefore not "virtues" in anything like the traditional sense. Second, they do appear to have cross-situational predictive validity (weak, but not nonexistent) (Moskovitz 1994; Carducci 2009: 298–301). So, for example, a test used to determine how introverted or extraverted a person is might ask a question such as, "When the phone rings, does it make you feel excited or anxious?" Knowing how a person answers this and other related questions would actually be useful in predicting how that person is likely to behave in other situations, such as entering a room full of strangers. This is precisely the type of cross-situational predictive validity that was claimed for the classic Aristotelian moral virtues, but which empirical research has subsequently failed to demonstrate.

The second major blow to virtue theory came from a raft of studies showing that, while we tend to focus on personality traits that have no predictive value when explaining individual behavior, we often ignore factors that are of major importance in determining whether people act morally. Perhaps the most celebrated discovery is the power of conformity, imitation, and the associated set of social expectations, in determining conduct.[6] Whatever particular habits or scripts people may have, these can quite easily be overridden not just by what

---

stable behavioral dispositions. It rests on the observation that a large set of morally significant dispositions are uncorrelated across different types of situation.

[6] The most influential set of experiments on conformity were by Solomon Asch. For summary, see Asch 1955, also Turiel 2002.

the group does, but even by perceptions or hints as to what the group may do, or what the expectations are. For example, with respect to student cheating, a major literature survey concluded that "although both individual and contextual factors influence cheating, contextual factors, such as students' perceptions of peers' behavior, are the most powerful influence" (McCabe, Treviño, and Butterfield 2001). This poses an obvious challenge to virtue theory. As Annas put it, "we think of virtues as robust dispositions which can be compared across circumstances and even cultures, which makes it problematic to think that a disposition plastic enough to be pushed and pulled around by changing circumstances could be a virtue" (2011: 111).

Similarly, psychological studies have shown that people's "construal" of the situation—what kind of interaction they take themselves to be involved in, what role they ascribe to themselves and others, what they expect others to do, and to expect—is extremely important in determining how they will behave. To take just one example, Varda Liberman, Steven Samuels, and Lee Ross were able to show that in an experimental public goods game, subjects were almost twice as likely to act cooperatively if the interaction was labeled "the community game" rather than "the Wall Street game." One version of the study was carried out on Israeli air force trainees, along with their instructors, who were asked to predict the rates of cooperation of different trainees in the two games. The instructors did no better than chance at predicting the trainees' choices ("predictions about individual players proved valueless, as did trait ratings" [2004: 1181]). At the same time, the trainers failed to anticipate that the different labels attached to the game might result in different behavior. Here one can see the classic pitfall, of overestimating factors that are unimportant (i.e., believing that one's knowledge of the person will help one to predict what he or she will do) while underestimating factors that turn out to be quite important (i.e., ignoring social cues that affect how people will think about the interaction).

The most common response to this among virtue ethicists has been to appeal to the "multi-track" character of virtues. The dispositions in question are complex, they argue (Russell 2009: 330). The individual possessing a particular virtue only has a disposition to act in a particular way in relevantly similar situations (Sreenivasan 2012: 301). But whether one situation is similar to some other depends on that agent's construal of it (Solomon 2003: 52). So, for example, if an individual acts honestly when given incorrect change, but dishonestly when an item scans wrong, this might be because she does not construe the latter situation as one that calls for honesty.

This is no doubt true, but notice what it does to virtue theory. Most obviously, it runs the risk of making the psychological portion of the theory unfalsifiable. If you try to find evidence of character traits like "honesty" using standard construals (e.g., by assuming that the agent interprets the situation in the way that a typical person would interpret it), then you cannot find any evidence of these sorts of traits. So the only way to rescue the theory is to posit an idiosyncratic construal. But this means that any time a putatively virtuous agent fails to act virtuously, one can always say that it is because the situations were not relevantly similar, according to that agent's construal. And who knows how agents are construing situations?

This strategy implicitly turns virtue into a non-operationalizable psychological concept. By severing the connection between the hidden "disposition" being posited and the agent's behavior, ascription of that disposition becomes compatible with any set of observations. Furthermore, this defense of virtue theory winds up leaving the agent's construal of the situation doing all the work in determining if he behaves morally or not. But then what is the point of positing a virtue like "honesty" if, in the end, whether or not the person acts honestly is determined by whether he conceives of the situation as calling for honesty, and there are no discernable differences between individuals in their level of commitment to honesty, only differences in their tendency to construe situations as calling for honesty? One could just as easily (and with greater psychological plausibility) posit a desire to "do the right thing," and then explain all behavior as determined by the agent's construal of what "the right thing" to do is, and which situations call for it.

In other words, if this analysis is correct, then it suggests that business ethics should be focused on construal, not on character traits. As psychologist Roy Baumeister has observed, "When the line between right and wrong is clear, most people will consistently do what is right" (1997: 255). Thus any factor that tends to obscure the line, or raise doubts about which side is which, has a propensity to generate immoral action. The good news for business ethics is that the question of how employees construe their actions is one that managers typically have a great deal of control over. It may be too late to go back and make sure all of one's employees were raised right, but it is not too late to try to influence how they think about their job-related activities, and how they interact with others both inside and outside the firm. The resulting approach to ethics, however, would be far more cognitivist than traditional virtue theory.

Finally, some virtue theorists have been inclined to defend the theory by emphasizing its normative status, and pointing out that according to the traditional conception very few people actually manage to achieve virtue. The medieval Christian conception of a "saint," for example, was simply that of a virtuous individual, and it was clearly understood that not many people could realistically aspire to this status.[7] From this perspective, one would not expect

---

[7] It is common among philosophers, when teaching Aristotle, to claim that there is no good English translation for the word *eudaemonia*. This is not exactly true—it is just that many modern Aristotelians do not like the correct English word, and so choose to use "happiness" as a poor approximation. The Latin word *beatitudo* was used to translate *eudaemonia*, from whence we get the English word beatitude (beatification, beatific, etc.). Beatitude perfectly captures the sense of *eudaemonia.* The fact that it has religious overtones reflects the real history of Aristotelianism in the Christian West.

to find a detectable level of virtuous character traits in a random sample of undergraduate students (Sreenivasan 2012: 296). Of course, it is then incumbent upon the virtue theorist not only to show that such rare individuals do exist, but that other people are able to identify them reliably (so that emulation of these persons can serve some kind of useful function in everyday morality). In any case, even if this could be done, the overall argument has the effect of undermining the way that virtue theory has traditionally been employed in business ethics. For example, there is no point advising companies to cultivate virtuous character traits among their employees if no more than a few saints have any chance of achieving them, while for the rest of us the exercise has no detectable influence on behavior.

12.3. The Sociological Critique

The second major problem for virtue ethics arises in a relatively neglected area of the theory, namely, the account of vice. Virtue theory attempts to provide an account of why people act morally, but it is easy to forget that it also has an account of why people act immorally. Many books on virtue ethics say surprisingly little—sometimes nothing—about vice. In fact, it is not clear that there is anything resembling a consistent view among virtue ethicists about the nature of vice. There are at least two major positions. The first regards vice as the mirror image of virtue, as a character trait that disposes individuals to seek evil. Jean Baechler, for instance, describes vice as emerging "where values undergo an inversion, and evil is substituted for good and vice versa" (Baechler 1992: 29).

On this view, the vicious person is one who subscribes to Milton's injunction of despair: "Evil be thou my Good" (1824: 223). This theory provides a fairly natural interpretation of what we have in mind when we talk about certain vices like cruelty, but in many other cases it is quite unrealistic. Most obviously, it fails to provide a natural interpretation of the most common moral failings, which typically involve individuals acting in a mildly egotistical or self-interested fashion. Absent belief in the Devil, it is difficult to summon up the outrage to denounce this kind of mundane selfishness as "evil."

Thus the mainstream modern view seems to be that vice is simply the absence of virtue, or the result of a failure to cultivate a firm disposition to seek the good. As Annas puts it, vice is "a failed or misguided commitment to goodness" (2011: 104). Dissipation, rather than perversity, is taken to provide the most plausible model of vice.

For convenience we may refer to dissipation as the "modern" view and perversity as the "medieval." Yet regardless of the specific conception, the virtue ethicist is clearly committed to the claim that immoral behavior follows from a certain defect of character. "Bad people do bad things," as the saying goes. This is, of course, an empirical claim, one that is not so difficult to test. An obvious way of investigating the question is to study criminals, particularly those who have violated laws that coincide with moral constraints (against theft, battery, murder, etc.).

Since Émile Durkheim's pioneering work (1895/1937: 71–72), the study of crime has been a major preoccupation of sociologists (it is due to Durkheim's influence that criminology, to this day, remains largely a subspecialty of sociology). Durkheim was the first to show how population data and statistics could be used to test various hypotheses about the motives underlying individual actions (most famously, in his study of suicide). The same set of techniques can be applied to crime, and indeed, one of the reasons that Durkheim was interested in these issues stemmed from his desire to operationalize various moral-philosophical theories (particularly utilitarianism and Kantianism), in order to put these to the empirical test.

When one adopts a sociological perspective on crime, one of the most striking features that shows up are its "social dimensions," namely, the way that it tends to cluster in particular neighborhoods, demographic groups, social networks, or in the case of white-collar crime, corporations or sectors of the economy. (Indeed, sociologists are sometimes criticized for seeking to find the "root causes" of crime in factors outside the individual. It is suggested that this is driven by a desire to diminish personal responsibility. The actual explanation is that it stems from a failure to find such causes inside the individual.) Now one of the often-cited advantages of virtue theory is that it is a social theory of morality (Solomon 1992: 192–193). The account of habituation suggests that you cannot be moral all by yourself, you need to learn how to be moral from an appropriate role model. The concern raised by criminological studies is that virtue theory, despite being social, may be social in the wrong way. It emphasizes the "vertical" dimension of behavioral transmission, from parents or authority figures to children, instead of the "horizontal," from one peer to another. It also assigns greater importance to interactions in the past—on the grounds that they produce habits, which sediment to form character—over interactions in the present. Criminological research suggests that this puts the emphasis on the wrong set of social interactions. When it comes to determining criminality, horizontal interactions appear to be far more important than vertical ones.

In a highly regarded meta-analysis of the factors that are positively correlated with subsequent violence and "serious" delinquency, Mark Lipsey and James Derzon found that broken homes and abusive parents were relatively poor predictors of subsequent misconduct (1998: 98). The most significant predictors were a past history of delinquency (even if minor or nonviolent), lack of social ties, and having peers engaged in antisocial activity (1998: 97). Although this is a complex and difficult literature, one of its persistent themes is that the importance of home life is easily overestimated (Lane and Davis

1987: 135). Similarly, interventions aimed at improving the quality of home life seem to have negligible effects (for discussion, see Nisbett and Ross 1991: 5–6). This is why criminologists, when studying social-environmental factors in order to better understand crime, tend to focus more on the immediate, peer environment, and not the past, background environment.

Even when one focuses on peer relations, however, it can be difficult to figure out what it is about an environment that makes it criminogenic. One of the earliest hypotheses was that members of a particular social group shared a deviant set of values, and that individuals came to share those values by interacting with, or being inducted into, the group. For example, whereas mainstream society condemns violence, it might be thought that members of a criminal subculture condone and perhaps even celebrate it, leading members of that group to act more violently.

This "subcultural" theory of crime bears some similarity to the medieval view of vice, in that it posits an inversion of values—people do bad things because they regard these things as good. The hypothesis has been abandoned with respect to crime, however, because it is contradicted by all of the evidence. Criminals, far from having a deviant set of values, seem to have quite conventional moral values (Sykes and Matza 1957). With respect to violence, for example: "One of the most decisive reasons the theory was dropped was that it proved impossible to find groups or subcultures that held positive or favorable attitudes toward violence. . . . The theory of the subculture of violence also lost ground as studies showed that people who behaved violently did not seem to regard these actions as a means of gaining prestige or making a positive impression on others" (Baumeister 1997: 273). Violence will often be used to intimidate, or to cultivate a reputation for violent retaliation in order to deter others from attacking, but there is no inversion of values. Even within violent subcultures, it is simply not the case that violence is regarded as a good thing.

The rejection of the subcultural theory of crime was consistent with a great deal of rethinking of the traditional concept of "evil" that occurred after the Second World War, as people tried to come to terms with the realization that psychologically normal men and women were, under the right conditions, capable of perpetrating heinous crimes. Hannah Arendt's observations, attending the trial of Adoph Eichmann—one of the key architects of the Holocaust—were enormously influential. Although wanting to believe that Eichmann was a monster, Arendt found herself incapable of avoiding the conclusion that he was just an ordinary German, a relatively unexciting bureaucrat going about his work. Most importantly, he bore no personal animus toward Jews. (Several psychiatrists who examined Eichmann declared him to be psychologically "normal," with one adding that he was "more normal, at any rate, than I am after having examined him" [Arendt 2006: 25].)

Arendt's observations about "the banality of evil" generated considerable outrage when first published. The Milgram experiments, it is worth recalling, were undertaken with the goal of disproving Arendt's hypothesis, although in the end they wound up providing its most powerful confirmation. All of this suggests that there is actually an element of wishful thinking in the idea that "bad people do bad things." Since the people who say this typically do not conceive of themselves as bad people, adherence to this theory is a way of putting some distance between themselves and those who do perpetrate such acts, and thereby of avoiding the disquieting suggestion that they too are perfectly capable of inflicting great suffering on others.

Of course, the fact that criminals are not positively evil does not exclude the possibility that they have some other sort of character defect that leads them to engage in antisocial behavior. One of the earliest applications of the newly emerging science of personality theory was to study criminals, in order to see what qualities of character might be responsible for their conduct. Again, it is worth noting that the background expectation was that there would be such personality differences, and that they would be relatively easy to find. The dominant late nineteenth-century view of criminals was that they were degenerates, both morally and intellectually. Even among progressive writers such as L. Gordon Rylands (1889), it was assumed that crime had a significant hereditary dimension, and that it would correlate with easily detectable psychological characteristics.

And yet from the 1940s through to the end of the 1970s, psychologists were unable to devise any personality test that showed any significant difference between criminals and the general population (Schuessler and Cressey 1950; Waldo and Dinitz 1967; Tennenbaum 1977). And even where some correlation was found between certain personality variables and criminal conduct, the range of variation within the criminal population greatly exceeded the difference between criminals and the general public. What this meant in practice was that if one were given a large stack of personality assessments and asked to guess which ones belonged to the incarcerated criminals, one would have a hard time doing much better than chance. (One would have a much easier time, for instance, guessing which social class the individuals came from.)

These results should not be entirely surprising, given the problems of personality theory outlined in the previous section. If there were such a thing as a general personality trait like "honesty," then it would be reasonable to suppose that many criminals would lack it (or that they would possess a negative trait, such as "dishonesty"). But since there is no such trait, it is not surprising to discover that you cannot sort the population very successfully using such criteria. Even a trait like "aggressivity" is less helpful than one might think, not because criminals do not possess it, but because large segments of the general public do as well.

To the extent that more recent work has succeeded in picking out personal qualities that are correlated with criminality, they tend to be very high-level dispositions with no specific moral valence. For example, among the "big five" personality traits, criminals often show up as having, on average, lower levels of conscientiousness (which seems to have a general relation to social deviance [Salgado 2002]). Also, one personality trait that tends to show up in the prison population is impulsivity, for reasons that are not difficult to imagine.

At the same time that criminologists were struggling to find any personality traits that were conducive to crime, they encountered increasing evidence that the way individuals construe their actions does play an important role. In the 1950s David Matza and Gresham Sykes suggested that the reason deviant subcultures (such as youth gangs) are criminogenic is not that they encourage primary deviance with respect to the moral norms and values of society, but that they facilitate secondary deviance with respect to the cognitive and epistemic norms governing the way that situations are construed. So, for example, when contemplating a particular act of violence, instead of maintaining that violence itself is good, members of the group may instead convince themselves that they had no choice but to act as they did, or that the victim had done something to deserve it (and so the aggression was not actually aggression, but rather punishment), or that the consequences of the action were not actually harmful, etc. What distinguishes the criminal, according to this view, is not a motivational defect or an improper set of values, but rather a willingness to make self-serving use of excuses, in a way that neutralizes the force of conventional values.

It should be noted that all of the theories discussed so far were developed in an attempt to understand street crime. When it comes to business ethics, the type of "vices" likely to be encountered are the ones that manifest themselves in the form of white-collar crime. And it is perhaps not surprising to discover that white-collar criminals are even more ordinary than street criminals (Coleman 1989: 200–204). They subscribe to conventional moral values, they do not come from broken homes, they are not high-school dropouts, and they typically were not drunk or high when they committed the crime. And yet white-collar crime still shares many features in common with street crime, including, for example, the peculiar fact that the overwhelming majority of perpetrators are male (Simon and Ahn-Redding 2005: 14–15).

Perhaps the best way to sum up the state of play in the literature on crime is to say that we have achieved the "Socratic wisdom" with respect to its causes, namely, of knowing that we do not know. Yet while a single, general theory of criminal motivation remains elusive, we have managed, over the course of the twentieth century, to do an enormous amount of debunking. In particular, the idea that criminals "don't know right from wrong," or that they were "not raised right," or that they don't "share our values" have all been decisively rejected. Furthermore, these claims have themselves become the object of suspicion, because they all have the effect of "othering" the criminal, suggesting that there is some kind of an intrinsic or essential difference between criminals and law-abiding citizens. This seems more likely to be a motivated belief imposed by the social psychology of punishment than an accurate characterization of the underlying structure of criminal motivation.

Again, it is always possible to design a more sophisticated, cognitivist version of virtue theory, which would avoid any commitment to these discredited common-sense ideas about bad upbringing and deviant values. Yet this is an uphill battle, because the traditional Aristotelian-Thomistic theory of virtue is the primary source of these common-sense ideas in Western societies. Rather than trying to rehabilitate the old-fashioned vocabulary, there is a lot to be said for starting from scratch, with the empirical evidence, and developing a vocabulary that is better suited to accounting for what we know and understand about norm-conformity and social deviance.

12.4. The Political Critique

The third major challenge to virtue theory is a consequence of the widely held view that virtues are oriented toward the achievement of some good. In the ancient and medieval world, it was widely believed that careful deliberation would tend to generate convergence upon a particular conception of the good (or a short list of goods). Experience, however, seems to suggest otherwise. To paraphrase John Rawls, the exercise of human reason, under conditions of freedom and equality, seems to generate more, not less, disagreement about the ultimate ends of life (1993: p. xvi). Liberal societies—societies that permit open and free discussion—are therefore likely to confront what Rawls referred to as "the fact of reasonable pluralism," namely, persistent disagreement about the nature of the good life, which cannot be expected to go away simply through persuasion, better education, or improved deliberative conditions.

The major lesson that has been drawn from this observation is that virtue theory, whatever its merits as an ethical theory, is not appropriate as a public philosophy. In particular, the ancient Greek view that political life should be organized in such a way as to promote the good life, above all by cultivating virtue in its citizens, has now largely been rejected. The problem is that the law, unlike morality, is coercively imposed. If the purpose of law is to promote the good, and yet there is widespread and persistent disagreement about the nature of the good, then such an arrangement has the potential to generate a considerable amount of violence, repression, and straightforward disorder.

In the seventeenth century, responding to the problem of religious strife in Europe, early liberals decided that it was inadvisable to have the state take sides on religious questions, for precisely these reasons. Contemporary liberals have generalized the lesson, arguing that in modern societies, especially ones marked by significant ethnic and linguistic pluralism, the state should try to remain neutral with respect to all "big-picture" views, be they religious, metaphysical, or philosophical. This requires neutrality toward any "system of values" that purports to describe, once and for all, the purpose of existence, the nature of human flourishing, or any other substantive conception of the good (Kymlicka, 2002: 220–221). This is not a rejection of virtue theory per se, but rather of its political expression—the doctrine often referred to as perfectionism.

From a liberal perspective, a large part of the success of the market economy, electoral democracy, and rights-based legal regimes is that they are all institutional arrangements that are neutral with respect to a wide range of particular goods, and thereby permit concerted collective action despite an underlying heterogeneity of individual preference. Political philosophers have invested considerable time and energy debating the normative principles underlying these institutions, in an attempt to articulate principles that might reasonably claim to be neutral in the requisite sense. The principles of efficiency, equality, and liberty have emerged from these debates as particularly privileged, because each is thought to allow for a persuasive ranking of aggregate outcomes without anyone having to judge the value of the particular projects that individuals are pursuing. (Thus a more efficient outcome is one in which everyone does better, by his or her own lights. A more egalitarian outcome is one in which there is less of a difference between persons, in terms of each one's ability to carry out his or her own projects. And an enhancement of liberty occurs when each individual has greater freedom to carry out his or her projects, whatever those projects may be.)

The upshot of these debates has been the widespread conviction that it is possible to formulate robust normative principles while at the same time remaining neutral with respect to controversial questions of value. Of course, people may still object to these principles. Liberalism is, and always has been, a "fighting creed," as Charles Taylor has observed (1994: 62). Nevertheless, there is clearly a difference between public policy governed by, say, the principle of Pareto efficiency, which recommends solving collective action problems wherever they can be found, and public policy governed by a shared conception of the good, since the latter requires a much greater level of agreement than the former. So while there remain many outstanding controversies, the general force of the liberal position holds considerable sway (leading prominent virtue theorists, such as Martha Nussbaum, to accept the superiority of Rawls's "political liberalism" over its "perfectionist" rivals).

