Here it is, almost three years after the fall of Enron, and the trials of some of the top executives involved in the corporate scandals are still proceeding or yet to begin. All too often the same pattern emerges. Significant senior executives, like a company's chief financial officer (CFO) and underlings, plead guilty to the misdeeds that brought down a firm and then act as government witnesses against the chief executive officer (CEO). The CEO resists vigorously, with the best defense counsel that money can buy, alleging that the scandal was solely the work of the CFO and his unscrupulous cohorts. In each scandal, the debate in the courtroom and the business press is about who is the individual ultimately responsible for the scandal. Generally, if a group is singled out in the discussion at all, it is the corporation's board of directors. But the board is not regarded as central to the scandal, except in a
kind of negative manner, as being too inactive and indifferent in its oversight to have detected the misbehavior.

Examples abound. The drama of the responsibility for WorldCom’s demise has just played out in court in a trial pitting its former CEO Bernie Ebbers (who was found guilty on numerous charges) against its former CFO Scott Sullivan (who has already pleaded guilty to securities fraud) as the prime mover in the fraud.¹ In another courtroom, former HealthSouth CEO Richard Scrushy is defending himself against fraud charges, with the chief opposing witnesses being the former CFOs (who all have pleaded guilty to securities fraud).² Kenneth Lay, former Chairman and CEO, and Jeffrey Skilling, former CEO, of Enron, have yet to come to trial.³ Arrayed against them will be the notorious Andrew Fastow, former CFO, and a host of lesser executives, who have all entered into plea agreements with the government. For the most part, board members of these companies (other than inside board members) have not been criminally charged, but are the subject of civil suits on account of their inattention that allowed the scandals to go on for so long.⁴

It is no surprise that the criminal trials are proceeding in this way, given that our criminal law and justice system are designed to determine an individual’s guilt. Nor is it a surprise that the business media, which is in many cases really only a step up from the tabloids, strives to gain and maintain readers by emphasizing the personal stories behind the corporate scandals. Both of these reactions to the scandals reveal a fundamental human tendency to attribute complex misdeeds to individuals.⁵ This is the “bad apples” understanding of the corporate scandals or problems. This attribution error applies to complex positive outcomes as well.⁶ For example, the

⁴ See, e.g., Settlement Motion, In re Enron Corp., et al. (No. 01-16034 (AJG) (Bankr. S.D.N.Y. Jan. 26, 2005)) (discussing settlement with outside directors of Enron).
⁵ This erroneous reasoning seems to be based on a general human tendency to simplify causation. See generally ROGER BROWN, SOCIAL PSYCHOLOGY 169-94 (2d ed. 1986) (discussing this human cognitive mistake).
⁶ See generally NASSIM N. TALEB, Fooled By Randomness (2001)
business media is only too ready to lionize corporate executives for the achievement of their firms: one has only to go back to the late 1990s to find article after article extolling Kenneth Lay and Bernie Ebbers. This is not to say that individuals matter for nothing in scandals or in success. But to focus on individuals blinds us to the complex causes of misbehavior (to consider the focus of this conference) and keeps us from making reforms that could prevent the recurrence of this misbehavior.

The principal purpose of the conference, “Corporate Misbehavior by Elite Decision-Makers: Perspectives from Law and Social Psychology,” is to offer an alternative to the understanding that corporate misconduct and scandals are due to the work of a few “bad apples” among corporate executives and directors. From information available on the scandals, it appeared that inner circles of top executives, corporate advisors (accountants, bankers and lawyers) and board members formed coherent social structures that engaged in illegal or unethical behavior that destroyed firm value. A possible determinant of the scandals, in other words, was group, not individual, behavior. If this was the case, a number of interesting questions are posed and research avenues opened that could have important consequences for legal policy-making. What is perverse group behavior and what distinguishes it from positive group behavior? How did this group misbehavior arise in so many publicly-traded firms, which suggests that corporate governance structure contributed to it? How could individual members of these circles or groups engage in behavior that, on some levels, they knew was improper, but nevertheless accepted from a group perspective?

To answer these questions, it is necessary to turn to the social sciences, for one of the goals of social psychology and organizational theory is to understand and explain group behavior, including deviant group behavior. My co-organizer, Professor Larry Solan, and I thought that it would be useful to ask what researchers in these human sciences could tell legal scholars about the social psychological and other organizational causes of the corporate elite’s misbehavior. We decided that the

(discussing tendency of individuals to take credit for random positive outcomes).

I develop this argument in greater detail in my article, Whistleblowing and the Public Director: Countering Corporate Inner Circles, 83 OREGON L. REV. 435 (2004).

Director of Brooklyn Law School’s Center for the Study of Law, Language and Cognition, which sponsored the conference together with the Alfred P. Sloan Foundation.
best way to achieve this purpose was to bring together at a conference prominent social psychologists and organizational and management specialists to present and to discuss their theories and research about group misbehavior that could help explain the corporate scandals. We planned to have corporate law scholars, as discussants, comment upon the implications of these findings and research for policy making on the legal regulation of corporate governance and decision-making.

One goal of the conference was pragmatic: to find solutions to the misconduct of elite corporate decision makers from a social science perspective that legal policy makers, fixated on individuals as primary causes of the scandals, had overlooked. Yet the dialogue of conference participants was not addressed directly to policy makers as such, and it would have that pragmatic effect indirectly by contributing to a new direction in corporate legal scholarship. A complementary goal was then to stimulate and promote interactions and research between social psychologists and organizational theorists, on the one hand, and corporate scholars, on the other.

These interactions and possible joint research would help correct a tendency in the dominant direction of current corporate law scholarship to ignore group causes of corporate governance problems, which tendency has led to reforms that can only incompletely prevent corporate misbehavior. So much of this scholarship is grounded in the law and economics tradition that bases its policy prescriptions upon rational, self-interested economic actors. When faced with the corporate scandals, scholars in this tradition react in ways not unlike business reporters, members of Congress or federal prosecutors: they focus on the individual. For example, if a CEO, like Dennis Kozlowski, took too many benefits from his former company, Tyco International, it was because appropriately designed incentives had not been in place for all the corporate actors involved: his compensation package was not correctly keyed to his performance, and the members of the Tyco’s board did not have adequate personal incentives to check his natural self-interested behavior. From this

---

9 There have been notable exceptions. We were fortunate to have at our conference as discussants, Donald Langevoort and Lynne Dallas, two legal scholars who have for some time used social science methods and research in their analysis of corporate and securities law issues and problems.
perspective, corporate scandals can be addressed only by reforms aimed at these individuals.10

Our focus in this conference, as would be our focus in the resulting research, is different. For us, individual behavior within groups can be understood as determined by group dynamics. This means more than that an individual is conscious of and affected by others, although this is certainly a part of the analysis. It is, rather, an account of how individuals as group members assume a social or group identity, which in turn influences their behavior when they are in the group. The influence comes when the group identity defines, among other things, the roles and appropriate behavior for individuals within the group. The study of group dynamics examines the formation of this group identity and the way in which it shapes thought, perception and behavior. The study should explain how groups with perverse purposes can form and how individuals participate in them and come to find their actions within the group entirely natural and proper, even though when viewed from a perspective outside the group, the individuals acknowledge them to be completely improper.

To encourage discussion as to the group nature of the misbehavior in the corporate scandals, we asked conference participants to consider the following related questions in making their presentations and preparing their articles:

1. Do corporate scandals reveal problems of group misbehavior among corporate elite decision-makers and advisors? Can social psychology and organizational theory help us identify the nature and causes of the problems?

2. Are there differences between the ways that people perceive themselves as individuals on the one hand and as members of groups on the other that can explain, at least in part, corporate misconduct?

3. Can corporate reformers do anything about these problems? Can social psychology and related human sciences offer reformers any guidance? In this connection, significant reforms aimed at elite corporate decision-makers and advisors have already occurred. From a social psychological perspective, how well-designed are they to address the group problems and how effective are they likely to be?

We think that each of the articles in this volume answers one or several of the above questions and together the articles represent the work that we wanted to foster and to share with a wide scholarly audience.

In his article, *The Cognitive and Social Psychology of Contagious Organizational Corruption*, Professor of Psychology John Darley appropriately begins the volume with a focus on the first group of questions by taking issue with the “bad apples” theory of corporate corruption, which, in his view, allows people to ignore the complex causes of corruption in corporations and other organizations. These causes include the phenomenon that Professor Darley has identified in his previous work: that corruption begins gradually, often with a small act that is morally ambiguous or barely improper, and proceeds by small steps until the corruption is monumental. This gradual descent into the immoral helps explain why so many ordinary individuals can participate in corruption. In a related point, Professor Darley observes that the acts leading to corporate corruption are the product not of moral deliberation but of quick, intuitive judgments, and these judgments are inherently self-interested (either for the individual, the group or both). He then discusses how these corrupt, self-interested acts spread or are imitated in an organization. This propagation occurs because those who are disturbed by the acts feel pressured to remain silent since the organization sends the message that the acts are acceptable. Even more significantly for the purposes of the conference, Professor Darley points out that group loyalty prevents employees from complaining about the progressively corrupt practices. The foundations of this loyalty, as discussed in more detail by Professor Hogg in his article, lie in an individual's self-identification with the group, which transforms him or her into a group member with the group's values.

The article, *Out of Touch: The CEO’s Role in Corporate Misbehavior*, by Professor of Organizational Behavior Linda Treviño, appropriately follows Professor Darley’s, because she argues that a CEO of a firm is critical for the development of its ethical culture, which can help prevent scandals and abuses. For her, the importance of the CEO appears first in the design and implementation of a firm ethics program: is it window dressing, or is the program formal, value-oriented and integrated into the life of the firm? These latter attributes make the ethics program effective, i.e., make employees more likely to behave ethically and report ethical violations because
they feel that ethics matters to the firm and the firm will support them and will punish ethical violators. She observes that the CEO must also foster an ethical culture informally, for the CEO’s everyday behavior (which employees observe) reinforces the firm’s formal systems. The CEO does this by providing a visible example of a moral individual (one who personally cares about ethics) and a moral manager (one who shows in his or her management that ethics matters). Professor Treviño further argues that leaders must make their ethical decisions visible, provided that they are otherwise ethical persons and not hypocrites, because, unfortunately, there is a large social distance between the CEO and most employees. Professor Treviño then uses this distance also to explain why many CEOs fail to understand the ethical problems in their organization. Situated at the top of the firm, insulated from much of its activities and yet personally identified with it, a CEO cannot see many of its ethical problems and tends to adopt a rosy view of the organization. She also points out that because CEOs socialize almost exclusively with others of similar status they are unaware of issues and problems at the ordinary employee level.

Then, in his comment on Professor Darley’s and Professor Treviño’s articles, entitled Discussing Corporate Misbehavior, Professor of Law Daniel Greenwood points to the disturbing outcome in U.S. corporate law: the narrowing of the purposes of the large complex organization, that is the public corporation, to a maximization of shareholder wealth. He insightfully observes that, while enormous CEO pay has been justified by the need to motivate self-interested individuals to perform well, CEOs state that they have little connection to and awareness of the enterprise whenever they are accused of being responsible for a scandal. The outsized CEO compensation, he contends, thus has no justification in light of the CEO’s great distance from other employees, especially since, as Professor Treviño shows, the CEO’s example and connection with the firm are critical to the development of its ethical culture. Commenting on Professor Darley’s article, Professor Greenwood emphasizes how it underscores the weakness of a corporate law that ignores the organizational complexity of firms. For Professor Greenwood, employees’ pursuit of organizational goals and purposes, which, according to Darley, lead them into scandals, points to a contradiction lying at the heart of firms: the good for which individuals are asked to sacrifice their time and even their liberty is only the
self-interest of the firm, which is an impoverished parody of morality and of the value of the common good and which, paradoxically, ends up driving the firm out of business. Professor Greenwood concludes his comment by setting forth broad ideas for reforms that might reintroduce values into corporate law debates.

Professor of Social Psychology Michael Hogg squarely addresses, in his article, *Social Identity and Misuse of Power: The Dark Side of Leadership*, the second conference question (Are there differences in social and individual self-conceptions?) by using social identity theory (of which he is a major exponent) to explain group features that could have led to the corporate scandals. Professor Hogg initially sets forth the basics of social identity theory, which explains how individuals categorize others and themselves in terms of their identity in significant groups (such as work groups), i.e., they “depersonalize” themselves and others in their group membership. He then describes how leadership makes sense in this theory: a leader is a prototypical member of a group, yet also with the ability to experiment in his or her behavior so as to lead the group in new directions (a member with “charisma”). But this account points to potential group problems. As Professor Hogg explains, if group norms defining group and individual member behavior do not include ethics, then group members, especially leaders, are likely to act unethically as part of the group. Moreover, given the deference of group members, the leader can move the group gradually towards unethical behavior. Indeed, as he explains, the leader can even use his or her status to begin to isolate himself or herself and a small coterie from the rest of the group (as demonstrated by CEOs in many scandal-ridden firms). Professor Hogg then introduces another important aspect of social identity theory, a group’s reduction of uncertainty, to explain unethical corporate behavior. He outlines a situation that can lead to scandals: individuals in corporations form groups with strong social identity and with powerful leaders because of the uncertainty of competitive corporate life. The leader and his or her minions may, in turn, become isolated in the organization and, if unchecked, they may lead it into corruption and disaster.

In their co-written piece, Professor of Organizational Behavior Rakesh Khurana and graduate student Katharina Pick pose the individual vs. the group problem in contending that a board of directors cannot be understood only as a
collection of individuals. They first survey the understanding of the board under the dominant agency theory as a collection of individuals contracting with the firm and argue that the agency perspective does not capture the experience of board members or the facts about the corporate scandals. (But they warn that agency theory may create boardroom reality by encouraging board members to think of themselves as only self-interested individuals.) In an approach similar to those of Professor Darley and Hogg, they then discuss the social attributes of boards that make them similar to other groups and that may have led to the scandals. Boards act cohesively pursuant to well understood group norms, but, as Khurana and Pick explain, these norms can prevent directors from questioning critically CEOs and other senior executives, which would have revealed problems in firms. They also point out that board norms are long lasting and difficult to change, and that these norms reinforce board members’ conformity to the group’s perspective and generally discourage dissent of any kind. The passivity is understandable in situations of uncertainty, which is usually that of a board, where experts, such as senior executives, offer authoritative views and where non-expert board members are discouraged from ever becoming active in raising issues. Khurana and Pick make the interesting additional observation that board norms of one firm are often similar to the norms of other firms, since board members constitute a small, closed population in U.S. society. Moreover, the authors observe that, like all groups, boards have developed routines for dealing with their tasks, which routines, occasionally, enable them to overlook critical issues and information (such as the Enron board’s routine approval of conflict of interest transactions). Khurana and Pick conclude their article, however, with an optimistic observation about the social nature of boards: increasingly norms of professionalism for board members are being developed, which may improve overall board performance and board norms.

In a response to the third set of questions about the contribution of the social sciences to corporate reform, Professor of Psychology Tom Tyler poses the question how businesses with well-meaning executives can achieve employee compliance with laws and ethical norms. At the beginning of his article, Promoting Employee Policy Adherence and Rule Following in Work Settings: The Value of Self-regulatory Approaches, he contrasts two methods of firm ethical governance: the “command and control” model, which relies on
external sanctions and rewards, with the “self-regulatory” model, which relies on internal employee motivations. Professor Tyler points out that businesses have traditionally used the “command and control” model, which ensures compliance with policies through incentives and sanctions but which is costly and not particularly effective. The self-regulatory model, on the other hand, ensures compliance by activating in work settings employees’ own ethical values, which in turn legitimates the organization’s values and rules. Professor Tyler discusses the empirical support for the self-regulatory model, which also demonstrates its utility and effectiveness in business settings. He next explains that a firm can best activate employees’ ethical values by having employees perceive that fair procedures are used in firm decision-making and other workplace events. His message here is that employees feel comfortable about working for an organization and find its rules to be legitimate if they perceive that it treats them fairly. This finding gives companies a clear guideline: if they want employees to comply with a company’s rules and otherwise to act ethically, the company must establish fair procedures for workplace decisions. This raises the question of what constitutes a fair procedure, and Professor Tyler offers definitions and examples involving four important procedural components: the quality of decision-making, the quality of peoples’ treatment by organizational authorities, the rules of the organization and an employee’s experience with a supervisor(s).

In his article, *Structural Holes, CEOs, and Informational Monopolies: The Missing Link in Corporate Governance*, Professor of Law Lawrence Mitchell argues that the reform focus of improving the independence and monitoring ability of corporate boards misses an important origin of the corporate scandals in the power of CEOs and other senior executives. To explain this power, Professor Mitchell uses the theory of “structural holes” from economic sociology, which posits gaps between networks of individuals and groups and the importance and advantages that accrue to people who can fill the gaps. Unlike classical economics, the theory sees the actor as a social creature, i.e., a part of various social networks. After discussing the theory and situating it in economic and sociological discourse, he uses it to point to the gaps in the governance of public corporations that corporate actors can use opportunistically and that reforms should (but don’t) address. For Professor Mitchell, the theory suggests that the current
effort to control CEO power by emphasizing boards with independent directors may have the opposite result, for in firms with these boards the CEO alone would fill the structural holes of information about the corporation for the directors (as opposed to the situation of a board with inside directors who would be part of other informational networks in the corporation). Even more interestingly, Professor Mitchell uses the theory to explain scandals involving senior executives, other than the CEO, because the CEO, consciously or not, may allow structural holes to develop beneath him in the organization, which are filled by senior executives who could engage in unscrupulous behavior (if so inclined) and which give the CEO “deniability.” Professor Mitchell further justifies his application of the theory by pointing to the movement in business organization from hierarchical to flatter organizational structures. The latter allow for more structural holes, because more managers report to the CEO, and thus more possibilities of CEO or senior executive opportunism (depending upon who fills the holes). He thus argues that typical reforms aimed exclusively at enhancing the board’s monitoring ability miss structural holes below the board as an important determinant of the corporate scandals.

Although she was not a presenter at the conference, Professor of Philosophy Margaret Gilbert offers her views on the subject of corporate misbehavior in her paper *Corporate Misbehavior and Corporate Values*. As is customary in philosophical discourse, she gives precision to the concepts of group beliefs and group values, to distinguish them from their personal counterparts and to emphasize their strength (through a group member’s “commitment” to these beliefs and values) in crowding out conflicting individual beliefs and values. As she argues, “collective beliefs, values and goals are apt to induce people to disguise their contrary opinions—however morally perspicacious—and to abstain from any active effort at their diffusion.” She joins with Professors Treviño and Tyler in asserting that it is critical to prevent corporate scandals for a firm to make moral values a part of its group beliefs.

These articles clearly fulfill the goal of our conference of promoting new avenues of scholarly research on business firms and their pathologies. At the end of the conference, moreover, its participants met in a planning session to discuss ways of promoting research between organizational theorists and social psychologists, on the one hand, and corporate scholars, on the
other. This session led to the following projects: In the near future, there will be launched an electronic journal to be published by the Social Science Research Network entitled “Business Associations and Financial Law: An Interdisciplinary Journal,” which is designed to circulate papers in the business law area with a social science focus. Participants in the conference as well as other interested scholars are also working to create a model interdisciplinary course with readings from the social sciences and the law. We hope that these will be the first of many collaborative projects between social scientists and legal scholars, and that the projects may improve legal policy-making on the regulation of corporations and financial institutions.
The Cognitive and Social Psychology of Contagious Organizational Corruption

*John M. Darley*

I. CORPORATE CORRUPTION: THE NATURE AND MAGNITUDE OF THE PROBLEM

Seen in the clear light cast by hindsight, several related puzzles emerge about the recent cases of corporate corruption. First, although such scandals may begin small, they often grow to huge and blatant proportions. Second, because of their blatant character, they seem suicidally stupid. They will eventually be detected, with the inevitable disgrace that this will bring about for participants. Third, in their later stages, they come to involve a number of people in the organization who are busily involved in committing complicit actions that forward the corruption. How all of this certainly unethical and generally stupid thought and behavior comes about is the problem that I will examine in this article. In sum, people seem more recruitable into corrupt practices than we would think.

Another puzzle that has come to our attention is what seems to be the high frequency with which these problems occur. One social scientist, looking back over the past few decades, has produced a chapter on thirty-six different and major cases of what he calls corporate “crime and violence.” Over the past few years, many major U.S. corporations have been caught in acts of corruption of quite startling magnitude.

---

* © 2005 John M. Darley. All Rights Reserved.
† Warren Professor of Psychology and Public Affairs, Princeton University.
B.A., Swarthmore College; Ph.D., Harvard University.

1177
To emphasize the question that this raises, why are so many of these incidents taking place in organizations that we would have thought were staffed by morally good people who were also prudent enough to realize that corrupt practices are frequently detected? That is, these corrupt incidents so often seem to involve corrupt, rule-breaking actions by people who we would have assumed were moral, prudential actors.

One conventional answer that resolves the puzzlement is to retrospectively decide that the assumption that we made about the people in question, that they were moral and prudential actors, was wrong; more specifically, that they were persons searching for corrupt opportunities, and were blinded to the probabilities of detection by their greed. The reader will recognize this stance as a variant of the “few bad apples” theory that has been cited to explain recent acts of corporate corruption such as the mutual fund scandals or the organizational corruption that led to the torture of Iraqi prisoners of war at Abu Ghraib.

I want to suggest that the bad apple theory is at best a factually incorrect reading of what has happened. In fact it is simply a useful fiction that enables those who hide behind it to avoid the more thoroughgoing implications of recent transgressions. Specifically, clinging to the myth enables us to avoid the realization that the world of corporate or governmental ethics requires more attention and more painful redesign than the minor housekeeping implied by the course of action involving the elimination of already discovered malefactors from a system that we assume is otherwise working perfectly. For those that hide behind the bad apple myth, the sole remedy is to be more careful at the recruiting and training end of the organizational world; perhaps checking the credentials of job candidates better, perhaps by the technological fix promised by the quest for the modern “lie detector” that will ensnare the wrongdoer on the way to his wrongdoing.

In this article, I will attempt to answer two questions that come to the mind of a psychologist who thinks about organizational corruption. First, why are so many “initial corrupt acts” taken in organizations? The answer cannot be that it is simply already-corrupt people who take these actions. Part of the answer is that some of the people who launch these corruption-initiating acts do not scrutinize these contemplated acts from an ethical perspective. Strange as it may seem, they do not see them as unethical.
The second question then becomes “why is it so easy to recruit other members of the organization to take the actions that amplify, extend, and continue these initiating actions to produce more and more corrupt outcomes?” What causes the organization to turn itself into one that works together to produce full-blown ethical transgressions? To foreshadow what I will suggest, the answer to this second question is threefold. First, because these others often accept the implied definition that the first actions were ethical in nature, the distance between that first act and the next one that amplifies it are not easily recognizable. Second, these follow on acts are perhaps seen as ethically grey and further are produced out of considerations of group loyalty and commitment. Third, when one is a committed member of an organization, social identity theory\(^2\) points out that we experience an alteration in personality. We “become” the prototypic member of the group, and the cues around us are that the prototypic group members are engaging in the corrupt actions. Thus we do so also. Finally, it is a little noticed truth that our society offers alternate identities to citizens, and some of them allow for acting in ways that, from the perspective of another identity the person could assume, are unethical.

To arrive at a better explanation of the apparently “infectious explosion” of these acts of corporate and organizational corruption, we need to consider several sources of information. First, the narratives that have emerged from first party participants in episodes of corruption, second, a new perspective that is emerging from judgment and decision making research about how it is that human decisions are made, and third we must take a closer look at the choices faced by individuals in an organization as the corruption begins to impinge on them.

II. THE INITIATION OF CORRUPT SEQUENCES

How does the first corrupt act, the one that starts the process in the wrong direction, come about? Let’s look at the narratives first. Sometimes the stories of corruption are simple. The organizational leaders deliberately act to bring about corrupt or otherwise immoral actions by the organizations they

lead. Thus, Film Recovery Systems, Inc. hired workers who could not read English so that they would not be able to read warnings on the containers of dangerous chemicals they were using in the deliberately dangerous processes of recovering silver from used photographic plates. Perhaps the easiest explanation for how the company is able to enlist the organizational members, in this case the foremen who gave the workers their orders, in carrying out the actual immoral actions is that the superiors who are determined to carry out corrupt practices simply recruit subordinates who will be willing to engage in corrupt practices.

But we should also consider some more disquieting narratives, which seem to suggest that the corrupt practices are somehow stumbled into, without exactly being intentional. This is a disturbing perspective, one that challenges the notion that corruption begins in corruption, that the source of corrupt acts is those individuals who are corrupt and extract corruption from their followers.

From this perspective, acts that start a chain of other actions that ultimately result in full blown corrupt actions often have their origins in actions that are not themselves corrupt, or at least not perceived as so by the original actors.

If this is so, then we lose the comfort of being able to deny that we ourselves would ever be enmeshed in corrupt acts. We deny the message given by the frequency with which we discover that the actors enmeshed in corrupt activities are anguished individuals, frequently individuals who saw very clearly that detection was, if not inevitable, at least highly likely. They simply could not see a way to escape from the meshes of the collective processes that were ongoing.

In our conventional way of thinking about ourselves, we are confident that we would know in advance that to do some set of actions would be morally wrong, and that this realization, occurring prior to the actions, would prevent us from taking them.

