The Uses and Abuses of Agency Theory in Business Ethics

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.[1] 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 recent wave of corporate scandals has 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. For every employee at Enron who lost a job, shareholders lost at least US$4 million.[2] Furthermore, employees escaped with their human capital largely intact. Creditors and suppliers continue to pick over the bones of the corporation (which still exists, under Chapter 11 bankruptcy protection, and continues to liquidate assets in order to pay off its debts).[3] But as far as shareholders are concerned, their investments have simply evaporated, beyond any realistic hope of retrieval. (In fact, 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 sort 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

1 See Henry Hansmann, The Ownership of Enterprise (Cambridge, MA: Harvard University Press, 1992), p. 11.

2 In the year 2000, Enron had 19,000 employees, and a peak market capitalization of $80 billion. Also, presumably not every Enron employee had to search for work (hence the "at least").

3 It is estimated that creditors will get about 20 cents on the dollar. See Bethany McLean and Peter Elkind, *The Smartest Guys in the Room* (New York: Penguin, 2003), p. 410.

described a self-dealing transaction conducted by Richard Perle, a member of the firms' 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.[4] 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."[5] At the time of writing Black is under criminal indictment; it is not known whether he will be convicted for his role in several dubious transactions at Hollinger. Thus there is still an open question as to whether he broke the law. It seems evident, however, that he conducted himself unethically.

One of the central tasks of theoretical business ethics is to provide a conceptual framework that allows us to articulate more precisely the intuitive sense we all have that "nest-feathering" and similar forms of conduct are unethical, so that we can state more clearly 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*.[6] After all, the relationship between owners and managers is a textbook example of a principal-agent relationship[7] 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 Enron and elsewhere.[8] One might also expect business ethicists to insist that greater

---

4 "Perle Asserts Hollinger's Conrad Black Misled Him" *New York Times* (Sept. 6, 2004).
5 "Panel Says Conrad Black Ran a 'Corporate Kleptocracy'," *New York Times* (Aug. 31, 2004).
6 For a complete technical overview of this theory, see Jean-Jacques Laffont and David Martimort, *The Theory of Incentives: The Principal-Agent Model* (Princeton: Princeton University Press, 2002).
7 See Paul Milgrom and John Roberts, *Economics, Organization and Management* (New Jersey: Prentice Hall, 1992), p. 170; Donald E. Campbell, *Incentives* (Cambridge: Cambridge University Press, 1995), pp. 79-86.
8 Allen Buchanan has provided what is perhaps the most sophisticated development of this approach. See "Toward a Theory of the Ethics of Bureaucratic Organizations," *Business Ethics Quarterly*, 6 (1996): 419-440.

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.[9] 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. 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."[10] Along similar lines, Brian Kulik 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.[11]

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

9 See, e.g., Marjorie Kelly, "Why all the Fuss about Stockholders?" reprinted in her *The Divine Right of Capital* (San Francisco: Berrett-Koehler, 2001); Max Clarkson's introduction to *The Corporation and its Stakeholders* (Toronto: University of Toronto Press, 1998); Margaret M. Blair, "Corporate 'ownership': A Misleading Word Muddies the Corporate Governance Debate," *The Brookings Review* (Winter 1995): 16-19.

10 Rakesh Khurana, Nitin Nohria and Daniel Penrice, "Management as a Profession" in Jay W. Lorsch, Leslie Berlowizt, and Andy Zelleke, eds., *Restoring Trust in American Business* (Cambridge, MA: MIT Press, 2005). Some theorists, such as Dennis P. Quinn and Thomas M. Jones, "An Agent Morality View of Business Policy," *Academy of Management Review*, 20 (1995): 22-42, simply equates agency theory with the doctrine of shareholder primacy, which leads then to the suggestion that anyone committed to stakeholder theory must reject agency theory.

11 Brian W. Kulik, "Agency Theory, Reasoning and Culture at Enron: In Search of a Solution," *Journal of Business Ethics,* 59 (2005): 347–36.

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.