None of this applies directly to business ethics, since the latter is concerned almost exclusively with issues that arise in the private sector. Indeed, one of the major differences between business ethics and political philosophy is that the former remains strongly embedded in a "personal values" framework, as a result of which neutrality does not serve as an important constraint on normative theorizing. Standard business ethics textbooks (e.g., Boatright 2009; Beauchamp, Bowie, and Arnold 2009) still start out with a survey of normative ethical theories: utilitarianism, Kantianism, and virtue theory, then consider the question of how these should be applied. The fact that these theories are all mutually incompatible and the subject of intractable disagreement among philosophers is seldom explicitly thematized, much less addressed as a problem. And yet the aspiring management student might reasonably ask how an organization is supposed to encourage "ethical" behavior if different people have completely different values and moral theory is powerless to resolve any of their disputes.

The central insight of contemporary liberal political philosophy is that despite such fundamental disagreements, there is still a lot that can be said about what sort of normative principles should govern our institutions. However, the first step to discovering the merits of such principles lies in recognizing that one must move beyond one's "personal" moral commitments, in recognition of the fact that other people have different, incompatible commitments. Yet seldom do business ethicists take this step. This is puzzling. Virtue theory has been almost unanimously rejected as a suitable basis for running a modern state, primarily because no one can come up with a list of substantive goods that is not hopelessly controversial. And yet it is still seriously proposed as a basis for managing a modern, publicly traded, multinational corporation (e.g., Dobson 1997; Koehn 1995; Moore, 2008, 2005). What could possibly justify the discrepancy?

There are some arguments to be made here. Although there are different accounts of liberal neutrality in the literature, and why it should count as an important constraint in the political realm, the most common view is that it is tied to the state's powers of coercion. For instance, since membership in the state is mandatory, intuitively it seems unfair (or inadvisable) to privilege one group's conception of the good over another's. If it were a private association, people could just quit and join some other group whose values more closely reflected their own. But because it is the state, they do not have a feasible option in this regard. Perfectionism is however fine for private associations. Churches are obviously allowed to have a narrow conception of the good, and have, for example, some freedom to violate norms of equality in ways that would be unacceptable if undertaken by the state (e.g., the refusal of the Catholic Church to ordain women). Individuals who do not share this conception of the good are free to go find themselves another church where their values are more widely respected.

Corporations, however, present a somewhat more complicated case. Rawls himself excluded the corporation from the "basic structure" of society (1999: 126), and so did not consider it subject to the same constraints of justice that he applied to the state. He later expressed some ambivalence about this (Rawls 2001: 10). Nevertheless, there is a clear sense in which a corporation is a voluntary association. Employees can quit their jobs, shareholders can sell their shares, and so on. This suggests that managers should be able to adopt a "love it or leave it" attitude toward whatever values they seek to promote within the organization. Yet there are several features of the corporation that militate in favor of greater neutrality. These include the fact that:

(a) Corporations are created through enabling legislation. Since there are a variety of benefits conferred through incorporation that cannot be achieved through ordinary contract, the state may reasonably demand some sort of *quid pro quo* in the form of responsibility to public norms.
(b) Many public corporations are very widely traded and owned, by pension funds and other institutional investors. Thus they are not like "clubs" that can pursue the particular projects of their owners or founders. In many cases, the diversity of the society is fully reflected in the ownership group of the firm.
(c) In the post-Lochner era, the associative character of the corporation has been subject to innumerable legal constraints, from labor and civil rights law to health and safety and consumer protection regulations. Thus the law does not treat the firm as a purely private concern or as a strictly voluntary association.
(d) Many firms exercise quasi or regional monopolies, as the only employer in town, or as the owner of de facto public space. A suburban shopping mall, for instance, may be the only place where people congregate in a particular region, and a prohibition on leafleting in the mall may impair freedom of speech (McKenzie 1994: 157–159). Other firms have significant market power, creating real costs for those who want to dissociate from them. This may warrant a greater measure of neutrality.
(e) Many firms are large enough that their employees come to fully reflect the pluralism of the background society. Thus showing partiality for a particular set of traditions or values may be either discriminatory or unrealistic, given the composition of the workforce. (For example, Walmart's communal corporate culture worked best when they could hire from a large pool of rural Christian employees in the American South and Midwest; but it met with resistance and push-back when the company expanded to other regions and countries.)
(f) It is not always as easy or costless to leave one's job as it is to leave a club or church. An employee may have invested a significant amount of time developing firm-specific skills that are not transferable (Blair 1999). Facing voluntary unemployment can cause great

> hardship for both the employee and his or her family; in the United States this may even entail losing health insurance. So while the corporation differs from the state in that it cannot literally coerce employees, it may still exercise considerable power over them.

Some business ethicists may come to the conclusion that the corporation is similar enough to the state that a shift toward neutral principles should be adopted. In other cases, they may decide that the level of value-pluralism is sufficiently great that it already characterizes the situation of most firms, and so neutral principles provide the only set of feasible options for management. In that case, they may want to look to deontological theories of justice for inspiration, or for suggestions on how to respond to the fact of pluralism. Even if one does not accept Rawls's institutionalism, there are serious questions about whether business ethicists should be appealing to "thick" values and moral intuitions when they make substantive recommendations or criticisms, or whether the discipline should instead focus on principles that are broadly accepted in the public political realm. Furthermore, there is no need to accept Rawls's "difference principle," or any of his more controversial ideas about distributive justice. One of the central attractions of the "market failures" approach to business ethics is that it shows how a robust deontology can be developed by applying just the Pareto principle.

Of course, virtue ethicists are quite aware of the fact that, if they were to propose a specific list of virtues, or substantive conceptions of the good, in the way that Aristotle did, the list would immediately be buried under an avalanche of objections, from people who simply reject one or more items on the list (or point to some other culture where those items are not valued). This leaves the virtue theorist with two options. The first is to declare these objections mistaken and try to find some basis for asserting the superiority of her own preferred list. The second is to develop a more abstract formulation of the relevant virtues or goods, in order to make them less controversial (or less parochial).

It is this second strategy that is typically regarded as the more promising one. It runs the risk, however, of turning perfectionism into just a rhetorically misleading way of presenting liberalism. After all, it is always possible to take liberal principles, redescribe them in the language of "goods," then posit as "virtues" whatever dispositions are required in order to achieve those goods. This is the strategy that Daniel Weinstock refers to as "neutralizing perfection" (1999). Some virtue theorists have taken it so far that their conception of the good winds up having even less content than traditional liberal principles, such as efficiency and equality. George Sher, for example, after writing an entire book criticizing liberal neutrality, goes on to propose a list of "goods" that are so abstract that they are compatible with almost any life project. For example, he takes "the formation and execution of reason-based plans" as an "inherent good" for all persons (1997: 204). But this "good" is just another way of describing the exercise of practical reason; subscribing to it says nothing at all about which plans one should pursue (only that they should be based on reasons).

So here is the dilemma facing the modern virtue theorist: either formulate your list of virtues as dispositions to seek substantive goods, in which case they will be controversial, or else redescribe them in more abstract terms, in which case the position risks becoming just a rhetorically misleading way of presenting what amount to liberal principles of justice. A far more plausible approach, I would argue, would be to approach the corporation as essentially a "cooperative venture for mutual advantage" (Rawls 1999: 4), or an institution designed to reduce transaction costs. Cooperation does not require a shared conception of the good. And yet even if each participant hopes to benefit in a different way, they may all benefit from having a set of rules that constrain their conduct. Liberal theories of justice are essentially an attempt to articulate these rules for political society, but one can easily see how the same approach can be fruitful when thinking about the corporation and the obligations of its individual members.

12.5. Conclusion

Criticizing virtue ethics may come across as somewhat mean-spirited, particularly when put forth in an uncompromising way, as I have done here. That is because virtue ethics has an important hortatory aspect. It encourages us to become better people, to live up to our ideals of integrity, courage, and compassion. Solomon, for example, recommends virtue theory to business ethicists on the grounds that it is not about narrow definitions of right and wrong: "What is important is rather the place of a virtue (along with other virtues) in the living of a meaningful, fulfilling life, and what is important for a business virtue is its place in a productive, meaningful life in business. And this does not simply mean, How does it contribute to the bottom line? but rather, Does it contribute to the social harmony of the organization?"(1992: 193). David McPherson recommends his "vocational virtue ethic" on the grounds that "it offers a different criterion of success in our work life than simply the maximal obtainment of external goods such as wealth, power, and fame: viz., success is to be measured according to how we live up to our general calling to pursue, through the practice of the virtues, the good life for ourselves and for others" (2013: 295).

These are all no doubt laudable sentiments. How could anyone be opposed to people living meaningful, fulfilling lives, or improving the harmony of business organizations? Even if there are some doubts about whether one will succeed, it surely can't hurt to try.

Virtue ethics, however, does more than just encourage us to become better people. It introduces an entire vocabulary for thinking about moral questions, tied to a set of empirical claims about human psychology, all of which have real implications for how we go about trying to improve corporate conduct. For example, if we took seriously the claim that people's "ethical character is already firmly set" by the time they enter the workforce, then the best way to create an "ethical organization" would be for corporations to invest resources in screening prospective employees, in order to avoid hiring those who have a vicious character. And yet modern psychological research suggests that these sorts of initiatives would be useless. The underlying idea is based on a set of empirical claims about moral motivation that have essentially been shown to be false. Thus virtue theory encourages us to invest time and energy in projects that are likely to have no positive effect. At the same time, it tends to distract attention away from interventions that could prove useful.

Suppose, for example, that you are a manager at an investment bank. You notice one day that your traders routinely describe making money at the expense of a client as "ripping his face off" (Johnson and Kwak 2010: 114). It turns out that this has become part of the office culture, a standard way of talking. Given the delicate issues that are raised by banks taking positions opposite their clients (and executing both sides of the trade), you might worry that this kind of talk could lead to unethical or even illegal behavior. Or you might not. You might even defend it, perhaps along the following lines:

> I'm not particularly concerned, my traders are fundamentally good guys, decent family men, they know where to draw the line.
>
> Don't worry about the way they talk. That's superficial, they're just having some fun, blowing off steam. What really matters is their character, their underlying sense of what's right and wrong.
>
> Our firm has been providing outstanding service to our clients for over a century, and the people that we hire share the values of the organization.

If you were to say any of these things, you would be making a terrible mistake. Modern moral psychology tells us that in dismissing concerns in this way, you would be ignoring something that makes a huge difference and appealing to all sorts of things that actually make no difference. The question is, does thinking about the issues in terms of virtue tend to highlight the features of the situation that are the most important from a moral perspective? Or does it tend rather to distract us from them? And therefore, would thinking about business ethics in virtue-ethical terms make you more or less likely to make this mistake?

Whatever our sentimental attachment to the theory, it is time to recognize that Aristotle's psychology is just as outdated as his biology and his physics. It is perverse to hold on to an outdated theory, when doing so prevents us from actually taking effective action to create a more ethical culture. As Doris has observed, "Rather than striving to develop characters that will determine our behavior in ways substantially independent of circumstance, we should invest more of our energies in attending to the features of our environment that influence behavioral outcomes" (2002: 146). This is particularly wise counsel for business managers, who are in the privileged position of actually controlling many important features of the work environment, and are therefore able to have a significant impact on "behavioral outcomes." As Lynn Sharp Paine has argued, "Rarely do the character flaws of a lone actor fully explain corporate misconduct. More typically, unethical business practice involves the tacit, if not explicit, cooperation of others and reflects the values, attitudes, beliefs, language, and behavioral patterns that define an organization's operating culture. Ethics, then, is as much an organizational as a personal issue" (1994: 106). Furthermore, in order for managers to figure out how they should structure the work environment, they need not appeal to any "deep" conception of the good. They need only consider the normative principles that are already implicit in the basic institutional structure of a market economy and of the corporation as an economic actor.

13 }

Reasonable Restrictions on Underwriting

There are very few issues in business ethics that are as polarizing as the practice of risk classification and underwriting in the insurance industry. Not everyone who seeks indemnification against a particular loss faces the same probability of suffering that loss, or faces a loss of equal magnitude. Thus insurers typically try to ascertain both the magnitude and probability of the loss for which a prospective policy-holder seeks indemnity, in order to determine an appropriate premium level. The ideal is to charge each policy-holder the so-called "actuarially fair" premium, which represents the anticipated cost of compensating that individual for the loss, multiplied by the probability that the loss will occur during the term of the policy. In reality, insurers are often unable to determine the risks that each individual faces. Thus they use a system of more-or-less broad classification, in order to determine which "risk pool" or class an individual falls into. This is used to determine a base premium, which is then "topped up" to cover transaction costs, commissions, and possibly—but not necessarily—profit.

Theorists who approach these insurance practices from a background in economics or business often start from the assumption that the actuarially fair premium represents the most "equitable" (Bossert and Fleurbaey 2001) or "just" arrangement, such that any deviation from actuarial fairness requires justification.[1] Rates are considered "unfair" if "the insured is overcharged for the loss exposure in comparison with another similar loss exposure" (Outreville 1998: 149), but there is no question that individuals who present different loss exposures should be charged different premiums. On the other hand, theorists who approach the question from a background in philosophy or civil rights law often begin with a presumption against actuarially fair premiums

[1] In their paper "Equitable Insurance Premium Schemes," Bossert and Fleurbaey write: "The basic assumption underlying our analysis is that the most equitable insurance premium scheme is one where the premium paid by each agent is equal to the expected value of the payout of the insurer to this agent" (2002: 114). This assumption is not defended; it is taken as the point of departure.

and in favor of so-called "community rating," in which everyone, no matter what their background risk profile, has access to the same insurance policy at the same price (Daniels 1991; Austin 1983). Deviations from this baseline are then regarded as standing in need of justification. As a result, the entire practice of underwriting is presented as morally suspect. Tom Baker, for instance, describes the idea of actuarial fairness as "a watered-down form of liberalism that privileges individual interests over the common good and that privileges, above all, the interests of insurance institutions organized on its terms" (2003: 277).

Critics of risk classification (or more tendentiously, "statistical discrimination") have derived considerable support from a series of Supreme Court decisions in the United States that disallowed categorization according to sex in defined-benefit pension schemes. The insurance industry also suffered a series of public-relations disasters associated with its underwriting practices, particularly in the United States where the absence of comprehensive public health insurance has made conditions of access to private health insurance a highly charged moral and political issue. This became especially apparent in the early stages of the AIDS epidemic, when insurers began refusing coverage not just to individuals who had contracted the HIV virus but also those with a record of having been tested for it (on the grounds that only people who engaged in high-risk behavior would elect to test themselves). A similar furor erupted in the 1990s when it was discovered that over half of American insurers routinely denied health, life, and disability coverage to battered women, on the grounds that victims of domestic abuse had an adverse claims history. These episodes resulted in legislation in several American states imposing restrictions on the "freedom to underwrite" of insurers, preventing insurers from requesting certain types of information from prospective policy-holders, or else directly prohibiting them from charging different premiums to members of different groups (Hellman 1997; Austin 1983).

The result has been the development of considerable inconsistency in public policy. Certain forms of risk classification are prohibited for certain types of insurance, but not others. No general legal principles have been developed to govern the practice either, in part because the ideal of actuarial fairness is rejected by many as inherently discriminatory or unjust. My goal in this paper will be to critically evaluate the latter claim. I begin by examining three different arguments that have been given, purporting to show that the practice of charging actuarially fair premiums is inherently unjust. I will try to show that each of these arguments is informed, in one way or another, by an essential misunderstanding of the mechanism through which insurance serves as a source of cooperative benefit. I go on to consider the two primary arguments that have been offered by those who wish to defend the practice of risk classification. These arguments, I will argue, overshoot their target, by demanding a "freedom to underwrite" that is much greater than the level of freedom enjoyed in most other commercial transactions. Thus, I conclude by presenting an outline and defense of a somewhat more limited "right to underwrite," one that grants the legitimacy of the central principle of risk classification, but permits specific deviations from that ideal when other important social goods are at stake. This in turn allows us to develop relatively precise criteria for determining what constitutes a reasonable restriction on underwriting.

13.1. Risk pooling

The world is full of risk. Knowing the probability of various events is extremely useful when it comes to engaging in practical deliberation. Unfortunately, what matters to most of us when we make our plans is not the background probability of an event, but the actual frequency with which it occurs. We know that a fair coin has a 50 percent probability of landing heads, but we also know that flipping it 10 times is quite unlikely to produce exactly 5 heads and 5 tails. As a result, we need to be concerned not just with the mean, but also with the variance—how far individual outcomes can be expected to deviate from the mean, and how often. However, it is also well known that as the number of tosses increases, the frequency will tend to converge with the probability (a phenomenon often referred to as "the law of large numbers"). In other words, increasing the number of trials induces *statistical stability* (Hacking 2002: 190–192); it decreases the variance in the distribution. This increase in stability is the central mechanism through which insurance schemes are able to produce welfare benefits for their members.

To see how a group of individuals can benefit from the law of large numbers, it is important to remember that individuals are often risk-averse. Consider a farmer who under normal conditions is able to produce 10 tons of grain—enough to feed his entire family well throughout the winter. However, his land is also subject to a highly localized blight, which sometimes wipes out the entire crop. Suppose that the chances of this blight striking his field in a given year are 20 percent. Although the expected annual output of his field is therefore 8 tons, he would gladly swap a guaranteed revenue of 8 tons for the gamble that he faces between 10 tons or nothing. That way, his family would have a bit less to eat, but they would never risk starvation.

On his own, this is something that he cannot achieve. Suppose, however, that there are 100 small farmers who find themselves in identical circumstances, all facing the danger of this highly localized blight. They might agree to a "risk-pooling" arrangement, under which farmers who lose their crop in a given year are compensated by those who do not. Under this arrangement, the objective risk of blight does not diminish: 20 of the 100 farmers can, on average, expect to lose their crops. However, with the risk-pooling arrangement, each farmer can expect a revenue that will be, with 95 percent probability, between

7.2 and 8.8 tons (see Moss 2002: 28–31). Because the farmers are risk-averse, this gamble has greater *subjective utility* than the gamble that gives each individual farmer an 80 percent chance of getting 10 tons and a 20 percent chance of nothing (even though the two gambles have the same mathematical value).

Thus what insurance offers is a form of superior "risk management," but not necessarily "risk abatement." It does not eliminate the loss but just redistributes it. Of course, the insurance arrangement has the agreeable consequence of preventing anyone in the community from starving. But it is important to keep in mind that this is not why people buy the insurance. Insurance is not charity. They buy insurance in order to reduce their own uncertainty. If some farmers happened to like the gamble, and thought that it was worth risking starvation in order to get a shot at the full 10 tons of grain, then they would have no incentive to join the insurance scheme.

It is important to note as well that the risk-pooling arrangement is neither a gain from trade nor a straightforward economy of scale, but rather a *sui generis* source of collective benefit. Theorists sometimes mistakenly assimilate the gains that come from *trading* risks with those that come from *pooling* risks (e.g. see Barr 1998: 111–112; Easterbrook and Fischel 1991: 53). In the former case, two individuals with different levels of risk aversion can generate efficiency gains by exchanging a risk—specifically, the one who is most risk-averse can pay the other, in return for a promise of indemnity in the case of an outcome that it is too far from the mean. Here, the welfare gain is possible only because one person is less risk-averse than the other. In the case of insurance, however, people with the same level of risk aversion and risk exposure can still benefit from the "law of large numbers" mechanism, by agreeing to hold the risk in common. This is the typical arrangement within a mutual society, which was the dominant model of (non-commercial) insurance in the twentieth century (Hansmann 1992). Here there are no investors or stockholders, the company is owned by the policy-holders, and all of the money paid out in claims is simply levied from the policy-holders in the form of premiums.

The mutual society also provides the best model for examining the merits of actuarially fair premiums. Because there are no investors, the controversial issue of "profit" is taken off the table. Premium levels are determined by "the insurance company," but the company in this case is simply a group of managers appointed to act as agents of existing policy-holders. In this case, it is not difficult to determine how premiums should be calculated. Take the example of 100 small farmers above. One can start by imagining a meeting at which all the potential policy-holders get together in order to determine the terms of the insurance arrangement. (This is not so fanciful, since many early mutual societies did originate in this way, as witnessed by the fact that publicans were often the founders and record-keepers for early "friendly societies" [Neave 1991: 51].) Together the farmers can be confident of producing close to 800 tons of grain, while losing 200 tons to the blight. Each farmer who joins the pool can therefore receive a guarantee of 8 tons of grain, in return for a commitment to contribute all that he is able to grow to the pool (either 10 tons, or 0 tons, depending upon how things work out). This is equivalent to keeping one's own crop and paying a premium of 2 tons into the insurance pool. (When 80 farmers contribute 2 tons each, it generates the 160 tons needed to indemnify the 20 farmers who lose their crops.) Thus what each farmer pays in the way of a premium is equal to the expected loss that he brings to the pool, namely, a 20 percent probability of drawing 8 tons while contributing nothing.[2]

Of course, the assumption so far is that each farmer is identically situated—having the same amount of land, the same level of productivity, and experiencing the same probability of suffering from the blight. But what if one of the farmers happened to have a plot of land that was twice the size of anyone else's? If he sought to insure his entire crop (i.e., sought a guarantee of 16 tons), it stands to reason that his premium should be higher. Indeed, the natural thing would be to charge him a premium that was twice as high, in reflection of the fact that the magnitude of the loss that he may impose upon the other members of the insurance pool is twice as large. But similarly, if a farmer had a plot of land that for some reason was twice as likely to be struck by the blight, then it would also be natural to charge him a higher premium. Indeed, the natural thing would be to charge him a premium that was twice as high. A 40 percent probability of drawing 8 tons while contributing nothing represents the same expected loss as a 20 percent probability of drawing 16 tons while contributing nothing.