These comforting thoughts turn out to be not true. Instead, people habitually commit actions that are self-serving, or unduly favorable to the organizations in which they are situated. On careful examination by a non-biased individual, these actions would be judged to be morally dubious or morally

---

3 A succinct account of this corporation and its misdeeds is given in Nancy Frank, Murder in the Workplace, in CORPORATE VIOLENCE (Stuart L. Hills ed., 1987).
just wrong. But they don’t receive that scrutiny. These actions are often the ones that set in motion a cascade of further corrupt actions; that set up what we might think of as the tornado of corruption that gathers force and pulls in more of the organization’s members. So let us call these initiating actions the generative actions of corruption.

An example of this is useful, and there is one that is often cited in the literature. The circumstances that bring it about are the constant and high pressures for a for-profit organization to show a steady rise in earnings in each successive reporting quarter. But what counts as “a sale” that can be counted as earnings in a particular quarter? There is often judgment required in answering this question. But if the sale can be counted in the present quarter, and it will move earnings to a higher level, then the temptation is to “recognize the revenue” in the present quarter.

The example involves the practice of tobacco companies at the end of the business quarter.

Loading wasn’t unique to Reynolds, every tobacco company did it to some extent. Just prior to its regular semiannual price hikes, Reynolds regularly offered huge volumes of cigarettes to its customers—customers and supermarket chains—at the old prices. Customers loved it because they could sell low-cost cigarettes at the new, higher prices. Reynolds loved it because it cleared away unwanted inventory, kept the factories humming, and, most important, produced large, artificial, end of the quarter profits.4

The problem with this was that the distributors were free to return the cigarettes to Reynolds a month or two into the new quarter, after they had served to create the fictitious “profits” at the end of the last quarter.

III. THE UNFORTUNATE CASE OF AUTOMATIC INTUITIVE JUDGMENTS

Earlier I said that some of these acts that initiated further corrupt practices were not decided upon in any very thoughtful way. This needs explication. Recently, psychologists have summarized5 a good deal of research and thinking about decision processes as requiring us to make a distinction

---

between several rather independent systems that inform us about the world. For some time, we have known that we need to distinguish the human perceptual system from the human reasoning system. The perceptual system presents us with what we “see.” We know in fact from countless demonstrations that perception is in fact a decision process, in the sense that it involves a good deal of past learning, often confirms stereotypes and generally sees what we expect to be there. However, partially because perceptual processes are overlearned, partially because they are what we now call automatic, we are misled about the truth of our perceptions. What we “see” must be what is true, a stance that is generally called naïve realism. What we see is unproblematically true.

On the other hand, when we engage the reasoning system, which we sometimes use to make decisions, we are aware that reasoning is in progress because that reasoning is controlled and effortful. It often involves deliberately engaging problem solving rules that we have learned before. Therefore it is often cued into action by the conscious recognition of “what kind of problem that it is.”

It is Kahneman’s recommendation that it is worth distinguishing a third cognitive system that shares components of both of these other systems, and exists intermediate between them. This we will call the intuitive system. More will be said about it in a minute, but let me tell you the use I will make of it this discussion of corporate corruption. Recent research demonstrates that it often the case that the acts that can originate unethical chains of occurrences arise from the quick decisions that are products of the intuitive judgment system. As one consequence, these acts often are not subject to the scrutiny by the actor that we apply to action decisions that we know are the product of the more deliberative reasoning system.

Here is the cash value of this realization. It pinpoints the attributional mistake we make when we think people who commit unethical actions are characterologically unethical. We expect that all good people, and we ourselves, scrutinize acts that we are contemplating taking from an ethical perspective. Therefore we do not take unethical courses of action. We then reason that if an ethically wrong act is committed, a person who is morally corrupt has committed it—we have returned to

\[6\] Id. at 697-99.
the “bad apple” theory. The way to deal with corruption is to screen out individuals who are corrupt.

The disturbing message from those that study decision-making is that these reassuring thoughts are untrue. Many of the actions that begin cycles of corruption are the products of the intuitive judgment system, which means that they are rapidly arrived at, less than consciously considered, and unintentional in their ethical dubiousness. Further, they are often the product of pressure to make fast decisions. And under this condition, they are not subject to the monitoring of the decision, which is done by the reasoning system. As Kahneman’ comments, “the monitoring is normally quite lax and allows many intuitive judgments to be expressed, including some that are erroneous.” The suggestion that emerges is that the “natural” intuitive decision is likely to be a self interested one. To quote researchers on this topic:

[S]elf-interest is automatic, viscerally compelling, and often unconscious. Understanding one’s ethical and professional obligations to others, in contrast, often involves a more thoughtful process. The automatic nature of self-interest gives it a primal power to influence judgment and makes it difficult for people to understand its influence on their judgment, let alone eradicate its influence.8

This decision may be overridden by the more deliberate thinking of the reasoning system, but only if something triggers that system into action. Thus, in sum, corrupt actions are often committed by people who are not themselves corrupt.

A. Self-Interested Intuitive Judgments

Let us trace this out at the level of personal decision-making. A doctor orders perhaps unnecessary tests for a patient from a testing laboratory in which he has a financial stake. He knows he did not make a self-interested decision because he knows that “he didn’t even think” of his stake in the laboratory while he was making the decision. A human relations person hires a member of her ethnic group for a job for which there were many candidates, and is sure that the decision is a fair one because she “examined the credentials of all of the candidates with an open mind.” An auditor examines

7 Id. at 699.
the accounts of a corporation his firm is engaged to audit and is sure his judgments of the acceptability of various decisions that the corporation has taken are appropriate according to a fair reading of the auditing standards. But in all of these cases, it is possible that in fact these sorts of decisions are frequently biased by self-interest,\textsuperscript{9} in-group favoritism,\textsuperscript{10} egocentricism,\textsuperscript{11} or conflicts of interest.

B. A Biased Intuition and its Entrapping Consequences

A well-known example from corporate life is useful here. It is driven by the previously mentioned desire of the organization to produce smooth patterns of earnings across the quarterly reporting periods. It involves what becomes “improper revenue recognition.” The problem it solves is enabling sales of product to be sufficiently concrete to be bookkept in the present quarter rather than the next one.

The famous organizational example involves the Kurzweil Applied Intelligence Company.\textsuperscript{12} First, the CEO allowed sales persons to post sales that in fact came in a few days after the quarter closed. This seems a rather harmless practice, but it creates a slippery slope problem that is well-described by Tenbrunsel and Messick.\textsuperscript{13} By allowing the act, the CEO authoritatively stamped that act as ethically allowable. But if it is acceptable to “count” orders a little after the quarter closed, then why is it not ethical to count orders that came in a day or so after the late orders that were allowable?

Eventually, the company went so far down this path that they counted orders far into the next quarter as revenue, and then salespersons began forging customer signatures on orders that they thought would be forthcoming and counted those. And so on.

Notice two things that were happening. First, eventually a line was crossed from ethically grey actions to blatantly illegal ones; eventually when the auditors wrote to

\textsuperscript{9} \textit{Id.}


\textsuperscript{12} Mark Maremont, \textit{Anatomy of a Fraud}, BUS. WEEK, Sept. 16, 1996 at 90-94.

customers to verify sales contracts, bogus responses from “the customers” were also forged. And second, notice that the eventual fraudulent endpoint was a consequence of the first grey actions. Robbing sales from the next quarter to pad results for this quarter made it more likely that the next quarter would be short even more sales. Thus the company was in some very real sense committing itself to an increasingly morally wrong and desperate set of escalating acts. But that commitment was unlikely to be apparent to the actors who initially claimed a few sales from the beginning of the next quarter. The slippery slope was inevitable but unforeseen.

The moral is that an initial ethically grey act can later been seen as committing the corporation to further and further actions, and these later actions were more and more clearly across any ethically boundaries that could be imagined when the first steps were taken.

IV. ENTRAINMENT

I have given a brief sketch of psychologists’ current thinking of the two rather different stances in which decisions get made. One way of drawing the implications of this for the present problem is to say that people are ethical, but only intermittently so. Whether we will be ethical depends on whether events in the past or the present trigger the reasoning system to generate a checking ethical perspective on courses of action that are generated from other more intuitive perspectives. This solves the problem of why it is that so many ethically bad actions are authored by individuals who are not themselves chronically unethical.

The next problem to solve is why these initial unethical actions so often seem to capture others in the organization, who build on, add to, and amplify the continuing chain of unethical actions. This past example of counting non-existent orders as revenue is one such example.

Let me give you a metaphor for what I want to suggest. Entrainment is a concept that comes from early twentieth century that I want to borrow for an organizational process. Originally it referred to a perceptual phenomenon: an object is moving in one direction. As it passes other stationary objects, those objects themselves begin to move, and they move in the same direction as the original object. This seems to be to be a useful visual metaphor for the way that a corrupt act seems to affect an organization. Often it spreads in the following senses.
More and more people commit similar acts, often ones “triggered” by the original acts, and those subsequent acts often grow more extreme in their wrongness. How this comes about is the next question to answer.

A. Imperceptible Differences

If an action is committed, and is not criticized, punished, or otherwise labeled as wrong, it becomes “the standard.” It may not be criticized, even though many in the organization think that it is wrong, because their insecurity or their lower position in the organizational hierarchy makes them unwilling to say publicly what they really believe, which is that it is wrong. But when this happens, psychological research demonstrates an interesting process called “pluralistic ignorance.”\(^{14}\) Rather than realize that the other silent individuals are being silent for exactly the same reasons that he is, the individual tends to conclude that these others think that the act is an acceptably moral one and are keeping silent for that reason.\(^{15}\) The individual then, is the deviant, and under this pressure, comes to think that the act is more normal and more ethical than he previously thought. It is now the standard for what is allowable in this organizational context.

But then, a slightly more unethical action becomes possible, and the then relevant question is the distance of this next possible act from the act that is now the standard. Tenbrunsel and Messick have a useful term for this, which is the “induction mechanism.”\(^{16}\) “This mechanism uses the past practices of an organization as a benchmark for evaluating new practices. If the past practices were ethical and acceptable, then practices that are similar and not too different are also acceptable.”\(^{17}\) In small steps, an organization moves from ethical actions, to ones that are ethically grey to ones that are simply immoral.

It is possible that by progressing in these small steps, the organizational group never becomes aware of the moral


\(^{16}\) Tenbrunsel & Messick, supra note 13, at 228.

\(^{17}\) Id.
wrongness of the procedures with which they end up. Recent business pages are full of reports of how insurance brokers got into the pattern of taking what were essentially kickbacks from insurance firms to whom they brought clients. The stories of the final stages of this process seems so prototypically corrupt that it is hard to believe that the perpetrators could code them as anything but unethical, but that they did so is not yet clear. However, it is difficult to think that at least some of those involved did not at some point see the wrongness of their actions.

Here are some descriptions of the patterns of actions once the system was in full swing. Apparently charades were staged, in which some insurance providers were solicited to put in bids for insuring the broker’s clients, but the bids were organized so that they would be higher than the bid of the provider who was to be the eventual winner of the insurance contract. The purpose of this was to provide “proof” to the client that the broker had solicited bids and was giving the contract to the lowest bidder, as was proper.

This is an interesting process, since it engages the high-bidding insurance providers in the charade, with the incentive that they someday will allowed to be the “lowest bidder” and win a contract. One frequently sees mechanisms for recruiting other organizational units into a corrupt system, and here we have identified one.

These patterns, described as I have done, from the perspective of the final stages of the system, seem to so clearly be corrupt that it is hard to believe that they can be anything other than the consciously immoral acts of conspirators. However, if we think of a person being recruited into the system, it is possible that she would simply see it “as the way we do things around here.” We will return to this theme later.

B. Loss Aversion

Recent psychological research has conclusively demonstrated that people will go to great lengths to avoid losses. See Daniel Kahneman et al., Prospect Theory: An Analysis of Decision under Risk, 47 ECONOMETRICA 263, 263-91 (1979).
example of quarterly profits. The stock prices of American companies are significantly dependent on the company slightly increasing its earnings on a quarter-by-quarter basis. This apparently is taken as the sure sign of a steadily more-profitable company, one that one should invest in. However, on reflection, there are many reasons why good companies would not produce that patterning of profit: seasonal sales profiles for one, high expenditures in one quarter for research and development costs is another. Companies are led to accounting practices that allow for “earnings management” to enable them to produce the preferred steady rise in earnings even when more standard accounting practices would produce variable quarter by quarter gains.

Suppose that you are in control of the accounting process in such a company and see that some perfectly justifiable expenses will bring you to earnings that fall just below those of the last quarter. And suppose that you correctly think that the “increased earnings every quarter” criterion is a stupid one. Yet you know that if you show reduced earnings in this quarter, the price of the company stock will drop, and research reports may comment about “disappointments at company X.” If you are one of the company’s executives that has a “pay for performance” plan, you realize that you may lose considerable sums of money, money that you had counted as already in your pockets. Would you stretch accounting rules to produce increased earnings? Perhaps not. But would you have been clever enough to “stash” some earnings from a previous highly profitable quarter that could now be pulled out of the “cookie jar” to produce those earnings in this quarter? They were, after all, perfectly legitimate. Would you be morally wrong to ensure that your company was buffered from the ups and downs in stock prices caused by the essentially stupid focus on increases in quarter-by-quarter earnings?

Now recognize what is often the case, which is that the company CEO receives a good deal of added incentive pay if the corporation “makes its numbers” on a quarter-by-quarter basis. This means that there will be a good deal of pressure on the auditing group to make the audit output conform to the “steady growth in earnings” numbers. So the real question is not whether you yourself would independently produce numbers

---

that you think it might be morally justified to produce, but whether you would resist pressures from corporate superiors to do so. And thinking about this as the CFO would, you should realize that there are a number of cases in which CFO’s have been dead-ended or fired for refusing to go along with these directives from above. Loss aversion might be less of an abstract concept, and more a realistic fear of loss of job.

V. GROUP LOYALTY AND COMMITMENT

We now see how an individual in an organization can impulsively take an action that is, from a perspective that was not apparent to the actor, wrong in the sense of being an action with morally flawed outcomes. Eventually, this action becomes known to other members of the organization. The question is how they react to it. We would hope that they would repudiate it, both because it is wrong and because it is likely to commit the organization to a bad course of action. However, there is one problem with people following this path. The action has already been taken. It is done. The pollutants have already been dumped into the river, or the quarter’s profits have been overstated, or the member of my in-group has been hired. Often the consequences of these actions are irreversible. And even in those few occasions when the action consequences can somehow be reversed, it is still the case that there are likely to be records around that they were the actions initially taken by the organization.

Previously we considered the possibility that the performance of actions that from some outside perspective would be considered wrong might instead convince others within the organization that those actions were right in the context in which they were committed. They were, in other words, the way that my company does things. Now I want to consider another branch of the situation; the branch in which the other actors in the organization realize that the act is bad, either intrinsically bad or bad because of the consequences that will follow. Since the act has already been committed, the decision is not about making it disappear—that cannot be made to happen. The decision is between making others in the

---

organization aware that bad actions have been taken and letting them continue, or abetting their continuation.

It is likely, for reasons of loyalty to the group, loyalty to the person who made the bad decision, and a feeling that the commitment to the course of action is irrevocable, that others in the group will allow or abet the continuation of the actions. This is particularly true when the actors who have become aware take actions that seem to temporize, and keep open the possibilities of later actions that halt the bad practices. These actually often allow the bad course of action to continue to develop because if one does not intervene when one first becomes aware, it is very difficult to find an exact time when one should intervene later.

A famous example of this comes from the first person account of the fraud that the Goodrich Company backed into committing when they were pursuing a design for aircraft brakes that could not possibly work. Briefly, an engineer had made calculational mistakes in designing a brake assembly for an airplane. The plans called for brake lining pads that were too small to provide the braking friction to stop the plane in the required distance. “The brake was too small. There simply was not enough surface area on the disk to stop the aircraft without generating the excessive heat that caused the linings to fail.”

From the point of view of our analysis, a critical incident then occurred. “New menaces appeared. An engineering team from LTV (the primary contractors) arrived at the plant to get a good look at the brake in action. Luckily, they stayed only a few days, and Goodrich engineers managed to cover the true situation without too much difficulty.” What I suggest is that the visit of outsiders caused the Goodrich personnel, although aware of the eventual guaranteed failure of the brake assembly, to rally to the support of their fellow engineers to conceal this critical fact. By doing so, many of them became complicit and caught up in perpetuating the fraud.

A different case, with the same ultimate consequence, occurs when the individual who first committed the bad action shifts perspective, and sees the potential bad consequences of the bad action that he launched. Here he needs to make sense of his own past actions. The true reason that the person

22 Id. at 148.
23 Id. at 150.
committed the action is that he did not think at the time about the potential bad consequences of the action. From the intuitive perspective he adopted at the time of the decision, it was the right decision—or at least not the wrong one. This is the whole message of the previous excursion into the intuition-based decision system. However, one of psychology’s more interesting discoveries is that people do not grant themselves this sort of charity. Instead, hindsight causes them to think that it was a well-considered decision, made by the conscious, reasoning system. This retrospective perspective leads them to go into a sort of decision-hiding mode, in which they seek to deny their involvement in the decision, or to experience the dissonance they feel, and think of the reasons that the decision was the right one. To do the latter, they have to think about themselves in different ways.

A. Social Identity Considerations

A theory developed in the last two decades has made and validated a very important point. When an individual is a member of a group, in the sense that she is committed to the purposes of the group and that a group has tasks to do, the task of the individual is to first become a prototypical member of that group, and then help the group as best she can in reaching its goals. Among other things, this may mean adopting the moral perspectives of the group. And recall what we said earlier. Because of pluralistic ignorance, she may not be aware that others in the group consider the initial act an unethical one. The signal that silence conveys to her is the incorrect but persuasive message that the group regards the initiating act as a morally appropriate one. The task of the individual group member is to accept that decision and move the group forward. This may mean taking actions that conceal the prior transgressions, but these may also be the actions that continue the bad course of action. In the example from Why Should My Conscience Bother Me? the loyalty-driven actions of the Goodrich workers in assuring the visiting team from the contractor that “everything is going along ok,” contributed to

26 See Vandivier, supra note 21.
the continuation of the doomed fraud. The contractor then did not raise questions that could have headed off the final bad outcome.

B. Alternate Identities Are On Offer

The concept of identity can be made to do more work. There are identities in which it is part of the role enactment to adopt a different moral code from the one we usually espouse. The violence endemic in hockey and American football is an example. One inflicts violence on others in ways that would normally involve morally unacceptable acts. Of course, the allowable violence is constrained by rules, but there are two interesting things to note about that. Some acts of violence, like “late hits” in football and “slashing” in hockey are against the rules but a second set of rules is in existence to assign penalties for those rule infractions, which in some sense brings the rule violations “within” a broader domain of “rule acceptable” actions. Second, as has been commented on by those who follow sports, team members often collectively adopt a “persona” that makes the goal of inflicting harm on the other side acceptable and even desired. “Let’s get out there and knock them dead” is an injunction that brings violence within the somewhat ambiguous orbit of legitimacy in game settings. Other roles contain elements that legitimate morally dubious actions against others.

And all of us can give a reasonable performance in at least many of the roles. Let me give you an example of a situation in which a person who had detected corruption and is set to denounce it, is sent away instead with an offer of a role. If he accepts the role, he will embrace the deception and play his part in continuing and expanding it. He has entered the situation as an upstanding person of high moral rectitude, which too is a role. But how will he continue on when that role is challenged? Notice in this specific situation, considerable pressure exists to accept the new, deception-embracing role.

The dilemma arose for Michael Lewis, and is described in his book Liar’s Poker, in which he summarizes his experiences as a bond salesman in training with Salomon Brothers. An experienced trader had advised him that AT&T

27 See id.
bonds would be good ones to sell to his customers and he sold about three million dollars worth to one of his clients. The bonds rapidly dropped in value, harming the customer and harming the relationship of trust that Lewis had built with the customer. A more experienced salesman clued Lewis that the reason that these bonds were “good ones to sell” was that they were in Salomon Brothers’ inventory, and the firm was quite sure that their value was going to decline. Thus they wanted to unload the bonds, and did so on unwary clients.

Lewis protested to the trader, arguing that they had quite badly harmed the client and behaved in a way that contradicted their high-flown ethics codes about duties to customers. “Look,” he (the trader) said, losing his patience, who do you work for, this guy or Salomon Brothers?” At this point, Lewis realized that the real practice of Salomon Brothers was to mistreat clients for the good of the firm. If he were to stay at Salomon Brothers, he would need to adopt the identity of “the jammer,” a person who was willing to take these sorts of trust-violating actions. But the only other identity available to him at the moment of decision was that of a naïve fool, who did not know the ways that the real world worked. And he was surrounded, in the close confines of the trading floor, by many others who would certainly have contempt for a naïve fool and publicly express that contempt. These are the circumstances in which people adopt identities that enable them to act unethically toward certain groups of others.

VI. CONCLUSION

I have attempted to solve two puzzles. First, why so many acts that generate bad moral consequences are begun by people we would think are proper moral actors. The answer is that they are what most of us are, which is intermittent moral actors. They take a moral perspective if the reasoning system is engaged, but otherwise can be driven by quite intuitive and automatic thinking to “solve the immediate problem” which will often be done quite pragmatically.

The second problem might be called the “lemmings” problem. Why is it that other people in the organization so often seem to aid, abet, and advance the morally bad course of action? My answer here was more complicated, and involved a

---

29 See id. at 167.
set of independent but generally correlated processes. Some of the processes lead to the prior corrupt act being perceived as ethically appropriate “within the organizational context.” Others work by the engagement of a group loyalty or commitment, which causes people to work to conceal the prior corrupt actions from public view. This often entails further corrupt actions, either “covering up” the previous actions, or continuing them. It is sometimes the case that actors who previously were careful to act in moral ways, are now recruited into adopting a persona that goes along with, and even becomes an independent origin of corrupt practices.
Out of Touch

THE CEO'S ROLE IN CORPORATE MISBEHAVIOR

Linda Klebe Treviño

I. INTRODUCTION

The last few years have brought an endless parade of headlines and “perp walks,” raising questions about who is to blame for an apparent spike in corporate misbehavior. In this paper, I rely on social-scientific theory and empirical research to focus on the role of the CEO in corporate misconduct. I demonstrate, first, that an active CEO role in ethics management is essential because the CEO's commitment to ethics influences key characteristics of formal ethics and legal compliance programs. In addition, as Chief Ethics Officer, CEOs must create and maintain the ethical culture in their organizations. Both of these types of influence can have a powerful impact on employee behavior. However, research also suggests that many CEOs are out of touch with the importance of their ethics management role. Senior managers tend to view the firm’s ethical climate in “rosy” terms compared to lower-level employees. In addition, many CEOs become far removed from the ethical realities in their organizations simply because they rarely interact with lower-level employees. As a result, their organizations and employees are left to flounder without a strong rudder to guide the organization in an ethical direction.
One very pragmatic reason to focus on the CEO role has to do with the U.S. Sentencing Guidelines compliance standards. The guidelines were first adopted in November 2001 to reduce judicial discretion and to provide a “carrot and stick” approach to sentencing corporations convicted of crimes. In the “carrot” part of the approach, the original Sentencing Guidelines called for leniency in sentencing organizations that can demonstrate that they had made a strong effort to prevent employee misconduct. The guidelines listed seven standards for judging what would be considered an “effective” legal compliance program, including: high-level executive oversight of the firm’s efforts to insure legal compliance, the exercise of care in delegating this authority to others, communication of conduct standards through dissemination, and regular employee training. The guidelines also included requirements regarding the establishment of systems to monitor employee behavior, including systems that allow employees to report misconduct they observe as well as consistent discipline for misconduct when it occurs, and responses to misconduct that are designed to prevent its reoccurrence. A 2003 survey found that most large organizations had formal ethics or legal compliance programs. The study found that the larger the organization, the more likely it was to have formal codes, ethics training, ethics offices or advice lines, and anonymous reporting systems.

Ideally, the CEO’s role should be important in guiding the establishment and implementation of these programs. However, in practice, most large firms that implemented legal compliance programs in the 1990s delegated authority for these formal programs to an “ethics or compliance officer”—the Chief Legal Counsel or another executive appointed to manage internal ethics and legal compliance programs. Many of these officers belong to the Ethics Officers Association (EOA), a professional organization that has grown through the 1990s to its current size of over 1000 members. Although members of the EOA meet regularly to benchmark and discuss best

---

3 Id.
4 See Ethics Officers Association (EOA), About the EOA, at http://www.eoa.org/AboutEOA.asp (last visited April 20, 2005).
practices in ethics and legal compliance management, research suggests that the large majority of ethics/compliance officers have little regular contact with the CEO, calling into question how active CEOs are in guiding ethics and compliance management efforts.

A number of observers became concerned over the years that some of these formal ethics/legal compliance programs were little more than “check-off” efforts that allowed organizations to say that they were in compliance with the Sentencing Guidelines while, in fact, the programs were seen by employees as little more than window dressing. Perhaps because of these concerns, changes to the U.S. Sentencing Guidelines, as of November 1, 2004, further highlighted the role of senior executives in creating a strong ethical culture in the firm in addition to a formal ethics or legal compliance “program.” These changes require that the “governing authority” be knowledgeable about and exercise reasonable oversight regarding the implementation and effectiveness of the ethics or compliance program, and that the organization “promote an organizational culture that encourages ethical conduct and commitment to legal compliance.” In addition, Sarbanes-Oxley has increased the accountability of both senior executives and the board for the oversight of financial reporting. As a result, CEOs and boards have taken more interest in the implementation of ethics/compliance programs in their firms and are asking more questions about what their role should be in promoting an organizational culture that encourages ethical conduct and commitment to legal compliance.

Fortunately, empirical research conducted over the past ten years provides some guidance. First, such research demonstrates clearly that CEOs matter. Their personal commitment to ethics influences characteristics of formal ethics and legal compliance programs. In addition, their leadership has a powerful influence on the creation and maintenance of ethical cultures in their organizations.