1. What is 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, viz. "situations in which one individual (the agent) acts on behalf of another (the principal) and is supposed to advance the principal's goals."[12] This is already a potential source of confusion, since the term "agent" is used differently here than in certain other contexts, such as corporate law, where the "law of agency" assigns a much narrower meaning to the term.[13] But these disputes over the use of the term "agent" are not where the real controversy arises. It is the use of game theory that makes agency theory 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

12 Milgrom and Roberts, *Economics, Organization and Management*, p. 170.

13 See Robert C. Clark, "Agency Costs Versus Fiduciary Duties," in John W. Pratt and Richard J. Zeckhauser, *Principals and Agents* (Boston: Harvard Business School Press, 1985), p. 56. In part because of this potential for confusion, some game theorists have begun to speak more generally about "the theory of incentives." See, e.g. Laffont and Martimont, *The Theory of Incentives*, also Campbell, *Incentives*.

rationalizing it through the 'discovery' and ascription of some underlying non-moral incentive.

Thus 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 the preferences of individuals, and preferences reflect whatever desires individuals happen to have, egoistic or altruistic.[14] David Gauthier made the point most succinctly, when he observed that, in 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."[15] 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.[16] 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 they affect that agent's goals, or ability to satisfy these goals.

On these grounds, many business ethicists have concluded that agency theory is perfectly anodyne. Allan Buchanan articulates this view well when he writes:

If, in applying principal/agent theory, it were necessary to assume that motivation is exclusively

14 For an example of a typical – and typically acrimonious – exchange between a business ethicist and a game theorist on this point, see Robert C. Solomon, "Game Theory as a Model for Business and Business Ethics," *Business Ethics Quarterly*, 9 (1999): 11-30, and Ken Binmore "Game Theory and Business Ethics," *Business Ethics Quarterly*, 9 (1999): 31-36.

15 David Gauthier, *Morals by Agreement* (Oxford: Clarendon, 1986), p. 7. See also Daniel M. Hausman and Michael S. McPherson, *Economic Analysis and Moral Philosophy* (Cambridge: Cambridge University Press, 1996), pp. 52-53.

16 J. Gregory Dees, "Principals, Agents and Ethics," in Norman E. Bowie and R. Edward Freeman, eds. *Ethics and Agency Theory* (New York: Oxford University Press, 1992), pp. 37-38.

> 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.[17]

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.[18] 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

17 Buchanan, "Toward a Theory of the Ethics of Bureaucratic Organizations," p. 421.
18 J. Gregory Dees, "Principals, Agents and Ethics," p. 29.

of social interactions.[19] 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 others' preferences 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 others' actions, and so are able to 'communicate' in this sense).[20] Furthermore, insofar as they *are* able to communicate with one another, standard game-theoretic solution concepts, like Nash equilibrium, do not apply.[21] This non-trivial restriction on game-theoretical 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 whether the light is best. For instance, in their widely-used management textbook on organizational theory, game theorists Paul Milgrom and John Roberts 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.[22] 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

19 Jack Knight, *Institutions and Social Conflict* (Cambridge: Cambridge University Press), p. 18.
20 See John Nash, "Non-cooperative Games," *Annals of Mathematics*, 54 (1951): 289-295.
21 Joseph Heath, *Communicative Action and Rational Choice* (Cambridge, MA: MIT Press, 2001), pp. 73-78.
22 Milgrom and Roberts, *Economics, Organization and Management*, pp. 179-192. In this respect, their discussion falls significantly below the level of sophistication exhibited in older classics such as James G. Marsh and Herbert Simon, *Organizations* (New York: John Wiley & Sons, 1958), pp. 65-81.

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 occlusion of motivational issues 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.[23] 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.[24]

Once again though, this emphasis on external incentives is not a necessary consequence of the commitment to the economic conception of rational action. There is nothing intrinsic to agency theory that prevents people 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 way to model information states, and so chose to 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 a 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

23 Milgrom and Roberts, *Economics, Organization and Management*, p. 350.
24 Compare this view to the discussion in Francis Fukuyama, *Trust* (London: Penguin, 1995), pp. 185-193, 255-266.

the chance to do so.[25] For example, it is routinely assumed 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.[26] 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 opportunistic behavior, 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 requires 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 agents 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

25 For definition and discussion of this term, see Oliver E. Williamson, *The Economic Institutions of Capitalism* (New York: The Free Press, 1985), pp. 47-49.
26 Milgrom and Roberts, *Economics, Organization and Management*, p. 128.

to make commitments is one of the defining postulates of non-cooperative game theory (and again, all of the standard solution concepts do not apply in cases where that assumption is relaxed).[27] 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.[28] 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...[29]

"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.

27 Heath, *Communicative Action and Rational Choice*, pp. 86-92.
28 Drew Fudenberg and Jean Tirole, *Game Theory* (Cambridge, MA: MIT Press, 1991), pp. 72-74
29 Eric Rasmusen, *Games and Information*, 3rd edn. (Oxford: Blackwell, 1989), p. 95.