Thus the reasoning that leads to higher premiums for people seeking indemnity for losses of greater magnitude directly parallels the reasoning that leads to higher premiums for people who present a greater risk of loss. Those who expect to take more out of the pool should be required to put more in. This is the principle that underlies the idea of actuarial fairness, which simply stipulates that the premium paid by an individual should be equal to that individual's expected loss (magnitude multiplied by the probability).

This proposal appears simple, and one can certainly imagine it serving as a basis for agreement in an initial meeting at which individuals get together to form an insurance pool. There are a number of "real world" complications that arise, however, when it comes to determining just what the probability of a given loss is for a given individual. Many philosophers in fact think there is no "fact of the matter" as to whether a particular event, taken all by itself, can be said to occur with some probability. Either it happens or it does not. Probabilities belong only to classes of events, as a function of the frequency with which they occur, or as a function of our ability to predict them, based upon the frequency with which they occur (Hacking 1975). Thus the only way

[2] For a more formal calculation of the "pure premium," see Outreville (1998: 156).

to determine the probability of an individual's loss is to pick out some sort of frequency to which it belongs. With respect to certain events, the individual's own history may provide a sufficient record (so that the insurer is able to use so-called "experience rating" to determine that individual's premium). More often, the individual's own history is inadequate, and so insurers seek to establish a "class rate," by finding a larger group to which that individual belongs (Outreville 1998: 150–151) and seeing what the loss frequency is within that group. For example, a person who has just learned how to drive has no safety record, and so no basis for estimating his or her chances of having an accident. An insurer may notice, however, that young men have an accident rate that is significantly higher than that of young women, or that single men have a much higher accident rate than married men (Dahlby 1983). Thus the insurer might respond by classifying individual policy-holders into such groups and using the frequency of losses among members of the group as a way of determining the expected loss that the individual brings to the insurance scheme.

This is where things get controversial. Class rating may seem like a form of guilt by association. Just because other young men are terrible drivers does not mean that this particular young man is going to be one. But of course, this is not a very helpful observation. If we knew how things were going to turn out in the end, then there would be no need for insurance. All that we have to go on in designing insurance contracts is the ex ante perspective. The way to think about the fairness of premium schemes is to imagine all of the policy-holders getting together in an initial meeting, in order to create the insurance pool. In principle, they are not obliged to do business with anyone, and so are not obliged to admit into the pool anyone that they do not want. Furthermore, the terms under which individuals are to be admitted are entirely up for negotiation. Under such conditions, if the best available information indicates that young men pose a particularly elevated risk, then these young men are going to have to offer more in the way of premiums in order to secure admittance.

Of course, there are a number of important "second-best"[3] problems that arise when it comes to implementing actuarially fair premium schemes. In most cases, insurers will not have all the information that is needed to determine the expected loss that an individual brings to an insurance pool, and therefore cannot actually calculate the actuarially fair premium. They are left having to approximate that premium, using the best information available. Yet one cannot assume that refining the partition of the insurance pool using new information will necessarily bring the premiums that all individuals pay closer to the actuarially fair level. Thus there is room for significant injustice to arise

[3] The term comes from the celebrated paper by Lipsey and Lancaster 1956, which showed that if it was impossible to satisfy one of the conditions needed for a competitive market economy to achieve Pareto efficiency, then strict adherence to the remaining conditions would almost certainly make the outcome less, rather than more, efficient.

out of attempts to implement an actuarially fair premium scheme in a world of imperfect (and asymmetric) information (Abraham 1986: 86; Promislow 1987: 216). For example, when insurers decided to deny health and disability insurance to battered women, most did not distinguish between those who continued to live with their abusive partner and those who had ended the relationship (Hellman 1997: 361). Members of the latter group may well have been disadvantaged by the fact that information injurious to their risk rating was easily available (police reports, hospital records, etc.), while potentially exculpatory information (present living arrangements) was either unavailable or unverifiable.

Yet these types of "second-best" problems are not where critics of actuarial fairness have focused their energies.[4] The most important arguments have all been directed against the principle itself. Critics argue that actuarially fair premiums are inherently unjust, and thus not even what insurers should be aiming for. The discussion that follows will therefore focus upon these sorts of principled objections, setting aside all "second-best" considerations. This is not to suggest that problems arising at the level of the "second-best" are not serious—in many cases they are enormously so. The discussion will be restricted to the principle of actuarial fairness simply because the arguments about "second-best" problems cannot really begin before it is established what a "first-best" solution looks like, and this question is still subject to enormous controversy.

13.2. Straightforward discrimination

The most damning criticism of risk classification and "class rating" is the claim that it represents plain old-fashioned discrimination. It penalizes certain individuals, not for their individual characteristics, but merely because of their membership in a group. This is, in effect, what the US Supreme Court decided in *City of Los Angeles Department of Water and Power v. Manhart* (435 U.S. 702 [1978]), with respect to the use of sex-segregated actuarial tables in the determination of contribution levels to a defined-benefit pension plan. These types of pensions are essentially insurance products (with the payroll contributions being the premiums). Like life annuities, they generate a stream of fixed payments until the death of the beneficiary, thus providing insurance against the risk of outliving one's retirement savings. (Otherwise put, they represent an arrangement under which multiple individuals pool their retirement savings in order to reduce uncertainty.) Since women on average live longer than men,

---

[4] For a sober discussion of the basic "second-best" problems that arise, and some principles to guide insurers, see Abraham (1986: 64–100).

the expected value of a typical pension of this type is of greater value to women than to men at the time of retirement. Thus the employer in this case created a system of differential contribution levels for its employees, with women paying a higher premium than men. The Supreme Court ruled that this was a violation of Title VII of the Civil Rights Act of 1964, which prohibits discrimination against "any individual because of his race, color, religion, sex, or national origin" by employers.[5]

In order to reconstruct the court's reasoning, it is helpful to introduce a couple of distinctions, implicit in the judgment, but drawn out more explicitly in an influential gloss on that judgment by Brilmayer, Hekeler, Laycock, and Sullivan (1980). In civil rights law, there is an important distinction between permissible and impermissible grounds for discrimination, and between disparate treatment and disparate impact. It is permissible for employers to treat employees differently, according to some characteristic that they possess, provided that they are able to demonstrate a "business necessity" for so doing. For example, it has been deemed permissible for employers to require that candidates pass a weight-lifting test in order to be considered for certain heavy manufacturing jobs (*Bowe v. Colgate-Palmolive Co.*, 416 F.2d 711 [7th Cir.1969]). This, of course, will have a disparate impact upon women, since women are on average able to lift less than men. But this does not count as disparate treatment of women, because the effect is indirect, and is a consequence of a system of discrimination based upon permissible grounds. If, however, the employer were to refuse to consider women for such jobs on the grounds that they are less likely to pass the weight-lifting test, this would constitute impermissible discrimination. The mere fact that some characteristic is statistically correlated with a characteristic that serves as permissible grounds for discrimination does not make it permissible for the employer to use the former as grounds for discrimination as well. In other words, the characteristic "being

---

[5] It is worth noting that Title VII discrimination is not the only form of discrimination that underwriting practices might run afoul of. The more obvious suggestion might be that the practice of charging different policy-holders different premiums was a form of price discrimination. This is prohibited in the United States by the Robinson–Patman Price Discrimination Act of 1936, which makes it unlawful "to discriminate in price between different purchasers of commodities of like grade and quality." This is not a powerful argument, however, because it is not difficult to make the case that insurers are selling a different product to clients in different risk-classes—since the actuarial value of each policy is different, depending upon the risk profile of the individual and the magnitude of the loss. Furthermore, it should be noted that there is considerable tolerance for price discrimination in the market. The Robinson–Patman Act only targets forms of price-discrimination that have anticompetitive consequences. Practices such as price-skimming—when new goods are introduced at inflated prices, and then dropped over time—are considered quite normal, even though they amount to forms of price discrimination. They are usually regarded as morally unproblematic, simply because those who wind up paying the higher prices are generally those who derive the largest welfare benefit from the purchase. This is also clearly true in case of insurance, since high-risk individuals get more value out of the policies that they purchase.

permissible grounds for discrimination" is not preserved through probabilistic inference. This is as it should be, since an arrangement under which women were excluded from certain manufacturing jobs on the grounds that women in general are less likely to be able to lift heavy weights is clearly unfair to those women who are able to lift such weights, and thus do possess the relevant job qualification (however unlikely this may be ex ante).

According to the Supreme Court's reasoning in *Manhart*, charging women more for their pensions on the grounds that they are less likely to die young is like excluding women from certain classes of employment on the grounds that they are less likely to be able to lift heavy weights. According to Brilmayer et al.:

> American women as a group currently live longer than American men as a group, just as they are able to lift less weight as a group. But some women will die earlier than some men, just as some will be able to lift more weight. An employer who pays annuities on the basis of integrated tables in effect distinguishes among his employees on the permissible basis of longevity, for those individuals who live the longest will collect the most periodic payments and thus the largest total sum. Of course, the employer's practice may have disparate impact on men, for as a group they may not live to collect as many periodic payments as women. If he tries to avoid this disparate impact by using segregated tables—making larger periodic payments to all men as a group—he distinguishes on the basis of sex. This would be disparate treatment, for individual men and women of equal longevity would be treated differently: both periodic benefits and total benefits will be greater for a man than for a woman of equal longevity. (1980: 510–511)

Central to this argument is the idea that "longevity" in this case constitutes the permissible basis for discrimination. Thus equality requires that all employees receive the same ex post net benefit from the pension scheme unless they differ with respect to longevity (just as all job applications for a position in heavy manufacturing must be considered equally unless they differ with respect to weight-lifting ability). Naturally, using longevity as grounds for discrimination has disparate impact on men, just as a weight-lifting test has disparate impact on women. This is permissible. However, the employer is not entitled to use sex as a predictor of longevity, in order to determine pension benefits, because the former is merely statistically correlated with the latter. According to the Court's ruling, even though women are more likely to receive more periodic payments after retirement, "there is no assurance that any individual woman working for the Department will actually fit the generalization on which the Department's policy is based. Many of those individuals will not live as long as the average man. While they were working, those individuals received smaller paychecks because of their sex, but they will receive no compensating advantage when they retire" (*Manhart*, 435 U.S. 702, 708 [1978]). Thus sex-segregated actuarial tables violate equality, by creating a situation in which a man who lives to the same age as a woman would pay less for the pension benefits received.

This argument is ingenious, and at first glance also seems compelling.[6] And even though it is limited in scope from a legal point of view (to relations between employers and employees, and with respect to only the enumerated categories of discrimination), the moral implications of the argument are much broader. If sound, the argument shows that actuarially fair premiums in general violate the principle of equality. It suggests, for instance, that automobile insurance companies that charge drivers of red cars higher premiums than drivers of beige cars (of the same make and model) violate equality. In this case "having an accident" constitutes permissible grounds for discrimination. Yet risk-rating in accordance with color means that people who drive beige cars and do have accidents pay less for the same benefit received as drivers of red cars who have accidents. Why should we reward people just for driving beige cars?

Yet the analogy that the argument depends upon is clearly strained. First of all, one can see the sense in which individuals are "rewarded" for their weight-lifting ability by being given access to an employment opportunity, but it is odd to think of a defined-benefit pension scheme as "rewarding" individuals for longevity (as thus of longevity as a "permissible grounds for discrimination"). What is the point of discriminating on this basis? It is difficult to avoid the impression that the purpose of the pension scheme is being misdescribed. Second, there is the fact that, in the case of weight-lifting, it is the negative correlation between being female and the property that is being rewarded that motivates the discrimination against women. Yet in the case of pensions, there is a positive correlation between being female and the property that is being rewarded. So if the goal was actually to use sex as a predictor of longevity, and people were being paid more for living longer, then that should have led to an arrangement under which women were charged lower premiums than men. Here we can see the most serious problem with the framework that the Court used to describe the issue: it is unable to make any sense of the idea that sex is being used as a predictor of longevity. This in turn generates a serious misunderstanding of the purpose of segregated actuarial tables.

The employer in *Manhart* unfortunately muddied the waters by suggesting that the rationale for the sex-segregated tables was fairness to its male employees as a class. Instead of having each individual who lives a given number of

---

[6] It is far more persuasive than the arguments of Austin 1983, who merely points to the fact that risk classification will have disparate impact as a way of showing that the practice involves unjust discrimination.

years receive a net benefit of equal magnitude, they suggested that the goal was to have men as a class receive net benefits that were of equal (average) value to those received by women as a class. But this is clearly a terrible argument. It suggests, as the Court rightly saw, that proponents of sex-segregated actuarial tables did not want to use sex as a predictor of longevity, but rather wanted to add sex to longevity on the list of legitimate bases for discrimination. Under such an arrangement, each individual would get the same net benefit as each other individual, unless they differed in longevity or sex. But what could possibly motivate adding sex to the list of discriminators, since it obviously results in unequal treatment of individuals? The suggestion that it was being done in order to achieve fairness to classes—so that, in the aggregate, men receive the same average net benefit as women—is a strange rationale. Under such an arrangement, women would be forced to contribute more to the pension scheme, not because they were expected to live longer (this is the rewarded property!), but merely because they were women. This is plain old-fashioned discrimination. The Court quite rightly observed that the goal of civil rights legislation is to protect individuals from this sort of treatment, and that the language of the statute explicitly prohibits it.

The problem with the judgment lies in the framing of the question. It starts with the way that the principle of equality is applied (both by the Court, and in the more perspicuous argument of Brilmayer et al. [1980]). The conflict is not one between equality for individuals versus equality for classes. The relevant contrast is between equality ex ante and equality ex post. Consider the situation in which one uses a randomizing device, like a coin toss, in order to allocate an indivisible good between two individuals. A proposed distribution that gives each individual a 50 percent probability of getting the indivisible good creates a situation that is equal ex ante, but of course, the distribution that results from the coin toss (i.e. ex post) seems quite unequal, since one person gets the entire good and the other gets nothing. What makes this final distribution acceptable is the fact that the expected value of the lottery ex ante was exactly (or roughly, depending upon the *equalisandum*) the same for both individuals. The Supreme Court's approach in *Manhart,* on the other hand, would have us saying that the distribution is actually equal ex post, except that "winning a coin toss" represents a permissible ground for discrimination.

Numerous critics of *Manhart* have pointed out that the analogy between weight-lifting (or height) and longevity is faulty, because in the case of weight-lifting one can simply do a test to see how much a person can lift, and so there is no reason to rely upon the statistical correlation between sex and weight-lifting ability. In the case of longevity, on the other hand, there is no way of checking to see when a person will die (Kimball 1979: 118). Indeed, if it were possible to do so, there would be no reason for insurance in the first place—each person could simply save exactly as much as he or she required for retirement (Kimball 1979: 133). Yet these critics have failed to articulate the full force of this objection. What the element of uncertainty means, in the case of insurance, is that the principle of equality must be applied ex ante. Since there is no way to guarantee that the *ultimate benefit* of entering into an insurance scheme will be the same for all individuals (if there were, people could just save), we must ensure that the *expected benefit* be the same for all.[7] This is precisely what the actuarially fair premium represents—each individual pays a premium sufficient to cover the expected loss that her participation in the insurance scheme brings. Thus actuarially fair premiums are not motivated by a commitment to equality for classes rather than for individuals, but rather by a commitment to ex ante equality for individuals.

Thus, contrary to the Court's ruling, the point of risk classification and underwriting is not to ensure that each risk class receives an equal benefit, but rather to ensure that each individual receive an equal expected benefit. The strange idea that "longevity" constitutes permissible grounds for discrimination in pension schemes should be rejected (as should the idea that "having an accident" constitutes permissible grounds for discrimination in automobile insurance). Both ideas are a consequence of the mistaken attempt to apply the principle of equality ex post rather than ex ante.

Of course, there are a number of complicated "second-best" issues that arise with respect to the use of crude partitioning devices, like sex, to estimate the expected loss that an individual brings to an insurance arrangement. These are not at issue here. What is noteworthy about the reasoning of *Manhart* is that it attacks the basic principle of actuarial fairness, claiming that such premium schemes violate equality even under optimal conditions. The discussion here is intended to show that this argument is based upon a misapplication of the principle of equality. There may still be cogent arguments to be made against the use of sex-segregated actuarial tables for defined-benefit pension schemes. It is, however, a mistake to think that the use of such tables is a case of plain old-fashioned discrimination.

13.3. Choice and circumstance

The second major argument against risk classification and underwriting is based upon the moral intuition that it is unfair to penalize individuals for

[7] Kimball (1979) overstates the case somewhat, arguing that what employees receive in return for their contributions is simply risk-protection (i.e., the welfare benefits associated with the insurance scheme), and so it does not matter what benefits, if any, they ultimately receive. This involves a rather excessive disregard for the benefits. Insurance allows individuals to exchange one gamble for another, less risky one. It is, however, the (at least rough) mathematical equivalence of the two gambles that makes the individual willing to enter into the exchange. Thus the person who acquires an annuity is not just purchasing risk-protection, he is also purchasing an income stream that must be approximately equal to that achievable through savings (i.e., self-insurance).

circumstances that are outside of their control, or for things that are not their fault (Daniels 1991; Hellman 1997; Abraham 1986: 89–92). According to this view, it is acceptable for insurers to penalize a driver with a history of moving violations by charging him a higher premium—after all, he has the option of improving his driving habits—but it is unacceptable to penalize a young man with higher premiums merely because young men in general have bad driving habits. Similarly, it is thought reasonable to penalize smokers by charging them higher premiums for home insurance, but not people who live in high-crime neighborhoods. And, of course, since people have no control over their sex, race, age, or for the most part, health status, insurers should be prohibited from charging differential premiums on the basis of such characteristics.

Thus, Norman Daniels, in an influential article on health insurance, claims that the argument for actuarially fair premiums rests upon a "controversial premise," namely, "that individuals should be free to pursue the economic advantage that derives from any of their individual traits, including their proneness to disease and disability" (1991: 504). (The idea that individuals should not be penalized for their circumstances, in an insurance context, is equivalent to the idea that individuals should not be advantaged by their circumstances, since it is the relative premium level that determines what counts as a penalty or an advantage.) Thus Daniels (1991) imputes the following argument to proponents of actuarial fairness:

1. Individual differences—any individual differences—constitute some of an individual's personal assets.
2. People should be free, indeed are entitled, to gain advantages from any of their personal assets.
3. Social arrangements will be just only if they respect such liberties and entitlements.
4. Specifically, individuals are entitled to have markets, including medical insurance markets, structured in such a way that they can pursue the advantages to be derived from their personal assets (1991: 505).

Daniels goes on to point out that this argument constitutes a direct statement of the basic premises underlying Robert Nozick's (1974) libertarianism, which is a highly controversial political philosophy. Many others, including John Rawls (1971) and Ronald Dworkin (2000), believe that the outcome of the "natural lottery" is an effect of brute luck, not desert, and so individuals have no moral entitlement to benefit from their natural endowment. According to G. A. Cohen's influential formulation of this thesis "a large part of the fundamental egalitarian aim is to extinguish the influence of brute luck on distribution" (1989: 931). Anyone who shares this intuition should be unmoved by the argument for actuarial fairness, Daniels claims. Indeed, this is reflected to some degree in current employment legislation, he argues, where "we believe that justice requires us to sever consideration of race, sex, or handicaps from deliberations about hiring, firing, and reimbursement for services performed, although in practice we fall far short of what justice demands.... Thus we reject, in its most general form, the view that all individual differences can be a moral basis for advantage or disadvantage" (Daniels 1991: 506).

This is, however, a very odd reading of current anti-discrimination law. Daniels is essentially claiming that actuarially fair premiums are unjust because they conflict with luck egalitarianism, which is the most widely shared liberal conception of justice. The prima facie difficulty with this argument is that it appears to hold insurers to a higher standard of ethical conduct than any other business enterprise. Individuals routinely benefit from their natural endowments (intelligence, beauty, creativity, etc.) or from brute luck (plentiful rain, an early frost, a change in interest rates, etc.) in market transactions, and we think nothing of it. The idea that individuals should only be penalized for their choices and not their circumstances may be part of some luck-egalitarian ideal, but it is not part of what Christopher McMahon (1981) has called "implicit morality of the market."

The reasons for this are not hard to find. The task of carrying out the luck-egalitarian project of indemnifying individuals against the effects of bad brute luck will in many cases require pure redistributive transfers—that is, win–lose transformations. Thus any business arrangement (including an insurance scheme) organized along luck-egalitarian principles could leave individuals worse off than if they had never chosen to participate (or bought insurance) at all. Such an arrangement might not even offer them the *prospect* of being better off. Thus in the absence of altruistic preferences, such an arrangement cannot emerge as a result of private contracting. Private contracting is only feasible when there is at least an ex ante Pareto improvement. Why would individuals sign up to pay for someone else's misfortune? This is Good Samaritanism. There may be a moral case to be made for such behavior, but to propose such principles as a basis for the legal regulation of the marketplace is essentially to argue that there should not be a marketplace.

Thus what Daniels has actually produced is not really an argument for restrictions on underwriting by private insurance companies, but rather a general egalitarian argument for social insurance in the health care sector. He observes that "the design of health-care systems throughout most of the world rests on a rejection of the view that individuals should have the opportunity to gain economic advantage from differences in their health risks" (1991: 507). But this is precisely why the health care systems that he refers to are operated in the public sector. The fact that there is a strong case to be made for the state to deliver a particular type of service, in accordance with certain principles of distributive justice, does not mean that there is an equally strong case to be made for the state to compel the private sector to deliver that service under the same terms. Most welfare states also offer defined-benefit pension plans that are financed in accordance with principles that impose non-actuarially fair premiums (often aimed at producing a more progressive distribution of retirement income). But this does not mean that the state would be justified in imposing progressive payment schedules on private pension plans.