---

II. CEO INFLUENCE ON FORMAL ETHICS/COMPLIANCE PROGRAMS

Research has found that the CEO’s “commitment to ethics” influences the scope, orientation, and integration of the formal ethics/compliance program. Ethics and legal compliance programs can be conceptualized as organizational control systems that aim to control employee ethical and legal conduct. As suggested above, these programs generally include some or all of the following elements: ethics or legal compliance officers, formal codes of conduct, training programs, systems for reporting misconduct, and disciplinary mechanisms for handling unethical or illegal behavior. Previous studies had generally documented the existence of such programs and elements, but had not attempted to differentiate among them in terms of their “scope.”

Programs have also been discussed in terms of their control orientation. Programs may rely on a coercive approach to controlling employee behavior that is based upon rules, monitoring for rule compliance, and discipline for rule infraction—a compliance-based approach. Alternatively, programs may attempt to control employee behavior in a more aspirational manner by creating commitment to shared ethical values—a values-based approach. These approaches need not be mutually exclusive. In fact, in many organizations they are designed to work together. Programs can work to develop shared values, support and encourage employees whose behavior is consistent with those values, while holding others accountable for behavior that is inconsistent with the values.

Finally, programs can be differentiated in terms of the level of their integration with daily organizational activities. Some may be highly integrated programs that affect everyday decisions and actions in the organization, while other programs

---


7 Scope of a program is defined as the number of different program elements that a particular organization includes in its formal program.

are perceived to be little more than window dressing. Organizational researchers refer to the latter as “decoupled” programs because they operate in a way that has little influence on daily decisions and behavior. These programs may look good from the outside, but members of the organization recognize that they have little impact on daily organizational functioning. So, for example, an organization may develop and distribute a code of conduct, but do little to enforce it. As a result, the code gathers dust or ends up “filed” in the circular file.

If they were asked directly, most CEOs would likely say that they are highly committed to ethics. But the more important questions are whether employees perceive that commitment, and whether the CEO’s commitment to ethics is seen as high relative to the executive’s commitment to other, more bottom-line oriented concerns. In my research with colleagues, we hypothesized that the CEO’s commitment to ethics would be associated with increased program scope (the existence of more formal ethics/compliance program elements), a higher likelihood that the program would be values-oriented (rather than compliance-oriented), and the likelihood that the program would be integrated into daily organizational activities such as performance appraisal systems. We asked ethics/compliance officers to rate the CEO’s commitment to ethics relative to other operational and strategic concerns, and we found that, as proposed, the CEO’s commitment to ethics was associated with program scope, orientation, and integration. In organizations with strong CEO commitment to ethics, we found more formal program elements, a stronger values orientation in those programs, and greater integration of the formal program into daily organizational life.\(^9\)

It is particularly important to focus on these program characteristics because research has also found that they make a difference in employee outcomes. First, employees who work for organizations that have formal ethics and legal compliance programs with multiple program elements are more likely to say that they would report misconduct and are less likely to report feeling pressure to compromise ethical standards or to say that they have recently observed misconduct in their organization.\(^10\) Even more important, however, than formal

\(^10\) See generally SURVEY, supra note 3.
program elements are a values orientation and integration of the program into daily organizational life. Research has found that these aspects of ethics and compliance management have a stronger impact on employee attitudes and behaviors than do the existence of formal program elements.\footnote{See Managing Ethics, supra note 8.} Employees respond best to a formal program that has a primary values orientation backed up by a system of accountability. Attention to shared values (e.g., integrity, respect, etc.) creates norms and behavioral expectations, but it may also help to create shared trust and a perception of organizational support. In addition, employees want to know that when values and rules are violated, the organization will hold the violator accountable.\footnote{See id.; Compliance, supra note 6.} Accountability contributes to perceptions that the organization is fair and means what it says with regard to ethics.\footnote{See Linda Klebe Treviño, The Social Effects of Punishment: A Justice Perspective, 17 ACAD. MGMT. REV. 647, 647-76 (1992).}

In empirical research, employee perceptions that the ethics/compliance program was primarily values-based were positively associated with, among other outcomes, employees’ awareness of ethics at work, their willingness to seek ethical advice in the organization, their commitment to the organization, willingness to report misconduct, and lower-observed misconduct. Although a perceived compliance focus was also associated with positive outcomes, a values orientation was the more important influence in every case.\footnote{See, e.g., Linda Klebe Treviño & Gary Weaver, Managing Ethics in Business Organizations: Social Scientific Perspectives 211 (2003); Managing Ethics, supra note 8; Compliance, supra note 6.}

Finally, employee perceptions that ethics and legal compliance programs are integrated into the daily life of the organization are also important. For example, if employees perceive consistency between formal policies and programs and organizational practices, and believe that the organization follows up on ethical concerns reported by employees and works hard to detect misconduct, employee outcomes are more positive.

In sum, we see that senior leadership is important because it influences the scope, orientation, and integration of formal ethics and compliance programs. To the extent that the CEO is highly committed to ethics, the organization includes more formal program elements in its ethics and legal
compliance management, the program is more values-oriented in its focus, and is more integrated into the daily life of the organization. These program characteristics are all associated with positive employee outcomes, including reduced levels of misconduct and higher willingness to report misconduct when it is observed. But the emphasis on integration also suggests that the best outcomes are achieved when ethics are perceived to be integral to the overall organizational culture and not just another program (because programs come and go in organizations).

III. CEO INFLUENCES ON AN ORGANIZATION’S ETHICAL CULTURE

We have now seen that senior executives influence the characteristics of formal ethics management. Yet creating a formal ethics or legal compliance program, by itself, does not guarantee effectiveness. Recall that Enron had an ethics code and other aspects of a formal program. Not surprisingly, research suggests that employees must perceive that formal policies are consistent with the real ethical culture of the organization. For formal systems to influence behavior, they must be part of a larger, coordinated cultural system that supports ethical conduct every day.

Culture can be defined as a body of learned beliefs, traditions, and guides for behavior that are shared among members of a group.15 This idea of culture has been used extensively to understand work organizations and the behavior of organizational members.16 Organizational culture is thought to be important because it has a powerful impact on employee behavior. Leaders influence culture by portraying a vision, by paying attention to, measuring, and controlling certain things, by making critical policy decisions, by recruiting and hiring personnel who fit the vision and values of the organization, and by holding people accountable for their actions.17 So, for example, CEOs who care about ethics will include ethics in their vision of the organization. They will design a reward

system that values and measures both means and ends, and, because they will be concerned about sending the right messages to employees, they will use discipline wisely.

Ethical culture can be thought of as a component or slice of the overall culture of the organization. Ethical culture provides informal as well as formal systems in a complex interplay that either supports ethical or unethical conduct. Formally, messages from executive leadership, organizational structure, selection systems, orientation and training programs, rules and policies, formal reward and performance appraisal systems, and decision-making processes all contribute to ethical culture creation and maintenance. The ethics programs discussed earlier can be thought of as part of the formal cultural systems. Employees are introduced to codes of conduct, ethics training programs, and systems for reporting misconduct. But how they think about and respond to these is highly dependent upon other, mostly informal, cultural systems. Informally, the culture’s norms of daily behavior, heroes, rituals, stories, and language keep the ethical culture alive and indicate to both insiders and outsiders whether the formal systems are actually implemented or are merely a façade. I will not discuss all of these ethical culture components separately because they are extensively covered elsewhere. Rather, I will focus on the ethical leadership role of the CEO because CEOs set organizational priorities (including funding priorities), contribute to the design of organizational systems, and send powerful messages about valued behavior through their ongoing communications and actions. Messages about ethics flow from the top down in organizations and the CEO is the source of many of those messages.

So, what do we know about CEOs and ethical leadership? Recent research suggests that executive ethical leadership in large business organizations is a reputational phenomenon. Most employees observe senior executives from a distance rather than through direct interaction. As a result, they form impressions of the senior executive’s ethical stance

18 LINDA KLEBE TREVINO & KATHERINE A. NELSON, MANAGING BUSINESS ETHICS: STRAIGHT TALK ABOUT HOW TO DO IT RIGHT 255 (3d ed. 2004).
19 See id.
from afar. And, to develop a reputation for ethical leadership, executives must be perceived to be both “moral persons" and “moral managers.” These two dimensions combine to create an ethical leadership matrix with ethical leaders being high on both dimensions (see Figure 1).

Figure 1: Executive Ethical Leadership Reputation Matrix

<table>
<thead>
<tr>
<th>Moral Manager</th>
<th>Weak</th>
<th>Moral Person</th>
<th>Strong</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>Hypocritical Leader</td>
<td>Ethical Leader</td>
<td></td>
</tr>
<tr>
<td>Weak</td>
<td>Unethical Leader</td>
<td>Ethically neutral leader</td>
<td></td>
</tr>
</tbody>
</table>

The moral person dimension represents the individual traits ascribed to the executive by employees. CEOs who are seen as moral persons are thought to be honest, open, trustworthy, concerned about people, personally moral, as well as fair and principled in their decision making. Because they realize that employees are watching them for cues about appropriate behavior, those who are also moral managers make it a point visibly to role model ethical conduct and to communicate an ethics and values message. In addition, moral managers use the reward system to support ethical behavior and discipline unethical behavior.

In fact, because a reward system lets employees know what is truly important, it is one of the most important cultural systems that can be influenced by the CEO's ethical leadership. Employees know that observing who gets ahead (and who doesn't) and how rewards and discipline are allocated in an organization is probably the best indicator of what really matters. Simply put, what is rewarded is what gets done. If salespeople are rewarded on commission only, it should be no surprise that salespeople will lie to make a sale. Or, if commissions are higher for some products than for others, salespeople can be expected to push those products even if they don't fit customer needs. Although people don't expect to be
rewarded for being ethical, some ethical leaders are sending important cultural messages by rewarding ethical behavior. For example, at Lockheed Martin, the chairman instituted a “Chairman’s Award” for exemplary ethical conduct. The award is given annually at a meeting of 250 senior managers. Each of these senior leaders is expected to nominate someone each year—meaning that, at the highest organizational levels, senior leaders are looking for exemplary ethical conduct to reward. The award ceremony has become a cultural ritual and stories about the winners and runners up are distributed to all employees via the company newsletter.  

How senior leaders react to unethical behavior is also extremely important. In his book, Thomas Watson, Jr., the son of IBM’s founder, told a story about the importance of disciplining unethical behavior and the message it sends to employees. Under his leadership at IBM, a group of managers started a chain letter that eventually found its way to employees who felt pressure to join so that managers would get their payoff. When Watson learned about it, he wanted heads to roll, but he couldn’t convince the division head to fire any of the managers involved. A couple of years later, the company fired a low-level employee for stealing engineering drawings and selling them. Unfortunately, the firing was handled poorly and the fired employee made Watson’s life miserable for years based upon the fact that the company had failed to fire anyone in the earlier chain letter situation. Watson learned his lesson, saying that after this experience, he always fired managers who failed to act with integrity, and that included very senior managers. He often had to overrule other managers who preferred lesser punishment. In the end, though, the company was better off because the clear message that was sent to everyone was that integrity really does matter. 

According to the matrix, a leader who is strong on both the moral person and moral manager dimensions is perceived to be an ethical leader. Founders are often credited with establishing a culture that continues in the organization long after they are gone. Interestingly, Arthur Andersen, the founder of the now defunct auditing firm, was an exemplar of

---

21 See Treviño & Nelson, supra note 18.
23 See id; Treviño & Nelson, supra note 18.
24 See supra figure 1.
ethical leadership. Among other attributes, Andersen was known as a highly ethical person who also led his employees to operate with strong ethics and values. Stories about the founder’s personal ethics were told and retold. For example, a young Arthur Andersen, at the age of twenty-eight, confronted a railway executive who demanded that his books be approved or he would pull his business. Andersen said, “there’s not enough money in the city of Chicago to induce me to change that report.” The railway company later filed for bankruptcy and Arthur Andersen became known as a firm one could trust. Andersen also taught employees his mother’s challenge to “think straight—talk straight” and the phrase became a corporate mantra. Partners said proudly “that integrity mattered more than fees.” That culture was maintained by subsequent ethical leaders for decades.

A leader who is low on the moral person and moral manager dimensions is perceived to be an unethical leader. While at Sunbeam, Al Dunlap lied to Wall Street about the firm’s financial state and became known for his “emotional abuse” of employees. He also pressured employees to use questionable accounting and sales techniques in order to meet bottom-line goals. He crippled the company before the board fired him in 1998.

Leaders who talk the ethical talk (they are moral managers), but don’t walk the walk (they are not moral persons), are seen as hypocritical leaders. Such leaders talk about integrity, but their conduct tells a different story. Hypocritical leadership is about ethical pretense—putting on a good show. As the founder of PTL Ministries, Jim Bakker preached about doing the Lord’s work while he raised funds for lifetime memberships in his Heritage USA Christian theme park. He diverted millions of dollars in donations and memberships to support PTL operating expenses and a lavish lifestyle for family and associates. PTL went bankrupt in 1987 and Bakker spent eight years in prison. Hypocritical leaders create cynicism in employees. Why should an employee reject a gift from a vendor (as the code of ethics requires) when the
CEO regularly sits in an expensive stadium seat that is provided by a key client?

The final category, ethically neutral leadership, is the most controversial with executives. It applies to executives who fall into what employees perceive to be an ethically neutral zone. They may be ethical persons, but followers aren’t really sure because the executive fails to “lead” in the ethics arena. The ethically neutral leader isn’t unethical, but he isn’t visibly ethical either. He fails to be a conscious ethical role model, and tends to focus on bottom-line goals without equal attention to how these goals are achieved. Essentially, the leader is silent about ethics and that silence is interpreted to mean that the top executive must not care as much about ethics as she or he does about issues that get more attention. Business leaders don’t like to think that their employees perceive them as ethically silent or neutral. They think, “I can’t be ethically neutral—I’m making tough ethical decisions all the time!” And they are. The problem is that most employees don’t know what they’re thinking or how those decisions are made unless they choose to communicate with employees.

A Fortune magazine writer referred to Sandy Weill, former CEO of Citigroup, as “tone deaf” on ethics issues. Weill’s management philosophy led him to decentralize and delegate management of the firm’s many business units. It appears that ethics management was delegated along with everything else. Arguably, the firm was lacking a strong rudder to guide it in a highly competitive business environment. As a result, Citigroup was implicated in a number of scandalous allegations and has spent a great deal of time and energy over the past few years responding to ugly headlines.30

Interestingly, the current CEO, Chuck Prince, who is beginning his second year as chief executive, has been much more proactive in the moral manager role. He seems to understand the importance of ethical leadership in a way that his predecessor did not. Prince has a sign on his desk that says “No Excuses.” He is strengthening formal risk and compliance systems and has vowed to be “ruthless” with rule-breakers. At a minimum, he has asked employees to know the rules that govern their own work. But he has also begun talking about values and the need for employees to internalize good ethics.31

31 Mitchell Pacelle, Citigroup CEO Makes “Values” A Key Focus; Prince Veers
So, being an executive who is perceived to be an ethical leader requires more than strong personal character. In order to be an effective ethical leader, executives must demonstrate that they are not only ethical themselves, but they must make their expectations of others’ ethical conduct explicit and they must hold all of their followers accountable every day. Research has found that executive ethical leadership is critical to employee behavior. In a recent study, firms that had an ethical culture characterized by top executives who represented high ethical standards, regularly showed that they cared about ethics, and were models of ethical behavior had a lower incidence of unethical behavior. Further, employees in these firms were more committed to their organization, more ethically aware, and more willing to report problems to management.\footnote{See Managing Ethics, supra note 8.}

As suggested earlier, some CEOs neglect this important aspect of their responsibility and, as the following quotes from interviews with senior executives show,\footnote{See Qualitative Investigation, supra note 20.} reject the notion that they could possibly be perceived as ethically neutral:

\begin{quote}
I don’t think there is such a thing as ethical neutrality \ldots because I think \ldots we are forced to make judgments and decisions that, whether we like it or not, have a moral dimension.
\end{quote}

\begin{quote}
You cannot be ethically neutral. No you can’t because you decide every day and ultimately people start to understand. You decide \ldots what disciplinary action you’re going to take because someone else did not act ethically and everyone’s in the room when you make that decision. So how can you be ethically neutral? You decide.\footnote{Linda Kebe Treviño et al., A Qualitative Investigation of Perceived Executive Ethical Leadership: Perceptions From Inside and Outside the Executive Suite, 56 HUM. REL. 5, 25-26 (2003).}
\end{quote}

Although the last quote states that “everyone’s in the room,” the reality is that there is only a discrete group of individuals in the room when a CEO is making the tough decision about how to discipline unethical conduct. Frequently, only other executives—and certainly not rank-and-file employees—are present. In fact, employees in most organizations will never learn about disciplinary action taken because such actions are considered private personnel matters. As a consequence, senior executives must become much more
sensitive to the view from the bottom of the organization and the fact that, to be perceived as an ethical leader, their communications and actions must speak loudly about the importance that everyone in the organization behave ethically.

IV. CEOs OUT OF TOUCH

We have now established that the CEO influences the characteristics of formal ethics and compliance programs and the ethical culture of the firm through ethical leadership, and that both of these significantly impact employee attitudes and behaviors. These findings are consistent with the general understanding that CEOs set the “ethical tone at the top.” So why aren’t CEOs more directly involved in the management of ethics in their organizations? Research suggests some preliminary answers that help us understand that what one sees and knows is determined by where one sits in the organization. First, because of their inclination to identify closely with the organization and its image, top managers have a “rosier” view of their organization’s ethical climate than do lower-level employees. Further, due to fear and futility concerns, employees are unlikely to report ethical problems up the chain. As a result, CEOs are unlikely to know about ethical problems in their organizations. Finally, because most CEOs interact primarily with others of high status, they are likely to be out of touch with the daily realities of their own organizations and employees, including the ethical climate.

Because their own identities are tied to the organization’s identity and image, employees tend selectively to perceive the good, ethical side of their organizations more readily than the bad, unethical side. But, as the most senior leader of the organization, CEOs’ personal identities can be expected to be linked even more closely with the identity and image of their organization than are the identities of average employees. The CEO is intensely involved with the organization and its interests, represents the organization to the outside world, and serves as the organization’s agent with

37 Kimberly D. Elsbach & Roderick M. Kramer, Members’ Responses To Organizational Images and Member Identification, 39 ADMIN. SCI. Q. 442, 447 (1996).
multiple stakeholders. Therefore, there is a greater tendency for CEOs to perceive their organization in a positive light compared with lower-level employees. In fact, research has found that senior managers have significantly more positive perceptions of organizational ethics when compared to rank-and-file employees. Senior managers are less likely to see ethics initiatives cynically and are more likely to perceive the internal ethical environment to be supportive of ethical conduct in the organization. They are also more likely to believe that employees will raise ethical issues and report ethical problems to management.

Despite their powerful role and place at the apex of the organization, many CEOs base their perceptions of their own organization on highly limited information. Normal organizational communication processes can insulate senior managers from negative perceptions of the organization, keeping them out of touch with lower-level employees on matters of organizational ethics. Research on information processing and upward communication in organizations suggests that senior managers may be somewhat naïve about and protected from the realities of organizational ethics. Upward communication in organizations is frequently filtered and distorted, with information gaps growing larger as the number of intervening hierarchical levels increases. Research on voice and silence in organizations also suggests that important information, especially negative information, is often withheld from executives. Employees are hesitant to relay unfavorable information up the organizational hierarchy because they fear retaliation or because they believe such efforts to be futile. Thus, accurate information, especially

---


39 Linda Klebe Treviño et al., Comparing Senior Managers’ and Employees’ Perceptions of Organizational Ethics, in ACADEMY OF MANAGEMENT ANNUAL MEETING BEST PAPER PROCEEDINGS (2000).


information about organizational problems, is unlikely to find its way up through multiple organizational layers from lower-level employees to senior managers. As a result, many CEOs simply do not have their fingers on the ethical “pulse” of their organizations.

Finally, executives’ association patterns may contribute to this general lack of information from lower-level employees. Some executives choose to engage nearly exclusively in associations with high status communication partners (e.g., other CEOs and other elites) while minimizing associations with parties of low status, especially lower-level employees.\(^4\) Such an interaction pattern influences the information that executives have available because executives’ interactions are an important source of social information that influences their interpretations and decision making.\(^4\) In order to access unfiltered information about the ethical climate and culture, executives must reach out directly and regularly to rank-and-file employees. If they do not, the information they have available about the ethical climate will be highly limited and they may miss information about ethical breaches. Executives can only improve their access to important information from lower-level employees by finding ways to interact with them directly and in regular two-way communication about ethics. Such efforts are more likely to encourage honest input to the CEO about the organization’s ethical climate and culture.

V. Conclusion

CEOs contribute to corporate misbehavior in a number of ways. First, CEO commitment to ethics has a powerful impact on the scope, orientation, and integration of formal ethics programs. But many, if not most, CEOs delegate

---

\(^4\) G. Chen et al., CEO Elitist Association (2005) (unpublished manuscript, on file with author).

responsibility for these formal programs to other executives with whom they rarely interact. Second, through their ethical leadership, CEOs are essential to the development and maintenance of a strong ethical culture and climate in the organization. But, again, many CEOs devote too few corporate or personal resources to this effort. Many are more concerned with “checking off” requirements of the U.S. Sentencing Guidelines or the Sarbanes-Oxley Act. Finally, many CEOs are out of touch with the ethical realities of daily life in their organizations because of their close personal identification with the organization, typical organizational communication patterns that block information flow, and their own interaction choices that limit the availability of information from lower-level employees. CEOs who wish to contribute to “good” corporate behavior must commit to being the firm’s Chief Ethics Officer, recognizing the importance of a strong reputation for ethical leadership, and taking responsibility for the development and maintenance of a solid ethical culture.
Discussing Corporate Misbehavior

THE CONFLICTING NORMS OF MARKET, AGENCY, PROFIT AND LOYALTY

Daniel J. H. Greenwood

I. INTRODUCTION

Corporate law remains thin, but corporate law scholarship is thickening. This Symposium is both a symbol of and a major contribution to that process. We are stepping beyond the narrow models of rationally maximizing fictional shareholders and purely self-interested managers competing in an evolutionarily determined and purely individualistic market inevitably maximizing social wealth through the pursuit of private profit. Instead, new scholarship is taking a richer perspective infused with the insights of group and individual psychology, recognitions of institutional realities, and broader conceptions of the social good.

American corporate law restricts itself to a limited view of the public corporation. In state corporate law, a corporation consists of little more than directors and shares,1 with the

---

1 © 2005 Daniel J. H. Greenwood. All Rights Reserved.
1 Professor of Law, S. J. Quinney College of Law, University of Utah. My deepest thanks to Jim Fanto, Larry Solan, Brooklyn Law School and the Sloane Foundation for organizing and making possible this fascinating Symposium and to the participants and Leslie Francis, Daniel Medwed and Manuel Utset for helpful comments.
1 I use the term “share” rather than the more common “shareholder” because corporate law and scholarship alike normally ignore the portfolios and people who own the shares (i.e., the shareholders), instead focusing on a purely imaginary creature with no views, interests or desires other than maximizing the value of the particular corporate stock at issue. In corporate law, then, the term “shareholder” while misleadingly invoking images of a human being, actually refers only to a role. See generally Daniel J. H. Greenwood, Fictional Shareholders: “For Whom is the
occasional cameo appearance of creditors of a firm near bankruptcy, or managers as the secret doppelgangers of the inside directors. The issues of central concern to the law are similarly restricted: the formal voting rights of shares, the ultimate power of the directors to manage the corporation and the limited exception granted to shares to sue derivatively, the directors’ limited fiduciary duties to the corporation and its shares, and some cameo appearances of other legal values when shares and directors are at odds over takeovers. Even in these areas, corporate law is famously “enabling,” “towering skyscrapers of rusted girders, internally welded together and containing nothing but wind.”

When corporate law has entered the normative thicket, it has usually been to enforce the thin view of its purposes: to define shareholder interests as the interests of the role, rather than the human beings who inhabit it, and to force managers to restrict their view of the corporation’s interests to those of these legally constructed fictional shareholders. ERISA and the fiduciary and agency rules regulating the decision-makers for most institutional shareholders (that is, the holders of most shares), often require them to act as if their only concern were maximizing returns to undiversified shareholding in the particular corporation. Moreover, corporate law gives directors and shares the right to sell corporate control without consent of other corporate constituencies. Combined with the anonymous market for publicly traded stock, this creates vast market pressure to run the firm in the manner most likely to be rewarded by the stock market. And (at least since the demise of the conglomerate fad of the 1960s) the stock market has generally bid up the stock prices of corporations that

Corporation Managed,” Revisited, 69 S. Cal. L. Rev. 1021 (1996). Moreover, voting in a corporation is on the basis of one share one vote, rather than one shareholder one vote; here too the term “shareholder” tends to mislead, giving the appearance of democracy where there is at best plutocracy.


3 See Dodge v. Ford Motor Co., 170 N.W. 668 (Mich. 1919) (declaring that the purpose of the corporation is to earn returns for its shares, regardless of the expressed views of the majority shareholder); State ex rel. Pillsbury v. Honeywell, Inc. 191 N.W. 2d 406 (Minn. 1971) (declaring illegitimate attempt of shareholder to cause corporation to consider values other than profit-maximization); Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173 (Del. 1985) (requiring managers, under limited circumstances, to pretend that shareholders have no interests other than the value of their shares, and to run the corporation in those legally defined interests). More often, however, corporate law’s business judgment rule removes the issue of corporate purpose from judicial purview.
demonstrate a decent respect for the opinions of institutional shareholders and show a keen focus on identifying corporate interests with stock market interests.