Central to this conception is the consequentialism postulate, which states simply that the value of an action is a functions 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.[30] 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. 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,"[31] he is only half wrong. 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.[32]

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 moral skepticism. 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

30 Or what Robert Nozick refers to as "side constraints," *Anarchy, State and Utopia* (New York: Basic Books, 1974), pp. 28-32.
31 Eric Noreen, "The Economics of Ethics: A New Perspective on Agency Theory," *Accounting, Organizations and Society*, 13 (1988): 359-369.
32 Most of these examples are drawn from Milgrom and Roberts, *Economics, Organization and Management*, p. 170.

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 behavior, why should that be any cause for alarm among ethicists? And how could the mere teaching of such methodological tools in business schools be blamed for the recent spate of corporate scandals?

2. What's the problem with agency theory?

Here is the question, simply put: 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: more than one might think.

The first step to understanding this answer lies in an appreciation of the fact that, as a positive theory of the firm, 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.

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 groups ranging from 4 to 80 members, and with a variety of different monetary rewards.[33]

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 per cent, while offers below 20 per cent are rejected 40 to 60 per cent of the time. Experimental evidence from non-industrialized societies reflects greater variability – including examples of mean offer rates above 50 per cent, combined with frequent rejection of such offers. Yet in spite of these variations, no experiment has ever come close to

33 Robyn M. Dawes and Richard H. Thaler, "Anomalies: Cooperation," *The Journal of Economic Perspectives*, 2 (1988): 187-197. See also Friedrich Schneider and Werner W. Pommerehne, "Free Riding and Collective Action: An Experiment in Public Microeconomics," *Quarterly Journal of Economics*, 96 (1981): 689-704, Oliver Kim and Mark Walker, "The Free Rider Problem: Experimental Evidence," *Public Choice*, 43 (1984): 3-24, Mark Isaac, Kenneth F. McCue, and Charles R. Plott, "Public Goods Provision in an Experimental Environment," *Journal of Public Economics*, 26 (1885): 51-74.

conforming to the expectations of standard game theory.[34]

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 Oliver Williamson calls "information impactedness," simply would not function if the only tools that managers had at their disposal were external punishments and rewards.[35] Bengt Holström showed very early on how imperfect observability could make it impossible to devise efficient incentive schemes for individuals working in teams.[36] George Baker 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.[37] 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.[38] 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

34 Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath, "Cooperation, Reciprocity and Punishment in Fifteen Small-scale Societies," *American Economic Review*, 91 (2001): 73–78. The offers above 50 per cent are typical in societies with "gift" economies, or legacies thereof.

35 Oliver Williamson, "Markets and Hierarchies: Some Elementary Considerations, " *American Economic Review*, 63 (1973): 316-325, at 318.

36 Bengt Holström, "Moral Hazard in Teams," *Bell Journal of Economics*, 13, (1982): 324-340.

37 George P. Baker, "Incentive Contracts and Performance Incentives," *Journal of Political Economy*, 100 (1992): 598-614.

38 For an overview, see Robert Gibbons, "Incentives in Organizations," *Journal of Economic Perspectives*, 12 (1998): 115-132.

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:

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 sort of internal constraint at work, these theorists assume that the external incentives must be there, but that they simply have 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."[39]) The commonsense explanation would be to suppose that Ford tapped into an underlying norm of reciprocity.[40] 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 (or the old Soviet variant "we pretend to work, and they pretend to pay us").[41] So when the "boss" agrees to pay you a rate that it is, by common admission, far in excess of what he is obliged to pay, he has in essence done you a favor. And since

39 Henry Ford, *My Life and Work* (Garden City, NY: Doubleday, 1922), p. 147.
40 For example, see E. Fehr, S. Gachter and G. Kirchsteiger "Reciprocal fairness and Noncompensating Wage Differentials," *Journal of Institutional and Theoretical Economics*, 152 (1996): 608-640. George Akerloff, "Labour Contracts and Partial Gift Exchange," *Quarterly Journal of Economics*, 84 (1982): 488-500.
41 Hausman and McPherson, *Economic Analysis and Moral Philosophy,* pp. 55-56.