Daniels acknowledges that his argument may simply militate in favor of public insurance (1991: 518). However, he also wants to suggest that the basic luck-egalitarian principles, which require "community rating" in private health insurance, are not entirely foreign to the marketplace, and that many other enterprises are subject to similar restrictions. This is why he claims an analogy between restrictions on underwriting and anti-discrimination law in other areas of private contracting, and why he argues that anti-discrimination law is based upon a rejection of "the view that all individual differences can be a moral basis for advantage or disadvantage" (1991: 506). But in order to see the problem with this claim, one needs look no further than the *Manhart* judgment. The court had no trouble with the idea that height, or weight-lifting ability, or longevity, could count as permissible grounds for discrimination or advantage, even though individuals have very little control over these characteristics. Even sex has been ruled a permissible grounds for discrimination if the employer can show that being of one sex or the other is itself actually necessary for performance of the job. Thus anti-discrimination law does not follow a luck-egalitarian logic. What the law restricts employers from doing is using criteria that are not relevant to job performance as a basis for discriminating against individuals. And this is clearly not what insurance companies are doing, when they practice risk classification and underwriting.

Daniels actually sums up the problem with his own view when he writes that the argument for actuarial fairness and the practice of denying coverage to high-risk individuals "is persuasive only if the important function of health insurance is risk management. Because health insurance has a different social function—protecting equality of opportunity by guaranteeing access to an appropriate array of medical services—then there is a clear mismatch between standard underwriting practices and the social function of health insurance" (1991: 514). This may be true, but one could just as easily argue that current pricing practices in the grocery industry are acceptable only so long as one thinks that the function of that industry is to sell food to people—however, since the true social function of the grocery industry is to protect equality of opportunity by guaranteeing adequate nutrition for all citizens, there is a clear mismatch between the practices of the industry and its social function. As one can see, this is not an argument for changing pricing practices in the grocery industry; it is an argument for socializing the grocery industry.

When the argument is formulated with respect to groceries, the problems with it become immediately apparent, in a way that they do not when it is formulated with respect to insurance. This is because of a widespread misunderstanding of how the insurance industry works, which leads many people to think that the industry does have a "social function" that extends beyond mere risk management. In particular, it is widely thought that the goal of insurance is not merely to socialize risk, but rather to socialize the actual losses against which individuals seek indemnity. Even very knowledgeable commentators are prone to confusion on this score. Carol A. Heimer, for example, writes that "at its most basic, insurance is a social arrangement to reduce the effects of losses by employing the resources of the group to cushion individuals. The key task of insurers is to organize the insurance pools, turn them into groups with a common fate, and act as agents of these groups" (2003: 288). Deborah Hellman argues that legal restrictions on underwriting are desirable, on the grounds that they represent "a first step toward treating the misfortunes of poor health and disability as communal responsibilities" (1997: 359).

But the goal of private insurance is not to pool losses. That would require altruism as an economic incentive on the part of a large number of the participants in the insurance pool. Why would one person want to take on someone else's loss? The fact that socialization of losses generates a group benefit does not mean that it generates a benefit for each individual. The reason people sign up for insurance is because of risk aversion, and because they seek to reduce subjective uncertainty. The reduction in uncertainty is achieved through a socialization of losses, but the latter is instrumental to the former, it is not the objective of the arrangement. Thus participants in an insurance scheme will not accept any sort of socialization of losses, only ones that are conducive to the management of certain risks, and on terms under which the welfare benefits stemming from the reduction in subjective uncertainty outweighs the cost of having to indemnify others for their losses.

Of course, there are cases in which there is an argument to be made for socialization of certain losses. An epidemic disease such as AIDS or SARS provides a powerful example. Daniels himself describes the creation of mandatory insurance for high-risk drivers to be a case where "our social interest in guaranteeing a public good . . . is allowed to overrule otherwise sound (and actuarially fair) underwriting practices" (1991: 510). I believe that health insurance represents a similar case, in which a particular public policy objective trumps the argument for industry practices. It is, however, misguided to transform this into an argument against risk classification in the private insurance industry. Insofar as there is a strong case to be made for socializing losses, rather than just socializing risks, then there is a strong case to be made for the involvement of the public sector.

What then should we say about our intuition that when people are charged high premiums as a consequence of events or circumstances that are not their fault—such as being abused by their husbands (Hellman 1997)—that they are being treated unjustly? The first thing to note is that insurance arrangements in general tend to wreak havoc with our intuitions about responsibility and desert. Much of this has to do with a simple tension between our moral reasoning, which is firmly governed by the language of free will, and the perspective that one must adopt when making statistical generalizations. Indeed, François Éwald (1986) has argued that the development of social insurance in the nineteenth century required a fundamental break with the central concepts of rights and responsibility that determined the structure of nineteenth-century "liberal" capitalism. Central to this development was the discovery that the rate of industrial accidents was highly predictable, regardless of who was responsible. Since sometimes a worker would be at fault, and sometimes the owner, the most socially efficient arrangement was simply to have both groups set aside a certain amount of money to indemnify the victims under a "no fault" arrangement. Yet even if this is better for everyone involved, it does mean that we can no longer expect the operations of the insurance system to track our intuitions about responsibility. (One can see the same thing with no-fault automobile insurance. Under such arrangements, some people will clearly get benefits that, from a strict view of personal responsibility, they are not entitled to. Does that make the arrangement unacceptable?) Similarly, the way that the insurance industry handles fraud offends the moral intuitions of some people, because it is governed more by a concern over loss-ratio security than by the binary opposition of guilt and innocence. As a result, the insurance industry tolerates a lot of behavior that the criminal justice system would regard as felonious (Ericson, Doyle, and Barry 2003: 340–346).

Second of all, our intuition that it is somehow more legitimate for an insurance company to penalize individuals for the consequences of choices they have made, rather than the circumstances they find themselves in, is perfectly cogent, but it does not count against the principle of actuarial fairness. When dealing with a person who voluntarily runs a risk, or chooses to act in a way that increases the risk of a particular loss, there are strong moral hazard arguments in favor of insurance schemes that penalize or deter such behavior (e.g., increasing premiums). Naturally, in the case of circumstances that are outside the individual's control, there is no moral hazard argument for increasing premiums. But this does not mean that there are no other arguments for charging higher-risk individuals higher-premiums. Most obviously, there is an adverse selection argument (discussed in section 13.6), and this argument applies regardless of whether one is dealing with the individual's choices or circumstances.

Thus, the luck-egalitarian argument against actuarial fairness fails. It is based upon the plausible intuition that it is permissible, from the standpoint of justice, to penalize individuals when the high risks that they bring to an insurance scheme are the product of their own voluntary choices—since it at least gives them the option, in cases where they find the insurance too expensive, of changing their own behavior. Yet this does not make it impermissible for insurers to charge individuals higher premiums merely because they are high risk.

13.4. Inequality

Finally, it is often suggested in the debates over actuarial fairness that the problem with risk classification and underwriting is that it leaves high-risk individuals unable to afford insurance (Daniels 1991; Ericson et al. 2003). This is often felt to be unjust, because it leaves certain individuals excluded from a beneficial social arrangement that the rest of the population is able to enjoy. Yet though this argument occurs with enormous frequency, it is based on confusion. Naturally, insurance costs less for high-risk individuals when they are pooled with a group of low-risk individuals in a "community rating" scheme. That is because they are using the insurance scheme, not just to secure the benefits of reduced uncertainty but also to externalize some of the costs associated with their losses onto other members of the insurance pool. The problem with this arrangement is that it can easily make insurance unaffordable for the low risks—since the inclusion of high-risk individuals within the insurance pool can drive premiums to a level where low risks are better off "self-insuring" (e.g. putting their own money into a rainy-day fund, or else just tolerating the uncertainty). It is precisely this exclusion of low-risk individuals from the insurance market that constitutes the classic efficiency loss associated with adverse selection (Akerlof 1970).

One might be inclined to think that segregating the insurance scheme, so that higher-risk individuals pay a higher premium than low-risk individuals, simply creates the opposite problem, excluding high-risk individuals from the market. But the two situations are not parallel. With pooling, the reason that low-risk individuals drop out is that the premium level becomes so high that it is no longer worth their while to buy insurance (because the premium significantly exceeds the "actuarially fair" rate). But with a segregated pool, it will still be worthwhile for high-risk individuals to buy insurance. Their premiums are high in reflection of the fact that the loss exposure they bring to the insurance scheme is high, but it will still be better for them (assuming risk aversion) to buy insurance than to face the loss without indemnity.

Nevertheless, some commentators do speak as though there were an asymmetry in the position of low- and high-risk individuals, such that insurers are less willing to deal with the latter, even when they have the ability to pay. This appears to be based on another misunderstanding of how most insurance markets function. There are in fact two models of insurance (Hacking 2003: 28). The first is the mutual society described in the first section of this paper. Under such an arrangement, a group of individuals with identical preferences and levels of risk aversion can nevertheless benefit by agreeing to pool their losses and gains. The benefits in this case stem from the reduction of uncertainty thanks to the law of large numbers. The second model, which Hacking refers to as the "Lloyd's of London" model, essentially involves a trade between a risk-averse and a risk-tolerant individual. Lloyd's rich "names" bet on outcomes, much as gamblers bet on horses. The names make money by getting lucky, betting that losses do not occur—collecting the premium but not having to pay out a claim. High-risk individuals, in this case, are like racehorses with terrible odds. There needs to be a huge potential payoff in order to persuade anyone to bet on them, and once the odds get bad enough, they cannot attract any bettors.

The Lloyd's model of insurance is particularly well-suited for dealing with risks for which there is little or no actuarial knowledge (e.g. uncommon, low-risk, events). Thus high-risk individuals may be unable to secure insurance from companies operating on the Lloyd's model. But this is not the case with the mutual society model, which is the one that predominates in the standard categories of health, life, home, and automobile insurance. This model is structurally neutral with respect to high- and low-risk individuals, since both groups can benefit equally from forming their own insurance pools. Thus there is no reason, in principle, that risk classification should leave high risks any less able to purchase insurance than low risks in the standard run of cases. Generally speaking, if people can afford a loss, then they can afford the insurance to cover that loss. Of course, if they cannot afford the loss, then they may not be able to afford the insurance either. But there is no independent issue of whether they can afford the insurance and thus no special question of justice that arises with respect to the cost of insurance.

Of course, there is considerable evidence to show that high-risk individuals are less likely to purchase insurance when they are put into a high-risk pool and charged a higher premium (e.g., with automobile insurance, see MacAvoy [1977: 38]). The important point is that these people drop out of the insurance market, not because the insurance policy no longer has value to them (as is the case with low-risk individuals who drop out of a community-rated pool), but for some other reason, such as an inability to pay. The way that individuals discount the future is also likely to become a more significant factor as premiums increase, leading many to forego insurance because of a hypertrophied valuation of the present cost of the premium versus the future benefit of the potential indemnification. These, however, are "second-best" problems, which do not speak against the principle of actuarial fairness.

Consider a person who is, by genetic predisposition, almost guaranteed to get a particular form of cancer and is thus facing the prospect of a significant financial loss for private medical care. Since there is very little uncertainty in this outcome, what this person needs to do is start saving in order to cover the cost of future cancer treatment. Of course, insofar as there is some uncertainty and this individual is able to find other people who are similarly situated, there is no reason that they cannot get together to pool their savings, and thereby achieve the efficiency gains of an insurance arrangement. Setting aside transaction costs, this insurance is guaranteed to cost less than the cancer treatment, simply because the savings realized by the individuals who happen not to require that treatment are distributed out to all members of the insurance pool. Naturally, some people may not be able to afford the treatment, in which case they may not be able to afford the insurance. But that is not the fault of the other participants in the insurance scheme; it is a problem of bad brute luck, or perhaps injustice in the distribution of income. The important point is that the anticipated loss is what the high-risk individual is unable to afford, not the insurance to cover that loss.

Thus what many people are articulating, when they worry that risk classification will leave high-risks unable to purchase insurance, is not an objection to the practice of risk classification, but rather a desire to see the losses to which high-risk individuals are exposed socialized (and an attempt to use private insurance as a way of achieving this objectives). For example, what they object to is not that some people cannot afford health insurance, but that some people cannot afford cancer treatment. Yet rather than arguing that the latter costs should be directly socialized, by having the state pay for cancer treatment for everyone, they seek to socialize it indirectly (and partially), by externalizing a large segment of the cost onto other policy-holders in the high-risk individuals' insurance plan by imposing restrictions on underwriting (MacAvoy 1977: 39). Thus the concern that risk classification will leave some individuals unable to afford insurance is often just a misleading way of arguing that individuals should not have to bear the burden of certain losses.

There is nothing intrinsic to the nature of being high risk that makes a person any less able to participate in an insurance plan than being low risk. Wherever there is uncertainty, people can benefit from risk-pooling arrangements. Problems arise only when there is an information asymmetry that makes the insurer, or other policy-holders, unable to distinguish between low and high risks. Thus it is simply false in many cases to claim that actuarially fair premiums leave high-risks unable to purchase insurance. In fact, insurers are sometimes eager to insure high-risks, because such accounts generate a larger flow of premiums, and therefore potentially larger investment returns. Ironically, it is often restrictions on underwriting that leave high-risk individuals unable to buy insurance. While insurers can be legally prevented from setting higher rates for certain classes of individuals, they are seldom obliged to sell a policy to anyone who comes along (except in special cases, such as automobile insurance or health insurance in certain jurisdictions). Thus they are often better off not selling policies at all to high-risk individuals (Pauly 1984). Restrictions on underwriting can therefore motivate "cream-skimming" on the part of insurers, which in turn may leave high-risk individuals unable to buy insurance, even if they are willing and able to pay an actuarially fair premium.

13.5. The case for actuarial fairness

Surveying these arguments against actuarial fairness, it is sometimes difficult to avoid the impression that critics have failed to appreciate the full consequences that the rejection of this principle would entail. Most of the cases that have attracted controversy involve members of groups who are already stigmatized or subject to unjust discrimination being denied insurance or charged higher premiums by insurance companies. Furthermore, given that disadvantaged individuals tend to be exposed to higher levels of risk (by virtue of living in high-crime neighborhoods, driving less safe vehicles, eating a poorer diet, working in less safe conditions, suffering more ill health, etc.), community rating tends to be progressive with respect to income and social class (Abraham 1986: 76). But if one is to abandon the principle of actuarial fairness, one must do so across the board. And there are many cases in which doing so will not have progressive consequences. With respect to men and women, for instance, while women are the primary beneficiaries of the use of community rating for annuities, men stand to derive an equally large benefit from the use of community rating for life insurance. Similarly, men benefit considerably from the use of community rating in automobile insurance. To select just one example, during a period of intense debate over the use of sex-segregated actuarial tables in the calculation of automobile insurance premiums in Canada, it was calculated that community rating would see the average premium of a single woman between the ages of 21 and 22 rise by over 61.3 percent, while the average premium of a man in the same category would drop by 11.6 percent (Dahlby 1983: 130). The result would be a significant redistribution of wealth from female to male drivers.

While the unfairness of this arrangement seems palpable, one must be very careful when seeking to articulate the complaint. The natural inclination is to say "Why should women be forced to subsidize the terrible driving habits of young men?" Yet to formulate the argument in this way is to buy into precisely the sort of "fairness to groups" argument that the justices of the US Supreme Court so effectively dissected in *Manhart.* After all, there is no requirement that women as a group come out the same as men as a group. In order to demonstrate unfairness, it is necessary to show that individual women are treated unfairly by the community rating arrangement. This is the challenge confronting those who would like to show that deviations from actuarial fairness are unjust.

The first thing to note is that, in an idealized "mutual society" insurance arrangement of the type considered here, premiums are always actuarially fair in the aggregate. This is a simple function of the fact that total claims paid out are equal to total premiums taken in. Now, as we saw in the first section, the most obvious argument for actuarial fairness in premiums is the principle that those who expect to take more out of the pool should be expected to put more in. In cases where each individual's premium is equal to the expected loss that he brings to the pool, each individual derives a pure welfare benefit from participating in the insurance arrangement—the mathematical value of participation is equal to the mathematical value of the gamble that he faces without insurance, it is just that the former has higher subjective utility because it is less risky. In cases where an individual's premium deviates from the actuarially fair level, it means that this person derives not just a welfare benefit from participating in the insurance arrangement, but also a monetary benefit, since the mathematical value of participating in the insurance pool now exceeds the value of the uninsured gamble (Bossert and Fleurbaey 2002: 114). Furthermore, since premiums are actuarially fair in the aggregate, the fact that one policy-holder derives a monetary benefit from participating in the insurance scheme means that some other policy-holder must suffer a monetary loss. As a result, when premiums deviate from the actuarially fair level there will be cross-subsidization within the insurance pool, or an implicit transfer of wealth between policy-holders. Another way of putting it is to say that when an insurance company fails to charge actuarially fair premiums, it allows high-risk policy-holders to externalize some of the costs they face onto other policy-holders. Thus the insurance pool, which is intended to be a source of mutual benefit in the form of welfare gains, is used as a way of effecting implicit redistributive transfers between individuals (Kimball 1979: 106).

This sort of cost externalization is widely regarded as contrary to the basic principles of justice, even by theorists whose work is often appealed to in defense of the principle of community rating. Dworkin, for instance, states that one of the central virtues of his "resource egalitarian" auction is that it forces individuals to take into consideration the full cost that their choices impose upon others (2000: 70; see also Gauthier 1986: 225). When a high-risk individual joins an insurance pool governed by a community rated premium, it generates a negative externality for all the other participants. It is like the person who orders an expensive drink or appetizer when dining in a large party at a restaurant, knowing that the bill is going to be divided up evenly between everyone. When the cost of each diner's meal is shared collectively, it allows those with more expensive tastes to externalize the cost of their actions onto others (which in turn erodes the value of the communal eating arrangement). Thus there is an argument to be made for "internalizing" the externality, by giving each diner an individual bill.[8] The same moral intuition, when

[8] This analogy is due to Promislow (1987: 217). He uses it to dramatize some of the "second-best" problems that may arise with risk-classification under conditions of imperfect information. Consider a situation in which a waiter, while refusing to give individual bills, offers to split the bill into two: one for diners on the north side of the table, another for diners on the south. Assuming that diners should pay their own way, under what conditions would this and would this not be a more just arrangement?

applied to the case of insurance, suggests that individuals should be charged an actuarially fair premium.

This argument is, in my view, sufficient to establish a general presumption in favor of actuarially fair premiums, from the standpoint of justice. This is, however, just a presumption. It has not been shown that deviations from actuarial fairness are necessarily unjust. This is because the high-risk individual who joins an insurance scheme does not merely generate a negative externality. She also creates a slight positive externality for all other participants in the insurance pool, via the "large numbers" effect, by virtue of having increased the size of the pool. Thus, she is not a pure free-rider. Her inclusion in the pool generates a welfare benefit for all the other policy-holders, even if it also imposes a slight monetary loss upon them. Furthermore, there will be a region in which the welfare benefit generated by expansion of the pool outweighs the welfare loss occasioned by the monetary disadvantage imposed upon the other policy-holders. As a result, it can be in the interest of everyone in the pool to accept new members, even when these new members are charged less than the actuarially fair premium. (Just as it can be in the interest of firms to hire more workers, even when doing so depresses average output. It is only when the marginal gains in net output reach zero that the firm should stop hiring.)

There is no question that when premium schemes offer a monetary advantage to some individuals and a disadvantage to others, it results in some people getting a better deal out of their insurance purchase than others. The question is whether this itself is unjust. According to some conceptions of justice, especially strictly egalitarian ones, it may turn out to be so. But there is no question that variations in the level of welfare benefit derived from commercial transactions are tolerated by the "implicit morality of the market." In the absence of price discrimination, for instance, consumers who are further in from the margin derive a larger welfare benefit from their purchases than those who are closer to the margin. A person who would have been willing to pay $2,000 for a flat-screen television may be able to buy it for $500, because the latter sum represents the most that other more price-sensitive consumers are willing to pay. Even with insurance, highly risk-averse individuals derive more benefit from their policies than less risk-averse individuals, and yet we do not take this to be an affront to justice.

This is not an accidental feature of the market. One of the fundamental features of capitalism is the repudiation of the notion of a "just price"—or of a principled division of the gain from trade—in favor of a competitive determination of price levels. The value of the efficiency gains associated with the establishment of market-clearing prices is taken to outweigh the value of a more egalitarian determination of price levels. The core criterion used to evaluate market transactions is therefore the Pareto principle, or the requirement that exchanges be mutually beneficial. The desire to achieve market-clearing prices, however, requires a willingness to tolerate transactions in which these mutual benefits are unequally divided. Thus the fact that some people get more out of their insurance purchases than others is not a special injustice, but rather an ordinary feature of commercial transactions in a capitalist economy. The insistence that any deviation from actuarial fairness is unjust, on the other hand, is essentially an insurance-specific version of just price theory.

As a result, the argument for actuarial fairness in premium levels does create a presumption in favor of the justice of such schemes, but it does not show that deviations from actuarial fairness are necessarily unjust. It therefore does not preclude restrictions on underwriting, especially when it can be shown that some other important social good is promoted through such restrictions. Governments do this routinely. As long as the insurance scheme remains mutually beneficial for all parties, there is nothing wrong with restrictions on underwriting that for some good reason (i.e., non-capriciously) give some people access to the scheme on preferential terms. The argument for actuarial fairness only precludes restrictions on underwriting that eliminate this benefit for some people (i.e., make it more attractive for low-risks to self-insure). Thus the demand for an unrestricted "freedom to underwrite" is a demand for a degree of freedom for insurers that no other type of business enjoys in a modern market economy.