Other areas of the law regulate other aspects of the public corporation, but generally without consideration of the specific characteristics of corporations as such. Thus, securities law, in general, protects securities holders as outsiders, consumers of a product produced by the corporation, creating rights to information in the manner of a truth-in-packaging law. Environmental law, constitutional law, criminal law, labor law and so on, generally regulate the corporation as a “person,” ignoring its collective and corporate character and subjecting it to norms created for citizens without much consideration of special issues of organizational behavior.

But public corporations are not individuals. They are large bureaucratic organizations, no more likely to respect individual rights or needs than the large bureaucratic organizations of the state, and generally a good deal less responsive to the views of the majority or those they affect. General American law starts from a basic distinction between state and citizen, public and private—but we have placed our large public business corporations on the private, citizen side of the divide, as if we needed rights for them rather than rights against them, or as if we existed for their sake rather than they for ours. Since the fundamental foundation of the individualist liberal political theory on which our polity is based is the recognition that groups and organizations often do not act as the individuals in them would like them to, this conflation of corporations with people is strange, to say the least.

---

4 Even the Williams Act, which edges into regulation of the firm’s internal decision-making processes while retaining the form of consumer protection by disclosure regulation, strikingly avoids substantive discussion of how, by and for whom corporations are run.

5 The corporate income tax scheme also treats the corporation as a person or entity, taxing the corporation’s income just as it taxes any other person’s income. The current attack on so-called “double taxation” however, seems to be premised on the claim that the corporation can be reduced completely to its shareholders, so that its income is theirs even though they have no right to receive or control it. This revisionist view rejects both the person/entity view most commonly seen in regulatory statutes in favor of corporate law’s currently dominant view of the firm as a “moment in the market” with no institutional existence. As should be apparent, neither of these views are compatible with a perspective that takes the institutional reality of the firm seriously. A serious discussion of taxation of public corporations would have to begin by discarding rhetorical claims that the corporation is a person like any other or doesn’t exist and instead to confront the near impossibility of determining the actual impact of corporate income tax on corporate participants in increased prices or reduced payments (salary, dividend, interest or prices).
Corporate law scholarship in the last quarter of the millennium both celebrated the limited protection the law offers and narrowed the scope of its own concerns even further.

Its psychological theory began, and usually ended, with a model of rational profit maximizers borrowed from neo-classical economics. Shareholders were modeled as one-sided fictions with no interests or values other than increasing the value of the stock they hold in a single corporation—as if they were undiversified aliens or colonial occupiers with no interest in the society they sought to exploit. Directors and managers were reduced to self-interested cynics who must be coerced or paid obscene amounts to do their jobs or observe minimal professional norms. Other employees simply disappeared from view altogether, except perhaps as tools to be maximally exploited.

Its normative concerns sometimes seemed restricted to no more than a debate between advocates of idolatry—viewing every idiosyncrasy of legally regulated markets as sacred—and plutocracy—seeking to ensure, market or no, that an ever increasing slice of the corporate pie was served to the capital markets.

Sociologically, in the leading models, the firm itself seemed at times to lose its corporeality, as “nexus of contracts” models made the institution invisible, no more than a collection of individuals meeting in a collection of moments to exchange and then depart again into the “woods of America . . . perfectly in a state of nature.”

In the last decade, however, we have moved into a new era. Corporate law scholars have begun to recognize that we must take into account the learning of real psychology, that we must understand group interactions outside the narrow bounds of neo-classical individualism, that the organization has

---

6 John Locke, Two Treatises of Government 295 (Peter Laslett ed., Cambridge Univ. Press, 1988) (1689). Locke here follows Thomas Hobbes’ description of the state of mutual disinterestedness—war—between individuals in America and anywhere else “where there were no common Power to feare.” Thomas Hobbes, Leviathan 65 (Prometheus 1988) (1651). Hobbes’ portrait of man in the state of nature is closely related to the usual models of motivation used in classic late twentieth century corporate law. As he describes it, in the absence of relationship or government: To this warre of every man against every man, this also is consequent; that nothing can be Unjust. The notions of Right and Wrong, Justice and Injustice have there no place. Where there is no common Power, there is no Law: where no Law, no Injustice. Force, and Fraud, are in warre the two Cardinall vertues.

Id. at 66
behaviors and meanings that can no more be reduced to the individuals in it than government can be reduced to the governed, that ethical issues may be more complex than simply not stealing from shareholders (and stealing as much as possible for them).

The essay proceeds as follows. Part II uses Professor Linda Treviño’s contribution to this Symposium, which details the important ways in which a corporate CEO influences the rule-abidingness of his subordinates, as a reminder that corporate law must consider the sociological institution of the firm, not merely the limited roles which we usually emphasize. Part III explores Professor John Darley’s contention that people in firms seem unduly “recrutable into corrupt practices” and discusses how people in corporate roles often seem to conclude that the right thing is the wrong thing to do. Part IV takes the idea of role based norms one step further: The corporation, I contend, functions at the intersection of radically different market and fiduciary norms and is inherently a locus of normative conflict. Thus, corporate wrongdoing is as likely to result from the wrong norm in the wrong place (including but not limited to the team spirit, internal culture and loyalty that Professors Treviño and Darley discuss) as from selfishness, corruption or other forms of explicitly bad behavior. Part V discusses the ways in which corporations can mediate these normatively conflicts, successfully or not. Part VI concludes with some preliminary suggestions for further research and law reform.

II. THE ROLE OF THE CEO

Professor Treviño and others have demonstrated that CEO behavior critically affects firm ethics. When the CEO acts ethically and creates institutional structures that make clear to other employees that the firm values ethical behavior, the firm acts more ethically.

Over the course of the last three decades, CEO pay has risen astronomically in the United States, from perhaps 24 times the average employee’s pay in 1965 to 300 times by the turn of the century. With our CEOs now paid so much more

---

than the rest of us (and even than comparable CEOs in other countries"), a cottage industry has sprung up defending the high compensation. Institutional shareholders were persuaded by agency theory arguments that high, stock-based pay for CEOs would cause higher CEO productivity: By tying CEO incentives to share price performance, high pay would lessen CEO shirking, thereby increasing returns for shares. After the pay increases, the standard neo-classical model of wage setting for employees provided an equal and opposite justification: A profit-maximizing company would hire someone only if their pay is lower than their contribution to the firm’s profit. On the “best of all possible worlds” view of this perfect market theory, it seems to follow that high CEO pay must result from high CEO productivity. Combined, the theories contend that American CEOs are paid more than their foreign counterparts and their predecessors because they are doing a better job. But with high compensation should come high responsibility. All that pay must be for something. It has been

CEO compensation was 531 times the pay of the average blue collar worker. Another way to see the same phenomenon is that average hourly wages were up 10% in real terms between 1989 and 2000, while average CEO compensation increased 342% in that period. MISHEL, supra, at 113 tbl. 2.1, 213 tbl. 2.46.

American CEO pay is almost as unequal as American income generally. Accordingly, average CEO compensation is vastly higher than median CEO compensation (in 2000, $11.194 million vs. $3.101 million). MISHEL, supra, at 213. Median CEO compensation of $3.6 million in 2003 was 336 times the $10,712 a full time worker would earn at minimum wage ($5.15/hr, 8 hours/day, 260 days/year) and 101 times average 2002 wages of $35,424. MISHEL, supra, at 113.

8 Stabile, supra note 7, at 121 n.22; MISHEL, supra note 7, at 214, 215 tbl. 2.47 (stating that US CEOs are paid roughly three times as much as their counterparts in other developed countries).

9 See, e.g., Michael C. Jensen & Kevin J. Murphy, CEO Incentives— It’s Not How Much You Pay, But How, 68 HARV. BUS. REV. 138 (1990). Agency theory suggested that tying CEO pay to stock performance would make CEOs more entrepreneurial. In my view, the changes in CEO pay probably did increase shareholder gains, but the incentive theory seems unduly mechanical. Instead, high pay alone, as well as vastly increased stock holdings, has changed CEO views of the team for which they play. The modern CEO is far less likely to identify with the bureaucracy he heads and far more likely to identify with his peers heading other corporations. Moreover, as CEOs become increasingly wealthy and increasingly large stockholders (both in their own company and the market generally), they are less likely to think of themselves as professionals and more likely to think as investors or owners. These reorientations make it less likely for CEOs to identify the interest of the firm with the interest of its employees. They, however, also make it more likely for CEOs to view their role inside the firm cynically and, in some instances, to abandon team playing altogether. For further discussion, see Daniel J.H. Greenwood, Enronitis, 2004 COLUM. BUS. L. REV. 773, 802 (2004).

10 See, e.g., Randall S. Thomas, Explaining the International CEO Pay Gap, 57 VAND. L. REV. 1171, 1172 (2004) (noting that US executives are paid more than their foreign counterparts and explaining it as reflecting greater productivity).
startling, then, to see the alacrity with which the CEOs on the “perp walk” (as Professor Treviño terms it) have adopted a defense based on ignorance. Apparently, we are supposed to believe that CEOs earn their enormous compensation through their ability to make the organization more productive, while simultaneously accepting that they really have no idea what is going on in their organization, even at the basic level of whether the organization is actually producing or just faking it. On this view, ignorance is a defense, even for those whose business is knowledge. As in My Lai or Abu Ghraib, the upper echelons deny any connection to crimes committed by their underlings: Those designing institutions insist that they should be entitled to assume that those below them will act appropriately in all circumstances regardless of institutional pressures, temptations or norms.

Professor Treviño shows that, at least in the corporate world, we should just say no. Whether CEOs like it or not (and whether or not it fits into the thin theories of corporate behavior based on rational maximization), institutional ethics are largely under CEO control. CEO behavior matters. CEO talk matters. CEO silence matters. Organizational structures implemented by CEOs matter.

Professor Treviño maintains that most corporate actors are not fully autonomous rule-following ethicists unaffected by the norms and expectations of those around them. Instead, like most adults, they tend to conform their behavior to the norms they believe are expected of them. Since CEOs are instrumental in creating those expectations and the institutions that back them up, they are instrumental in determining how employees will behave.

---

11 For a recent corporate law discussion of this issue, see In re Caremark International, Inc. Derivative Litigation, 698 A.2d 959, 970-71 (Del. Ch. 1996). Caremark limited the old Delaware rule, set out in Graham v. Allis-Chalmers Mfg. Co, 188 A.2d 125, 130 (Del. 1963), that "absent cause for suspicion there is no duty upon the directors to install and operate a corporate system of espionage to ferret out wrongdoing which they have no reason to suspect exists," Graham, 188 A.2d at 130, holding that directors have a “duty to attempt in good faith to assure that a corporate information and reporting system, which the board concludes is adequate, exists” In re Caremark, 698 A.2d at 970, and that liability may be found if directors “utter[ly] fail[] to attempt to assure” such a system exists. Id. at 971 By focusing on “information and reporting,” id., Caremark continues to operate on a background assumption that wrongdoing occurs independent of corporate structure or internal corporate norms. Note however that Caremark involves board liability. Boards might reasonably be expected to have less responsibility for the firm’s behavior than CEOs, who unlike directors are full-time employees of the firm with primary responsibility for creating and operating its structures.
Corporate law generally treats corporate ethics as outside of its scope, as if the structures created and regulated by corporate law had no significant influence on ethical behavior or as if those structures were outside the realm of corporate law. Companies and their employees may follow the rules or they may not. For corporate law (like criminal law), the firm is black box. These areas of the law look just to the results—if things do not work out, the firm will not be competitive or the regulators will stop it. But Professor Treviño’s work, like the other work presented at this Symposium, suggests that we needn’t be so agnostic. We know how to influence the degree of corporate ethical behavior, and we could mandate better processes than the ones we allow.

III. CORPORATIONS COMPOSED OF SHARES VS. CORPORATIONS COMPOSED OF PEOPLE

Corporate law, in my view, generally regards the corporation as a commonwealth of dollars. Shares, each representing an equal equity investment in the firm, are the citizens of this polity; it is they and they alone who are granted the political right to vote, entitled to demand equality, or barred from immigration and emigration without consent of the whole. Shares and shares alone are given the right to invoke the assistance of the law to insist that the firm consider their interests; in extreme versions of corporate law ideology (and in Revlon mode\textsuperscript{12} as a matter of corporate law) only the interests of shares ought to be considered.

In this picture, even the shareholders, in their full humanity as pensioners, employees, CEOs or progeny of the robber baron elite, citizens or aliens, parents and children, and holders of various religious, ethical or aesthetic values, ultimately disappear. The people who are the corporation in the ordinary course—employees—never appear in the first place. They are the concern of some other area of law.

When corporations go bad, however, the firm itself comes back into focus, not as a evanescent nexus of contracts or as a commonwealth of shares, but as a sociological entity—a group of people—with actions and values that cannot be reduced to those of its component parts.

\textsuperscript{12} Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 182 (Del. 1985) (creating duty to maximize share value in narrow circumstances, when the company’s sale has become inevitable).
This is where Professor Darley starts: From the firm as a group of people working together, not a pot of money managed on behalf of its “owners.” His puzzle is that people in firms seem “more recruitable into corrupt practices” than one would anticipate. His explanation is a combination of normative drift, group loyalty and social roles that value “playing rough.”

The normative drift argument is a variant on the famous claim that a frog won’t jump out of a pot if you heat it gently enough. Because it can only identify incremental change rather than absolute states, it feels perfectly comfortable until it cooks. Here, the argument is that many scandals begin with a small deviation and proceed in small increments, so that at each point the participants can see the next step as merely an insignificant addition to a commitment already made. They start out with the typical lawyer’s rationalization—if “x” is permissible, then “almost x” must also be, since the two are so close as to be indistinguishable. By the point this rationalization no longer works, participants shift to the criminal’s commitment: I’m in so deep already that a little more won’t hurt.

This story is important, especially for law students, since it is often the lawyer’s role to turn the heat up gently by pressing the interpretative limits of regulatory norms—but also to remember that when you press far enough, even in increments that are each justifiable, you generally end up cooked. Too many clever arguments without enough grounding in extra-professional norms ultimately lead to scandal, corporate collapse or even jail.

But it is on the latter explanations that I want to focus here. Here, Professor Darley is telling us not about people who have lost their normative way but about specific choices that seem correct to them, but wrong to more detached observers. Sometimes, as he points out, they are even consciously understood as moral dilemmas, not invisible at all, but great and traumatic, in which the employee must choose between loyalties. Then, corporate criminality partakes of classic tragedy. Pity the person who must choose between friend and country; only doom can result.

IV. THE CORPORATION’S CONFLICTING NORMS

As a lawyer rather than a social psychologist or student of management, my own concerns are centrally normative.
Corporations exist on the border between two conflicting sets of norms, and the issues raised by this Symposium can be seen as resulting from the conflicts between those norms. The problem, or at least part of it, is not violation of norms so much as their inappropriate application, following the right norm at the wrong time or in the wrong place. The normative conflict is irresolvable, but by highlighting it we can work towards a better mediation of our contradictory normative intuitions. Conventional corporate law scholarship, however, has largely concealed the conflict.

On the one hand, we have the market, governed by contract law. The basic norms are of John Locke’s state of nature and Adam Smith’s market: disinterested strangers treating each other as means to their own ends, not as Kantian ends in themselves. The principles of loving your neighbor—or even your child—as yourself are out of place here; a marketplace can’t function if the bargaining parties view their opposites as parts of themselves. This is the world of every man for himself, not one for all and all for one.

To be sure, trade and the division of labor ultimately result in more stuff for everyone, and cooperation is usually the only way to achieve private aims. However, any given bargain is inherently competitive: The more you get, even of the gains from cooperation, the less I do. An invisible hand may assure that my selfishness works to the common good (at least under the right conditions), but in the norms of this role, the common good, or indeed your personal happiness, is important to me only to the extent that it makes me more likely to get what I want.

Even norms as seemingly fundamental to the market as honesty are justified only in the self-centered utilitarian

\[13\text{Hobbes, of course, is the locus classicus of the fully self-centered psychology. I choose Smith and Locke precisely because these authors have a clearer sense of the common good and allow for human sympathy, and yet still model a state of nature characterized largely by mutually disinterested (if not necessarily hostile) people. See Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations (Univ. of Chicago Press 1976) (1776); Locke, supra note 6. Rawls, for all his emphasis on the Kantian view, describes parties behind the veil of ignorance in a similar fashion, as disinterested rather than mutually concerned. Respect for others is a constraint, not a consequence of love or community. This disinterestedness, which to me seems implicit in and necessary to the contract view that individuals join society to gain benefits for themselves, is central to Rawls's description of the original position and, therefore, undermines any commitment to treating others as ultimate ends. Were the bargaining parties allowed true mutual concern (or even envy, a closely related emotion), it seems likely that they would reject the difference principle as insufficiently egalitarian. See John Rawls, A Theory of Justice (1999).} \]
calculus of personal interest. Good bargainers never tell the whole truth about their intentions or concern themselves with the needs or expectations of the other side, except as tools to their own ends. The person who spots a value that others have missed and takes advantage of it—buying low, selling high—is an entrepreneurial hero, not a deceitful cheat.

The market is a world of symmetrical autonomy. In the market, we, or at least our dollars, are all the same in relevant part, all equal and all presumed to be able to take care of ourselves. Markets presume that their participants are self-sufficient adults, aware of what they want, able to balance their needs, desires and capacities, negotiating in their own interest from a position of reasonable independence. Equally important, markets price products, not status or person. Equal products ought to command equal prices; price discrimination is presumptively improper. Money is green regardless of pedigree; opportunities should be taken regardless of tradition; cooperation is bought and sold. The anonymous stock market—where buyers and sellers never even learn the other’s identity—is the paradigm; the ideal of blind meritocracy its extension. 14 The market normative system is radically disrespectful of persons, status and relationship.

In the market, obligation stems mainly from contractual promise, and extends only so far as the promise did. This is not a world of solidarity or natural obligations. Your humanity or fellow citizenship has only a limited negative claim on me, that I not violate your individuality by violence or fraud, narrowly understood. Market participants are free, of course, to care for others, but to do so in public is nepotism or favoritism—a violation of market norms even when not illegal. In the public sphere, if you take the needs of your bargaining counterpart as values in themselves rather than tools to be used to overcome him, you are either discriminating or just a mark, a fool or a sucker. The goal of arms-length bargaining is to get as much as you can while giving as little as you can, limited only by the rules of the game.

But if firms hire, fire, buy and sell in the world of the market, they produce in a different arena altogether. As Coase

14 For further discussion, see, Daniel J. H. Greenwood, Beyond the Counter-Majoritarian Difficulty: Judicial Decision-Making in a Polynomic World, 53 Rutgers L. Rev. 781, 813-15 (2001) (discussing the difference between market and majoritarian decision-making in a democracy).
pointed out, were firms just markets, they would not exist. Markets are better at being markets. Firms exist because they can do things that markets cannot do well. In particular, they replace disinterested contracting with cooperative planning.

Inside the firm, market norms disappear. Instead, the fundamental legal norm governing employees is agency, with its strongly asymmetrical fiduciary obligations.

An agent is required to work for her principal. Instead of maximizing what she can get from a stranger (and relying on an invisible hand to turn this seemingly selfish behavior into something socially useful), she must take the principal's ends as her own. Much as a mother is better off when her child is better off (without a contractual expectation of payback) or a patriot wins when the nation wins (even if he is killed in the process), the norms of agency demand that the employee see the employer's good as her own. She must put aside her self-interest (in the market, contractual sense) and concern herself only with the interests of the other.

Agency is profoundly opposed to the symmetrical anonymity and mutual indifference of market's strangers. The market-contract ideal is of blindness to personal characteristics and relationships—a common carrier open to all comers, an anonymous stock market, a medieval fair in which ancestral enemies meet momentarily to trade, charging the same price to lords and peasants alike, or a Weberian bureaucracy promoting and deciding on merit alone. In contrast, agents must always remember who is who. To her principal, the agent owes a fiduciary duty; that duty requires her to work for the principal and against (in the market sense) all others. Employees work for employers, not the other way around: The principal gives orders and the agent takes them; the employee must set aside her own interests but the employer need not. Inside the organization, you obey or you cooperate; with outsiders you compete. Relationships and status are essential. Until you know who is agent and who principal, who boss and who subordinate, who insider and who outsider, you know nothing at all. The market ideals of anonymity, autonomy and equality have no place in an agency relationship.

---

This world of agency is defined by group loyalty and team spirit. The firm should be a team, with each agent working for the good of the whole (otherwise, it is hard to see how it could out-compete a market). Instead of the morality of independence, equality and autonomy, here we have the morality of self-abnegation, sacrifice, caring and unity. We work together, to promote our collective good (against the others). This is a world not of strangers but of family-like groups, in which each person’s actions (and, as Professor Darley points out, self-understanding) is interrelated with others’.

We cannot privilege one value system over the other. Nor can we live by either alone. Market disinterestedness underpins bureaucratic regularity and offers liberation from feudal oppression, but threatens to descend into the Hobbesian war of all against all exemplified by the free market of post-communist Moscow or civil war Beirut. Agency status and mutual concern is the foundation of patriotism and social justice—as well as nationalist riots, mob violence, prejudice and group-think. Capitalist affluence and democratic liberal freedom alike depend on both existing and each restrained.

Professor Darley points out the “alternate identities on offer” in our business corporations, using the example of Michael Lewis’s experiences as a bond trader. Lewis was shocked to discover that in his employer's view, his client was the bond desk, not the customer. To modify Professor Darley’s terminology slightly, Lewis was confronted with a choice of roles and the morals associated with them. According to the bond desk culture he confronted, he was a member of a team. The team was the company, and the team contended that team loyalties trumped legal responsibilities (or alternative views of who he should have viewed as his team).

Note, though, that the issue here is not selfishness or greed (at least not Lewis’s). Lewis was being asked to be loyal and self-sacrificing—but to the company, not his client. The ethical breach was not normlessness, but the wrong norm; not selfishness, but the wrong team. The alternative identities available include not just the Commedia del’Arte trickster that Professor Darley discusses, but the market hard bargainer, the capitalist entrepreneur, the team player and the loyal professional.

As Lewis’s story demonstrates, when values slip out of their proper spheres or when the spheres overlap, scandal can result. In Lewis’s case, the issue was how to define his
fiduciary duties: Was he meant to be working for his employer or his client, should his loyalty to his (local) teammates trump loyalty to his client or the more abstract requirements of law?

Often, however, the problem is not defining the team or limiting its demands. Many corporate problems stem simply from the conflict between the fiduciary norms of agency within the firm, which demand self-sacrifice in support of the team, and the market norms outside it, which demand self-interestedness. The celebration of unrestrained market in American political ideology of the last couple of decades has accentuated the conflict. Thus, a CEO who bargains for as many stock option grants as he can get or who sells his stock when the market is up is acting according to market norms. When he announces that the company is doing well when in fact it is not, he is bluffing in the way that contractual bargainers regularly do. When he fires employees for no reason other than maximizing corporate profit, or treats their pensions not as a social good but as a cost to be cut in any way legally permissible, or encourages his traders to use the rules of a semi-deregulated market to generate private profits at the expense of the California public, he is acting quite properly if the proper norms to follow are those of the disinterested market. In a market, strangers may be exploited to the maximum extent permitted by law. Indeed, if he cooks the company’s books to make it appear to be worth more than it is, he may be acting as a faithful agent, putting aside his own interests (and integrity) in order to aid the firm and its shareholders.¹⁶

In these scandals, the problem is not lawless selfishness, corruption or poorly socialized sociopaths. The evil is not normlessness but the wrong norm in the wrong place at the wrong time. It is a complex problem that will require complex analysis and complex solutions.

V. MEDIATING THE CONFLICTING NORMS

Combining Professor Treviño’s contribution with Professor Darley’s, we can take the point further. Corporate

¹⁶ The court in *Kamin v. American Express* accepted precisely this explanation of a corporate manager’s decision to account for a transaction in a way that cost the company real money but made its profit appear greater, and therefore, if the stock market was deceived, would keep its stock price higher. See 383 N.Y.S. 2d 807 (N.Y. Sup. Ct. 1976).
wrongdoing often results from a conflict of norms, rather than a simple refusal to act ethically. In these cases, the problem is not “bad apples” rejecting the norms they are expected to apply, but institutions imposing inappropriate norms (as in Lewis’s case) or a corporate actor making what we—after the fact—view as inappropriate choices among conflicting normative demands each of which requires setting aside the employee’s own interests or desires. Importantly, when this happens, actors may experience their wrong-doing as self-sacrificing, ethical and externally constrained, not as an act of selfish evil at all.

To be sure, sometimes bad apple explanations are correct. Employees sometimes place their own self-interest above their obligations to others—Bering Bank failed when a trader made a bad trade and then covered it up rather than admit to his mistake, and many of the top executives who falsified their books profited directly from the artificially high stock prices resulting from their lies.

But the articles by Professor Treviño and Professor Darley suggest that this type of corruption—variants on embezzlement—is not the most useful paradigm for corporate wrongdoing. Without ignoring outright criminals, we should focus our primary attention on ordinary people caught in a web of conflicting norms—norms that are mediated, explained and ultimately enforced by corporate structures under the control of the CEO.

The contradictory norms are fundamental. First, we operate under a background regime that glorifies self-help, even at the expense of others. The basic market norm encourages hard bargaining, thinking of yourself and your needs without any consideration of others except as tools to your own ends: that is what we call arms-length bargaining, and success at it is success in the business world. Competing hard is a good thing, not a bad one.

Of course, at the same time that we glorify strong competitors, we expect them to observe certain limits. Athletes should play fair, not beat up their opponents or take performance-enhancing drugs. Businessmen should create better products for less, not lie about the product, falsify their books, or shake-down competitors. These rules act as constraints to a game otherwise structured by a different set of norms.