"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 commonsense appeal, there is significant empirical evidence to support this "norm of reciprocity" explanation of efficiency wages.[42] 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.[43] 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."[44] 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

42 For overview, see discussion in Uri Gneezy, "Do High Wages Lead to High Profits? An Experimental Study of Reciprocity Using Real Effort" (forthcoming).
43 Timothy C.G. Fraser and Robert G. Waschik, *Managerial Economics: A Game-Theoretic Perspective* (London: Routledge, 2002), p. 291.
44 John Boatright, *Ethics in Finance* (Oxford: Blackwell, 1999), p. 48. See also J. Gregory Dees, "Principals, Agents and Ethics," in Norman E. Bowie and R. Edward Freeman, Ethics and Agency Theory (Oxford: Oxford University Press, 1992), p. 35.

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."[45]

The other potential source of mischief is caused by 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 (or "self-interested") 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.[46] 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 there is a tendency among agency theorists to resist taking these obligations at face value, and to regard them instead as just legal "shorthand" for a certain set of implicit contracts, ones that are ultimately structured by external incentives.[47] 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 "self-interested"

45 Boatright, *Ethics in Finance*, p. 49. For further discussion, see papers collected in Bruno S. Frey and Margit Osterloh, eds, *Successful Management by Motivation - Balancing Intrinsic and Extrinsic Incentives* (Berlin: Springer Verlag, 2002).

46 For example, see Michael C. Jenson and William J. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3 (1976): 305. For critical discussion on this point, see Robert C. Clark, "Agency Costs versus Fiduciary Duties," p. 59-60.

47 Frank H. Easterbrook and Daniel R. 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 penalities for bank robbery rather than pat-down searches of everyone entering banks," *The Economic Structure of Corporate Law* (Cambridge, MA: Harvard University Press, 1991), p. 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.

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.”[48] 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, due to 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 be “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.[49]

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 a 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

48 Clark, “Agency Costs or Fiduciary Duties,” p. 65.
49 Lawrence Revsine, “The Selective Financial Misrepresentation Hypothesis,” *Accounting Horizons*, 5 (1991): 16-27 at 18.

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 neutralization" used by white-collar criminals to rationalize – and hence to grant themselves permission to engage in – illegal conduct.[50] 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, and so the potential for mischief that it creates merits emphasis.

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 can't 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 incentives schemes can exacerbate agency problems. Yet there is a more general problem that has been almost entirely ignored, namely, that even

50 James William Coleman, "Toward an Integrated Theory of White-Collar Crime," *American Journal of Sociology*, 93 (1987): 406-439, at 414.

a well-designed system of external incentives has the potential to undermine moral motivation, and thus to create agent costs where previously none existed.[51] Bruno Frey and Felix Oberholzer-Gee refer to this phenomenon as the "crowding out" of moral incentives. Their research highlights some of the ways in which pecuniary incentives can have the effect of undermining moral incentives. 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.[52] 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 per cent) 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 offer by adding fairly large sums of money as compensation (US$2,175-$6,525), support for the project plummeted to 24.6 per cent.

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 some 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

---

51 See Ernst Fehr and Simon Gächter. 2002. "Do Incentive Contracts Crowd Out Voluntary Cooperation?" *University of Zurich Working Paper 34*. http://www.iew.unizh.ch/wp/iewwp034.pdf

52 Bruno S. Frey and Felix Oberholzer-Gee, "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review*, 87 (1997): 746-755.

the relevant cost-benefit calculations. They approach the question from the standpoint of what John Rawls calls "the reasonable," rather than "the rational."[53] Furthermore, their consent may be *based* upon the fact that they do not enter into these calculations. (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 negligible.) 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. If the external incentives are inadequate, 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.

What this shows 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. 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 typically respond 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 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

53 John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), pp. 48-54.

unilateral and one-sided relationship. Thus the “framing effect” of agency theory tends to encourage essentially Taylorian management practices. There is nothing in the agency perspective, for instance, that discourages the principal from acting opportunistically with respect to the agent, or even speaks to this problem.[54]

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. Thus moral incentives usually develop within relations that are mutual and two-sided. These can be extraordinary 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. Thus the type of incentives schemes that tend to flow from an agency analysis 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 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.)

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 is that the basic approach has as its foundation a *normative* theory of practical rationality, one which 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

54 Dees, “Principals, Agents, and Ethics,” p. 49.

constraints. 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."[55] 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 anti-social behavior.[56]

There is some evidence to support this concern about instrumental rationality becoming a self-fulfilling prophecy. It was widely reported, for instance, that one of the only significant anomalies discovered in experimental trials of the "public good" game in North America occurred when the game was played among economics graduate students. There the rate of cooperation fell to only 20 per cent, whereas it remained over 40 per cent when played by students in other disciplines.[57] 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

55 Duska, "Why be a Loyal Agent? A Systemic Ethical Analysis," p. 149. Substitute "fickle" for selfish and the argument is quibble-proof.