13.6. Adverse selection

This brings us finally to the centerpiece of the insurance industry's defense of its underwriting practices: the adverse selection argument. There are both moral and non-moral versions of this argument. According to the non-moral version, it does not really matter whether it is just or unjust to charge actuarially fair premiums, it is necessary for insurers to do so if they wish to remain solvent. This is because community rating gives low-risk individuals an incentive to defect from the insurance scheme. Thus if the insurer charges a premium that represents the average loss in a pool that includes both high- and low-risk individuals, claims will wind up exceeding total premiums collected, simply because low-risk individuals will drop out of the pool (or high-risk individuals will join in great numbers). This is the adverse selection problem (Akerlof 1970; Rasmusen 1986: 230–235). In some cases, these low-risk individuals will defect to another insurer offering a lower premium and a pool with fewer high-risk individuals. But in other cases, the low-risks will drop out of the insurance market altogether. (Moral versions of the adverse selection argument then point to this deadweight loss as a case of injustice perpetrated against these low-risk individuals.)

There is, however, an ambiguity in the adverse selection argument, which can be clarified using the game-theoretic distinction between a Nash equilibrium and the core of a game. A particular insurance arrangement is in equilibrium if no individual participant has an incentive to drop out and "go it alone." The *feasible set,* therefore, represents the set of possible cooperative arrangements that offer individuals expected payoffs higher than those that could be obtained through the non-cooperative strategy of self-insurance. Depending upon how risk-averse individuals are, this set can be fairly large. The *core* of a game, on the other hand, represents an arrangement from which no individual *or coalition of individuals* has an incentive to defect (Ordeshook 1986). The core will tend to be much smaller (indeed, sometimes it will be nonexistent), because not only must the arrangement offer benefits that are superior to what the individual could achieve through self-insurance, it must also be superior to what any subset of insured individuals could achieve by defecting and forming their own insurance pool.

Consider the example of a three-person cooperative project, in which the gains from cooperation significantly exceed the returns to individual strategic action. There are various ways in which this "cooperative surplus" could be divided up between the players. In principle, any individual should be willing to accept even a tiny fraction of the cooperative surplus, with the lion's share going to the other two players, so long as the tiny fraction received leaves that individual with a better outcome than she could achieve through defection (i.e., dropping out of the cooperative scheme altogether). Thus highly inequitable divisions of the cooperative surplus will still be within the feasible set, as long as every player gets at least something (Gauthier 1986: 178). Such divisions may not be in the core, however, because the player who is most disadvantaged may be in a position to make a "divide and conquer" offer to one of the other players, promising to exclude the third player from the cooperative agreement altogether in return for larger payoffs for them both. Obviously this is only possible if the cooperative project is subject to decreasing returns to scale (Gauthier 1993: 46), but this tends to be the case with insurance. Thus interactions with a risk-pooling structure will tend to have cores, simply because the "large numbers" effect diminishes as the pool grows larger.

An insurance arrangement that deviates significantly from the principle of actuarial fairness in premiums is unlikely to be in the core of the relevant game, simply because the low-risk individuals (assuming they are sufficiently numerous) could all defect and form their own insurance pool. It is sometimes argued, on this basis, that adverse selection problems will begin to show up the moment that premiums take individuals outside the core (Daniels 1991: 513). This is, however, dubious as an empirical contention. The defection of a coalition is a form of collective action and is thus much more difficult to organize than the defection of an individual. It would be difficult to find any workplace, for instance, in which the division of labor and reward was genuinely in the core of the underlying interaction. The most talented group of employees could almost always benefit by defecting from the firm and setting up their own shop. This does happen, but just as often it does not. Thus it is not adequate, when considering insurance schemes, to suggest that the mere fact that a premium scheme takes policy-holders outside of the core will necessarily generate an adverse selection problem. There are too many other complicating factors. The only thing that can be said with confidence is that a premium scheme that takes some policy-holders outside of the *feasible set* (i.e. makes self-insurance a more attractive strategy) is likely to generate an adverse selection problem, because defection in this case does not involve collective action, individuals simply cancel their policies.

Thus the adverse selection argument does not provide a powerful justification for the principle of actuarial fairness. If premiums had to be in the core, then there would be a strong case to be made on adverse selection grounds for actuarial fairness. But if premiums only need to be within the feasible set, then the adverse selection argument only shows that they must not depart too radically from that principle. There is, however, likely to be a considerable zone of tolerance for deviations. In this context, it is worth recalling that many flat-rate "friendly societies"—which essentially eschewed any actuarial basis for the calculation of premium levels—survived well into the twentieth century (Emery 1996). Furthermore, it was seldom bottom-up pressure from policy-holders that led to greater risk classification and differentiated premiums, but rather the aggressive lobbying efforts of actuaries, which date back to the beginning of the nineteenth century (see Ansell 1835; Baker 2003). Although there are certain important exceptions, risk classification has seldom been a defensive response to bottom-up adverse selection problems and has more often been used as a competitive tactic by insurers against one another. In other words, rather than policy-holders using private information to purchase insurance at prices below the actuarially fair level, thereby creating losses for insurers, historically it has been more common to see insurers instituting risk classification schemes as a way of creating low-premium pools, which could then be used to entice clients away from rival insurers (Baker 2003). It is worth recalling that in order for adverse selection to cause a serious problem, absent these competitive tactics, policy-holders must be at least roughly aware of the expected loss that they bring to an insurance pool. Yet individuals are seldom in such a position, simply because they have neither the interest nor the ability to analyze the relevant data. With automobile insurance, for instance, an insurer typically knows a lot more about the accident risk posed by a particular driver than the driver ever will. Thus in practice adverse selection has often turned out not to be the powerful force that economic theory predicts it to be.

As a result, a premium scheme that merely takes a group of policy-holders outside the core is unlikely to provoke defection. Under real-world conditions, not only are there significant transaction costs and collective action problems associated with such a defection, but individuals themselves usually lack the information needed to determine whether defection would be advantageous (i.e., whether they are in the core or not). Thus the real danger is not "classical" adverse selection, but rather the prospect that rival insurers, sensing an opportunity, will try to identify and recruit low-risk individuals from the community-rated pool, in part by informing them that they are implicitly subsidizing other participants in their existing insurance pool. In other words, the major problems with community-rated premiums are the various forms of cream-skimming that they encourage.

But because the most important problems arise out of the competitive behavior of insurers, and not strategic behavior on the part of policy-holders, the argument fails to provide a very powerful defense of the idea that risk classification is somehow "necessary" to ensure the financial solvency of insurers. Naturally, if one company is engaging in aggressive risk classification and underwriting, then rival firms may be forced to respond in kind in order to remain solvent. But so long as legal restrictions on underwriting apply equally to all firms in an industry, then this is not a concern. For example, in jurisdictions where sex-segregated automobile insurance policies are not permitted, it is possible in principle for a group of female drivers to defect from existing insurance schemes and form their own pool, with much lower premiums. But since the restrictions on underwriting make it impossible to exclude men from this new pool, or to charge them higher premiums, any such move would be quickly undermined by the number of male drivers who would be attracted to the lower premiums as well.

Thus the argument from adverse selection fails to show that an unrestricted freedom to underwrite is a business necessity. There are simply too many ways in which legal restrictions on underwriting can counteract the tendency, and ensure that insurers who refrain from engaging in a particular type of risk classification are not put at a competitive disadvantage by virtue of that fact. Of course, serious problems can arise from the incentives that are inadvertently created for insurers to find ways around the law. The biggest concern, mentioned above, is that insurers will simply drop entire classes of clients or red-line residential districts in order to avoid attracting high-risk clients (whom they do not have a right to charge extra). There is also the possibility that insurers will avoid certain clients, or charge them extra, in circuitous and indirect ways that are difficult to regulate. For example, although it is generally regarded as discriminatory to refuse individuals insurance coverage on the basis of their sexual orientation, it was suggested at one point that health insurers were refusing to provide coverage to male hairdressers (Aaron and Bosworth 1994: 269). Needless to say, this sort of behavior not only reproduces the injustice that the original legislation was intended to prohibit, but compounds it in various ways.

Thus restrictions on underwriting must be carefully weighed, and not undertaken lightly, as they do have a strong tendency to generate perverse effects. The argument from adverse selection is important in that it draws attention to the strategic context in which insurance decisions are made. One cannot simply legislate changes in premium levels or underwriting practice without taking into consideration the changes that this will cause in the purchasing decisions made by policy-holders, along with the competitive tactics used by insurers. On the other hand, it is far too simplistic to say that any deviation from actuarial fairness is bound to generate such perverse consequences. There is far too much friction in the real-world marketplace for that to be the case. Thus the argument from adverse selection adds very little to the basic argument from justice when it comes to supporting the practice of risk classification. Restrictions on underwriting certainly have redistributive effects, but so long as the outcome remains within the feasible set of policy-holders, these redistributions fall within the range that market institutions normally permit. Furthermore, such restrictions need not create significant adverse selection problems, unless they are so extreme as to drive low-risk individuals from the market entirely.

13.7. Conclusion

The argument so far has been focused entirely upon the principle of actuarial fairness in premiums. The conclusion has been that actuarial fairness represents a just ideal, but that restrictions on underwriting which move the premium scheme away from actuarial fairness are permissible when doing so is needed in order to achieve some important social good, and when doing so will not create significant Pareto inefficiencies as a result of low-risk individuals dropping out of the insurance market. Thus I defend a somewhat more limited "right to underwrite" than those who insist upon actuarial fairness, or require that premiums be kept within the core. Neither of these is necessary in order to ensure that the insurance scheme remain advantageous for all parties involved. Furthermore, mutual advantage is all that is required in order for the transaction to satisfy "the implicit morality of the market." Thus while it remains permissible for firms to charge actuarially fair premiums, it is not necessary, and there is nothing in principle wrong with statutory restrictions on underwriting that take certain policy-holders outside the core. It is wrong, however, for restrictions on underwriting to take any policy-holder outside his or her feasible set.

BIBLIOGRAPHY

Aaron, Henry J. and Barry P. Bosworth (1994) *Economic Issues in Reform of Health Care Financing.* Brookings Papers on Economic Activity (Brookings Institute, Washington, DC).

Abraham, Kenneth S. (1986) *Distributing Risk* (Yale University Press, New Haven).

Agnew, Robert (1994) "The Techniques of Neutralization and Violence," *Criminology*, 32/4: 555–580.

Agnew, Robert and Ardith A. Peters (1986) "The Techniques of Neutralization: An Analysis of Predisposing and Situational Factors," *Criminal Justice and Behavior*, 13/1: 81–97.

Ainslie, George (1992) *Picoeconomics* (Cambridge University Press, Cambridge).

Ainslie, George (2001) *Breakdown of Will* (Cambridge University Press, Cambridge).

Akerlof, George (1970) "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84/3: 488–500.

Akerlof, George (1982) "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 84: 543–569.

Akers, Ronald L. (1990) "Rational Choice, Deterrence, and Social Learning Theory in Criminology: The Path Not Taken," *Journal of Criminal Law and Criminology*, 81/3: 653–676.

Akers, Ronald L. (1998) *Social Learning and Social Structure: A General Theory of Crime and Deviance* (Northeastern, Boston).

Akers, Ronald L., Marvin D. Krohn, Lonn Lanza-Kaduce and Marcia Radosevich (1979) "Social Learning and Deviant Behavior: A Specific Test of a General Theory," *American Sociological Review*, 44/4: 636–655.

Alcian, Armen A. and Harold Demsetz (1972) "Production, Information Costs, and Economic Organization," *American Economic Review*, 62/5: 777–795.

Anderson, Elizabeth (1999) "What is the Point of Equality?" *Ethics*, 109/2: 287–337.

Annas, Julia (2011) *Intelligent Virtue* (Oxford University Press, Oxford).

Ansell, Charles (1835) *A Treatise on Friendly Societies* (Baldwin and Cradock, London).

Applbaum, Arthur (1999) *Ethics for Adversaries* (Princeton University Press, Princeton).

Aquinas, St. Thomas (1947) *Summa Theologica*, Vol. 2 (Benziger Brothers, New York).

Archer, Simon (1996) "The Ethics of Creative Accounting," *Science and Engineering Ethics*, 2/1: 55–70.

Arendt, Hannah (1969) *On Violence* (Harcourt Brace Jovanovich, San Diego).

Arendt, Hannah (2006) *Eichmann in Jerusalem*, rev. edn (Penguin, London).

Argyris, Chris (1973) "Some Limits of Rational Man Organizational Theory," *Public Administration Review*, 33/3: 253–267.

Ariely, Dan (2008) *Predictably Irrational* (HarperCollins, New York).

Armour, John, Henry Hansmann and Reiner Kraakman (2004) "What is Corporate Law?" in John Armour, Reiner Kraakman et al., *The Anatomy of Corporate Law* (Oxford University Press, Oxford), 1–34.

Armstrong, J. Scott (1977) "Social Irresponsibility in Management," *Journal of Business Research*, 5/3: 185–213.

Arnott, Dave (2000) *Corporate Cults* (Amacom, New York).

Arnsperger, Christian (1994) "Envy-Freeness and Distributive Justice," *Journal of Economic Surveys*, 8/2: 155–186.

Arrow, Kenneth (1973) "Social Responsibility and Economic Efficiency," *Public Policy*, 21/3: 303–317.

Arrow, Kenneth (1974) "Limited Knowledge and Economic Analysis," *American Economic Review*, 64/1: 1–10.

Arrow, Kenneth and Gerard Debreu (1954) "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, 22/3: 265–291.

Asch, Solomon E. (1955) "Opinions and Social Pressure," *Scientific American*, 193/5: 31–35.

Austin, John L. (1979) "A Plea for Excuses," in *Philosophical Papers*, 3rd edn (Oxford University Press, Oxford), 175–204.

Austin, Regina (1983) "The Insurance Classification Controversy," *University of Pennsylvania Law Review*, 131: 517–583.

Ayres, Ian and John Braithwaite (1992) *Responsive Regulation* (Oxford University Press, Oxford).

Baechler, Jean (1992) "Virtue: Its Nature, Exigency and Acquisition," in John W. Chapman and William A. Galston (eds), *Virtue, Nomos 34* (New York University Press, New York), 25–48.

Baier, Annette (1985) *Postures of the Mind* (University of Minnesota Press, Minneapolis).

Baier, Kurt (1958) *The Moral Point of View* (Cornell University Press, Ithaca).

Bainbridge, Stephen M. (2002) *Corporation Law and Economics* (Foundation Press, New York).

Bakan, Joel (2004) *The Corporation* (Penguin, Toronto).

Baker, George (1992) "Incentive Contracts and Performance Measurements," *Journal of Political Economy*, 100/3: 598–614.

Baker, John S., Jr. (2004) "The Sociological Origins of 'White-Collar Crime,'" *Heritage Foundation Legal Memorandum* 14 (<http://www.heritage.org/research/legalissues/>).

Baker, Tom (2003) "Containing the Promise of Insurance: Adverse Selection and Risk Classification," in Richard V. Ericson and Aaron Doyle (eds), *Risk and Morality* (University of Toronto Press, Toronto), 258–283.

Barkan, Steven E. and George J. Bryjak (2011) *Fundamentals of Criminal Justice: A Sociological View*, 2nd edn (Jones & Bartlett, Sudbury, MA).

Baron, Marcia (2005) "Justifications and Excuses," *Ohio State Journal of Criminal Law*, 2/2: 387–413.

Barr, Nicholas (1998) *The Economics of the Welfare State*, 3rd edn (Stanford University Press, Stanford).

Barry, Brian (1989) *Theories of Justice* (University of California, Berkeley).

Bastiat, Frédéric (1880) *Essays on Political Economy*, trans. David A. Wells (G. Putnam's Sons, New York).

Bator, Francis M. (1958) "The Anatomy of Market Failure," *Quarterly Journal of Economics*, 72/3: 351–379.

Baumeister, Roy F. (1997) *Evil* (W. H. Freeman, New York).

Baumol, William J. (1974) "Business Responsibility and Economic Behavior," in Melvin Anshen (ed), *Managing the Socially Responsible Corporation* (MacMillan, New York), 59–71.

Baumol, William J. (1986) *Superfairness* (MIT Press, Cambridge, MA).

Baumol, William J. and William G. Bowen (1966) *Performing Arts: The Economic Dilemmas* (20th Century Fund, New York).

Beauchamp, Tom L. and Norman E. Bowie (2001) *Ethical Theory and Business*, 6th edn (Prentice Hall, Upper Saddle River, NJ).

Beauchamp, Tom L., Norman E. Bowie and Dennis G. Arnold (eds) (2009) *Ethical Theory and Business*, 8th edn (Pearson/Prentice Hall, Upper Saddle River, NJ).

Bebchuk, Lucien A. and Jesse M. Fried (2003) "Executive Compensation as an Agency Problem," *Journal of Economic Perspectives*, 17/3: 71–92.

Bebchuk, Lucien Arye and Jesse M. Fried (2004) *Pay Without Performance* (Harvard University Press, Cambridge, MA).

Bebchuk, Lucien Arye, Jesse M. Fried and David I. Walker (2002) "Managerial Power and Rent Extraction in the Design of Executive Compensation," *University of Chicago Law Review*, 69/3: 751–846.

Benhabib, Seyla (1986) *Critique, Norm and Utopia* (Columbia University Press, New York).

Berle, Adolf A. and Gardiner C. Means (1932/1991) *The Modern Corporation and Private Property* (Transaction, London).

Binmore, Ken (1999) "Game Theory and Business Ethics," *Business Ethics Quarterly*, 9/1: 31–35.

Bittlingmayer, George (1988) "Property Rights, Progress, and the Aircraft Patent Agreement," *Journal of Law and Economics*, 31/1: 227–248.

Blair, Margaret M. (1995) "Corporate 'Ownership': A Misleading Word Muddies the Corporate Governance Debate," *The Brookings Review*, 13/1: 16–19.

Blair, Margaret M. (1999) "Firm-Specific Human Capital and Theories of the Firm," in Margaret M. Blair and Mark J. Roe (eds), *Employees and Corporate Governance* (Brookings Institute, Washington, DC), 58–90.

Blair, Margaret M. and Lynn A. Stout (1999) "A Team Production Theory of Corporate Law," *Virginia Law Review*, 85/3: 248–328.

Blake, John, Oriol Amat and Jack Dowds (1998) "The Ethics of Creative Accounting," in Catherine Gowthorpe and John Blake (eds), *Ethical Issues in Accounting* (Routledge, London), 24–40.

Blass, Thomas (1991) "Understanding Behavior in the Milgram Obedience Experiment: The Role of Personality, Situations, and their Interactions," *Journal of Personality and Social Psychology*, 60/3: 398–413.

Blaug, Mark (1985) *Economic Theory in Retrospect*, 4th edn (Cambridge University Press, Cambridge).

Blickle, Gerhard, Alexander Schlegel, Pantaleon Fassbender and Uwe Klein (2006) "Some Personality Correlates of Business White-Collar Crime," *Applied Psychology*, 55/2: 220–233.

Boardman, Anthony E. and Aidan R. Vining (1989) "Ownership and Performance in Competitive Environments: A Comparison of the Performance of Private, Mixed, and State-Owned Enterprises," *Journal of Law and Economics*, 32: 1–33.

Boatright, John R. (1994) "Fiduciary Duties and the Shareholder Managements Relation: or, What's So Special about Shareholders?" *Business Ethics Quarterly*, 4/4: 393–407.

Boatright, John R. (1999) *Ethics in Finance* (Blackwell, Oxford).

Boatright, John R. (2000) *Ethics and the Conduct of Business*, 3rd edn (Pearson/Prentice Hall, Upper Saddle River, NJ).

Boatright, John R. (2002) "Ethics and Corporate Governance: Justifying the Role of the Shareholder," in Norman E. Bowie (ed), *The Blackwell Guide to Business Ethics* (Blackwell, London), 38–60.

Boatright, John R. (2000) *Ethics and the Conduct of Business*, 3rd edn (Pearson/Prentice Hall, Upper Saddle River, NJ).

Boatright, John R. (2006) "What's Wrong—and What's Right—with Stakeholder Management," *Journal of Private Enterprise*, 21: 106–130.

Boatright, John R. (2009) *Ethics and the Conduct of Business*, 6th edn (Pearson/Prentice Hall, Upper Saddle River, NJ).

Boehm, Christopher (1999) *Hierarchy in the Forest* (Harvard University Press, Cambridge, MA).

Bossert, Walter and Marc Fleurbaey (2001) "Equitable Insurance Premium Schemes," *Social Choice and Welfare V*, 19/1: 113–125.

Bowie, Norman (1995) *Business Ethics: A Kantian Perspective* (Blackwell, Oxford).

Bowie, Norman E. and R. Edward Freeman (eds) (1992) *Ethics and Agency Theory* (Oxford University Press, New York).

Bowles, Samuel and Herbert Gintis (2011) *A Cooperative Species* (Princeton University Press, Princeton).

Boyd, Robert and Peter J. Richerson (2005a) "Punishment Allows the Evolution of Cooperation (Or Anything Else) in Sizeable Groups," in Robert Boyd and Peter J. Richerson (eds), *The Origin and Evolution of Culture* (Oxford University Press, Oxford), 166–188.