Predictably, a certain number of people will fail to observe the limits, getting so caught up in the game that they
forget that winning is not everything. Basketball players foul, seemingly patriotic commanders condone or order torture, publicists and advertisers spin, and accountants stretch. But they do so to benefit the team, not necessarily to benefit themselves. Perhaps they are “bad apples” in the sense that they have lost their sense of where playing hard meets playing fair, but more importantly, they may also be altruistic team players—they are cheating for the glory of the team, not themselves. Perhaps this is why we often seem so conflicted about punishing them.

Second, corporations create their own internal norms of accepted and expected behavior. To be a professional requires, in most instances, putting aside your own beliefs and adopting the goals of your client. Agency law principles (and their ordinary morality equivalents) usually say the same thing for employees: An employee, on the job, acts ethically and properly if she puts aside her values and adopts her employer’s. Enron famously took these potentially conflicting rules to the paradoxical extreme, telling its employees that the way to be team players—to work for the firm—was to compete as hard and ruthlessly as possible—for themselves as individuals. 17

Corporations, like other groups and bureaucracies, can be structured to maximize group solidarity or not. In this age of the imperial CEO, the corporate world seems especially intent on creating solidarity in the work force. Managers train in the techniques of creating team spirit, team norms, team ethics and team loyalty. Whether by Professor Tom Tyler’s procedural justice, tent-revival meeting style hortatory, or simple “us against them” competition, well-run firms work to create and maintain norms of employee loyalty to the firm.

Group loyalty has well-known problems—most importantly, for our purposes, group-think. Tightly knit groups tend to develop their own internal norms, as Professor Darley points out, with most group members recreating themselves to fit their understanding of the demands of the group. But a group that adopts a uniform and closed analysis of the world will be poorly equipped to deal with changing external challenges. When everyone thinks the same way, no one will see the errors in the standard thinking, whether the problem is

17 Enron was famous for its harsh intra-firm competition, in which employees were regularly required to rank each other, with the winners receiving bonuses and promotions and the losers being fired.
false analysis of the external world (seeing weapons of mass destruction where there are none, or not seeing competitive threats where they are clear) or normative drift.

With the collapse of the unions, the main external countervailing force to corporate loyalty is the corrosive individualism of the market itself. Employees are not only members of the team influenced by norms of team spirit and mutual responsibility, but also potentially out for themselves in a market governed by self-interest. Thus, market norms suggest, as the Wall Street Journal's career management column regularly does, that the intelligent employee will always focus on creating the appearance of team loyalty without succumbing to its reality, and, as Dilbert teaches each day, a cynical distrust of anything “the Boss” might say. These market norms limit the power of group loyalty and group-think, but may not do so in particularly useful ways. Cynical self-interestedness leads to conformity, not to the sort of brave dissent that functional institutions need.

One aspect of corporate wrongdoing then, is the difficulty of mediating fundamentally contradictory norms. We demand that corporate agents cause the corporation to compete hard, treat the people they contract with at arms length as tools to the end of profit, set aside their personal political and moral beliefs while on the job in order to do the work they are paid for. We demand that employees be team players and make clear that the team is the firm, regardless of the needs of the greater society. Paradoxically, that is, we demand that employees altruistically and selflessly serve—but the cause they are to serve is simply the self-interested firm, which is free to treat employees as competitive opponents rather than teammates. It is not surprising that some fail at this difficult game.

Moreover, one key way in which we identify ethical behavior in ourselves and others is self-denial. Kant claimed that the highest form of moral behavior is that which does not benefit the actor at all. I suspect he was wrong, but the kernel

---

18 The Kantian formulation, in my view, fails to capture the importance of other-directed team behavior, which is deeply satisfying to the actor. Mothers do sacrifice for their children as patriots do for their country, but the sacrifice is not the essence or the test of the morality of their actions. Rather than sacrificing our own good, within a team we treat the other as ourself, the good of any team member as our own good. Moreover, emphasizing sacrifice suggests, as I note in the text, that selfishness is the main cause of evil, missing the important point that each form of good self-sacrifice has an accompanying evil form. People are often happy to sacrifice
of truth is clear. The clearest instance of unethical and unprofessional behavior is pure self-interested selfishness: stealing from the client.

This agency understanding that the ethical thing to do is to set aside your own views and instead adopt the profit-maximizing norm of the corporation is so strong that it may cause decision-makers to act contrary to their views as citizens even when profit is not at stake. It is often difficult to tell what the right thing to do is and even more often difficult to tell what the profitable thing to do is. But the Kantian understanding that “morality hurts” allows a quick (if often misleading) heuristic. The one time that I know I am acting as a good agent, setting aside my own views in the interest of the team, is when I do something I know that I (in my citizen role) would disapprove of. The easiest way to show that I am acting properly is to act improperly. And in the corporate context, this means I should do what is profitable, regardless of whether it is socially useful.

Moreover, I believe the heuristic is commonly taken one step further. Often I may be unsure whether something is profitable, but confident that it is wrong. The agency understanding powerfully (if illogically) suggests that I should pick the wrong action. Virtue hurts; agents are supposed to set aside their own beliefs. By picking something I know I would disapprove of in my private role, I know I am acting as an agent, even if profit remains elusive.

It may be hard to prove whether the strictures of the corporate social responsibility movement increase or decrease long term profit. But it is easy to see that as citizens, we will find many of them attractive. Rejecting them, even in the absence of any actual evidence regarding their costs or benefits to the firm, is an easy and psychologically clear way of proving that the decision-maker is acting in role, as a team member. Perversely, then, actors attempting to do the right thing (in their role as agents) may end up making decisions they know are wrong (in their role as citizens) and which do not even maximize profit.

Team players act on behalf of the team, not themselves. When the subcontractors described by Professor Darley decided to falsify test results, they knew they were risking their

themselves for evil causes, as every student of the virulent forms of nationalism, religion, fascism and communism knows.
DISCUSSING CORPORATE MISBEHAVIOR

2005] 1231

personal careers and freedom. They were setting aside their own interests in favor of the interests of a whole greater than themselves. Professor Darley describes this as adopting a social role and proposes various villains. But in a very simple and clearly understandable sense, the role adopted is simply the good and loyal team player. This is ethical behavior—self sacrifice for a cause. Good employees don’t shirk and they don’t tattle either.

The problem, then, is not (only) bad apples, selfishness or insufficiently socialized individuals who do not understand how to play fair. It is also morality in the cause of the firm—a type of local patriotism. Self-sacrifice in the cause of profit. Setting aside your own moral views to do what the job requires, distasteful as that may be. Enron’s traders viewed themselves as heroes, and on this level they were right: heroes fighting to oppress the ratepayers on behalf of the firm and its shareholders.

We have then this paradox. If the purpose of the corporation is to make a profit, as many have claimed, then the way that an employee fulfills her ethical obligation on the job is by promoting the corporation’s profit. But despite Milton Friedman’s famous claim that the ethical obligation of business is profit, placing profit above all is precisely the definition of unethical behavior in the corporate context. Ethical behavior in the corporate context means understanding that loyalty, honesty and relationships with your customers, your suppliers, your employers, your country, even the earth itself, are sometimes more important than short-term (or even long term) profits. Ethical behavior on the individual level means knowing when to buck the group norms, when to stand up for one normative system in opposition to another, when to violate the agency norm of setting aside your own sense of right and wrong and to selfishly follow your own lights. On the institutional level, ethical behavior requires building in safeguards against group-think, limits to the pursuit of short-term profit,

---

19 Even Friedman acknowledges that pursuit of profit must be restrained by external legal norms of “deception and fraud,” although he does not explain why these violations differ from the additional ethical obligations ones he rejects. “There is one and only one social responsibility of business—to use it resources and engage in activities designed to increase its profits so long as it stays within the rules of the game, which is to say, engages in open and free competition without deception or fraud.” Milton Friedman, The Social Responsibility of Business is to Increase its Profits, N.Y. TIMES, Sept. 13, 1970 (magazine).
commitments to particular relationships and focus on products and services rather than stock market prices.

When an entire institution goes bad, it is not the result of a “bad apple.” Rather, it is the result of employees or corporate agents doing exactly what they think they are supposed to do under the circumstances, even when they have strong misgivings about whether the actions are proper from a societal point of view. As Professor Darley notes, in several of the most famous cases of corporate wrongdoing, the employees who did the bad acts were perfectly aware that they were acting improperly according to social norms, and even worried about criminal liability. The point is, however, that as they took actions they knew were wrong—falsifying test results, falsifying financial data, distorting the California electric market to “steal” from Aunt Millie—these employees saw themselves as acting properly within their job. Much as soldiers learn to kill in the cause of the nation, Enron’s employees learned to pillage in the cause of the firm. The moral thing to do, at least so long as you have not resigned from your job, appeared to be to set aside individual morality and act in the interests of the team, to do what was necessary to win.

Team loyalty and internal corporate norms trumped national patriotism and national norms. Indeed, in the end, they even trumped profit: Profit, like happiness, is one of those things best achieved by aiming elsewhere. The exclusive emphasis on short term profit in each of the institutions caught up in the great turn of the century scandals is not accidental. In the end, a corporation can only out-compete markets by being good at things markets are not good at, and the most important of those is long term relationships. But short term profit demands treating all relationships instrumentally, and that, in turn, destroys them, and with them, the basis of future profits. Laser-like focus on profits, as Enron demonstrates, is fundamentally incompatible with actually earning them.

The issue, then, is corporate culture, and as Professor Treviño has pointed out, corporate culture starts at the top. CEOs have enormous influence on the way in which their corporations respond to these normative demands, through modeling as well as their decisions about how to structure the corporation, about what measures of success to emphasize, about how and who to promote or not, about how to resolve conflicts between the alternative norms and identities “on offer.” Even if corporate culture is never entirely under CEO control, the contributors to this Symposium make clear that in
fact CEOs do matter, that it is not enough to simply dismiss them either as rapacious kleptocrats or as absent minders unfairly held responsible for the wrongdoings of subordinates.

VI. CONCLUSION

From a regulatory perspective, several consequences follow from this discussion of normative conflict. First, we should not fall into the mistake of thinking that the problem is centrally one of sociopaths, or to be dealt with by criminal law. Soldiers kill but they are not murderers (unless they are soldiers for a particularly evil state). Many of our corporate wrongdoers, particularly at the lower levels, are not motivated by selfishness but by selflessness. This means that obtaining convictions under conventional criminal law is going to be difficult: Some of the worst malefactors will be able to demonstrate that they were not thinking of themselves as they did their destruction. In the manner Hannah Arendt called the “banality of evil,” they were good bureaucrats, doing their jobs as they understood them, according to the norms of the job. Criminal law looks for corruption—self-interest where loyalty is required—but the problem here may often be loyalty where rebellion was needed.

Relatedly, and more tentatively, it is time for further examination of whether criminal law and regulatory control of the corporation ought to focus more on the CEO as an individual and less on the firm as a legal person. Currently, the corporation itself is often the target of regulatory law: fines are imposed on it, injunctions are entered against it, and so on. Unfortunately, when firms are penalized, many humans who were not in positions of authority are likely to suffer. Thus, for example, when Enron went bankrupt due to institutional wrongdoing, thousands of lower level employees lost their jobs and pensions. Professor Treviño’s research suggests that we might do better—causing less disruption to the economy and less harm to innocent corporate participants—if regulatory and criminal law took more account of the internal workings of corporations. If CEOs are important, then we should be able to reform corporations by changes at the top before writing off the particular corporation as incorrigible. To the extent that CEOs

20 Warren Buffett’s reform of Salomon Brothers following its bond trading scandal might be an example.
can affect the corporate culture, then they, rather than the institution, ought to be held liable for its misdeeds.

Second, both at the legal level and within the organization, we need to think harder about what works and what does not.

Corporate law is silent on ethics. The state law that creates our corporations and governs their basic decision-making processes does not require any consideration of the claims of stakeholders, society as a whole or even the real human beings who own shares. We rely on the price mechanism and external regulation to control our corporations, but build nothing into the firms to prevent them from seeking to subvert those systems or to induce them to explicitly consider any countervailing values.

For several decades, corporate law scholars, even more than the case law, have taught that the only ethical course of behavior for a corporation is the unmitigated pursuit of profit. This laser-like focus on profit is precisely our problem, not the solution.

Corporate law scholarship can most usefully aid this project of creating a more profitable and more committed corporate sector by exploring the ways in which the law structures our existing markets, to their benefit and detriment, rather than by pretending that markets somehow exist abstracted from a legal framework. We have no agreement on how to resolve the conflicts among our normative systems (and perhaps such agreement is impossible). Still, we can have a fuller discussion of what it means to be ethical, of when market pressures ought to be resisted, of short term and long term conflicts, and so on, if we escape the trap of market determinism and the unfounded faith that in the end markets will automatically take us where we want to go.

CEOs, it is now clear, are critical. They, along with corporate law and both legal and business scholarship, can begin to engage the challenge of a richer understanding of corporate purpose, or they can insist that the stock ticker must rule. They can reinforce the cultural influences suggesting that the only proper role of a corporate employee is to help report a quarterly profit, or they can counter them. They can treat their employees as disposable means to a corporate end, or as the corporation itself. They can, by their choices about their own compensation, further the cynical view that the corporate world is only about getting your own, or they can join the corporate team and make it a more genuine enclave in the market.
Corporate law itself could start by abandoning the internal affairs doctrine, so that a genuine competition between the states, seeking to affect market incentives instead of simply pandering to them, can develop. Once the larger states begin to take responsibility for the governance of their largest economic entities, a thicker corporate law might move us in the direction of a more functional set of norms, practices and incentives.

Specifically, corporate law should begin exploring mandatory internal compliance mechanisms. Although courts eventually concluded that it might be a breach of fiduciary duty for a bank to hand cash to tellers with no safeguards, they have imposed no parallel requirement that firms have an ethics manager. We know that group-think is a problem in all bureaucratic organizations. Now legislatures need to consider mandating ombudsmen or similar parallel reporting systems to lessen the likelihood of group-think and to catch outright corruption. Similarly, courts need to consider whether corporate law fiduciary norms require more than avoiding unprofitable illegality while maximizing profit—whether, indeed, it isn’t a corporate law requirement that firms observe the spirit and not merely the letter of our other regulatory schemes.

More fundamentally, the profit maximization norm itself is dysfunctional even in the narrowest sense. Part of our problem is that too close a focus on profit maximization focuses firm actors on the wrong parts of our normative systems and, in the end, is not profit maximizing at all. While courts generally do not require anything resembling a strong version of short-term profit-maximization, the market structure in which publicly traded corporations operate, and the ethos of our business and law schools, often do press managers in that direction.

We need countervailing power structures within the corporation. Profit is best pursued indirectly, by commitment to the products and services that create demand, together with commitments to the relationships that make a corporation a viable alternative to markets. Currently, the corporate decision-makers—upper management—are answerable to a board elected only by shares. The short-termism and narrow focus of the stock market become overly influential almost automatically as a result. Boards need to have built-in pressures to respond to other values and other commitments as well, including both representatives of those values and interests and answerability to them.
We need an external reporting system, modeled after successful practices of the airline accident reporting system or the CDC’s emerging hospital error reporting systems, to generate systematic knowledge about which ethical breaches regularly repeat and under what circumstances.

We need restrictions on corporate interventions in the regulatory system. An organization that has as its central mission the pursuit of profit restrained only by the limits imposed by law will always have a tendency to subvert the restrictions on it. Business corporations by their nature need to be policed; we need to keep them out of the business of eliminating their policemen.

We need to seek ways to increase team behavior inside the corporation: mutual commitment and joint enterprise. This is mainly an issue for management and those who educate them, rather than the law, but not entirely. The current structure of the public corporation invites too close an alliance between short-term profit oriented institutional shareholders (reflecting the problematic dynamics of any market and our particular stock market) and managers bribed and beaten into a similar view. 21

We need limits on CEO compensation relative to ordinary employees for the simple reason that too big an income gap automatically creates distance and the “out of touch” CEO identified by Professor Treviño as the source of ethics problems and bad business decisions alike. While there is strong evidence that in at least some cases CEO

21 In theory, of course, institutional shareholders, being permanent investors, ought to be concerned about the long term prospects of the corporations they invest in, and quarterly results should be important only to the extent that they accurately predict those. If the newer theories are correct that excessive focus on short-term results is bad for long-term results, institutional investors ought to see this and correct their behavior. Indeed, the success of Berkshire-Hathaway and the current popularity of hedge funds may reflect those investors’ ability to take long term views. But more often, institutional investors are structurally incapable of ignoring short-term fluctuations. Mutual funds, for example, must respond to the short-term vagaries of their own investors. Even institutions with more stable investment pools are typically staffed by decision-makers whose careers depend on beating the market every quarter. Our longest term investors, therefore, are famous for too often acting as if they were the shortest-term ones.

Moreover, there is little reason to believe that this market can self-correct. It is notoriously difficult to distinguish good investment managers (including in the underlying corporations) from poor ones, a problem made more difficult by the end-game problems posed by the fact that both on Wall Street and in the chief executive offices most actors are already contemplating retirement by the time they arrive. Thus, high reliance on quarterly results is not obviously irrational from an individual perspective even though it is harmful from a social one.
compensation has become so large that it materially affects corporate profits directly, that problem seems to me less likely to demand urgent legal intervention: When the market concludes that CEO compensation is hurting the prospects of future returns to stock ownership, stock price drops are a fairly effective method of enforcing its will.

And because any reform that empowers employees or relationships relative to the stock market will tend to reduce the constant revolution of commitment-less capitalism, we need to work to lessen some external sources of friction in our system. We need to allow our industrial unions to organize by industry rather than by plant, so that they can stop being a lobbying force for obsolete plants and so that managers do not have a constant incentive to churn physical capital simply to union bust. We need to separate health and retirement benefits from specific firms for exactly the same reason, so that employees can change jobs more readily and less traumatically and so that employers don’t function under constant stock market pressure to find new and creative ways to abandon past commitments.

But most of all, we need to recognize that market Darwinism, like the natural kind, can lead to success, to extraordinarily strange, creative and unexpected uses of existing resources (the panda’s thumb, the bat’s wing), or simply to extinction. Fitness in a Darwinian world means only that you’ve survived so far. Stock market success comes from maximizing stock price, not from maximizing ethics, human decency, or even wealth—except to the extent that we can figure out how to make stock prices reflect human values. If we are to harness the wonderful power of the market to our own good—to human values of well-being, justice and the good life—we must always remember that the invisible hand is ours, and that if it is leading us in directions we don’t like, we can redirect it elsewhere. The choice is not between market or law, but between legally regulated markets that mandate or encourage us to abandon our other commitments and different regulation that could better harness the power of the market to work for us. We are the masters; the question is how to use our mastery.
Social Identity and Misuse of Power

THE DARK SIDE OF LEADERSHIP

Michael A. Hogg

Corporations have enormous power over people’s lives—directly, through employment, and indirectly, through corporations’ relationships to and role in government and governance. In many respects the corporation is today’s dominant institution—replacing the role of Church, Monarchy and State in earlier times. It is, therefore, hardly surprising that people pay very close attention to how corporations conduct themselves—are they principled, are they moral, can they be trusted? Because corporations are hierarchical in nature, this attention is particularly focused on the behavior of senior management (CEOs, the Board, and so forth), and therefore on leadership. People worry about the motivations of senior management and, more generally, about the prevalence of “bad” corporate leadership. Corporations and corporate leadership are often viewed with profound suspicion, as is portrayed by Rachel Carson’s classic 1962 book *Silent Spring*,

© 2005 Michael A. Hogg. All Rights Reserved.

School of Psychology, University of Queensland, Brisbane, QLD 4072, Australia.

This article is based on a paper presented at the Brooklyn Law School and Alfred P. Sloan Foundation conference on Corporate Misbehavior by Elite Decision-Makers: Perspectives from Law and Social Psychology. New York, USA. November 12-13, 2004. I wish to acknowledge grant support from the Australian Research Council for my research on social identity and leadership.


Rachel Carson, Silent Spring (1962).
and, most recently, the Sundance Film Festival award-winning documentary, *The Corporation.⁵*

What is it about corporate leaders and corporate management that may produce unprincipled behavior and undesirable or ethically inappropriate outcomes? In this article I present a social-psychological perspective that views corporate leadership as a group process—a process in which individuals or cliques have a leadership role in a wider group that people identify with.

Social psychology, in common with lay psychology, has a long tradition of attributing aberrant and undesirable human behaviors to aberrant and undesirable human personalities—personalities that are formed early in life and remain resistant to change. Prejudice, aggression and so forth are reflections of prejudiced or aggressive personalities. In this vein, people who abuse power or succumb to corrupt practices do so because they cannot help it—they have personality dispositions to behave in this way. Behavior reflects individual differences in personality.

However, social psychology also has a long tradition of focusing on how responsive people are to the situations they find themselves in. Anyone can be prejudiced, aggressive, corrupt, and so forth if the situation constrains them to behave in this way. Behavior reflects differences in social context. In truth, most contemporary social-psychological theories subscribe to an interactionist metatheory—behavior reflects an interaction between contextual factors and what a person brings to the situation in terms of relatively enduring individual habits. Theories vary in their emphases on personal or situational factors.

In this article I describe an interactionist theory of group processes and intergroup behavior, social-identity theory, which places its emphasis squarely on contextual influences.⁶ I briefly introduce key features of the social-identity approach in order to focus on its analysis of group leadership, with a particular emphasis on processes that encourage or inhibit leaders from abusing their position of power. The emphasis is primarily on conceptual review and development. Because corporations and organizations are groups, the social-identity analysis can be readily applied to

---

⁵ *THE CORPORATION* (Zeitgeist Films 2004).
organizational and managerial contexts. However, there is, as yet, little explicit discussion in the social-identity literature of corporate misbehavior by elite decision-makers.

I. SOCIAL-IDENTITY APPROACH

The social-identity approach is a general social-psychological analysis of the role of self-conception in group membership, group processes, and intergroup relations. It explains the behavior of groups and of people in groups in terms of the interaction of social-cognitive (e.g., social categorization), motivational (e.g., self-enhancement), social-interactive (e.g., social influence), and macro-social (e.g., intergroup beliefs) processes. Group behaviors, whether desirable (e.g., loyalty) or undesirable, (e.g., prejudice) reflect the operation of these normal psychological processes rather than enduring individual predispositions to behave in certain ways.

People cognitively represent human groups and social categories in terms of prototypes—fuzzy sets of attributes (e.g., attitudes and behaviors) that define and evaluate one category and distinguish it from other categories in a specific context. The content and configuration of prototypes obey the meta-contrast principle and thus enhance entitativity (the property of a group that makes it appear to be a distinct and coherent entity).

When we categorize a person as belonging to a particular group, either one that we ourselves belong to (an ingroup) or one that we do not belong to (an outgroup), we

---


9 They maximize the ratio of inter-category differences to intra-category differences.
assign to that person, to varying degrees, all the attributes of our prototype of the group, and thus view him or her through the lens of that prototype. This is a process of depersonalization in which, rather than viewing someone as an idiosyncratic individual (with whom we may or may not have a close personal relationship), we view that person as “merely” a more or less prototypical member of an ingroup or an outgroup. We assign that person a group membership, social identity, and all the attributes associated with the identity. Because group prototypes are tied to specific intergroup relations, people in one group tend to have shared prototypes of their own and other groups. Thus, prototype-based depersonalization underpins the more commonplace term, stereotyping.

One of the key insights of the social-identity approach is that we categorize ourselves just as we categorize other people, and thus we assume a social identity and depersonalize ourselves. Since our perceptions and evaluations of other people are almost always comparative, and, generally speaking, we are concerned about locating ourselves and understanding who we are with respect to others, social-categorization processes almost always involve self—either directly or indirectly. Thus self-categorization is intricately intertwined with social categorization in general.

Since the groups and categories we belong to furnish us with a social identity that defines and evaluates who we are, we struggle to promote and protect the distinctiveness and evaluative positivity of our own group relative to other groups. This struggle for positive distinctiveness and positive social identity unfolds with the guidance of our understanding of the nature of the relations between our own and other groups, and what strategies and behaviors seem possible. Social-identity processes are also motivated by a basic human concern to reduce uncertainty about ourselves, the world we live in, and our relations and interactions with others. Distinctive, high-entitativity groups with clearly prescriptive and consensual prototypes are particularly adept at achieving this social-identity objective.

Social-identity effects occur when, in a particular context, a specific social categorization becomes the salient basis for social perception and self-conception. Categories become salient, in this sense, if they are chronically accessible in memory (because we use them often and they are important to who we are) and immediately accessible in the current situation, make good sense of people’s behavior and of
similarities and differences among people, and reduce uncertainty and reflect relatively positively on self.

The social-identity approach has become well established in social psychology and enjoys substantial empirical support.10

II. SOCIAL IDENTITY AND GROUP LEADERSHIP

The implications of this analysis for leadership are quite straightforward and have been formulated into a social-identity theory of leadership that has attracted solid empirical support for its main features.11 Critically, as people identify more strongly with a group, they increasingly base their evaluations and perceptions of fellow group members on how prototypical those members are. The bottom line is that in high-salience groups prototypical members find it easier to be effective leaders, and leaders are more effective if they play up their prototypicality credentials.

High-salience ingroups are ones with which people identify strongly. These include groups that are central to overall self-definition, groups that saturate one’s day-to-day life, and groups that in a particular context experience a real or anticipated threat to their status and prestige or their very existence as a distinct entity. Since the world of work takes up much of our time and is critical to our existence and everyday life, it is quite likely that the organizations and corporations we work for play an important role in our social identity.