56 See Gresham M. Sykes and David Matza, "Techniques of Neutralization: A Theory of Delinquency," *American Sociological Review*, 22 (1957): 664-670 at 668.

57 Gerald Marwell and Ruth E. Ames, "Economists Free Ride, Does Anyone Else?" *Journal of Public Economics*, 15, (1981): 295-310. Results of the study need to be taken with a grain of salt, for a variety of reasons. See Robert H. Frank, Thomas Gilovich and Dennis T. Regan, "Does Studying Economics Inhibit Cooperation?" *Journal of Economic Perspectives,* 7 (1993): 159-171.

'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."[58]

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.[59] 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.

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, that uses an instrumental model of

58 Marwell and Ames, "Economists Free Ride, Does Anyone Else?" p. 309.
59 John Doris, *Lack of Character* (Cambridge: Cambridge University Press, 2002).

rationality as the basis for the development of a normative theory.[60] 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 the 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.[61] 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."[62]

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 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.[63] As Noreen puts it, "agency theory can be used to provide a

60 See Joseph Heath, "Rational Choice as Critical Theory," *Philosophy and Social Criticism,* 22 (1996): 43-62.
61 See David Braybrooke, "The Insoluble Problem of the Social Contract," *Dialogue*, 15 (1976): 3-37.
62 David Gauthier, *Morals by Agreement* (Oxford: Clarendon, 1986).
63 Noreen, "The Economics of Ethics: A New Perspective on Agency Theory." For an example of the "invisible hand" critique, see David Gauthier, "No Need for Morality: The Case of the Perfectly Competitive Market," *Philosophical Exchange*, 3 (1982): 41-54.

series of instructive parables that illustrate the adverse consequences on social and economic systems of unconstrained opportunistic behavior,"[64] 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 (viz. 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.[65] The fact that family members are able to draw upon preexisting trust relations allows them 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"[66] – 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 critical agency perspective is also able to explain quite clearly why, as production becomes more knowledge-intensive, there is a need to shift away from more Taylorian management strategies, in favor of a focus upon organizational culture, team-building and "shared values." It is because agency problems are caused, fundamentally, by information asymmetries. As production becomes more

64 Noreen, "The Economics of Ethics: A New Perspective on Agency Theory," p. 360.
65 Francis Fukuyama, *Trust*, (London: Penguin, 1995), pp. 74-80.
66 Fukuyama, *Trust*, p. 10. See also James S. Coleman, "Social Capital in the Creation of Human Capital," *American Journal of Sociology*, 94 (1988): S95-S120.

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."[67])

Allen Buchanan has taken this style of analysis 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."[68] Agency theory tells us where the major stress lines are within these organizations, where cracks are most likely to appear. The 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 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 imposes 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

67 Peter Drucker, "Management as Social Function and Liberal Art," in *The Essential Drucker* (New York: HarperBusiness, 2001): 3-13.
68 Buchanan, "Toward a Theory of the Ethics of Bureaucratic Organizations," p. 422.

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, due procedure and preservation of the "chain of command."[69]

The agency perspective is similarly useful when it comes to analyzing the obligations of senior managers. While agency theory itself is neutral with respect to the doctrine of shareholder primacy – indeed, the vocabulary is sufficiently broad that the manager can be characterized as the agent of any of the firm's patrons – it is able to explain why, in a standard business corporation, managers bear special fiduciary obligations towards shareholders.[70] 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 risks are simply much greater in the relationship between management and the firm's owners. 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). Agency analysis is also able to demonstrate quite clearly the ineffectiveness of equity-based compensation arrangements, such as stock options, as a way of creating an external alignment of managerial and shareholder interests. Furthermore, 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.[71] This in turn helps to make the case for the claim that moral restraint on the

69 Buchanan, "Toward a Theory of the Ethics of Bureaucratic Organizations," p. 431.
70 Alexei Marcoux, "A Fiduciary Argument Against Stakeholder Theory," *Business Ethics Quarterly,13* (2003): 1-24.
71 See Gary Miller, *Managerial Dilemmas* (Cambridge: Cambridge University Press, 1992), pp. 171-176. Also Lucian Arye Bebchuck, Jesse M. Fried and David I. Walker, "Managerial Power and Rent Extraction in the Design of Executive

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.[72] 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, techs, 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, but also because of the psychological effects of ill health). 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).[73] In the same way, would be unsurprising to discover that the major "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. 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, have this structure – they all specify what subordinates owe to their superiors. In part for this reason, Buchanan is it pains to

Compensation," *University of Chicago Law Review*, 69 (2002): 751-846.