Boyd, Robert and Peter Richerson (2005b) "Solving the Puzzle of Human Cooperation," in Stephen Levinson (ed), *Evolution and Culture* (MIT Press, Cambridge, MA), 105–134.

Boye, Michael W. and John W. Jones (1997) "Organizational Culture and Employee Counterproductivity," in Robert A. Giacalone and Jerald Greenberg (eds), *Antisocial Behavior in Organizations* (Sage, London), 172–184.

Brady, F. Neil and Jeanne M. Logsdon (1988) "Zimbardo's 'Stanford Prison Experiment' and the Relevance of Social Psychology for Teaching Business Ethics," *Journal of Business Ethics*, 7/9: 703–710.

Braithwaite, John (1981–1982) "The Limits of Economism in Controlling Harmful Corporate Conduct," *Law and Society Review*, 16/3: 481–504.

Braithwaite, John (1989) *Crime, Shame and Reintegration* (Cambridge University Press, Cambridge).

Brandom, Robert (1994) *Making It Explicit* (Harvard University Press, Cambridge, MA).

Braybrooke, David (1976) "The Insoluble Problem of the Social Contract," *Dialogue*, 15: 3–37.

Brilmayer, Lea, Richard Hekeler, Douglas Laycock and Teresa A. Sullivan (1980) "Sex Discrimination in Employer-Sponsored Insurance Plans: A Legal and Demographic Analysis," *University of Chicago Law Review*, 47/3: 505–560.

Broome, John (1998) "Kamm on Fairness," *Philosophy and Phenomenological Research*, 58: 955–961.

Brown, Donald (1991) *Human Universals* (McGraw Hill, New York).

Buchanan, Allan (1985) *Ethics, Efficiency and the Market* (Oxford University Press: Oxford).

Buchanan, Allan (1996) "Toward a Theory of the Ethics of Bureaucratic Organizations," *Business Ethics Quarterly*, 6/4: 419–440.

Buchanan, James M. (1986) "Rights, Efficiency, and Exchange: The Irrelevance of Transaction Costs," in James M. Buchanan (ed), *Liberty, Market and State* (Harvester, Brighton), 92–107.

Buchholz, Rogene A. and Sandra B. Rosenthal (2005) "Toward a Contemporary Conceptual Framework for Stakeholder Theory," *Journal of Business Ethics*, 58/1–3: 137–148.

Buffalo, M. D. and Joseph W. Rodgers (1971) "Behavioral Norms, Moral Norms and Attachment: Problems of Deviance and Conformity," *Social Problems*, 19/1: 101–113.

Burrough, Bryan and John Helyar (1990) *Barbarians at the Gate* (HarperCollins, New York).

Campbell, Donald E. (1995) *Incentives* (Cambridge University Press, Cambridge).

Carducci, Bernardo J. (2009) *The Psychology of Personality*, 2nd edn (Wiley-Blackwell, Oxford).

Carneiro, Robert L. (1970) "A Theory of the Origin of the State," *Science*, 169: 733–738.

Carr, Albert Z. (1968) "Is Business Bluffing Ethical?" *Harvard Business Review*, 46/1: 143–153.

Chan, Marjorie (2003) "Corporate Espionage and Workplace Trust/Distrust," *Journal of Business Ethics*, 42/1: 45–58.

Chandler, Alfred D. (1962) *Strategy and Structure* (MIT Press, Cambridge, MA).

Chandler, Alfred D. (1977) *The Visible Hand* (Belknap Press, Cambridge, MA).

Clark, Robert C. (1985) "Agency Costs versus Fiduciary Duties," in John W. Pratt and Richard J. Zeckhauser (eds), *Principals and Agents: The Structure of Business* (Harvard Business School Press, Cambridge, MA), 55–80.

Clarkson, Max (1998) "Introduction," in Max Clarkson (ed), *The Corporation and its Stakeholders* (University of Toronto Press, Toronto), 1–9.

Clinard, Marshall (1983) *Corporate Ethics and Crime: The Role of Middle Management* (Sage, Beverly Hills).

Clinard, Marshall B. and Peter C. Yeager (1980) *Corporate Crime* (Free Press, New York).

Coase, Ronald H. (1937) "The Nature of the Firm," *Economica*, New Series, 4/16: 386–405.

Cohen, Albert K. (1955) *Delinquent Boys: The Culture of the Gang* (Free Press, Glencoe).

Cohen, Gerald A. (1989) "On the Currency of Egalitarian Justice," *Ethics*, 99/4: 906–944.

Cohen, Gerald A. (1991) "Incentives, Inequality and Community," in G. Peterson (ed), *Tanner Lectures on Human Values*, Vol. 13 (University of Utah Press, Salt Lake City), 263–329.

Cohen, Gerald A. (2000) *If You're an Egalitarian, How Come You're So Rich?* (Harvard University Press, Cambridge, MA).

Cohen, Gerald A. (2008) *Rescuing Justice and Equality* (Harvard University Press, Cambridge, MA).

Cohen, Gerald A. (2009) *Why Not Socialism?* (Princeton University Press, Princeton).

Coleman, James S. (1988) "Social Capital in the Creation of Human Capital," *American Journal of Sociology*, 94/S1: S95–S120.

Coleman, James W. (1987) "Toward an Integrated Theory of White-Collar Crime," *American Journal of Sociology*, 93/2: 406–439.

Coleman, James W. (1989) *The Criminal Elite: The Sociology of White Collar Crime* (St. Martin's, New York).

Commission for Fair Play (1993) *The Attitude of Canadian Youth and Coaches Concerning Fair Play* (Government of Canada, Ottawa).

Crawford, Vincent P. and Joel Sobel (1982) "Strategic Information Transmission," *Econometrica*, 50/6: 1431–1451.

Cressey, Donald (1950) "The Criminal Violation of Financial Trust," *American Sociological Review*, 15/6: 738–743.

Cressey, Donald (1953) *Other People's Money: A Study in the Social Psychology* (Free Press, New York).

Crook, Clive (2005) "Survey: Corporate Social Responsibility," *The Economist*, Jan. 20. Print.

Dahl, Robert (1985) *A Preface to Economic Democracy* (University of California Press, Berkeley).

Dahlby, Bev G. (1983) "Adverse Selection and Statistical Discrimination: An Analysis of Canadian Automobile Insurance," *Journal of Public Economics*, 20/1: 121–130.

Daly, Kathleen (1989) "Gender and Varieties of White Collar Crime," *Criminology* 27/4: 769–793.

Daniels, Norman (1991) "Insurability and the HIV Epidemic: Ethical Issues in Underwriting," *The Milbank Quarterly*, 68/4: 497–525.

Darley, John M. and C. Daniel Batson (1973) "From Jerusalem to Jericho: A Study of Situational and Dispositional Variables in Helping Behavior," *Journal of Personality and Social Psychology*, 27/1: 100–108.

Dawes, Robin H. and Richard M. Thaler (1988) "Anomalies: Cooperation," *The Journal of Economic Perspectives*, 2/3: 187–197.

Dayen, David (2013) "Turns Out Much-Hyped Settlement Still Allows Banks to Steal Homes," *Salon*, May 2, http://www.salon.com/2013/05/02/the_foreclosure_fraud_settlement_was_a_big_dud/.

de Girardin, Émile (1855) *La Politique universelle* (Librairie Nouvelle, Paris).

De Marchi, Neil (2003) "Visualizing the Gains from Trade: Mid-1870s to 1962," *European Journal of the History of Economic Thought*, 10/4: 551–572.

De Sousa, Raymond (2007) "Lessons from the Prosecution of Lord Black," *National Post*, Mar. 8.

Dees, J. Gregory (1992) "Principals, Agents and Ethics," in Norman E. Bowie and Edward Freeman (eds), *Ethics and Agency Theory* (Oxford University Press, New York), 25–57.

DeGeorge, Richard T. (1992) "Agency Theory and the Ethics of Agency," in Norman E. Bowie and Edward Freeman (eds), *Ethics and Agency Theory* (Oxford University Press, New York), 65.

Dennett, Daniel (1995) *Darwin's Dangerous Idea* (Simon & Schuster, New York).

DesJardins, Joseph R. and John J. McCall (2005) *Contemporary Issues in Business Ethics*, 5th edn (Wadsworth, Belmont, CA).

Dietsch, Peter (2010) "The Market, Competition and Equality," *Politics Philosophy and Economics*, 9/2: 213–244.

Dixit, Avinash (1997) "The Power of Incentives in Private versus Public Organizations," *American Economic Review*, 87: 378–382.

Dobson, John (1997) *Finance Ethics: The Rationality of Virtue* (Rowman & Littlefield, Lanham, MD).

Dobson, John (2008) "Utopia Reconsidered: The Modern Firm as Institutional Ideal," *Philosophy of Management*, 7/1: 67–75.

Donaldson, Lex (1990) "The Ethereal Hand: Organizational Economics and Management Theory," *Academy of Management Review*, 15/3: 369–381.

Donaldson, Thomas and Lee E. Preston (1995) "The Stakeholder Theory of the Corporation: Concepts, Evidence and Implications," *Academy of Management Review*, 20/1: 65–91.

Doris, John (2002) *Lack of Character* (Cambridge University Press, Cambridge).

Dow, Gregory K. (2003) *Governing the Firm* (Cambridge University Press, Cambridge).

Downs, Anthony (1957) *An Economic Theory of Democracy* (Harper & Row, New York).

Drucker, Peter (2001) *The Essential Drucker* (Harper Business, New York).

Dugatkin, Lee (1999) *Cheating Monkeys and Citizen Bees* (Harvard University Press, Cambridge, MA).

Durkheim, Émile (1893/1930) *De la division du travail sociale* (Presses Universitaires de France, Paris).

Durkheim, Émile (1895/1937) *Les Règles de la méthode sociologique* (Presses Universitaires de France, Paris).

Durkheim, Émile (1958) *Professional Ethics and Civic Morals*, trans. Cornelia Brookfield (Free Press, Glencoe).

Duska, Ronald F. (1992) "Why be a Loyal Agent? A Systemic Ethical Analysis," in Norman E. Bowie and Edward Freeman (eds), *Ethics and Agency Theory* (Oxford University Press, New York), 143–168.

Dworkin, Ronald (1980) "Is Wealth a Value?" *The Journal of Legal Studies*, 9/2: 191–226.

Dworkin, Ronald (1981) "What is Equality? Part 2: Equality of Resources," *Philosophy and Public Affairs*, 10/4: 283–345.

Dworkin, Ronald (2000) *Sovereign Virtue* (Harvard University Press, Cambridge, MA)

Dworkin, Ronald (2004) "Ronald Dworkin Replies," in Justine Burley (ed), *Dworkin and His Critics* (Blackwell, Oxford), 339–395.

Easterbrook, Frank and Daniel Fischel (1991) *The Economic Structure of Corporate Law* (Harvard University Press, Cambridge, MA).

Easterbrook, Frank H. and Daniel R. Fischel (1982) "Antitrust Suits by Targets of Tender Offers," *Michigan Law Review*, 80/6: 1155–1178.

Eisenhardt, Kathleen H. (1989) "Agency Theory: An Assessment and Review," *Academy of Management Review*, 14/1: 57–74.

Eliason, Stephen L. and Richard A. Dodder (1999) "Techniques of Neutralization used by Deer Poachers in the Western United States: A Research Note," *Deviant Behavior*, 20/3: 233–252.

Ellickson, Robert C. (1993) "Property in Land," *Yale Law Journal*, 102/6: 1315–1400.

Elster, Jon (1979) *Ulysses and the Sirens* (Cambridge University Press, Cambridge).

Elster, Jon (1995) *Local Justice* (Russell Sage, New York).

Elster, Jon (2000) *Ulysses Unbound* (Cambridge University Press, Cambridge).

Emerson, Michael and Christian Smith (2001) *Divided by Faith: Evangelical Religion and the Problem of Race in America* (Oxford University Press, New York).

Emery, Hebert J. C. (1996) "Risky Business? Nonactuarial Pricing Practices and the Financial Viability of Fraternal Sickness Insurers," *Explorations in Economic History*, 33/2: 195–226.

Ericson, Richard V., Aaron Doyle and Dean Barry (2003) *Insurance as Governance* (University of Toronto Press, Toronto).

Esping-Anderson, Gøsta (1990) *The Three Worlds of Welfare Capitalism* (Princeton University Press, Princeton, NJ).

Evan, William M and Freeman, R. Edward (1988) "A Stakeholder Theory of the Modern Corporation: Kantian Capitalism," in T. Beauchamp and N. Bowie (eds), Ethical Theory and Business (Prentice Hall: Englewood Cliffs, NJ).

Éwald, François (1986) *L'État providence* (Grasset, Paris).

Fama, Eugene F. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25/2: 383–417.

Feezell, Randolph M. (1988) "Sportsmanship," in William J. Morgan and Klaus V. Meier (eds), *Philosophic Inquiry in Sport* (Human Kinetics, Champaign, IL), 251–262.

Fehr, Ernst and Simon Gächter (2002) "Do Incentive Contracts Crowd out Voluntary Cooperation?" http://www.iew.unizh.ch/wp/iewwp034.pdf. University of Zurich Working Paper, 34.

Fehr, Ernst, Simon Gächter and Georg Kirchsteiger (1996) "Reciprocal Fairness and Non-compensating Wage Differentials," *Journal of Institutional and Theoretical Economics*, 1524: 608–640.

Feigenbaum, Harvey (1982) "Public Enterprises in Comparative Perspective," *Comparative Politics*, 15/1: 101–122.

Ferner, Anthony (1988) *Governments, Managers and Industrial Relations* (Basil Blackwell, Oxford).

Ferraro, Fabrizio, Jeffrey Pfeffer and Robert Sutton (2005) "Economics Language and Assumptions: How Theories can Become Self-fulfilling," *Academy of Management Review*, 30/1: 8–24.

Flanagan, Owen (1991) *Varieties of Moral Personality: Ethics and Psychological Realism* (Harvard University Press, Cambridge, MA).

Ford, Henry (1922) *My Life and Work* (Doubleday, Garden City, NY).

Fraleigh, Warren (1984) *Right Actions in Sport* (Human Kinetics, Champaign, IL).

Frank, Robert H. (1985) *Choosing the Right Pond* (Oxford University Press, Oxford).

Frank, Robert H. and Phillip J. Cook (1995) *The Winner-Take-All Society* (Penguin, New York).

Frank, Robert H., Thomas Gilovich and Dennis Regan (1993) "Does Studying Economics Inhibit Cooperation?" *Journal of Economic Perspectives*, 7/2: 159–171.

Fraser, Nancy (2003) "Social Justice in the Age of Identity Politics: Redistribution, Recognition and Participation," in Nancy Fraser and Axel Honneth (eds), *Redistribution or Recognition?* (Verso, London), 7–109.

Fraser, Timothy C. G. and Robert Waschik (2002) *Managerial Economics: A Game-Theoretic Perspective* (Routledge, London).

Freeman, Edward R. (1984) *Strategic Management: A Stakeholder Approach* (Pitman, Boston).

Freeman, Edward R. (1994) "The Politics of Stakeholder Theory: Some Future Directions," *Business Ethics Quarterly*, 4/4: 409–421.

Freeman, Edward R. (1998) "A Stakeholder Theory of the Modern Corporation," in Max B. E. Clarkson (ed), *The Corporation and its Stakeholders* (University of Toronto Press, Toronto), 125–138.

Freeman, Edward R. and William M. Evan (1990) "Corporate Governance: A Stakeholder Interpretation," *Journal of Behavioral Economics*, 19/4: 337–359.

Freeman, Samuel (1990) "Reason and Agreement in Social Contract Views," *Philosophy & Public Affairs*, 19/2: 122–157.

Frey, Bruno S. and Felix Oberholzer-Gee (1997) "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-out," *American Economic Review*, 87/4: 746–755.

Frey, Bruno S. and Margit Osterloh (eds) (2002) *Successful Management by Motivation: Balancing Intrinsic and Extrinsic Incentives* (Springer Verlag, Berlin).

Frey, Bruno S., Felix Oberholzer-Gee and Reiner Eichenberger (1996) "The Old Lady Visits Your Backyard: A Tale of Morals and Markets," *Journal of Political Economy*, 104/6: 1297–1313.

Friedman, Milton (1962) *Capitalism and Freedom* (University of Chicago Press, Chicago).

Friedman, Milton (1970) "The Social Responsibility of Business is to Increase its Profits," *New York Times Magazine*, Sept. 13.

Friedman, Milton (1982) *Capitalism and Freedom*, 2nd edn (University of Chicago Press, Chicago).

Fudenberg, Drew and Eric Maskin (1986) "The Folk Theorem in Repeated Games with Discounting or with Imperfect Information," *Econometrica*, 54/3: 533–554.

Fudenberg, Drew and Jean Tirole (1991) *Game Theory* (MIT Press, Cambridge, MA).

Fukuyama, Francis (1995) *Trust* (Penguin, London).

Funder, David C. (1999) *Personality Judgement* (Academic Press, San Diego).

Gadd, Jane (2005) "Is Ethics the New Bottom Line?" *The Globe and Mail*, Mar. 8, 2005, E6. Print.

Galbraith, John Kenneth (1952) *American Capitalism* (Houghton Mifflin, Cambridge, MA).

Galbraith, John Kenneth (1971) *The New Industrial State* (New American, New York).

Gates, Jeff (1998) *The Ownership Solution* (Basic Books, New York).

Gauthier, David (1982) "No Need for Morality: The Case of the Competitive Market," *Philosophic Exchange*, 3: 41–54.

Gauthier, David (1986) *Morals by Agreement* (Clarendon, Oxford).

Gauthier, David (1988) "Moral Artifice," *Canadian Journal of Philosophy*, 18/2: 386–418.

Gauthier, David (1993) "Fairness and Cores: A Comment on Laden," *Philosophy and Public Affairs*, 22/1: 44–47.

Geis, Gilbert (1968) "The Heavy Electrical Equipment Antitrust Case of 1961," in idem (ed), *White Collar Criminal* (Atherton, New York), 103–118.

Ghilarducci, Teresa (2000) "Myths and Misinformation about America's Retirement System," in Jeff Madrick (ed), *Unconventional Wisdom: Alternative Perspectives on the New Economy* (The Century Foundation, New York), 69–92.

Ghoshal, Sumantra (2005) "Bad Management Theories are Destroying Good Management Practices," *Academy of Management Learning and Education*, 4/1: 75–91.

Ghoshal, Sumantra and Peter Moran (1996) "Bad for Practice: A Critique of the Transaction Cost Theory," *Academy of Management Review*, 21/1: 13–47.

Gibbons, Robert (1998) "Incentives in Organizations," *Journal of Economic Perspectives*, 12/4: 115–132.

Gibson, Kevin (2000) "The Moral Basis of Stakeholder Theory," *Journal of Business Ethics*, 26/3: 245–257.

Gneezy, Uri (2003) "Do High Wages Lead to High Profits? An Experimental Study of Reciprocity using Real Effort," Chicago: University of Chicago Graduate School of Business Working Paper.

Goldberg, Lewis R. (1990) "An Alternative Description of Personality: The Big-Five Factor Structure," *Journal of Personality and Social Psychology*, 596: 1216–1229.

Goodin, Robert (1995) "Political Ideals and Political Practice," *British Journal of Political Science*, 25/1: 37–56.

Goodpaster, Kenneth (1991) "Business Ethics and Stakeholder Analysis," *Business Ethics Quarterly*, 1/1: 53–73.

Grant, Robert (1998) *Cases to Accompany Contemporary Strategy Analysis* (Blackwell, Oxford).

Green, Gary S. (1990) *Occupational Crime* (Nelson-Hall, Chicago).

Greenberg, Jerald (1989) "Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts," *Journal of Applied Psychology*, 75/5: 561–568.

Greenfield, Kent (2006) *The Failure of Corporate Law* (University of Chicago Press, Chicago).

Habermas, Jürgen (1985) *The Theory of Communicative Action*, Vol. 2, trans. Thomas McCarthy (Beacon Press, Boston).

Hacking, Ian (1975) *The Emergence of Probability* (Cambridge University Press, Cambridge).

Hacking, Ian (1990) *The Taming of Chance* (Cambridge University Press, Cambridge).

Hacking, Ian (2002) *An Introduction to Probability and Inductive Logic* (Cambridge University Press, Cambridge).

Hacking, Ian (2003) "Risk and Dirt," Richard V. Ericson and Aaron Doyle (eds), *Risk and Morality* (University of Toronto Press, Toronto), 22–47.

Hamilton, William D. (1963) "The Evolution of Altruistic Behavior," *The American Naturalist*, 97/896: 354–356.

Hammerstein, Peter (2003) "Why Is Reciprocity So Rare in Social Animals?" in Peter Hammerstein (ed), *Genetic and Cultural Evolution of Cooperation* (MIT Press, Cambridge, MA), 83–93.

Haney, Craig, Curtis Banks and Phillip G. Zimbardo (1973) "Interpersonal Dynamics in a Simulated Prison," *International Journal of Criminology and Penology*, 1: 69–97.

Hanks, James J., Jr. (1991) "Playing with Fire: Nonshareholder Constituency Statutes in the 1990s," *Stetson Law Review*, 21/1: 97–120.

Hansmann, Henry (1992) *The Ownership of Enterprise* (Harvard University Press, Cambridge, MA).

Hansmann, Henry and Reinier H. Kraakman (2003) "The End of History for Corporate Law," in Jeffrey Gordon and Mark J. Roe (eds), *Convergence and Persistence in Corporate Governance* (Cambridge University Press, Cambridge), 33–68.

Harford, Tim (2005) *The Undercover Economist* (Oxford University Press, New York).