A. Influence, Popularity and Compliance

Prototypical members, by definition, embody central and desirable aspects of the group more so than other members. As such, their behavior is the standard for others’ behavior, and they appear to influence the rest of the group. Influence processes in salient groups cause people to conform to the group prototype.\(^2\)

Prototypical members are consensually liked by the rest of the group. They are popular, in group terms,\(^3\) and this popularity allows them to be broadly influential because people tend to comply more with suggestions from people they like.\(^4\)

B. Trust and Innovation

People tend to trust ingroup members more than outgroup members.\(^5\) Furthermore, within the ingroup, prototypical members are trusted more than less prototypical members. Because the identity of prototypical members is tightly meshed with the life of the group, it is assumed that whatever prototypical members do, however bizarre, must be in the best interest of the group and thus is unlikely to harm the group.\(^6\) Paradoxically, it is this very trust that affords prototypical members greater latitude to diverge from group norms and thus to be innovative\(^7\)—an analysis that is


consistent with Hollander’s earlier notion that leaders who conform to group norms on the way up earn idiosyncrasy credits that can be spent when they reach the top. After all, a key feature of effective leadership is the ability to be innovative in order to transform the group and steer it in new directions. Trust plays a central role in this process. As Marar puts it, “If you want to lead . . . then you had better be someone people trust.”

C. Attribution and the Social Construction of Charisma

Finally, in salient groups people’s attention is drawn to highly prototypical members. People scrutinize prototypical ingroup members’ behavior closely because it is perhaps the most reliable and effective source of information about what the group stands for and how to behave as a group member. Because prototypical members are figural against the background of the rest of the group, their attributes (i.e., being influential, popular, innovative, and trustworthy) are more likely to be internally attributed to underlying dispositions that reflect invariant properties, or essences, of the individual’s personality, than externally attributed to situational or contextual factors. The fundamental attribution error, correspondence bias, or essentialism are more pronounced for individuals who are perceptually distinctive (e.g., figural against a background) or cognitively salient. There is evidence

31 Michael A. Hogg, All Animals are Equal but Some Animals are More Equal than Others: Social Identity and Marginal Membership, in THE SOCIAL OUTCAST: OSTRACISM, SOCIAL EXCLUSION, REJECTION, AND BULLYING (Kipling D. Williams et al. eds., forthcoming).
35 See Ralph Erber & Susan T. Fiske, Outcome Dependency and Attention to
that this tendency to make dispositional attributions is especially strong for attributions about leaders.\textsuperscript{26} In this way, a charismatic leadership personality is constructed for highly prototypical leaders, further fuelling their leadership effectiveness. Conger and Kanungo, for example, describe how followers attributionally construct a charismatic leadership personality for organizational leaders who have a vision that involves substantial change to the group.\textsuperscript{27} It should be noted that the social-identity analysis of charisma views it as a product of social-cognitive processes operating under conditions of self-categorization, not as an invariant personality attribute that determines leadership effectiveness.\textsuperscript{28} And it should be noted that charisma alone may not be a reliable predictor of group performance. For example, CEO charisma has been shown only to predict the size of the CEO's salary and, except for stock price, not the overall performance of the corporation.\textsuperscript{29}

\textbf{D. Managing One's Prototypicality}

Because prototypicality is critical for effective leadership in high-salience groups, leaders of such groups pay close attention to how prototypical they are perceived to be. They engage in prototypicality management strategies that rest on communication,\textsuperscript{30} or what can be called “norm talk.”\textsuperscript{31}

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{26} See James R. Meindl et al., The Romance of Leadership, 30 ADMIN. SCIENCE Q. 78, 78-102 (1985).
\item \textsuperscript{29} Henry L. Tosi et al., CEO Charisma, Compensation, and Firm Performance, 15 LEADERSHIP Q. 405, 405-21 (2004).
\item \textsuperscript{30} Scott A. Reid & Sik Hung Ng, Conversation as a Resource for Influence: Evidence for Prototypical Arguments and Social Identification Processes, 30 EUR. J. OF SOC. PSYCHOL. 83, 83-100 (2000).
\item \textsuperscript{31} See Michael A. Hogg & Scott R. Tindale, Social Identity, Influence, and Communication in Small Groups, in INTERGROUP COMMUNICATION: MULTIPLE
\end{itemize}
\end{footnotesize}
Language and communication play a key role in this type of prototype and identity management. In order to manage their prototypicality, leaders can talk up their own prototypicality and/or talk down aspects of their own behavior that are non-prototypical. They can identify deviants or marginal members in a manner that highlights their own prototypicality or constructs a particular prototype for the group that enhances their own prototypicality. They can secure their own leadership position by vilifying contenders for leadership and casting the latter as non-prototypical. They can identify outgroups that are most favorable to their own prototypicality as relevant comparison groups—that is, they can manipulate the social-comparative frame and thus the prototype and their own prototypicality. They can engage in a discourse that raises or lowers group salience. For highly prototypical leaders, raising salience is advantageous because it provides them with the leadership benefits of high prototypicality. For non-prototypical leaders, lowering salience is advantageous because it protects them against the leadership pitfalls of low prototypicality.

Reicher and Hopkins analyzed the rhetoric used by political leaders to show that such leaders are particularly prone to accentuate the existing ingroup prototype, pillory ingroup deviants, and demonize an appropriate outgroup. Furthermore, the use of these rhetorical devices is often viewed as convincing evidence of effective leadership. Reicher and Hopkins proposed that leaders are in this sense “entrepreneurs of identity”—they are experts in norm or prototype management through talk. In other research, Rabbie and Bekkers have shown that leaders whose positions are insecure are more likely to seek conflict with other groups, and Gardner and colleagues have shown that effective organizational leadership often rests on norm management.
through talk.\textsuperscript{36} Generally, leaders who feel they are not, or are no longer, prototypical, strategically engage in a range of group-oriented behaviors to strengthen their membership credentials.\textsuperscript{37}

III. LEADERSHIP AND MISUSE OF POWER

The previous section has shown how people who are viewed as highly prototypical tend to be more effective leaders in groups within which members identify strongly. These highly prototypical individuals are largely leaders who are trusted to be effectively innovative and therefore can lead through influence rather than coercion, fitting well the typical definition of leadership as “a process of social influence through which an individual enlists and mobilizes the aid of others in the attainment of a collective goal.”\textsuperscript{38} However, there are at least three paradoxical effects of prototype-based leadership in high-salience groups that can produce poor, and sometimes harmful, leadership. In addition, uncertainty can be a breeding ground for harmful leadership.

A. Dysfunctional Norms and Dysfunctional Leaders

Having good leadership skills is very useful in salient groups. However, while such qualities are critical in low salience groups, they are relatively less critical in high-salience groups. This can introduce a problem. Typically, elite decision-making groups are characterized by group norms that embody principles of ethical behavior and responsible leadership. In these instances, prototype-based leadership will be ethical and responsible if members identify strongly with such groups. However, if group norms do not embody principles of ethical behavior and responsible leadership, then increased salience and group identification may inhibit responsible and ethical leadership. This is one way in which social-identity-contingent leadership may be associated with poor leadership and corporate misbehavior. In salient groups, group norms not only


\textsuperscript{37} See Platow & van Knippenberg, supra note 17.

influence behavior directly (via conformity) but also indirectly by empowering as leaders those people who best embody those (deficient) norms. Shades of this process may be seen in Janis’s notion that groupthink may arise in highly cohesive groups that do not have strong norms for effective decision making.\textsuperscript{39} This is particularly so in groups where cohesion is based on group identification.\textsuperscript{40}

B. \textit{The Trust Paradox}

A second source of leadership deficiency in high-salience groups is, ironically, the strength of trust in and consensual liking for the leader. Although these processes allow the leader to be innovative (which is, of course, a positive attribute of leadership), these processes can also make it possible for the leader to “get away with anything” and lose sight of what is appropriate for the group and what is not. Pretty much whatever the leader does the group approves of, or at least does not openly disapprove of.

Under normal circumstances, leadership behavior which is too innovative will violate the limits imposed by the group’s identity (for example, the leader of a tight-knit vegan group advocating a shift to a purely carnivorous diet) and will quickly erode the leader’s prototypicality credentials and reduce his or her ability to influence. Reicher has used this analysis to explain the limits of crowd behavior—i.e., the way that collective behavior remains within the limits imposed by the social identity of the collective.\textsuperscript{41}

However, in extremely cohesive groups characterized by ultra-strong identification (e.g., cults), consensual liking for the leader is so strong, and attribution to charisma so complete, that dissent and criticism are unlikely. The leader’s leadership potential is literally unbounded—he or she has the power to do whatever he or she wants, with little or no normative framing to help decide which decisions are wise or ethical. Even in


relatively abstract laboratory settings, research has shown that, so long as the leader of a group is considered highly prototypical, group members are willing to endorse leaders who behave in ways that are not in the best interest of the group.\textsuperscript{42}

C. Hierarchy and Power

The third pitfall of prototype-based leadership in high-salience groups is the emergence of hierarchy and power-based leadership.\textsuperscript{43} Prototypical leaders do not need to exercise power over others (i.e., persuade, gain compliance, coerce, or resort to force) to have influence. In addition, it is possible that they may be “unable” to exercise power. High prototypicality is associated with strong ingroup identification; self and group are tightly linked prototypically and thus fellow group members are to some extent internalized as part of one’s self.\textsuperscript{44} Any harmful behavior directed against fellow members is effectively directed against self. There may exist an empathic bond between leader and followers that inhibits the leader from exercising coercive power over fellow group members.

However, there is a paradox. Occupying a highly prototypical position, particularly in an enduring and stable high-entitativity group with a focused and consensual prototype, makes one appear enduringly influential, consensually socially attractive, and essentially charismatic. Through structural role differentiation grounded in social attraction and attribution processes, there is a perceptual separation of the leader from the rest of the group. The leader


is gradually perceived as “other” rather than “one of us.” In corporate settings, this separation of the leader from the group as a whole can be strikingly evident. As Treviño puts it, it can be “eerily quiet at the top”—there is substantially more lateral communication within the leadership clique and between senior management of different corporations than vertical communication within the corporation itself. The leader can be markedly out of touch with the rank-and-file.

The person who originally embodied the essence of the group by being most prototypical has now become effectively an outgroup member within the group. There is an embryonic intergroup relationship between leader (along with his or her inner clique) and followers. This intergroup relationship is associated with a status differential that is perceived by the group to be consensual, stable, and legitimate—a potent mix that has potential for a conflictual intergroup relationship between leader(s) and followers in which the leader has most of the power. Although the seeds of autocracy are sown, they may not germinate. The relationship may still be viewed as a mutually beneficial role relationship in the service of superordinate group goals—everyone is on the same team, working for the same goals, but making different contributions to the greater good of the group. The leader may not be “one of us,” but he or she is certainly working with us and for us.

However, there are circumstances which may make power-based intergroup behavior a reality. A relatively inevitable consequence of role differentiation is that the leader realizes that he or she is being treated by followers as an outgroup member—a positive high-status exile, but nonetheless an exile who cannot readily share in the life of the group. The leader may try to re-establish his or her ingroup credentials by engaging in behaviors confirming his or her ingroup prototypicality. If this is unsuccessful, a sense of rejection by, and distance and isolation from, the group may occur (possibly also a recognition of reduced influence among followers). These feelings may then “embitter” the leader and, since the empathic bond is severed, allow the leader to gain compliance through the exercise of power over others. This may

---

involve coercive behavior: because the interests of the leader and the group have diverged, the leader is effectively exercising his or her will over others. The influence process essentially becomes one of coercion rather than attitude change.

This transformation of leadership into power is stronger in hierarchical extremist groups where the leader-follower role and power differentiation is more tangible, stark, and impermeable. The effect will also be stronger in groups where there is a leadership clique rather than a single leader. This is because a typical inter-group relationship has emerged, and thus the relationship between leader(s) and followers is an intergroup relationship where one group (the leaders(s)) has disproportionate legitimate power over the other group (the followers). Such a relationship will be competitive and potentially exploitative, a situation far removed from prototype-based leadership.

Leaders generally react unfavorably to perceived threats to their leadership position. Where a leader is prototypically influential and there is no pronounced intergroup differentiation between leaders and followers, threats to leadership largely come from prototype slippage—social-contextual factors may reconfigure the group prototype and thus reduce the leader’s prototypicality. We described above how leaders then strive to redefine the prototype to better fit themselves—they accentuate the existing ingroup prototype, pillory ingroup deviants, or demonize an appropriate outgroup. These tactics generally do not involve coercion.

However, where there is a pronounced intergroup differentiation between leaders and followers, perceived threats to leadership are automatically perceived in intergroup terms as collective challenges or revolts on the part of the followers. This makes salient the latent intergroup orientation between leader(s) and followers, and engenders competitive intergroup relations between leader(s) and followers—competitive relations in which one group has consensually legitimate and overwhelming power over the other. Under these circumstances leadership becomes coercion, based on the relatively limitless exercise of coercive power over others. The dynamic is similar to the way in which a power elite “reacts” to a perceived challenge to its privileged position, but, because it

occurs within the power-legitimizing framework of a common group membership, the “reaction” is potentially more extreme.

This analysis suggests a series of steps that transforms prototype-based leadership into power-based leadership. Highly prototypical leaders of salient groups, particularly newly-emerged leaders, provide leadership through influence—they do not need to exercise power over followers, and indeed may not actually be able to behave in this way. Enduring tenure renders leaders more influential and facilitates normative innovation—leaders still do not need to exercise power over followers because they now have the capacity to ensure that they remain prototypical and thus influential. Further tenure differentiates the leader(s) from the followers. It creates an intergroup differentiation based on widening, reified and consensually legitimized role and power differences—the potential to use power is now very real. The conditions that translate this potential into reality are ones that make salient the latent power-based intergroup relationship between leaders and followers—for example, a sense of threat to one's leadership position, a feeling of remoteness and alienation from the group, or a sense of becoming less influential in the life of the group.

The exercise of leadership through coercion rests on the psychological reality (based on self-categorization and social-identity processes) of a sharp role, status, and power discontinuity between leader(s) and followers that reconfigures cooperative intragroup role relations as competitive intergroup relations. Such intergroup relations within a group provide ideal conditions for unilaterally exploitative intergroup behavior. This is because the overarching common group identity and the diachronic process of leadership emergence strongly legitimize the status quo—there exists within the group what social-identity theory refers to as a “social change belief structure without cognitive alternatives.” Because power and leadership are attractive to some people, this belief system can be coupled with a belief in intergroup permeability that encourages followers to try, as individuals, to gain personal admittance to the leadership clique—a process that marshals support for the leader(s) and prevents the followers from forming a united front.
D. Uncertainty and Poor Leadership

One important motivation for social-identity processes is uncertainty reduction.\(^7\) People strive to reduce feelings of uncertainty about who they are, how they should behave, how they should interact with others, and how others will treat them. Social identity reduces uncertainty because prototypes specify one’s self-concept and regulate one’s behavior and interactions with others. Research has shown that people are more likely to identify with groups and identify more strongly with groups as a function of increasing uncertainty, especially self-conceptual uncertainty.\(^8\)

An extension of this idea argues that where uncertainty is extreme, people form, modify, or identify with groups that have prototypes that are simple, highly focused and consensual, and that have high entitativity and hierarchical internal structural arrangements—that is, extremist or totalist groups that have rigidly ideological belief systems.\(^9\) Uncertainty is potentially a significant force for autocratic leadership. The reason for this is that, all things being equal, members identify very strongly under uncertainty or under the threat of uncertainty. Thus, prototypicality is a very powerful influence on leadership and the processes described above are all so much stronger. Indeed, leaders may invoke the specter of uncertainty precisely in order to maintain their position of power within the group.

Uncertainty is a pervasive feature of corporate life because modern corporations often operate in high-risk environments\(^10\) that make employment insecure and raise uncertainty about the nature and viability of one’s organizational identity.\(^11\) At the very least, organizations and


\(^8\) For overviews of research see id.


\(^12\) See Kevin G. Corley & Dennis A. Gioia, Identity Ambiguity and Change in the Wake of a Corporate Spin-Off, 49 ADMIN. SCI. Q. 173, 173-208 (2004); Dennis A. Gioia et al., Organizational Identity, Image and Adaptive Instability, 25 ACAD. OF
their associated identities are often in flux as a result of takeovers, mergers and market forces, and there are strong mechanisms at play to make sense of identity uncertainty and change. Modern corporate life may be particularly, though not inevitably, prone to social-identity leadership processes.

IV. ORGANIZATIONAL AND CORPORATE LEADERSHIP

Organizations and corporations are groups, and therefore their leadership is subject to the social-identity processes described in this article. Typically, the nature of corporations is such that leaders are competent and moral individuals who have substantial leadership skills, qualities and experience. Furthermore, social-identity processes generate organizational identification, commitment and loyalty on the part of both leaders and non-leaders. Social-identity processes generate trust in and respect (consensual group membership-based liking) for the leader, construct a charismatic leadership personality for him or her, and provide an environment in which leadership-driven innovation can thrive. These are all good things—allowing the leader to lead rather than coerce, and to make wise consultative decisions about what the organization stands for and how it should conduct itself. Because of its grounding in consensus and accountability, this sort of leadership should conform to wider societal expectations for ethical conduct.

However, a number of problems may arise when members identify too strongly with an organization. These levels of identification are more likely to occur when organizational members invest too much of their lives in their organization and feel the threat of uncertainty about their future, their identity, and their future employment and organizational membership status. Under these circumstances, social-identification processes may create an atmosphere of unqualified trust and invest leaders with a sense of charisma that in turn makes them too consensually popular. Leadership...


becomes easy and normatively unbounded—it can be very difficult for a leader to choose between wise and unwise decisions and actions.

The problem can, however, become worse. A sharp power differential may exist between leader and employees that separates or isolates the leader from the rest of the group, and instantiates an intergroup orientation within the organization that gives the leader great power. Employees often view this arrangement as legitimate and unchangeable. Leadership can now mutate into coercion, liberating the leader from normative accountability. Such power makes misbehavior, cronyism, corruption and poor leadership a reality.

V. CONCLUSION

This has been a theory and overview article in which I have described the social-identity theory of leadership. Because organizations and corporations are groups, this analysis applies to them just as much as to other groups. The key point is that as people identify more strongly with a group, they increasingly base their leadership perceptions on group prototypicality—prototypical leaders are more effective than less prototypical leaders because they are popular, viewed as charismatic, and trusted to be innovative.

Within bounds, this kind of leadership is both desirable and highly effective. However, a group with which people identify too strongly has the potential to mutate into power-based coercion in which the leader is effectively unfettered by normative constraints and by accountability to the group. When this happens, the ground is ripe not only for poor decisions but also for unethical, exploitative and corrupt behavior.

Let us finish this article on a more positive note. From a social-identity point of view, corruption-based exploitative leadership is probably not that common, particularly in well-regulated Western organizations and corporations. Most organizations are not associated with extreme identification—employees have other aspects of their lives that provide them with a certain sense of self in their social world (e.g., family, ethnicity, recreational groups). Most organizations are not very salient and are not that cohesive. They have a diverse workforce in terms of ethnicity, gender, socio-economic status and so forth. They encourage, or at least do not severely punish, a degree of normative criticism—there are often formal
mechanisms in place to allow constructively critical discussion of normative practices. Employees have a voice, and thus organizational norms and prototypes can be grounded in common-sense principles of ethical conduct that reflect society’s values.

However, extreme uncertainty coupled with an all-embracing, highly cohesive, uniform and consensual organization will raise identification and set up a situation where the corporate leadership has too much power for its own or the organization’s good.
The Social Nature of Boards

Rakesh Khurana
Katharina Pick

I. INTRODUCTION

Still reeling from a series of high-profile and extremely costly corporate scandals involving dysfunctional board behavior, lawmakers and scholars are scrambling to make sense of the gaps in understanding that clearly exist in governance research. Inevitably, these scandals provoked a surfeit of Monday-morning quarterback explanations. Some argued that these problems should have been anticipated as the inevitable consequence of the proliferation of high-powered pay-for-performance plans. Others talked about investors' misplaced faith in a firm's stock price as an indicator of corporate governance quality. Ironically, the same fields of finance and law that are now offering these retrospective judgments previously advocated increased stock option grants and an unyielding faith in the efficiency of the stock market.

© 2005 Rakesh Khurana & Katharina Pick. All Rights Reserved.

1 Rakesh Khurana is an Associate Professor at Harvard Business School. He is also the author of SEARCHING FOR A CORPORATE SAVIOR: THE IRRATIONAL QUEST FOR CHARISMATIC CEOS (2002).

2 Katharina Pick is a Ph.D. Candidate in Organizational Behavior at Harvard University. She is conducting dissertation research on the effect of role identities on the behavior of directors.

Both authors would like to thank Professor Jay W. Lorsch for comments on an earlier version of this paper.


As well-intentioned and wise as many of these judgments are, we cannot help but feel that they perpetuate a false understanding as to the nature of boards. At the root of this finance and law perspective is the assumption that directors are fully motivated to act in the interests of the firm and its shareholders only when they have an individual interest to do so. Advocates of this perspective have argued, for example, that without financial inducements such as stock options or share grants, directors “have little incentive to extend themselves beyond relatively superficial oversight of their firms’ affairs. They mechanically fulfill their specified duties (certain approvals and audits) and watch for egregious derailments, but not much more.” Much of the corporate governance research and prescription of the last twenty or more years is rooted in this individualistic explanation of board and director behavior.4

We believe that the dominating focus on individual director incentives in governance scholarship misses a critical element of director behavior. As we will argue in this paper, a board is not a simple aggregation of individuals but is, in fact, a complex social system and must be understood as such. Financial considerations do not figure centrally in shaping a director’s behavior, particularly as it relates to boardroom culture. In fact, directors are highly cognizant of their membership, not only on a particular board, but also as members of a broader community.5 They have a clear sense of their own boundaries—how far their territory extends, who belongs, and who does not. The actions of a board, therefore, cannot be understood as being the aggregate product of each individual director’s behavior. Rather, their actions express the sum of connections and relationships of a group.

We know from decades of research on groups that they are enormously powerful social environments.7 Group


7 For a review, see J. Richard Hackman, Group Influences on Individuals in Organizations, in 3 HANDBOOK OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY 199 (M.D. Dunette & L.M. Hough eds., 1992).
influences on individuals, as well as factors that emerge purely at the group level and through the group’s situation in a wider social context shape members’ behaviors, beliefs, and attitudes. Our argument is that in order to understand the factors contributing to board culture and board outcomes—whether dysfunctional or functional—we must treat boards as complex social systems, and use the group as the basic unit of analysis when we study them. We must understand both the factors that are driven by group dynamics and culture and those that result from environmental influences on the board.

We acknowledge that several organizational scholars have indeed already approached boards as social systems, bringing a behavioral perspective to boards research. Lorsch and MacIver produced one of the earliest qualitative studies of directors describing the subtle mechanisms shaping behavior in board rooms and how the resulting conditions can yield both active and passive boards. William Ocasio explored how institutionalized action shapes board outcomes in CEO succession. Edward Zajac and James Westphal collaborated on several studies exploring how power dynamics, interlocks, director reputation, and demography shape board outcomes. Mark Mizruchi studied the power relations between boards of directors and management to explain how these manifested themselves in control over corporations. Finally, network studies conducted by Gerald Davis among others revealed how the interlocked nature of the director community shapes the diffusion of ideas and practices among boards.

Recognizing these important contributions, our aim here is simply to call attention to some of the significant ways in which boards function as social systems, and more specifically

---

8 See generally Jay W. Lorsch & Elizabeth MacIver, Pawns or Potentates: The Reality of America’s Corporate Boards (1989).
as groups. Because it is a review piece, our paper benefits from the flexibility to incorporate examples from recent corporate governance scandals, drawing together theory and practice.

We begin in Part II with a brief review and critique of agency theory, the dominant theoretical approach that now underlies corporate governance research in both finance and law. Because we find these individualistic conceptualizations problematic, in Part III we look outside the economic and legal literature and incorporate some of the tools developed in organization theory that explain board behavior. We apply these theoretical tools to some of the better-known cases of corporate malfeasance. Finally, in Part IV we discuss the implications of this approach for the future of governance research, particularly in light of recent environmental factors that now affect corporate boards.

II. AGENCY THEORY

A. What is Agency Theory?

More than any other approach, agency theory has focused on board room dynamics and the fundamental nature of the factors driving director behavior. Although this approach has spawned hundreds of articles and its conceptual language is now widely used by scholars and practitioners alike, our purpose here is to explore only two of its key assumptions regarding director behavior. First, agency theory conceptualizes the etiology of director behavior at the individual level. Second, even when an agency perspective considers the board as a unit, boards are treated as mere aggregations of individual director behavior as opposed to complex social groups.

In the United States, all large public companies have a board of directors that is approved by the shareholders. Legally, the board of directors is vested with enormous decision-making power over the activities of the company. In large companies, directors often delegate decision-making to corporate executives. Even when such delegation occurs, however, the ultimate power rests with directors. Almost all existing governance theory produced by economists, which has

---

had the dominant influence on law, defines the relationship between directors and shareholders as one instance of an agency relationship. The firm itself is described as a legal fiction that serves as a nexus of individual contracts.

This definition of the firm is critical to agency theory because it suggests the relationship between individuals and the firm is fundamentally a contractual one. An agency relationship is thus a contract under which one or more principal(s) (e.g., shareholders) engage an agent (e.g., directors) to perform some service on their behalf by delegating some decision-making authority to the agent. Because directors and shareholders are unique individuals, it is hypothesized that their respective interests will diverge. Consequently, shareholders will take the necessary means to ensure that directors will act in ways concomitant with their interests. Much of the research in this area involves identifying the means, and the efficacy of those means, by which this convergence of interests is achieved.