72 See Lucian Arye Bebchuk and Jesse M. Fried, "Excecutive Compensation as an Agency Problem," *Journal of Economic Perspectives*, 17 (2003): 71-92.

73 Ref. Bulletin of College of Physicians and Surgeons of Canada.

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, agency theory may serve a useful purpose in telling us where ethics is most needed within organizations, but morality has its own logic, and so we many not be able to develop an ethics code that is custom-tailored to resolve precisely this 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 domain, 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 an ethics code 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 some sort of magic bullet for resolving agency problems). The fact that a particular institutional arrangement generates an agency problem *in theory* is not in itself 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 calls "the strains of commitment."[74] A lot of problems would go away if people only behaved more ethically, but the fact is, people often don't behave all that ethically. Thus merely urging more ethical behavior

74 John Rawls, *A Theory of Justice*, 2nd edn. rev (Cambridge, MA: Harvard University Press, 1999), pp. 154-5.

upon them, beyond a certain point, no longer counts as offering a solution.

4. Agents for the greedy?

One might think that this sort of critical use of agency theory would be uncontroversial among business ethicists. Yet that would also be mistaken. There is another prominent line of objection to agency theory, this time one that condemns both positive and normative uses of the theory. It is centered 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, 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.

This view does have some *prima facie* plausibility. It is a well-known feature of conventional morality that promising to help a friend commit a crime does not generate a moral obligation on one's part to commit that crime (such that if one's friend were to fall ill on the eve of the bank robbery, one might be morally obliged to carry it out anyway on his behalf). 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.[75] 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. He remained "above the fray" throughout, by 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."[76] 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."[77] 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. 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

75 See Arthur Isak Applbaum, *Ethics for Adversaries* (Princeton: Princeton University Press, 1999), p. 77.
76 Applbaum, *Ethics for Adversaries*, p. 7.
77 Applbaum, *Ethics for Adversaries*, p. 8.

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."[78] 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."[79] 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."[80] Thus she refers to the moral framework encouraged by agency analysis as "theory-compelled irresponsibility."

There are two separate issues in this "agents for the greedy" critique, which need to be disentangled:

1. The nemo dat problem. In discussing principal-agent relations, Kenneth Goodpaster refers to the Latin proverb, *nemo dat quod non habet*, or "nobody gives what he doesn't have." He 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.[81] 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."[82] Yet since the agency relationship cannot be a source of moral permissions, it is then claimed, the fact that managers act, in some sense, as agents of shareholders, is simply lacking in moral significance.

78 Alex C. Michalos, *A Pragmatic Approach to Business Ethics* (Thousand Oaks: Sage, 1995), p. 45 [format altered].
79 Lisa Newton, "Agents for the Truly Greedy," in Bowie and Freeman, eds. *Ethics and Agency Theory*, p. 100.
80 Newton, "Agents for the Truly Greedy," p. 101.
81 Kenneth E. Goodpaster, "Business Ethics and Stakeholder Analysis," *Business Ethics Quarterly*, 1 (1991): 53-73, at 68 [slightly modified, text of the article is corrupt].
82 Richard T. DeGeorge, "Agency Theory and the Ethics of Agency," in Bowie and Freeman, eds. *Ethics and Agency Theory*, pp. 65-66.

Yet while these observations are correct, the conclusion does not follow. There is in fact some danger of equivocation in the talk about "ethical" and "unethical" conduct. With respect to agents, it is important to distinguish the deontic modalities of *permission* and *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 obligations of managers are to advance the *legitimate* interests of shareholders (not any-old-interests).[83] Even Friedman qualifies his defense of profit-maximization (in an unfortunately glib manner), 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."[84] 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."[85] 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

83 Buchanan, ""Toward a Theory of the Ethics of Bureaucratic Organizations," pp. 422-23.
84 Milton Friedman, "The Social Responsibility of Business is the Increase its Profits," *New York Times Magazine*, (Sept. 13, 1970).
85 Applbaum, *Ethics for Adversaries,* p. 259.

this permission generates an obligation to mount that defense.