Harman, Gilbert (1999) "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society*, 99/3: 315–331.

Harman, Gilbert (2003) "No Character or Personality," *Business Ethics Quarterly*, 13/1: 87–94.

Harrison, Jeffrey S. and R. Edward Freeman (1999) "Stakeholders, Social Responsibility and Performance: Empirical Evidence and Theoretical Perspectives," *Academy of Management Journal*, 42/5: 479–485.

Harsanyi, John (1963) "A Simplified Bargaining Model for the *n*-Person Cooperative Game," *International Economic Review*, 4/2: 194–220.

Hart, Oliver (1995) *Firms, Contracts and Financial Structure* (Oxford University Press, Oxford).

Hartman, Edwin M. (1998) "The Role of Character in Business Ethics Education," *Business Ethics Quarterly*, 8/3: 547–559.

Hartshorne, Hugh and Mark A. May (1929) *Studies in the Nature of Character*, Vol. 1 *Studies in Deceit* (MacMillan, New York).

Hasnas, John (1998) "The Normative Theories of Business Ethics: A Guide to the Perplexed," *Business Ethics Quarterly*, 8/1: 19–42.

Hauser, Mark (2006) *Moral Minds* (HarperCollins, New York).

Hausman, Daniel M. and Michael S. McPherson (1996) *Economic Analysis and Moral Philosophy* (Cambridge University Press, Cambridge).

Hayek, Friedrich A. (1940) "Socialist Calculation: The Competitive 'Solution,'" *Economica*, 7: 125–149.

Hayek, Friedrich A. (1945) "The Use of Knowledge in Society," *American Economic Review*, 35/4: 519–530.

Hayek, Friedrich A. (1976) *Law, Legislation and Liberty*, Vol. 2 (Routledge & Kegan Paul, London).

Hayek, Friedrich A. (1979) *The Counter-Revolution of Science* (Liberty Books, Indianapolis).

Hayek, Friedrich A. (2002) "Competition as a Discovery Procedure," trans. Marcellus S. Snow, *Quarterly Journal of Austrian Economics*, 5: 9–23.

Heath, Joseph (1996) "Rational Choice as Critical Theory," *Philosophy and Social Criticism*, 22/5: 43–62.

Heath, Joseph (1997) "Intergenerational Cooperation and Distributive Justice," *Canadian Journal of Philosophy*, 27/3: 361–376.

Heath, Joseph (2001a) *Communicative Action and Rational Choice* (MIT Press, Cambridge, MA)

Heath, Joseph (2001b) *The Efficient Society* (Penguin, Toronto).

Heath, Joseph (2004a) "A Market Failures Approach to Business Ethics," in Bernard Hodgson (ed), *The Invisible Hand and the Common Good* (Springer, Berlin), 69–89 (reprinted in this volume).

Heath, Joseph (2004b) "Dworkin's Auction," *Politics, Philosophy and Economics*, 3/3: 313–335.

Heath, Joseph (2006a) "Business Ethics Without Stakeholders," *Business Ethics Quarterly*, 16/4: 533–557 (reprinted in this volume).

Heath, Joseph (2006b) "The Benefits of Cooperation," *Philosophy and Public Affairs*, 34/4: 313–351 (reprinted in this volume).

Heath, Joseph (2007) "An Adversarial Ethic for Business: Or, When Sun-Tzu Met the Stakeholder," *Journal of Business Ethics*, 72/4, 359–374 (reprinted in this volume).

Heath, Joseph (2008a) "Business Ethics and Moral Motivation: A Criminological Perspective," *Journal of Business Ethics*, 83: 595–614 (reprinted in this volume).

Heath, Joseph (2008b) *Following the Rules* (Oxford University Press, New York).

Heath, Joseph (2009) *Filthy Lucre* (HarperCollins: Toronto).

Heath, Joseph (2011) "Three Normative Models of the Welfare State," *Public Reason*, 3/2: 13–44.

Heath, Joseph and Wayne Norman (2004) "Stakeholder Theory, Corporate Governance, and Public Management," *Journal of Business Ethics*, 53/3: 247–265 (reprinted in this volume).

Heimer, Carol A. (2003) "Insurers as Moral Actors," in Richard V. Ericson and Aaron Doyle (eds), *Risk and Morality* (University of Toronto Press, Toronto), 284–316.

Hellman, Deborah S. (1997) "Is Actuarially Fair Insurance Pricing Actually Fair? A Case Study in Insuring Battered Women," *Harvard Civil Rights-Civil Liberties Law Review*, 32/2: 355–411.

Henderson, David (2001) *Misguided Virtue: False Notions of Corporate Social Responsibility* (New Zealand Business Roundtable, Wellington).

Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath (2001) "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Review*, 91: 73–78.

Hertzman, Clyde and Arjumand Siddiqi (2009) "Population Health and the Dynamics of Collective Development," in Peter A. Hall and Michèle Lamont (eds), *Successful Societies* (Cambridge University Press, Cambridge), 23–52.

Heyman, James, and Dan Ariely (2004) "Effort for Payment: A Tale of Two Markets," *Psychological Science*, 15/11: 787–793.

Hillard, Kaplan, Kim Hill and A. Magdalena Hurtado (1990) "Risk, Foraging and Food Sharing among the Aché," in Elizabeth Cashdan (ed), *Risk and Uncertainty in Tribal and Peasant Economies* (Westview Press, Boulder), 107–143.

Hilts, Philip J. (2003) "Epilogue: Greed and Goodness," in Philip J. Hilts (e), *Protecting America's Health: The FDA, Business, and One Hundred Years of Regulation* (Alfred A. Knopf, New York), 338–344.

Hindelang, Michael J. (1970) "The Commitment of Delinquents to Their Misdeeds: Do Delinquents Drift?" *Social Problems*, 17/4: 502–509.

Hindelang, Michael J. (1974) "Moral Evaluations of Illegal Behavior," *Social Problems*, 21/3: 370–385.

Hollinger, Richard C. and John P. Clark (1983) *Theft by Employees* (Lexington Books, Lexington, MA).

Hollinger, Richard C., Karen B. Slora and William Terris (1992) "Deviance in the Fast Food Restaurant: Correlates of Employee Theft, Altruism and Counterproductivity," *Deviant Behavior*, 13/2: 155–184.

Holmström, Bengt (1979) "Moral Hazard and Observability," *Bell Journal of Economics*, 10/1: 74–91.

Holmström, Bengt (1982) "Moral Hazard in Teams," *Bell Journal of Economics*, 13/2: 324–349.

Holmström, Bengt and Paul Milgrom (1991) "Multitask Principal–Agent Analysis," *Journal of Law, Economics and Organization*, 7: 24–51.

Homann, Karl (1993) "Wirtschaftsethik: Die Funktion der Moral in der modernen Wirtschaft," in Josef Wieland (ed), *Wirtschaftsethik und Theorie der Gesellschaft* (Suhrkamp, Frankfurt), 32–53.

Honoré, Anthony M. (1961) "Ownership," in Anthony G. Guest (ed), *Oxford Essays in Jurisprudence* (Oxford University Press, Oxford), 107–47.

Hooker, Brad (2003) *Ideal Code, Real World* (Oxford University Press, Oxford).

Hsieh, Nien-hê (2005) "Rawlsian Justice and Workplace Republicanism," *Social Theory and Practice*, 31/1: 115–142.

Hume, David (1978) [1739] *A Treatise of Human Nature*, 2nd edn, ed. Peter H. Nidditch (Clarendon, Oxford).

Hurka, Thomas (2001) *Virtue, Vice and Value* (Oxford University Press, Oxford).

Hurka, Thomas (2012) "Desert: Individualistic and Holistic," in Serena Olsaretti (ed), *Desert and Justice* (Oxford University Press, Oxford), 45–68.

Isaac, Mark, Kenneth McCue and Charles Plott (1985) "Public Goods Provision in an Experimental Environment," *Journal of Public Economics*, 26/1: 51–74.

Jaffe, Adam B. and Josh Lerner (2004) *Innovation and Its Discontents* (Princeton University Press, Princeton, NJ)

Jensen, Michael C. (2002) "Value Maximization, Stakeholder Theory, and the Corporate Objective Function," *Business Ethics Quarterly*, 12/2: 235–256.

Jensen, Michael C. and William H. Meckling (1976) "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3/4: 305–360.

Johnson, Simon and James Kwak (2010) *13 Bankers* (Random House, New York).

Jones, Stephen (2005) *Criminology*, 3rd edn (Oxford University Press, Oxford).

Jones, Thomas M., Andrew C. Wicks and R. Edward Freeman (2002) "Stakeholder Theory: The State of the Art," in Norman Bowie (ed), *Blackwell Guide to Business Ethics* (Blackwell, Oxford), 19–37.

Jonsen, Albert R. and Stephen Toulmin (1988) *The Abuse of Casuistry* (University of California Press, Berkeley).

Kaplan, Benjamin J. (2007) *Divided by Faith* (Harvard University Press, Cambridge, MA).

Kavka, Gregory S. (1983) "When Two 'Wrongs' Make a Right: An Essay on Business Ethics," *Journal of Business Ethics*, 2/1: 61–66.

Kay, John (2003) *The Truth about Markets* (Penguin, London).

Kelly, Marjorie (1997) "Why All the Fuss about Stockholders?" *Business Ethics*, 11/1: 5.

Kelly, Marjorie (2001) "Why all the Fuss about Stockholders?" In Marjorie Kelly (ed), *The Divine Right of Capital* (Berrett-Koehler, San Francisco), 1–18.

Khurana, Rakesh, Nitin Nohria and Daniel Penrice (2005) "Management as a Profession," in Jay W. Lorsch, Leslie Berlowizt and Andy Zelleke (eds), *Restoring Trust in American Business* (MIT Press, Cambridge, MA), 43–62.

Kim, Oliver and Mark Walker (1984) "The Free Rider Problem: Experimental Evidence," *Public Choice*, 43/1: 3–24.

Kimball, Spencer L. (1979) "Reverse Sex Discrimination: *Manhart*," *American Bar Foundation Research Journal*, 4/1: 83–139.

Klockars, Carl B. (1974) *The Professional Fence* (Free Press, New York).

Knight, Jack (1992) *Institutions and Social Conflict* (Cambridge University Press, Cambridge).

Koehn, Daryl (1995) "A Role for Virtue Ethics in the Analysis of Business Practice," *Business Ethics Quarterly*, 5/3: 533–539.

Kolm, Serge-Christophe (1997) *Justice and Equity*, trans. Harold F. See (MIT Press, Cambridge, MA).

Kornai, Janos (1992) *The Socialist System* (Princeton University Press, Princeton).

Korsgaard, Christine (1996) "Skepticism about Practical Reason," in Christine M. Korsgaard (ed), *Creating the Kingdom of Ends* (Cambridge University Press, Cambridge), 311–334.

Koslowski, Peter (2001) *Principles of Ethical Economy*, trans. David W. Lutz (Kluwer, Dordrecht).

Kraakman, Reinier, Paul Davies, Henry B. Hansmann, Gerard Hertig, Klaus J. Hopt, et al. (2004) *The Anatomy of Corporate Law* (Oxford University Press, Oxford).

Kraut, Robert E. (1976) "Deterrent and Definitional Influences on Shoplifting," *Social Problems*, 23/3: 358–368.

Kulik, Brian W. (2005) "Agency Theory, Reasoning and Culture at Enron: In Search of a Solution," *Journal of Business Ethics*, 59/4: 347–360.

Kymlicka, Will (2002) *Contemporary Political Philosophy*, 2nd edn (Oxford University Press, Oxford).

Labaton, Stephen (2004) "Perle Asserts Hollinger's Conrad Black Misled Him," *New York Times*, Sept. 6.

Laden, Anthony Simon (1993) "Games Philosophers Play: A Reply to Gauthier," *Philosophy and Public Affairs*, 22/1: 48–52.

Laffont, Jean-Jacques and David Martimort (2002) *The Theory of Incentives: The Principal–Agent Model* (Princeton University Press, Princeton, NJ).

Laibson, David (1997) "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112/2: 443–477.

Landsheer, Johannes A., Harm't Hart and Willem Kox (1994) "Delinquent Values and Victim Damage: Exploring the Limits of Neutralization Theory," *British Journal of Criminology*, 34/1: 44–53.

Lane, Robert E. (1953) "Why Businessmen Violate the Law," *Journal of Criminal Law, Criminology and Police Science*, 44/2: 151–165.

Lane, Theodor W. and Glen E. Davis (1987) "Child Maltreatment and Juvenile Delinquency: Does a Relationship Exist?" in John D. Burchard and Sara N. Burchard (eds), *Prevention of Delinquent Behavior* (Sage Publications, Newbury Park), 122–136.

Langford, John and Ken Huffman (1988) "Air Canada," in Allan Tupper and G. Bruce Doern (eds), *Privatization, Public Corporations and Public Policy in Canada* (Institute for Research on Public Policy, Halifax, NS), 93–150.

Langtry, Bruce (1994) "Stakeholders and the Moral Responsibility of Business," *Business Ethics Quarterly*, 4/4: 431–443.

Leaman, Oliver (1988) "Cheating and Fair Play in Sport," in William J. Morgan and Klaus V. Meier (eds), *Philosophic Inquiry in Sport* (Human Kinetics, Champaign, IL), 277–282.

Lee, Ian B. (2006) "Corporate Law and the Role of Corporations in Society: Monism, Pluralism, Markets and Politics," *Canadian Bar Review*, 85/1: 1–28.

Lensberg, Terje (1988) "Stability and the Nash Solution," *Journal of Economic Theory*, 45/2: 330–341.

Leonard, William N. and Marvin G. Weber (1970) "Automakers and Dealers: A Study of Criminogenic Market Forces," *Law and Society Review*, 4/3: 407–424.

Lerner, Abba (1977) "Marginal Cost Pricing in the 1930's," *American Economic Review*, 67/1: 235–243.

Lerner, Abba P. (1937) "Statics and Dynamics in Socialist Economics," *The Economic Journal*, 47/186: 253–270.

Levin, Michael (1989) "Ethics Courses: Useless," *New York Times*, Nov. 25, section I.

Lewin, Arie Y. (1982) "Public Enterprise, Purposes and Performance: A Survey of Western European Experience," in W. T. Stanbury and Fred Thompson (eds), *Managing Public Enterprises* (Praeger, New York), 51–78.

Lewin, Christopher G. (2003) *Pensions and Insurance before 1800* (Tuckwell Press, East Lothian).

Liberman, Varda, Steven M. Samuels and Lee Ross (2004) "The Name of the Game: Predictive Power of Reputations Versus Situational Labels in Determining Prisoner's Dilemma Game Moves," *Personality and Social Psychology Bulletin*, 30/9:1175–1185.

Lipsey, Mark W. and James H. Derzon (1998) "Predictors of Violent or Serious Delinquency in Adolescence and Early Adulthood," in Rolf Loeber and David P. Farrington (eds), *Serious and Violent Juvenile Offenders* (Sage Publications, London), 86–105.

Lipsey, Richard and Kelvin Lancaster (1956) "The General Theory of Second Best," *Review of Economic Studies*, 24/1: 11–32.

Lukes, Steven (1975) "Political Ritual and Social Integration," *Sociology*, 9/2: 289–308.

MacAvoy, Paul W. (1977) *Federal-State Regulation of the Pricing and Marketing of Insurance* (American Enterprise Institute for Public Policy Research, Washington, DC).

Macintyre, Alasdair (1979) "Corporate Modernity and Moral Judgment: Are They Mutually Exclusive?" in Kenneth E. Goodpaster and Kenneth M. Sayer (eds), *Ethics and Problems of the 21st Century* (Notre Dame University Press, Notre Dame), 122–135.

Macintyre, Alasdair (1995) *Marxism and Christianity* (Duckworth, London).

MacIntyre, Alasdair (1984) *After Virtue*, 2nd edn (University of Notre Dame Press, Notre Dame).

MacIver, Robert M. (1922) "The Social Significance of Professional Ethics," *Annals of the American Academy of Political and Social Science*, 101/1: 5–11.

Maitland, Ian (1994) "The Morality of the Corporation: An Empirical or Normative Disagreement?" *Business Ethics Quarterly*, 4/4: 445–458.

Malkiel, Burton G. (2003) "The Efficient Market Hypothesis and Its Critics," *Journal of Economic Perspectives*, 17/1: 59–82.

Mandeville, Bernard (1705/1997) *Fable of the Bees*, ed. E. J. Hundert (Hackett, Indianapolis).

Manent, Pierre (1995) *An Intellectual History of Liberalism*, trans. Rebecca Balinski (Princeton University Press, Princeton).

March, James G. and Herbert A. Simon (1958) *Organizations* (John Wiley & Sons, New York).

Marcoux, Alexei (2000) "Business Ethics Gone Wrong," *Cato Commentary*, http://www.cato.org/publications/commentary/business-ethics-gone-wrong.

Marcoux, Alexei (2003) "A Fiduciary Argument against Stakeholder Theory," *Business Ethics Quarterly*, 13/1: 1–24.

Marens, Richard and Andrew Wicks (1999) "Getting Real: Stakeholder Theory, Managerial Practice, and the General Irrelevance of Fiduciary Duties Owed to Shareholders," *Business Ethics Quarterly*, 9/2: 273–293.

Mars, Gerald (1982) *Cheats at Work* (Unwin Wyham, London).

Martin, Dominic (2012) *Rivalité et marchés* (Université de Montréal & Université Catholique de Louvain, unpublished diss.).

Marwell, Gerald and Ruth E. Ames (1981) "Economists Free Ride, Does Anyone Else?" *Journal of Public Economics*, 15/3: 295–310.

Maxwell, John C. (2003) *There's No Such Thing as "Business" Ethics: There's Only One Rule for Making Decisions* (First Warner, New York).

McCabe, Donald L., Linda Klebe Trevino and Kenneth D. Butterfield (1996) "The Influence of Collegiate and Corporate Codes of Conduct on Ethics-Related Behavior in the Workplace," *Business Ethics Quarterly*, 6/4: 461–476.

McCabe, Donald L., Linda Klebe Trevino, and Kenneth D. Butterfield (2001) "Cheating in Academic Institutions: A Decade of Research," *Ethics and Behavior*, 11/3: 219–232.

McCrae, Robert R. and Paul T. Costa, Jr. (1996) "Toward a New Generation of Personality Theories: Theoretical Contexts for the Five-factor Model," in Jerry S. Wiggins (ed), *The Five-Factor Model of Personality: Theoretical Perspectives* (Guilford, New York), 51–87.

McGregor, Douglas (1960) *The Human Side of Enterprise* (McGraw-Hill, New York).

McGurn, William A. (1988) "Spotting the Thieves Who Work among Us," *Wall Street Journal*, Mar. 7.

McKenzie, Evan (1994) *Privatopia* (Yale University Press, New Haven).

McLean, Bethany and Peter Elkind (2003) *The Smartest Guys in the Room* (Penguin, New York).

McLean, Bethany and Peter Elkind (2004) *The Smartest Guys in the Room*, 2nd edn (Penguin, New York).

McMahon, Christopher (1981) "Morality and the Invisible Hand," *Philosophy and Public Affairs*, 10/3: 247–277.

McMahon, Christopher (1994) *Authority and Democracy* (Princeton University Press, Princeton, NJ).

McPherson, David (2013) "Vocational Virtue Ethics: Prospects for a Virtue Ethic Approach to Business," *Journal of Business Ethics*, 116/2: 283–296.

Menger, Carl (1985) *Investigations into the Methods of the Social Sciences* (New York University Press, New York).

Michalos, Alex C. (1995) *A Pragmatic Approach to Business Ethics* (Sage, Thousand Oaks, CA).

Milgram, Stanley (1974) *Obedience to Authority* (Harper and Row, New York).

Milgrom, Paul and John Roberts (1992) *Economics, Organization and Management* (Prentice Hall, Upper Saddle River, NJ).

Mill, John Stuart (2004/1848) *Principles of Political Economy* (Prometheus, Amherst).

Miller, Dale T. (1999) "The Norm of Self-Interest," *American Psychologist*, 54/12: 1053–1060.

Miller, David (1981) "Market Neutrality and the Failure of Cooperatives," *British Journal of Political Science*, 11/3: 309–329.

Miller, David (1989) *Market, State and Community* (Clarendon, Oxford).

Miller, Gary (1992) *Managerial Dilemmas* (Cambridge University Press, Cambridge).

Milton, John (1824) *The Poetical Works of John Milton*, Vol. 1, ed. Edward Hawkins (W. Baxter: Oxford).

Minor, William W. (1981) "Techniques of Neutralization: A Reconceptualization and Empirical Examination," *Journal of Research in Crime and Delinquency*, 18/2: 295–318.

Mintz, Mortin (1969) "F.D.A. and Panalba: A Conflict of Commercial, Therapeutic Goals," *Science*, 165: 875–881.

Mitchell, Lawrence (2001) *Corporate Irresponsibility* (Yale University Press, New Haven).

Mitchell, Ronald K., Bradley R. Agle and Donna J. Wood (1997) "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts," *The Academy of Management Review*, 22/4: 853–886.

Mokhiber, Russell (2006) "Top-100 Corporate Criminals of the Decade," *Corporate Crime Reporter*, http:// www.corporatecrimereporter.com/top100.html.

Moody-Adams, Michele (1997) *Fieldwork in Familiar Places* (Harvard University Press, Cambridge, MA).

Moore, Geoff (1999) "Tinged Shareholder Theory: Or What's so Special about Stakeholders?" *Business Ethics: A European Review*, 8/2: 117–127.