This research has translated into many individual-oriented reforms. Boards have adopted partial stock option compensation for directors in order to align directors’ interests with those of shareholders. Director independence requirements have been formulated primarily with regard to preventing individual directors from having conflicts of interest. Qualifications for committee membership, and audit committee membership in particular, have become clearer and more demanding so as to ensure that individuals possess the expertise and knowledge required for the position.

B. Why Agency Theory is Problematic

Is it realistic or useful to view the modern corporate board as comprising only, or even principally, a set of individual contracts? We think not, and we argue that the radical individualism embedded in this contractualist view is unreasonable. It blinds us to most of those features of modern boards that are distinctive and in accordance with directors’

---


15 See generally Jensen & Meckling, supra note 14.
own empirical experiences. For example, R.C. Clark notes that most of the particular rules that make up the relationships among corporate officers, directors, and stockholders—that is, the relationships that give operational meaning to the concept of the corporation—are not the products of individual contracts.

Some readers may concede that while a large firm is not a nexus of individual contracts in a strict definitional sense, the distinction is insignificant given that we can still analyze boards or firms usefully as if there were actual contracts in action. We believe such a response is inadequate. Viewing the board as an instance of individual contracting may have at least two objections. First, this individualistic and undersocialized view of boards, while consistent with agency theory and economic explanations of human behavior, has proven to be inadequate when evaluated vis-à-vis the autopsies of recent corporate misconduct. Moreover, it does not reflect the real nature of board behavior. Second, academic research could help a great deal in improving corporate governance, but a narrow individualistic approach may make it difficult to realize this potential contribution. Answering the question of how we are to understand directors’ behavior in boardrooms requires not only locating the etiology of their behavior, but also developing defensible premises about director behavior that can guide theory building and focus empirical investigation. As Jeffrey Pfeffer and his colleagues have noted, assumptions about human behavior tend to become self-fulfilling:

To the extent people believe in a particular theory, they may create institutional arrangements based on the theory that thereby bring the theory into reality through these practices and institutional structures. To the extent people hold a theory as true, they will act on the basis of the theory and expect others to act on that basis also, creating a normative environment in which it becomes difficult to not behave on the basis of the theory because to do so would violate some implicit or explicit expectations for behavior. And to the extent that people adhere to a theory and therefore use language derived from and consistent with the theory, the theory can become true because language primes both what we see and how we apprehend

16 See generally COLIN B. CARTER & JAY W. LORSCH, BACK TO THE DRAWING BOARD: DESIGNING CORPORATE BOARDS FOR A COMPLEX WORLD (2004); LORSCH & MACIVER, supra note 8.

the world around us, so that talking using the terminology of a particular theory also makes the theory become true.\textsuperscript{18}

In the next section we call attention to some of the important ways in which board outcomes may be shaped by mechanisms that operate within groups. Specifically, we will highlight the importance of group norms on individual director behavior, the effect of social influence on the way directors interpret information, the effect of group membership on directors’ attitudes, and finally the potential effect of habitual routines that develop at the group level and compromise mindful group decision-making. While these group mechanisms will be familiar to students of organizational theory, they have yet to penetrate the dominant law and economic perspective that continues to inform corporate governance research and legislation. Indeed, even within organizational research, many concepts specific to group behavior have not been applied to work on governance and boards of directors.

III. BOARDS AS SOCIAL GROUPS

Social influence \textit{in} groups and \textit{through} groups is extremely powerful in shaping the behavior of both members and non-members. A group environment influences how people behave, what they believe, and how they feel.\textsuperscript{19} The board environment is no different. Despite meeting episodically and infrequently, boards are groups in a truly psychological and sociological sense. Boards have clear boundaries, with membership that is stable over time and readily identifiable by both members and non-members. A board’s members are engaged in a common task that requires sharing information and making joint decisions, and this task is ongoing and does not end when the board is not in session. Finally, directors interact face to face at least part of the time.\textsuperscript{20}


\textsuperscript{19} See Hackman, \textit{supra} note 7, at 1.

A. Board Cohesiveness

Importantly, boards have several characteristics that make them highly cohesive groups. This is crucial because, as social psychologists have long recognized, social influence of the kind that we will be discussing throughout the remainder of this paper is stronger in groups that are cohesive.\(^2\)

First, boards are cohesive because they tend to be very homogenous. Although this is beginning to change with respect to gender and race, boards continue to be homogenous with respect to age, occupation, class, and status position.\(^2\) This gives boards the likelihood of shared mental models, attitudes, beliefs, and experiences that contribute to group cohesion.

Second, boards have high group distinctiveness, meaning that their membership is readily identifiable both to insiders and outsiders. Even when not face to face, the group membership is almost constantly salient given their names are often listed on websites and company documents and because the title “IBM Board of Directors” is synonymous with the list of names that comprise it. Contributing to the distinctiveness is the fact that membership requirements and tenures are clearly spelled out and that membership is stable. Indeed, it is rare that someone is expelled from the group.

Finally, boards are small, ranging from ten to twelve directors. This makes them small enough for directors and interactions to be highly visible. Anonymous action is nearly impossible, and directors are aware of each other both as directors and as individuals.

Cohesive groups, like boards, have the potential to realize benefits from diversity of ideas, skills and expertise, and to make process gains by operating more efficiently. However, they also have the potential for incurring costs and process losses.\(^2\) The gains and losses associated with group behavior are inherent in the very nature of boards, and must therefore be addressed in any explanation of board outcomes.


\(^2\) See *Khurana*, *supra* note 6, at 84.

In a recent Wall Street Journal article, Useem reiterated what many governance scholars have suggested before: although they are important, individual flaws like lack of expertise, prolonged tenure, and conflicts of interest (not to mention board level problems like inappropriate committee structures that governance critics like to harp on) will always be trumped by what ultimately happens when directors meet behind closed doors and are confronted with important decisions. What transpires there and over the course of meetings and telephone conversations cannot be reduced to individual or structural characteristics, but rather must be understood as the result of complex relationships of the board as a group.

B. How Groups Regulate and Shape Behavior

Reflecting on board decisions gone awry, it is always striking that individual directors were able to sit by and not pursue issues that later turned out to be consequential. The board at Hollinger International Inc. approved more than half of the $400 million worth of transactions that Lord Black and his colleagues improperly pulled from the company. They made these decisions in infrequent meetings that were described as “brief, casual affairs,” some lasting no longer than an hour and a half. The board adopted behavioral routines like rapidly shuffling through and approving Lord Black’s proposed transactions, discussing unrelated and trivial but intellectual affairs during lunch breaks, and allowing Black to dominate and flatter directors into acquiescence.

These norms, and the behavior they created, came back to haunt the Hollinger board. Though Lord Black and his

---

colleagues also consistently misled and lied to the board, the board passively accepted a process that prevented judicious oversight—a critical failure. The Hollinger board’s practices seem negligent in the extreme. But less obviously harmful behavioral patterns with potentially similar consequences exist on all boards.

Social norms, particularly in groups, are one of the most powerful forms of social control over people.\textsuperscript{27} Directors, like all members of groups, rely on norms as social indicators of what behaviors are, or are not, appropriate within the group context. Particularly within cohesive groups, which the Hollinger board had all signs of being, such norms are extremely difficult to challenge. As investigators discovered, Lord Black displayed consistent and profound loyalty toward his directors, offering generous donations to charities they supported, and often flattering them not only with perks but also with personal compliments.\textsuperscript{28} Having hand-picked the entire board, it was not difficult for him to mold the individuals into a group for whom the norm was to trust management, feel privileged by their membership among such an elite group, and not insult each other or waste valuable time with caution and skepticism.

Groups tend to create norms around behaviors that they consider to be important to effective group functioning and performance.\textsuperscript{29} On boards these behaviors are likely to include the content and flow of board discussions, the sharing of air time among directors and with management, the leadership and power balance on the board, the effective use of information, the proper availability and use of expertise, and the structure of meetings. It appears that the Hollinger board, for example, had developed a norm of brevity and cursoriness that likely undermined each individual director’s freedom to interrupt or hold up discussion no matter what the reason. Other boards with which we are familiar have norms governing the interruption of management presentations, or the appropriateness of asking questions in the last half hour of a meeting. Still others have norms regarding director participation; some discourage full-board discussions while others require each director to provide an opinion on any given topic of discussion.

\textsuperscript{27} See generally George C. Homans, The Human Group (1950); Mazafer Sherif, The Psychology of Social Norms (1936).

\textsuperscript{28} See generally Frank & Cherney, supra note 26.

\textsuperscript{29} See Hackman, supra note 7, at 235-36.
Whatever their nature, norms are extremely resilient once they have become established. Groups reinforce norms through sanctions and feedback to members who violate them. Ambient stimuli, the back-drop characteristics of the board’s surroundings and interactions (for example, how directors and the CEO are situated in the board room, what the board room looks like, and how directors speak to one another at a meeting) create a subtle but powerful normative inertia. In fact, ambient stimuli are often more powerful than the discretionary stimuli because they are “rarely noticed or discussed.” Instead, they function through directors’ implicit assumptions about what is appropriate based on the cues provided by the environment.

If a group environment is characterized by formality, with directors sitting stiffly around a conference table and interacting only in formal language at predetermined times, a director with an important but loosely formulated idea would be unlikely to present that idea because he could not do so in a sufficiently coherent and formal fashion. Note that his withholding of the idea may be based on his perception of what behavior is appropriate or desirable, however, and not necessarily on any real evidence about what is appropriate. Unfortunately, this may often prevent ideas and processes that would be beneficial to groups from surfacing and becoming part of the group norms. Instead existing norms are supported and perpetuated since members are more inclined to surface perceptions through behaviors from which they expect positive feedback. Members are unlikely to test behaviors that they think may yield negative responses. Thus individual members can quietly hold false assumptions about the group that cannot be revised because no real information is ever exchanged.

One director’s recent comments reflect the impact even assumed norms can have on board room behavior:

One of the good things, for all the unpleasantness associated with the post-Enron period, it is true that boards are much more assertive than they used to be. And it’s sort of like a natural change, they have to be, but they don’t have to do anything unclolike in order to be

---

31 Hackman, supra note 7, at 209.
that way. So that's a good outcome. It's expected, you know they've got to ask tough questions and expect answers. It's hard to describe precisely, but back prior to this period and certainly in periods in our history, you really had to kind of stick your neck out if you were going to really object, even if there were good substantial reasons for doing it. It's just that the norms were such. 33

By pointing to how external scrutiny has legitimated directors’ asking of tough questions, these comments highlight the extent to which being assertive in the boardroom can be perceived as undermining the board as a group or “club.” External influences, like those that resulted from recent corporate scandals, do not always occur and thus do not alter directors’ perceptions and experiences of norms. Absent these types of influences, norms usually persist over long periods of time.

C. Transmitting and Sustaining Norms

The norms developed in groups are not only stable over time, but often survive even membership turnover. Unless there is a shock to the group, either in the form of individuals deviating from the norm or a change induced by an external investigation, norms survive even as the composition of the group changes. This resilience is the result of a number of important factors, some of which are reinforced by characteristics specific to boards.

First, board membership usually changes incrementally, with no more than two or three new directors joining at a time. Second, new and inexperienced directors who have not been socialized into board culture and thus learn the behaviors befitting a board member primarily from those around them. One now very prolific director described his early board experience in the following way:

When I first started going on boards I was the youngest thing in the board room. And so it was very helpful to me to talk to experienced directors. You know when I first started going to board meetings I wouldn’t say anything. And obviously I’d listen to the conversation. Because my mother always taught me to . . . if you go to a big dinner and there’s a lot of silverware and a lot of crystal and you don’t know

33 Katharina Pick, The Adoption and Framing of a Corporate Governance Innovation: How Directors Make Sense of the Lead Director Position (2004) (unpublished Ph.D. qualifying paper, Harvard Graduate School of Arts and Sciences) (on file with author). All quotes from directors in this article are from directors of Fortune 500 companies to whom confidentiality was assured.
which one to pick up, just watch the hostess. So I spent my early meetings watching. And that’s a way to learn. And then you think “I would like to ask this question.” And then you see some guy who’s been there a long time who’s very smart and he asks the same question that you were thinking about. And that’s helpful to you. So that gives you some sense of confidence that you’re not exactly stupid. And then you hear a question that you really think is stupid and then at lunch some director will tell you “Wasn’t that a stupid question?”

Finally, because boards are so homogenously comprised of elites, the primary way in which people achieve status on a board is through tenure. This means that the people most able and likely to challenge norms (i.e., people of high status within the group), are the people who have been there the longest and are likely the most entrenched in the board’s norms.

In order to explore the nuances and consequences of norms we must first understand why they are so important to groups. The enduring norms of a group are functional in sustaining two critical features of groups: the diversification of roles within the group and the achievement of uniformity in the group. Although these can appear to be contradictory forces, they are both critical to how groups achieve organization, order, and predictability in a way that maintains the group. However, both the tendency toward diversification and the tendency toward uniformity also have importance consequences for behavior in groups.

1. Role Diversification

Diversity comes in the form of role differentiation, which is clearly visible in the formal structure of boards. Here, committees and committee chairs take on additional duties with respect to specific areas of expertise. This role differentiation is also reflected in a board’s informal structure. A recent study on Lead Directors showed that even where

34 All quotes from directors in this article are from directors of Fortune 500 companies to whom confidentiality was assured.


boards had not formally named a person to serve as Lead Director, there were individuals on the board who, in times of crisis or indecision, would be expected to assume responsibility or speak up on behalf of everyone else. Directors in most situations were confident that they could identify that person and that their selection would be consistent with the expectations of others. It is not surprising that many boards should have such informal role differentiation given their structure, where the independent directors usually have no formal leader with the exception of the CEO, whom the board is essentially evaluating. In these conditions an informal leadership structure is almost certain to emerge.

Subtle role differentiation can become problematic, however, because when that particular person does not speak up, it is less likely that others will, even if they feel uneasy about an issue. This is generally the case because: (1) the unwillingness of the person to speak up establishes, informally, that concern is not required; and (2) it feels inappropriate to violate the implicit roles that have been established. This dynamic undermines the true function of boards of directors, which is to allow an opportunity for individuals with different ideas to come together and share information. In effect, group dynamics undercut the positive effects of diversity.

2. Achievement of Uniformity

Another reason groups develop norms, and one reason they are so resilient and consequential to particular outcomes, is because groups have a tendency toward uniformity. Norms, by bringing individual members together into a behavioral pattern, are one means through which uniformity is created in groups.

On one hand, uniformity can be beneficial to groups. It produces order and predictability as well as a sense of harmony that enables the group to move toward its goals. It can prevent too much individualistic behavior from undermining productive discussions and make the group easier to maintain, for example when everyone in the group has bought into a common set of principles.

---

38 See Festinger, supra note 37, at 272-73; Nemeth & Staw, supra note 37, at 176-89.
39 See Nemeth & Staw, supra note 37, at 189.
40 See Hackman, supra note 7, at 214.
On the other hand, pressures toward uniformity also undermine the group’s willingness (and ability) to recognize and adapt to changing circumstances. In fact, we argue below that pressures toward uniformity actually can negatively impact board performance on a number of different levels.

D. Influences on Group Beliefs

A popular criticism leveled by legislators and scholars against boards involved in recent corporate scandals is that they did not have the courage or the conviction to challenge senior management and/or the CEO on important issues. In fact, some have accused directors of being “indifferent.” While some directors, no doubt, have been indifferent, this sweeping generalization overlooks the enormous pressures toward conformity in group environments. Groups have a strong tendency to conform around ideas and specifically to congregate around those held by a majority within the group, often relying on consensus to signify accuracy. Boards are subject to the same conformity pressures.

Observers say that even a lone dissenter can make a big difference in the board room. Bill George, former CEO and Chairman of the Board of Medtronic Inc., a leading medical technology company headquartered in Minneapolis, recently cited an instance in which all but one director on his board approved of a proposed acquisition. That one director telephoned George after the meeting had ended and made such a convincing argument for his position that George reconvened the board by telephone and together the directors decided to reverse their initial approval. According to George, the board eventually made the right decision. Unfortunately, social psychological research suggests that such persistence on the part of an individual member, and such deliberation and reassessment on the part of a group, are rare. In fact, studies show that, unless addressed deliberately through appropriate norms, the structure of group relations simply does not

---

41 See Nemeth & Staw, supra note 37, at 190. See generally Irving L. Janis, Groupthink, 5 PSYCHOL. TODAY 43 (1971).
42 See Useem, supra note 24.
43 See Nemeth & Staw, supra note 37, at 189.
encourage dissent and does not enable groups to handle it productively when it does occur.\textsuperscript{45}

1. Conformity

There are two types of influence that generate conformity: normative influence, which happens when people try to gain approval by conforming to group expectations, and informational influence, which happens when people accept information from other people as “evidence about reality” and thus conform to their view.\textsuperscript{46}

The power of conformity via these two mechanisms is best demonstrated in a classic study by Solomon Asch in which subjects were asked to match the length of one line to “one of three obviously unequal lines.”\textsuperscript{47} Each subject was situated in a group of eight with seven confederates who each, when asked to state their answer aloud, provided the same incorrect answer.

One third of the time, subjects responded in agreement with the confederates, providing the incorrect answer despite the fact that the correct answer was plainly obvious. When interviewed after the study, subjects provided three different explanations for their incorrect choices. Some said they were truly unaware that their estimates were distorted by the majority. Others said that the majority seeing it differently led them to doubt their own assessment and change their answer. Finally, some said they knew their own perception was correct but did not want to “appear different from or inferior to others.” Subsequent conformity studies have produced similar results. One striking nuance to the results is that the judgments groups converged around often tended to persist in people’s individual judgments even outside of the group setting.\textsuperscript{48}

Although we like to think of board members as highly experienced, independent, and empowered individuals, as

\textsuperscript{45} See Nemeth & Staw, \textit{supra} note 37, at 185-89.


\textsuperscript{47} See \textit{generally} Asch, \textit{supra} note 46.

members of groups they are susceptible to the same pressures toward conformity that all group members are. Moreover, there are some features of boards that render members particularly susceptible to conformist behavior. Research shows that there are several conditions under which informational influence can be very powerful.

First, members of groups are more likely to rely on each other in shaping their beliefs when the group’s environment or the stimulus for the group task is highly ambiguous.\(^49\) Certainly, most situations directors face are highly ambiguous. The issues they confront are usually complex and subject to interpretation. They rarely face an objective reality and clarity regarding outcomes and costs. Instead, decisions they make require judgment about tradeoffs required to achieve various outcomes and often must be based on incomplete, complex, and subjective information. Moreover, directors must make these decisions while being conscious of the fact that they must answer to shareholders, even when it is not clear whether or not short-term tradeoffs will benefit those shareholders and/or the company in the long run. This ambiguity is even more pronounced in the tumultuous post-Sarbanes-Oxley and post-Enron environment where directors are adjusting to new regulations, making sense of changing expectations, and are, in some sense, being cast in a newly conceived role that is more scrutinized than ever before.

The second condition under which informational influence is particularly powerful is when the relevant group providing the information is perceived to be credible and competent.\(^50\) Boards are usually made up of very smart, highly influential, and experienced people. Moreover, directors are often predisposed to respect those serving on the board with them, especially when such people are hand picked. Perceived expertise also plays a major role, as directors are likely to defer to each other on items where they believe others have greater expertise.

This relates to the final condition under which informational conformity is particularly strong: when a person


believes him or herself to be relatively unqualified to make a particular judgment.\textsuperscript{51} Despite the fact that directors are generally highly qualified individuals, the flood of information with which they must work likely leaves them never feeling completely informed. Directors are often inundated with pages of information prior to board meetings. Given time constraints, directors are usually unable to process this information and it is likely that this could make a director feel as though he or she is not entirely “qualified” to respond and/or act.

It is worth noting here that some characteristics which have not received sufficient scholarly attention might also work to encourage conformity and, thus, undermine the opportunity for dissent. Boards and board members are highly visible, despite the confidentiality of what happens in board meetings. The board has joint accountability and culpability in situations where the stakes are extremely high, where losses for shareholders can be in the millions and consequences for employees can be devastating. In addition, directors are also highly cognizant of the reputational effects, both within the director community and outside of that community, of misguided dissent or of any type of individual interference or failure in the board room.

2. Suppression of Dissent

It is not just the barrier to dissent that is problematic, but also what happens once dissent is expressed. Social psychological research shows a pervasive tendency for groups to follow the majority position, even when that position is erroneous,\textsuperscript{52} highlighting behaviors that the group may use to bring the views of the minority in line with those of the majority. These behaviors include increased communication with a deviant in a group,\textsuperscript{53} holding the deviant and his or her


\textsuperscript{52} See Nemeth & Staw, supra note 37, at 183.

\textsuperscript{53} See generally Leonard Berkowitz & R.C. Howard, Reaction to Opinion Deviates As Affected by Affiliation Need and Group Interdependence, 22 SOCIOMETRY 81 (1959).
position in disdain, and rejecting or cutting off communication with the deviant if necessary.

It is worth noting, however, that minority influence can be very beneficial to a group, especially when the group is engaged in non-routine tasks where flexibility of thinking is important. Some research suggests that “majority influence causes systematic, but convergent processing of a message, whereas minority dissent stimulates consideration of an issue from multiple perspectives, even perspectives beyond what the minority proposes” in a way that benefits group outcomes particularly in problem-solving situations. Even where minorities exert some influence, however, it is rarely outwardly manifested. Members of groups often privately shift their views in the direction of the minority influence. While such private reassessment is encouraging, a public shift does not usually follow and thus the group remains under the guise of consensus and uniformity.

3. The Costs of Conformity

The consequences of this sometimes superficial but always compelling uniformity are significant. One potential cost cited by Nemeth and Staw results from the “common assumption that truth is correlated with consensus.” Groups may stick to a consensus view, even in the face of changing information, because consensus assures them their assessment or decision is correct.

This process can lead to an escalating and eventually self-perpetuating cycle where critical thinking is swept aside by the momentum the group has developed toward uniformity. Janis labeled this phenomenon “groupthink,” where the pursuit of concurrence dominates group process and overrides realistic

---

55 See Nemeth & Staw, supra note 37, at 188. See generally Stanley Schachter, Deviation, Rejection, and Communication, 46 J. ABNORMAL & SOC. PSYCHOL. 190 (1951).
58 Nemeth & Staw, supra note 37, at 189.
and critical thinking. In his analysis of the Bay of Pigs “fiasco” undertaken by President John F. Kennedy and his cabinet, Janis discovered a group characterized by strong conformity pressures and self-censorship, direct pressure on members who dissented, illusions of unanimity, and failure to search for information or explore alternative courses of action. Based on his observations, Janis suggested that groups like this tended to rush to judgment, assume invulnerability and morality in their actions, and make poor decisions.

As mentioned above, the effect of conformity pressures and uniformity in general on group performance is especially consequential to boards because they are cohesive groups. Cohesive groups tend to exert more pressure toward even greater uniformity because members have positive feelings toward their groups and value the interpersonal rewards they get from membership. As a result, members usually do not want to jeopardize these feelings or rewards by dissenting. This “affiliation” aspect of the pressure toward uniformity in cohesive groups highlights another important mechanism by which groups shape individual behavior—social identity.

E. Influence on Attitudes via Social Identity

Group memberships comprise an important part of an individual’s social identity. Studies in social psychology show that even when the grounds for group membership are arbitrarily imposed, people often have affective, cognitive, and behavioral biases in favor of the group and its members. Individuals tend to feel more positive about people in the group than about those on the outside. In fact, board members will often go as far as to believe that other group members are similar to themselves by virtue of their membership in the

59 Janis, supra note 41, at 440.
60 See generally Janis, supra note 41.
61 See supra Part III.A.
62 See Hackman, supra note 7, at 252.
64 See Moreland, supra note 63, at 1173. See generally Jacob M. Rabbie & Murry Horwitz, Arousal of Ingroup-Outgroup Bias By a Chance Win or Loss, 13 J. PERSONALITY & PSYCHOL. 269 (1969).
same group.\textsuperscript{65} Individuals also behave in ways that favor other
group members over outsiders.\textsuperscript{66} In addition, people tend to
adjust their values and attitudes over time to fall in line with attitude norms of their membership groups.\textsuperscript{67} In other words, group memberships play a significant role in an individual’s self-definition.

Social psychologists suggest that individuals look to groups both to maintain a positive self-identity\textsuperscript{68} and to reduce subjective uncertainty about self-concept, which includes their beliefs, behaviors and attitudes.\textsuperscript{69} The extent to which directors have similar attitudes about their duties, about the legitimacy of various accountabilities asserted by external parties, about what makes an effective group process, and about the group membership they share, may impact how the board makes decisions together.

In thinking about group influences on attitudes and identity we must also consider directors’ memberships in other groups. Two groups in particular are likely relevant to the attitudes directors bring to the work they do on boards. First, in addition to being directors for specific boards, directors are also members of the wider population of directors. This population is densely connected through interlocks, with most directors serving on several boards at the same time.\textsuperscript{70} The changing environment for corporate governance has made this community even tighter as it becomes more bounded by opposition to external parties that scrutinize and attempt to exert influence.\textsuperscript{71} The broader population of directors now has a visible and distinct out-group, comprised of legislators, shareholders activists, and various other critics, against which to position itself.

Directors are also members of groups in their various professional roles. For example, there is a sub-group within the


\textsuperscript{70} See Davis, \textit{supra} note 12, at 592-98.