The relationship between the principal and the agent can be summed up in the following set of inferences. Consider the case where the principal instructs the agent to perform some action, *a*, that falls within the scope of a principal-agent relationship:

1. If *a* is obligatory for the principal then *a* is obligatory for the agent.
2. If *a* is permissible for the principal, then *a* is obligatory for the agent.
3. If *a* is forbidden for the principal, then *a* is forbidden for the agent.

Thus, from the standpoint of business ethics, insofar as shareholders are permitted to claim the residual earnings of the firm, and are permitted to seek maximization of those earnings, then managers are *obliged* to serve them loyally in this regard. This is morally salient, because the agency relationship creates that moral obligation, by transforming a permission into an obligation. Thus the *nemo dat* principle is misleading, and Goodpaster's observation does not undermine the significance of agency analysis. Agency relationships are not merely a distraction; they represent a genuine source of moral obligation.

2. Greed is not good. The second major issue concerns the moral status of shareholder interests, and the intuition that there is something morally dubious about the desire of shareholders to "make as much money as possible" (as Friedman so gracefully put it). Of course, 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 it 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." Certain actions may not be morally praiseworthy, but they are not, by virtue of that fact, morally impermissible.

Nevertheless, there is a persistent intuition among ethicists (and the broader public), that making money can't be the foundation of business ethics, regardless of whether that money is made for oneself or for someone else. Making money just doesn't seem ethical enough.[86] Naturally, there are those who disagree. Winston Churchill put it rather pithily, when he declared that "it is a Socialist idea that making profits is a vice: I consider the real vice is making losses." Kenneth Arrow, in his reflections on business ethics, adopted a similar point of view. "Profit," he said, "really represents the net contribution that the firm makes to the social good, and profits should therefore be made as large as possible."[87] Yet this view has never succeeded in carrying general conviction. (One need only consider the demonization of the profit motive that one finds in recent films, like *The Corporation*.)

There is a large constituency of business ethicists who are similarly skeptical about the profit motive. Duska, for instance, suggests that "profit" is merely a way of motivating managers, but that the real "purpose" of the business enterprise lies elsewhere. The actual goal of business, he claims, is "to help increase the production of goods and services." He then suggests that "it is a mistake to confuse incentives for good actions with the purpose of those actions. Good grades are incentives for study, but the purpose of study is primarily to learn, to expand the mind."[88] There are circumstances in which these two objectives can come apart. Students who engage in too single-minded a pursuit of grades often fail to expand their minds. Similarly, in a business context, there is potential for divergence between "first, the social purpose of providing goods and services; and second, the agent's altruistic commitment to the principal."[89] In this case, it is the "true purpose" of the business that should serve as the object of the manager's true loyalty, rather than the demands of the principal for profit. Thus any

86 Goodpaster formulates this intuition quite well in his discussion of the so-called "stakeholder paradox." See Goodpaster, "Business Ethics and Stakeholder Analysis," p. 63.
87Kenneth Arrow, "Social Responsibility and Economic Efficiency," *Public Policy*, 21 (1973): 303-317 at 305.
88 Duska, "Why be a Loyal Agent?" p. 161.
89 Duska, "Why be a Loyal Agent?" p. 161.

obligations arising out the agency relationship are easily trumped, in Duska's view, by considerations that pertain to the true purpose of the firm.

There is something right about this, but also something importantly wrong. What critics of the profit motive typically fail to take into consideration is the fact that the market economy is an adversarial institution.[90] Adversarial institutions achieve the results that they do by "playing off" various parties against one another, producing outcomes that correspond only accidentally to those intended or desired by any of the participants. Consider the adversarial structure of a criminal trial. The overall goal is to ensure that "justice is served" (and that conviction of the innocent is minimized). But this outcome is achieved, not by instructing all of the major parties to pursue the outcome that they consider to be the most just. On the contrary, the crown (or district) attorney is expected to exhibit "prosecutorial zeal" in seeking a conviction, regardless of what she may regard as the merits of the case. Similarly, the defense attorney is expected to mount a vigorous defense, independent of any opinions he may have about the defendant's guilt or innocence. The thought is then that these two forces cancel one another out, leaving the judge free (as much as possible) to assess the case on its true merits. Thus the prosecution and the defense, although they participate in a system that has as its goal the production of "justice," are not obliged to adopt that outcome as their objective. Indeed, they are often barred from doing so. Imagine a defense attorney with access to privileged information that confirmed the guilt of her client. It would be the height of professional misconduct for her to leak this information to the prosecution, or to sabotage her own client's case, simply because she wanted to ensure that justice was served.