Moore, Geoff (2005) "Humanizing Business: A Modern Virtue Ethics Approach," *Business Ethics Quarterly*, 15/2: 237–255.

Moore, Geoff (2008) "Re-Imagining the Morality of Management: A Modern Virtue Ethics Approach," *Business Ethics Quarterly*, 18/4: 483–511.

Moskowitz, Debbie S. (1994) "Cross-Situational Generality and the Interpersonal Circumplex," *Journal of Personality and Social Psychology*, 66/5: 921–933.

Moss, David A. (2002) *When All Else Fails* (Harvard University Press, Cambridge, MA).

Myles, Gareth D (1995). *Public Economics* (Cambridge University Press, Cambridge).

Nash, John (1949) "The Bargaining Problem," *Econometrica*, 18/2: 155–162.

Nash, John (1951) "Non-cooperative Games," *Annals of Mathematics*, 54/2: 289–295.

National Energy Board (2001) *Canadian Electricity: Trends and Issues* (National Energy Board, Calgary).

Neave, David (1991) *Mutual Aid in the Victorian Countryside* (Hull University Press, Hull, QC).

Newton, L. 1992 "Agents for the Truly Greedy," in Norman E. Bowie and R. Edward Freeman (eds), *Ethics and Agency Theory* (Oxford University Press, Oxford), 93–113.

Ng, Yew-Kuang (2004) *Welfare Economics: Towards a More Complete Analysis* (Palgrave Macmillan, New York).

Nichols, Shaun (2004) *Sentimental Rules* (Oxford, New York).

Nora, Simon (1967) *Rapport sur les entreprises publiques* (La documentation française, Paris).

Nord, Erik (1999) *Cost-Value Analysis in Health Care* (Cambridge University Press, Cambridge).

Noreen, Eric (1988) "The Economics of Ethics: A New Perspective on Agency Theory," *Accounting, Organizations and Society*, 13/4: 359–369.

Norman, Wayne (2012) "Business Ethics as Self-Regulation: Why Principles that Ground Regulations Should Be Used to Ground Beyond-Compliance Norms as Well," *Journal of Business Ethics*, 102 (S1): 43–57.

Norman, Wayne and Chris MacDonald (2004) "Getting to the Bottom of the 'Triple Bottom Line'," *Business Ethics Quarterly*, 14/2: 243–262.

Norris, Floyd (2004) "Panel says Conrad Black ran a 'Corporate Kleptocracy,'" *New York Times*, Aug. 31.

Nove, Alec (1983) *Economics of Feasible Socialism* (Allen & Unwin, London).

Nozick, Robert (1974) *Anarchy, State and Utopia* (Basic Books, New York).

Nussbaum, Martha (2011) "Perfectionist Liberalism and Political Liberalism," *Philosophy & Public Affairs*, 39/1: 3–45.

Okun, Arthur M. (1975) *Equality and Efficiency: The Big Tradeoff* (Brookings Institute, Washington, DC).

Ordeshook, Peter C. (1986) *Game Theory and Political Theory* (Cambridge University Press, Cambridge).

Orts, Eric W. (1992) "Beyond Shareholders: Interpreting Corporate Constituency Statutes," *George Washington Law Review*, 61/1: 114–135.

Orts, Eric W. (1998) "Shirking and Sharking: A Legal Theory of the Firm," *Yale Law and Policy Review*, 16/2: 265–329.

Orts, Eric W. and Alan Strudler (2009) "Putting a Stake in Stakeholder Theory," *Business Ethics Quarterly*, 88/4: 605–615.

Orwin, Clifford (2009) "Can We Teach Ethics? When Pigs Fly," *Globe and Mail*, Nov. 6.

Outreville, J. François (1998) *Theory and Practice of Insurance* (Kluwer, Dordrecht).

Paine, Lynn S. (1994) "Managing for Organizational Integrity," *Harvard Business Review*, Mar./Apr.: 106–117.

Paine, Lynn S. (2003) *Value Shift: Why Companies Must Merge Social and Financial Imperatives to Achieve Superior Performance* (McGraw-Hill, New York).

Palmer, John, John Quinn and Ray Resendes (1983) "A Case Study of Public Enterprise: Gray Coach Lines Ltd," in J. Robert S. Prichard (ed), *Crown Corporations in Canada* (Butterworths, Toronto), 369–438.

Parsons, Talcott (1937) *Structure of Social Action* (McGraw-Hill, New York).

Parsons, Talcott (1951) *The Social System* (Free Press, New York).

Parsons, Talcott (1970) *Social Structure and Personality* (Free Press, New York).

Parsons, Talcott, Edward Shils, Kaspar D. Naegele and Jesse R. Pitts (eds) (1961) *Theories of Society* (Free Press, Glencoe).

Pauly, Mark V. (1984) "Is Cream-skimming a Problem for the Competitive Medical Market?" *Journal of Health Economics*, 3/1: 87–95.

Pepper v. Litton 308 U.S. 295 (1939). Cited in Robert C. Clark. *Agency Costs versus Fiduciary Duties*. Division of Research, Harvard Business School, 1983.

Peterson, Martin (2009) "Mixed Solutions to the Number Problem," *Journal of Moral Philosophy*, 6: 166–177.

Philips, Robert, R. Edward Freeman and Andrew C. Wicks (2003) "What Stakeholder Theory Is Not," *Business Ethics Quarterly*, 13/4: 479–502.

Poff, Deborah C. (2004) *Business Ethics in Canada*, 4th edn (Prentice Hall, Scarborough).

Polanyi, Karl (1957) *The Great Transformation* (Beacon Press, Boston).

Porter, Michael E. (1990) *The Competitive Advantage of Nations* (Free Press, New York).

Posner, Richard A. (1973) *Economic Analysis of Law* (Little Brown, Boston).

Promislow, David S. (1987) "Measurement of Equity," *Transactions of the Society of Actuaries*, 39: 215–237.

Przeworski, Adam (2003) *States and Markets* (Cambridge University Press, Cambridge).

Putnam, Robert (2000) *Bowling Alone* (Simon & Schuster, New York).

Quinn Dennis P. and Thomas M. Jones (1995) "An Agent Morality View of Business Policy," *Academy of Management Review*, 20/1: 22–42.

Rachlin, Howard (2000) *The Science of Self-Control* (Harvard University Press, Cambridge, MA).

Railton, Peter (1995) "Made in the Shade: Moral Compatibilism and the Aims of Moral Theory," in Jocelyn Couture and Kai Nielsen (eds), *On the Relevance of Metaethics: Canadian Journal of Philosophy*, Supplementary Volume 21 (University of Calgary Press, Calgary), 79–106.

Ramanadham, Vemuri V. (1991) *The Economics of Public Enterprise* (Routledge, London).

Rasmusen, Eric (1986) *Games and information*, 2nd edn (Blackwell, Cambridge, MA).

Rasmusen, Eric (1989) *Games and Information*, 3rd edn (Blackwell, Cambridge, MA).

Rawls, John (1958) "Justice as Fairness," *Philosophical Review*, 67/2: 164–194.

Rawls, John (1971) *A Theory of Justice* (Harvard University Press, Cambridge, MA).

Rawls, John (1996) *Political Liberalism*, 2nd edn (Columbia University Press, New York).

Rawls, John (1997) "The Idea of Public Reason Revisited," *University of Chicago Law Review*, 64/3: 765–807.

Rawls, John (1999) *A Theory of Justice*, 2nd edn (Harvard University Press, Cambridge, MA).

Rawls, John (2001) *Justice as Fairness: A Restatement* (Harvard University Press, Cambridge, MA).

Revsine, Lawrence (1991) "The Selective Financial Misrepresentation Hypothesis," *Accounting Horizons*, 5/4: 16–27.

Richardson, John E. (2004) *Business Ethics*, 16th edn (McGraw-Hill, Dubuque).

Richerson, Peter J. and Robert Boyd (1995) *Not by Genes Alone* (University of Chicago Press, Chicago).

Ripstein, Arthur (1998) *Equality, Responsibility and the Law* (Cambridge University Press, Cambridge).

Ripstein, Arthur (2010) "Critical Notice: G. A. Cohen *Rescuing Justice and Equality*," *Canadian Journal of Philosophy*, 40/4: 669–699.

Robé, Jean-Phillipe (2011) "The Legal Structure of the Firm," *Accounting, Economics and Law*, 1/1: 1–86.

Robeyns, Ingrid (2008) "Ideal theory in theory and practice," *Social Theory and Practice*, 34/3: 341–362.

Roemer, John (1985) "Should Marxists be Interested in Exploitation?" *Philosophy and Public Affairs*, 14/1: 30–65.

Roemer, John (1994) *A Future for Socialism* (Harvard University Press, Cambridge, MA).

Roemer, John (2010) "Incentive Provision and Coordination as Market Functions," *Zeitschrift für ArbeitsmarktForschung*, 43/1: 6–16.

Rosen, Harvey S., et al. (2003) *Public Finance in Canada*, 2nd edn (McGraw Hill, Toronto).

Ross, Lee and Richard E. Nisbett (1991) *The Person and the Situation* (Temple University Press, Philadelphia).

Russell, Daniel E. (2009) *Practical Intelligence and the Virtues* (Oxford University Press, Oxford).

Rylands, L. Gordon (1889) *Crime: Its Causes and Remedy* (T. Fisher Unwin, London).

Salgado, Jesús F. (2002) "The Big Five Personality Dimensions and Counterproductive Behaviors," *International Journal of Selection and Assessment*, 10/1–2: 117–125.

Samuelson, Paul (1967) *Economics: An Introductory Analysis* (McGraw-Hill, New York).

Satz, Debra (2010) *Why Some Things Should Not Be For Sale* (Oxford University Press, Oxford).

Schelling, Thomas (1984) *Choice and Consequence* (Harvard University Press, Cambridge, MA).

Schmidtz, David (2011) "Nonideal Theory: What It Is and What It Needs to Be," *Ethics*, 121/4: 772–796.

Schneider, Friedrich and Werner W. Pommerehne, (1981) "Free Riding and Collective Action: An Experiment in Public Microeconomics," *Quarterly Journal of Economics*, 96/4: 689–704.

Schotter, Andrew (1981) *The Economic Theory of Social Institutions* (Cambridge University Press, Cambridge).

Schuessler, Karl P. and Donald R. Cressey (1950) "Personality Characteristics of Criminals," *American Journal of Sociology*, 55/5: 476–484.

Schultz, Walter J. (2001) *The Moral Conditions of Economic Efficiency* (Cambridge University Press, Cambridge).

Schumpeter, Joseph A. (1943) *Capitalism, Socialism and Democracy* (Routledge, London).

Scott, Joseph E. and Fahed Al-Thakeb (1977) "The Public's Perception of Crime: A Comparative Analysis of Scandinavia, Western Europe, the Middle East, and the United States," in C. Ronald Huff (ed), *Contemporary Corrections* (Sage, Beverly Hills), 78–88.

Selten, Robert (1975) "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4/1: 25–55.

Shankman, Neil A. (1999) "Reframing the Debate between Agency and Stakeholder Theories of the Firm," *Journal of Business Ethics*, 19/4: 319–334.

Sher, George (1992) "Knowing about Virtue," in John W. Chapman and William A. Galston (eds), *Virtue, Nomos 34* (New York University Press, New York), 91–116.

Shipman, Alan (1999) *The Market Revolution and Its Limits* (Routledge, London).

Simmons, A. John (2010) "Ideal and Nonideal Theory," *Philosophy and Public Affairs*, 38/1: 5–36.

Simon, Rita James and Heather Ahn-Redding (2005) *The Crimes Women Commit* (Lexington Books, Lanham, MA).

Simon, Robert L. (1988) "Good Competition and Drug-Enhanced Performance," in William J. Morgan and Klaus V. Meier (eds), *Philosophic Inquiry in Sport* (Human Kinetics, Champaign, IL), 209–214.

Simpson, Sally (2002) *Corporate Crime, Law and Social Control* (Cambridge University Press, Cambridge).

Skillen, Anthony (1998) "Sport is for Losers," in Michael J. McNamee and S. Jim Parry (eds), *Ethics and Sport* (Routledge, London), 169–182.

Smith, Adam (1776/1976/1776) *Inquiry into the Nature and Causes of the Wealth of Nations*, 2 vols., ed. R. H. Campbell and A. S. Skinner (Oxford University Press, Oxford).

Smith, Adam (1835) *Inquiry into the Nature and Causes of the Wealth of Nations*, 4 vols., ed. Edward Gibbon Wakefield (Charles Knight, London).

Smith, J. Maynard (1964) "Group Selection and Kin Selection," *Nature*, 201/4924: 1145–1147.

Smith, Michael (1994) *The Moral Problem* (Wiley-Blackwell, Oxford).

Snyder, Neil H. and Karen E. Blair (1989) "Dealing with Employee Theft," *Business Horizons*, 32: 27–34.

Solomon, Robert C. (1992) *Ethics and Excellence* (Oxford University Press, Oxford).

Solomon, Robert C. (1999) "Game Theory as a Model for Business and Business Ethics," *Business Ethics Quarterly*, 9/1: 11–30.

Solomon, Robert C. (2003) "Victims of Circumstances? A Defense of Virtue Ethics in Business," *Business Ethics Quarterly*, 13/1: 43–62.

Solomon, Robert C. and Clancy Martin (2004) *Above the Bottom Line,* 3rd edn (Wadsworth, Belmont, Calif.).

Spencer, John C. (1965) "White Collar Crime," in Tadeusz Grygier, Howard Jones, John Carrington Spencer and Herman Mannheim (eds), *Criminology in Transition* (Tavistock, London), 233–266.

Spieker, Ben (1999) "Habituation and Training in Early Moral Upbringing," in David Carr and Jan Steutel (eds), *Virtue Ethics and Moral Education* (Blackwell, London), 210–223.

Sreenivasan, Gopal (2002) "Errors about Errors: Virtue Theory and Trait Attribution," *Mind*, 111/441: 47–68.

Sreenivasan, Gopal (2012) "The Situationist Critique of Virtue Ethics," in Daniel C. Russell (ed), *The Cambridge Companion to Virtue Ethics* (Cambridge University Press, Cambridge), 290–314.

Stark, Andrew (1993) "What's the Matter with Business Ethics?" *Harvard Business Review* (May/June): 38–48.

Stark, Andrew (2010) *Drawing the Line* (Brookings Institute, Washington, DC).

Steenbergen, Johan, Albert Buisman and Ivo van Hilvoorde (2001) "Meanings of Fair Play in Competitive Sport," in Johan Steenberger, Paul De Knop and Agnes Elling (eds), *Values and Norms in Sport* (Meyer and Meyer Sport, Oxford), 137–156.

Stein, Janice Gross (2002) *The Cult of Efficiency* (Anansi, Toronto).

Stevenson, Garth (1988) "Canadian National Railways and Via Rail," in Allan Tupper and G. Bruce Doern (eds), *Privatization, Public Corporations and Public Policy in Canada* (Institute for Research on Public Policy, Montreal, QC), 45–92.

Stiglitz, Joseph E. (1994) *Whither Socialism?* (MIT Press, Cambridge, MA).

Stiglitz, Joseph, et al. (1989) *The Economic Role of the State* (Basil Blackwell, Oxford).

Stout, David (2006) "U.S. Orders New Training in Iraq after Civilian Deaths," *New York Times*, June 1.

Stout, Lynn A. (2008) "Why We Should Stop Teaching *Dodge v. Ford*," *Virginia Law and Business Review*, 3: 163–176.

Surowiecki, James (2005) *The Wisdom of Crowds* (Anchor, New York).

Sutherland, Edwin (1949) *White Collar Crime* (Dryden Press, New York).

Sutherland, Edwin (1968) "Crime of Corporations," in Gilbert Geis (ed), *White Collar Criminal* (Atherton, New York), 57–70.

Sutherland, Edwin and Donald Cressey (1978) *Criminology*, 10th edn (Lippincott, New York).
Sykes, Gresham M. and David Matza (1957) "Techniques of Neutralization: A Theory of Delinquency," *American Sociological Review*, 22/6: 664–670.
Taurek, John (1977) "Should the Numbers Count?" *Philosophy and Public Affairs*, 6/4: 293–316.
Taylor, Charles (1985) *Philosophy and the Human Sciences: Philosophical Papers 2* (Cambridge University Press, Cambridge).
Taylor, Charles (1994) *Multiculturalism and the Politics of Recognition*, exp. edn (Princeton University Press, Princeton).
Temkin, Larry (1993) *Inequality* (Oxford University Press, Oxford).
Tenbrunsel, Ann E. and David Messick (1999) "Sanctioning Systems, Decision Frames, and Cooperation," *Administrative Science Quarterly*, 44/4: 684–707.
Tennenbaum, David J. (1977) "Personality and Criminality: A Summary and Implications of the Literature," *Journal of Criminal Justice*, 5/3: 225–235.
The Economist (2005) "Survey: Corporate Social Responsibility," *The Economist*, Jan. 20.
The Economist (2011) "How Bad is BP?" *The Economist*, Jan. 22.
Thomsen, Steen (2001) "Business Ethics as Corporate Governance," *European Journal of Law and Economics*, 11/2: 153–164.
Thurow, Lester C. (1987) "Ethics Doesn't Start in Business Schools," *New York Times*, Jun. 14, section 4.
Tomasello, Michael (1999) *The Cultural Origins of Human Cognition* (Harvard University Press, Cambridge, MA).
Tomasi, John (2012) *Free Market Fairness* (Princeton University Press, Princeton).
Trebilcock, Michael (1993) *The Limits of Freedom of Contract* (Harvard University Press, Cambridge, MA).
Trebilcock, Michael J. and Roy Hrap (2003) "What Will Keep the Lights On in Ontario?" *C. D. Howe Institute Commentary*, 131: 6.
Tullock, Gordon (1970) *Private Wants, Public Means* (Basic Books, New York).
Tupper, Allan and G. Bruce Doern (eds) (1988) *Privatization, Public Policy and Public Corporations in Canada* (Institute for Research on Public Policy, Montreal).
Turiel, Elliot (2002) *The Culture of Morality* (Cambridge University Press, Cambridge).
Turow, Joseph, Lauren Feldman and Kimberly Meltzer (2005) "Open to Exploitation: America's Shoppers Online and Offline," *Annenberg School for Communication Departmental Papers*, http://repository.upenn.edu/asc_papers/35.
Tversky, Amos and Daniel Kahneman (1991) "Loss Aversion in Riskless Choice: A Reference Dependent Model," *Quarterly Journal of Economics*, 106/4: 1039–1061.
Valentini, Laura (2012) "Ideal vs. Non-ideal Theory: A Conceptual Map," *Philosophy Compass*, 7/9: 654–664.
Vanderschraaf, Peter (2011) "Justice as Mutual Advantage and the Vulnerable," *Politics, Philosophy and Economics*, 10/2: 119–147.
Vandivier, Kermit (1996) "Why Should My Conscience Bother Me? Hiding Aircraft Brake Hazards," in M. David Ermann and Richard J. Lundman (eds), *Corporate and Governmental Deviance* (Oxford University Press, New York), 146–166.
Vogel, David (2005) *The Market for Virtue* (Brookings Institute, Washington, DC).
Von Mises, Ludwig (1922) *Socialism* (Liberty Fund, Indianapolis).

Waldo, Gordon P. and Simon Dinitz (1967) "Personality Attributes of the Criminal: An Analysis of Research Studies, 1950–65," *Journal of Research in Crime and Delinquency*, 4/2: 185–202.

Walras, Leon (1954) [1874] *Elements of Pure Economics*, trans. William Jaffé (Richard D. Irwin Inc., Homewood, IL).

Ward, Benjamin (1958) "The Firm in Illyria: Market Syndicalism," *American Economic Review*, 48/4: 566–589.

Webb, Beatrice Potter (1891) *The Co-operative Movement in Great Britain* (Swan Sonnenschein, London).

Webley, Simon and Elise More (2003) *Does Business Ethics Pay?* (Institute of Business Ethics, London).

Weinstock, Daniel M. (1999) "Neutralizing Perfection: Hurka on Liberal Neutrality," *Dialogue*, 38/1: 45–62.

Weisburd, David, Stanton Wheeler, Elin Waring and Nancy Bode (1991) *Crimes of the Middle Classes: White Collar Offenders in the Federal Courts* (Yale University Press, New Haven).

White, Thomas (1993) *Business Ethics: A Philosophical Reader* (Macmillan Publishing, New York).

Williams, Scott D. and Todd Dewett (2005) "Yes, You Can Teach Business Ethics: A Review and Research Agenda," *Journal of Leadership and Organizational Studies*, 12/2: 109–120.

Williamson, Oliver E. (1973) "Markets and Hierarchies: Some Elementary Considerations," *American Economic Review*, 63/2: 316–325.

Williamson, Oliver E. (1975) *Markets and Hierarchies: Analysis and Antitrust Implications* (Free Press, New York).

Williamson, Oliver E. (1981) "The Economics of Organization: The Transaction Cost Approach," *The American Journal of Sociology*, 87/3: 548–577.

Williamson, Oliver E. (1985) *The Economic Institutions of Capitalism* (Free Press, New York).

Wilson, Timothy D. (2002) *Strangers to Ourselves* (Harvard University Press, Cambridge, MA).

Winterhalder, Bruce (1990) "Open Field, Common Pot: Harvest Variability and Risk Avoidance in Agricultural and Foraging Societies," in Elizabeth Cashdan (ed), *Risk and Uncertainty in Tribal and Peasant Economies* (Westview Press, Boulder), 67–87.

World Bank (2000) *Curbing the Epidemic* (World Bank, Washington, DC).

INDEX