One can see a similar structure in the competitive market. Profit itself is not the goal of the institution. On the contrary, the point of having a competitive market is to play various profit-seeking corporations off against one another (often on both the supply and the demand side) in order to achieve

90 [self-identifying note removed]

pressure in the direction of market-clearing prices. This in turn reduces deadweight losses, and thus ensures that all resources are put to their best employment. Under ideal conditions (i.e. in a perfectly competitive market), the strategies adopted by the parties completely cancel one another out – hence the absence of any profit in general equilibrium. (This is also why, in sectors where it is impossible to organize a competitive market, firms are usually subject to regulatory constraints that prevent them from adopting profit-maximizing strategies.) Thus John Kay gets things right when he claims that "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."[91] But one cannot infer from this, as Duska does, that the production of goods and services is therefore that "true" purpose of the corporation, and that the manager should assign priority to this objective whenever it conflicts with the interests of the principal. This would be like claiming that the "true" purpose of the defense attorney is to see that justice is served. Both managers and attorneys should respect certain moral constraints in choosing the strategies that they will employ. But there is an enormous difference between respecting certain moral constraints while still advancing the objectives of the principal, and substituting some other set of objectives for the principal's, on the grounds that one's own moral reasoning leads one to assign them greater priority.[92]

It is the adversarial nature of markets that explains the whiff of immorality that has always accompanied the profit motive. Those who demonize corporations for pursuing profits have much in common with those who denounce criminal defense lawyers for "defending rapists and murderers." In the case of doctors and patients, it's usually obvious that the health of the patient is a good thing, and therefore that the doctor is doing a good thing in helping the patient to secure it. In the case of managers and shareholders, on the other hand, it is not obvious that profit is a good thing, just as with defense attorneys and those who have been accused of a crime, it is not obvious that acquittal is a good thing.

91 John Kay, *The Truth About Markets* (London: Penguin, 2003), p. 351.
92 For further analysis, see [self-identifying note].

In order to see the full justification for the objectives that the agent is pursuing, it is necessary to situate them within the broader context of the institution within which they are pursued, and in particular, to take into consideration the countervailing forces that are set up within that same institution, which serve as a check upon the ability of any one party to achieve its objectives.

Thus the normative critique of the agency perspective is based upon a set of conceptual confusions. Of course, there is something genuinely counterintuitive about that 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 while the obligations that managers have toward the owners of a firm will most certainly not add up to *all* of business ethics, they will certainly form the *core* of any defensible conception of business ethics. When seen in this light, the big ethical question becomes, not *whether* managers should advance the interests of the owners of the firm, but rather *how far* managers should go in advancing the interests of the firm's owners. (In the same way, the major question in legal ethics is not whether the lawyer should serve the client, but rather how far the lawyer should go in advancing the interests of the client.) In the case of markets, this question admits of a very straightforward answer. Economists use the term "market failure" to refer to situations in which the competitive pursuit of profit by individual firms fails to achieve the purpose of the market economy, viz. the efficient production of goods and services. This provides a set of relatively straightforward guidelines for distinguishing permissible from impermissible profit-maximizing strategies. Taking advantage of a market failure in order to increase profits (e.g. through the creation of negative externalities, exploitation of information asymmetries, exercise of market power, etc.) are impermissible.[93] Anything else is fair game.

93 This is the core of Arrow's view ("Social Responsibility and Economic Efficiency,"), although he focuses only on the twin cases of pollution and information asymmetry.

5. Conclusion

The preceding discussion has examined two very different strategies for employing agency theory – the positive and the normative – 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 status of profit. With regard to the first, I have suggested that business ethicists have been fully 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.

With regard to the doctrine of shareholder primacy, and the extent to which agency theory encourages this perspective, I have tried to emphasize that there is no simply connection between the two. In this context, it is perhaps worth noting that R. Edward Freeman used principal-agent vocabulary as a way of articulating his commitment to a stakeholder approach to business ethics. Indeed, he even presented an "Agency Principle" as an element of his "Doctrine of Fair Contracts," intended to guide the management of stakeholder-focused firms.[94] Thus agency theory is neutral in this regard. Nevertheless, when employed cautiously, with due attention to the institutional context in which the

94 R. Edward Freeman, "A Stakeholder Theory of the Modern Corporation," in Clarkson, ed. *The Corporation and its Stakeholders*, p. 136.

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 shareholder is 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.