\textsuperscript{71} See \textit{infra} Part IV.
population of directors that is comprised of current CEOs. When this part of a director's identity becomes salient in the board context, it is likely to shape his or her behavior in a given situation. Given that a primary function of the board is to oversee and evaluate the CEO's performance, this particular role affiliation may strongly influence the director role.

While economic theory describes directors who are motivated primarily by the economic incentives of membership, social psychological research suggests that there are more important reasons for becoming a board member. In fact, some corporate governance research suggests most directors join boards not for money, but rather for the sake of learning from their peers and contributing their expertise and experience to the management of other companies. Thus, in order to properly study board outcomes, we must understand not just the economic incentives of directors, but the benefits and incentives derived from being part of the group and the broader director community.

F. Habitual Routines

A look inside one crucial board meeting at Enron Corporation, the now bankrupt energy company whose leaders concocted off-the-books partnerships, twisted accounting rules, and manipulated the energy market to inflate the company's profits and siphon money into their own accounts, shows how even the most alarming signs of wrongdoing can be missed. On June 25, 1999, three days prior to their next board meeting, each director of Enron received a proposal in their fax machines for suspending the ethics code of the company. CFO Andy Fastow had asked for the approval of a self-serving partnership which required the suspension of the code's mandate that "even an appearance of an improper transaction must be avoided" and that "no employee should gain separately from company service." The subsequent board meeting was conducted by phone in less than one hour and was "jam-packed" with important agenda items and topics requiring the board's approval. Directors proceeded expeditiously through transactions and proposals and, although no committee had

---

72 See generally Seymour Lieberman, The Effects of Changes in Roles on the Attitudes of Role Occupants, 9 HUM. REL. 385 (1956).
73 See Lorsch & MacIver, supra note 8, at 23-30.
74 See generally Useem, supra note 24.
vetted the suspension of the ethics code, it was approved by the group. This decision turned out to be a terrible mistake, as Fastow went on to add millions to his personal fortune by exploiting Enron and its shareholders.

Habitual routines, in the terminology of social psychologists, can undermine the activation of careful and scrutinizing decision-making in situations like the one just described. “A habitual routine exists when a group repeatedly exhibits a functionally similar pattern of behavior in a given stimulus situation without explicitly selecting it over alternative ways of behaving.” Habitual routines are distinct from norms because they do not emerge from group assumptions about appropriateness and their enforcement does not involve a response to someone who deviates from expectations. Rather, these behavioral patterns happen to the group as a whole.

Habitual routines can be both functional and dysfunctional for any group. First, because they function automatically in the absence of members’ conscious attention, they save the group time and energy, particularly in stable environments. Under such conditions, the group does not have to actively manage every situation but, rather, automatically engages in behaviors cued to a particular situation. Because the group is not actively assessing whether or not its habitual routine is appropriate in the given circumstance, however, it always runs the risk of failing to adapt to important changes in environmental stimuli. Specifically, the group may miscode a situation and proceed with a particular set of processes that are inappropriate to the situation and, thus, undermine the group’s performance.

The 1982 crash of Air Florida Flight 90, as analyzed by Gersick and Hackman, illustrates the disastrous outcomes that can result when habitual routines govern group behavior. The data collected from the cockpit voice recorder reveal that the crew carried out its ordinary takeoff routine, including a confirmation that the “anti-ice” indicator was “off” despite the fact that the current weather conditions required that it be on. The sad irony was that the crew continued to comment as to the frigid conditions and difficult weather while carrying out

76 Id. at 65-67.
their behavioral routine as if the conditions were dry and warm. The experiences of Flight 90 show how powerfully a group can be governed by its routines.

Habitual routines govern a good deal of group behavior and are likely are vital factor when it comes to explaining board behavior as well. Perhaps it was reasonable for the Enron board to expediently approve the agenda items that, while important, probably resembled transactions they had discussed countless times before. This time, however, the directors were approving a suspension of their code of ethics, certainly not a routine matter and most likely one that was qualitatively different from anything they had done before. Because of the board’s habitual routine of disposing of agenda items, this distinct undertaking was not cued as something that required a reassessment of how the group would discuss and process the decision.

Boards have been shaken up by the fallout from recent oversights like this. Experiencing failure or receiving an intervention are two of only a few factors that can make habitual routines salient and make group members aware of a need to recode certain situations. Indeed, even boards not beset by scandal are becoming aware of the pertinent issues at hand, and many of them are now reflecting on, and revising, their own organizational processes.

A prominent director of a Fortune 500 company has noted how executive session processes can be reconceptualized when stimulated by new conditions.

[T]his is where collegial process really worked, where somebody comes up with a thought that maybe is not even a prearranged thought in the person’s mind, he just says “well you know I’ve been thinking about this” . . . it sparks a comment from somebody else and then somebody else and then you come up with something. Sometimes it’s strategy, sometimes it’s nuance, sometimes it’s something we don’t like sometimes it’s something we do like.

Here, the director draws a contrast between an “old way” and a “new way” of doing things. The routines and processes directors once adhered to are currently being reshuffled and being actively shaped, most likely to the benefit

---


78 All quotes from directors in this article are from directors of Fortune 500 companies to whom confidentiality was assured.
of the process. As the external shock subsides, however, whatever processes boards adopt will reflect the same tendency toward routine that usually shapes group behavior.

IV. BOARDS IN A CHANGING ENVIRONMENT

As we consider the social influences that shape group behavior and board outcomes from within, we must not forget that boards are also embedded in a wider, social context—a complex and increasingly institutionalized environment. Corporate governance researchers in the 1990s recognized the importance of this environment to boards but focused primarily on the role of investors and their interests in driving board processes and outcomes. Although these issues are important, we believe there is a more subtle and more powerful environmental shift affecting boards.

There is no doubt that boards have evolved from being practically impenetrable groups to being more easily pressured to adopt certain forms and functions to have legitimacy. In fact, an increasingly organized environment has, and will, continue to affect boards through its ability to exercise authority over how boards should work. It does this by creating a set of broader norms that are thought to improve governance and lead to expectations that must be met to establish legitimacy. Sociologists would term such an evolution the “institutionalization” of the board, an evolution that certainly will penetrate the group context of the board and add another dimension of social influence to those which have been described above. By institutionalization, we mean a range of influences, controls, patterns, and tacit understandings that make up the whole corporate governance field. Institutional pressures emanate from the relational networks of organizations that arise in the broader societal context. This includes many elements beyond the boundaries of any single board. It includes the consensual notions held by ordinary investors about what a board ought to be like, how institutional investors believe board members ought to behave, and the network of governance rating agencies seeking to influence board behavior and governance outcomes, among others.

Concretely, the three domains effecting this institutionalization are the legal, the consultative, and the educational domains. First, legislators have just completed, through Sarbanes-Oxley, a prescriptive intervention more extensive than any that has come before, putting forth new
independence requirements, stipulating committee structures and frequency of meetings, and mandating director qualifications. Many of the recent actions boards have taken are efforts to demonstrate compliance to new legal standards. Organizational scholarship suggests that boards can easily go through the theatre of achieving legal compliance while doing little to improve the underlying quality of their governance. Compliance to legal mandates of what defines an independent director, for example, doesn’t guarantee psychological independence. The adherence to legal prescriptions does not necessitate that those principles are manifest in board behavior. Given the opaque nature of corporate governance activities, it is fairly easy for boards to adopt the structures that signal good governance on paper but are, in fact, loosely coupled to actual board activities. Requiring board committees to have a written charter does not ensure that directors will adhere to the spirit and not just the letter of that charter. As several commentators have noted, Enron’s board was upheld as a paragon of high quality, independent governance.

The various intermediary institutions that have evolved with respect to corporate governance have been even more influential in shaping the language and the expectations around boards of directors. Ratings agencies scrutinize boards based on checklists of desirable structures and processes, implicitly linking conformance to these standards to board performance. A whole industry has emerged around peddlers of governance best practices, with scores of consulting firms opening corporate governance practices and offering both consulting and training programs to directors. Even the public relations firms have gotten into the game, recognizing how important it is for boards to communicate to shareholders and the media in a way that is in line with best practice and restores trust. Several PR giants, including WPP Group PLC and Interpublic Group of Cos. Inc., have devoted business units to corporate governance.79

Finally, there is an emerging “professional” logic of the director position. Universities and business schools, in addition to consulting firms, have seen the need to develop executive programs for directors. Business schools are making an effort to include corporate governance courses in their MBA

---

curriculum. Ethics courses have appeared in response to the corporate scandals, discussing director roles in corporate governance outside of explicitly director focused courses. Built into the educational process itself is a system socializing students to a particular set of norms, values and dispositions that are created simply by living in, and responding to, institutional expectations and the routines of the course work.

As this continues, and as other influences including auditors and regulators contribute to professionalization, a stronger set of norms will shape board room behavior. Boards will be increasingly concerned with, and driven by, concerns for legitimacy. Thus far, and as evidenced in a recent study on the adoption of the Lead Director position, directors’ pursuits of legitimacy have been primarily concerned with the opinions of other directors on their boards and perhaps those of the wider population of directors, when it comes to signaling legitimacy.

As the environment around boards more effectively imposes its own set of norms, however, the ways in which boards function as groups will surely change to respond to what is seen as legitimate in that environment.

We end our essay where we began, calling for a reconsideration of the dominant law and economics perspective about the nature of boards. Radical theoretical individualists will undoubtedly disagree and contend that all corporate phenomena can be reduced to individual motivation and the firm to a nexus of individual contracts. But this hypothesis does not adequately account for corporate behavior today, as several autopsies of corporate governance failures have illustrated. We have suggested that there is much to be gained by examining boards not simply as an aggregation of individual contracts, but also as a singular social unit. This road may be more arduous than the dominant perspective. Its research and study requires a detailed analysis of the social structure of the board and recognition of the complexity of motivations that underlie behavior. We believe, however, that this route is more analytical and empirically defensible than the current theory. We also believe it opens up tremendous opportunities for new types of research methods, especially qualitative research that emphasizes field work and interviews.

\[80\] See Pick, supra note 33.
Promoting Employee Policy Adherence and Rule Following in Work Settings

THE VALUE OF SELF-REGULATORY APPROACHES

Tom R. Tyler

ABSTRACT

Securing employee adherence to workplace rules and company policies is one key antecedent of successful coordination and functioning within organizations. It is important for companies to be able to motivate effectively rule-following behavior among employees. This analysis highlights the value of identifying optimal approaches to securing such behavior. In this paper, two strategies for achieving policy adherence and rule following are compared. Those strategies are: (1) the sanction-based command-and-control model and (2) self-regulatory approaches that are linked to activating employees’ ethical judgments. Research findings suggest that, while command-and-control strategies influence employee
behavior, self-regulatory strategies have a stronger influence. Studies also explore the basis of these ethical judgments and find that the primary factor shaping them is the procedural justice that employees experience in their workplace. These results suggest that the roots of employee policy adherence and rule-following behavior lie in the procedural justice of the organization. Overall, this analysis highlights the important role ethical judgments play in motivating both rule following and policy adherence among employees in work settings and provides practical suggestions for shaping those judgments.

INTRODUCTION: CAN BUSINESSES EFFECTIVELY REGULATE EMPLOYEE CONDUCT?: THE ANTECEDENTS OF RULE ADHERENCE IN WORK SETTINGS

Can businesses effectively engage in the internal regulation of employee behavior, and if so, what strategies should they use to achieve best that objective? Recent corporate scandals have evoked a heightened concern among members of the public, government officials, and business leaders both about whether businesses can regulate the conduct of their employees and how to secure effectively employee adherence to corporate rules and policies. Such adherence is important in a wide variety of work settings and involves organizational policies that cover, among other things, accurate accounting, conflicts of interest, product or service quality, environmental safety, sexual harassment, and race, gender and/or sexual orientation discrimination. In these and many other ways, gaining adherence to organizational policies that control everyday employee behavior is critical for successful organizational functioning.¹

Unfortunately, there has long been extensive evidence that in many of these areas noncompliance within organizations is widespread.² Such issues of compliance and


noncompliance have been dramatically thrust into the public eye through recent highly-visible incidents of corporate misconduct. The prevalence and damaging consequences of such non-compliance underscores the importance of identifying an effective model of employee rule adherence. Businesses would benefit from such a model since it would allow them to shape employee conduct in desirable ways. Further, from a policy perspective, government agencies are more likely to feel that the active regulation of businesses is important if they believe that businesses lack an effective model for self-regulation.

Of course, it is also important to recognize that a wide variety of other issues are implicated in recent corporate scandals. In particular, in some cases the problem is linked to misbehavior among corporate leaders—i.e., CEOs. The focus of this paper is not on the leaders of corporations, but on employees within them. In particular, this paper does not consider the case in which leaders are creating an unethical climate within their companies so that they can break rules for personal profit. Rather, this paper begins with the assumption that the situation can be one in which the leaders of a company are motivated to encourage their employees to follow rules and are seeking to understand how best to do so.

Similarly, from the perspective of the law and legal institutions, this analysis assumes that legal authorities are interested in motivating employees to follow the law and are trying to understand the strategies that companies should be encouraged to follow to achieve this objective. In this case, the arguments outlined may well apply to corporate leaders as well as employees. Legal authorities need to create a strategy that will motivate corporate leaders to follow the law, and the arguments outlined here apply directly to that task.

I. BACKGROUND

My goal is to compare the utility of two approaches to employee regulation: the command-and-control model and the self-regulatory model. The command-and-control model represents a traditional approach to encouraging rule following insofar as it operates by drawing upon employees’ instrumental concerns and utility-maximization goals. Specifically, the

89 Cal. L. Rev. 917, 917-98 (2001); Yoav Vardi & Ely Weitz, Misbehavior in Organizations: Theory, Research and Management 3-4 (2004).
command-and-control model links employees’ motivation to follow rules to the manipulation of sanctions in the workplace. It is based on the view that people follow rules as a function of the costs and benefits they associate with doing so.

The command-and-control model reflects a strategy of **external regulation** whereby employee behavior is controlled by managers through their ability to implement sanctions and to punish undesired behavior. In contrast, the self-regulatory model is based upon the activation of **internal motivations**. This distinction develops from prior social-psychological research, which distinguishes between compliance based upon external contingencies and self-regulation linked to identification and internalization. The **self-regulatory model** represents an alternative approach to employee rule following. The model emphasizes the role that employees’ ethical values play in motivating rule following and, in particular, those ethical values that are related to—and developed in the course of interactions with—their work organization. That is, I focus on those ethical judgments that are linked to employees’ specific experiences at their work organizations. This can be contrasted to a focus on individual differences in ethical judgments—i.e., to those aspects of people’s personalities that shape how they judge particular ethical matters. My focus on organizationally-based ethical judgments is rooted in an interest in determining the characteristics of work environments—as opposed to individuals—that may shape employee rule following. This emphasis has the potential to be of particular utility to leaders and managers in their attempts to design workplace environments that foster rule-following among employees.

Two specific ethical judgments that are linked to organizational conditions are considered here: (1) the perceived legitimacy of organizational rules and authorities and (2) the congruence of those rules with an employee’s moral values. The self-regulatory model argues that the concerns embodied in these two ethical judgments have the potential to motivate employees to feel a personal responsibility for bringing their behavior into line with corporate rules and policies. It is based

---


4 This distinction is extended to organizational arenas, Herbert C. Kelman & V. Lee Hamilton, Crimes of Obedience 103-12 (1989) and to work settings, Charles A. O'Reilly & Jennifer A. Chatman, Organizational Commitment and Psychological Attachment, 71 J. APPLIED PSYCHOL. 492, 492-99 (1986).
on the assumption that people are motivated to align their behavior with the rules of organizations or groups they belong to when they view those groups as being legitimate and consistent with their own sense of right and wrong.

The first goal of this analysis is to compare the relative efficacy of the two distinct strategies outlined. While the use of sanctions represents a traditional management strategy to securing employee compliance with organizational rules and policies, I consider recent studies that directly examine whether activating employees’ ethical values is an effective management strategy for securing their compliance. The use of such a self-regulatory model has been long advocated within discussions of legal regulation of business,5 and has been advanced with particular frequency in recent years.6 The studies examined test whether employees’ ethical values can in reality—as hypothesized by self-regulatory models—provide a viable basis for encouraging employee policy adherence.

The second goal of this paper is to examine the antecedents of employee ethical values. To the extent that the self-regulatory model represents and describes an important influence on employee policy adherence, it becomes important to understand the factors that shape whether or not employees come to hold ethical values that encourage such adherence. Drawing upon the literature on procedural justice, it is hypothesized that employees’ ethical values will be activated and will be more salient in decision making when employees evaluate their organization as being governed according to fair procedures. This prediction is linked to one of the core hypotheses of the group engagement model: that procedural-

---

justice judgments are central to shaping employee cooperative behavior. This procedural-justice hypothesis has been supported by prior studies of rule following in legal and managerial settings, although it has not received universal support. If supported by research, this model provides a theoretical perspective within which managers can develop a strategy for activating employees’ ethical values in work settings and thus secure employee compliance with work rules and policies.

What are the behaviors we are interested in motivating employees to engage in?

There are several frameworks within which to conceptualize the ways in which employees may follow or break organizational rules, and this study will examine each of them. Two aspects of policy-related behavior are considered here: policy adherence and rule breaking. On the one hand, organizations want employees to adhere to organizational policies. Organizational rules and policies stipulate desired employee behavior, and the organization benefits when those policies are followed. For example, organizational rules often specify behaviors about how work should be carried out, when people arrive at work, etc. Such rules facilitate coordination between employees and ensure the smooth functioning of the organization. This aspect of rule following involves conformity to organizational policies since it encourages employees to align their behavior with organizational rules.

I further distinguish between two forms of policy-adherence behavior: conformity with organizational policies and voluntary deference to organizational policies. The roots of this distinction lie in the literature on obeying the law, which distinguishes between compliance with the law and voluntary, willing acceptance of the law. The same distinction is important in work settings.

---

10 See Cooperation in Groups, supra note 7, at 77-89.
12 See Kelman, supra note 3, at 51-60 (1958); Tyler, supra note 8; Tom R. Tyler & Yuen J. Huo, Trust in the Law 47-76 (2002).
13 See O'Reilly & Chatman, supra note 4, at 492-99.
The distinction between these two forms of behavior lies in the circumstances under which employees indicate that they follow rules. In terms of compliance, people indicate how often they follow the rules across all settings. With voluntary, willing acceptance, on the other hand, they indicate whether they follow the rules even when they do not have to, when no one is around, and when their behavior is not being monitored. In other words, when it comes to voluntary deference, people choose to follow the rules even when failing to do so will not be detected. Hence voluntary deference refers to rule following in that subset of situations in which issues of detection are largely or completely irrelevant.

On the flip side of conformity or deference to organizational policies lies deviant behavior by employees, or behaviors that are damaging and prohibited by organizational rules. For example, employees may use office supplies for personal use or use sick leave when not sick. More seriously, employees may steal or break organizational rules by lying and cheating. I refer to this deviant behavior as rule breaking because it involves the decision to ignore or violate organizational rules.

Naturally, companies want to reduce the degree of rule breaking that occurs among employees. For instance, a widely damaging form of inappropriate employee behavior is theft of business supplies and equipment. It is estimated that 30% to 50% of all business failures are linked to losses from employee theft, a problem that is ten times more costly than street crime in terms of loss to society, and whose costs are often estimated to be in the hundreds of billions of dollars in the United States alone.\(^\text{13}\) Again, the magnitude of these losses, and the suggestion that up to 75% of employees engage in theft in their workplace, indicates the challenge posed in trying to manage this problem.

II. MODELS OF MOTIVATION AND POLICY ADHERENCE

Command-and-control. The command-and-control perspective focuses on controlling people's behavior via the threat of punishments or sanctions for misbehavior. To the degree that employees are motivated instrumentally—and are

---

\(^{13}\) See Jerald Greenberg, The STEAL Motive: Managing the Social Determinants of Employee Theft, in ANTIMALICIOUS BEHAVIOR IN ORGANIZATIONS 85 (Robert Giacalone & Jerald Greenberg eds. 1997).
thus primarily interested in the resources and outcomes they receive from their organizations—some external authority, either the company or the government, needs to take an active role in enforcing rules regarding their conduct. In other words, to the extent that employees are extrinsically motivated, extrinsic forces are needed to regulate their behavior. In organizational settings, such extrinsic forces typically take the form of incentives (to encourage desired behavior) and sanctions (to discourage undesirable behavior). Incentives and sanctions in many ways represent two sides of the same extrinsic motivational coin—each is an organizational mechanism used to control employee behavior via employees’ concerns over the resources and benefits the organization provides them. There is already discussion in the organizational literature about problems with incentives, as well as a parallel discussion regarding the potential inadequacies and pitfalls of punishments as motivational tools. Many of the features of the modern workplace are the product of the use of command-and-control model. For example, the extensive use of surveillance techniques—such as the use of cameras, the monitoring of telephone calls and computer usage, etc.—is an artifact of the implementation of command-and-control techniques. Random drug testing, searching employees’ cars and lockers, and the use of time clocks and other performance-tracking devices similarly reflect the view that compliance develops from a credible fear of detection and ensuing sanctions. This instrumental strategy addresses the issue of employee motivation from the perspective of traditional economic theory—i.e., by assuming that employees are rational actors who are concerned primarily about maximizing their own outcomes in work settings. Studies generally support the suggestion that instrumental strategies do, as expected, shape people’s behavior, with some studies supporting this argument in work settings.

14 See generally Alfie Kohn, Punished by Rewards (1999).
17 See Daniel S. Nagin, Criminal Deterrence Research at the Outset of the Twenty-First Century, in 23 Crime and Justice: A Review of Research 1, 12-23
However, the use of instrumental strategies—and the command-and-control strategy in particular—requires the availability of resources. For sanctions and deterrence systems to work, organizations must be able (and willing) to devote significant resources to the surveillance needed to make detection of rule breaking sufficiently likely so that people are deterred. The cost of such surveillance should not be underestimated, since employees are inherently motivated to conceal their rule-breaking behavior and effective surveillance systems are essential for sanctioning systems to shape behavior. Incentive strategies do not have surveillance problems, but require the availability of resources for incentives as well as a system to define and evaluate performance.

In addition to their financial costs to the organization, there are also social costs associated with command-and-control systems. These systems have the potential to communicate a message of mistrust in employees, conveying a sense that the organization is an adversarial force to the employee. Significant repercussions on employee commitment and identification with the organization may thus result. Furthermore, interpersonal dynamics may often be affected, as employees that maintain surveillance systems are pit against those being scrutinized.

Perhaps most importantly, it is also not clear how effective command-and-control strategies are. For example, in legal settings sanction-based deterrent strategies are consistently found to have, at best, a minor influence on rule-breaking behavior. In his review of the deterrent effect of drug laws, for example, MacCoun finds that only about five percent of the variance in drug use is explained by deterrence factors. Based upon their workplace-based study, Tyler and Blader

---


20 See generally MacCoun, supra note 19.
estimate that around ten percent of the variance in employee behavior is shaped by incentives in the work environment. These results suggest that, while such systems are somewhat effective, they may only have a limited impact on employee behavior.

More generally, in recent years the limits of the command-and-control model have been noted. However, this increasing skepticism has occurred within the arena of legal regulation, and less so in discussions of work organizations. Thus, the managerial relevance of these critiques remains an open issue.

Of course, command-and-control strategies do not only exist within organizations. Organizations also function within a framework of government-imposed legal prohibitions and administrative requirements that are also based on incentive and sanction systems. Even at this more macro level, the utility of those systems has been increasingly questioned. For instance, they have been referred to as “ossified” systems that make “compliance difficult and impractical.” An additional difficulty often noted in this domain is the problem of monitoring behavior. Within the legal literature on government regulation, such skepticism about command-and-control strategies has lead to the flourishing of market-based models of regulation that emphasize economic incentive systems.

21 See Cooperation in Groups, supra note 7, at 38-42.
23 See Tyler & Huo, supra note 11, at 19-24.
25 See Spence, supra note 2, at 918.
Self-regulation. An alternative model of employee policy adherence is one in which the motivation to follow organizational rules resides in the employees themselves and not in extrinsic incentives or sanctions stipulated by the organization. According to such a model, employees can be intrinsically motivated to follow organizational rules—that is, they will do so out of their own desires and not in response to the regulations put in place by the organization to provide sanctions for employee misbehavior and/or incentives for desired employee behavior. The self-regulatory model tested in these studies specifically examines the role of employees' ethical values in shaping intrinsic motivation to follow rules. The success of this approach depends upon the power of employees' ethical values to motivate their rule and policy-following behavior in the workplace.

Calls for greater attention to ethics in business school curricula and for more attention to ethical issues in work cultures flow from the belief that employees' ethical values can be developed and activated within work settings. This belief, when combined with the assumption that ethical values can have an important role in shaping behavior, thus argues for the importance of corporate cultures that shape ethical values in ways that promote employee policy adherence. That is, to the extent that ethical values affect employee rule following, the challenge is to create organizational cultures that harness the motivational power of employees' ethical values.

Several types of evidence suggest that ethical values may shape employee behavior. Research suggests that ethical concerns motivate self-regulatory behavior in organizational settings. This includes studies focused on legitimacy, on

---


morality, and on the general role of fairness in shaping social behavior. Ethical values that encourage people to support the organization shape behavior when those people believe that the rules of their organization are legitimate (and hence ought to be obeyed) and/or that the values defining the organization are more congruent with their own moral values.

There is evidence of the importance of ethical values at the organizational level as well. Studies show that companies are reluctant to use their market power to lower employee wages during recessions because they believe such an action will be viewed by employees as unethical, that companies often forgo opportunities to press their market advantages when dealing with their customers due to ethical concerns, and that ethical issues shape wage determinations as well as other aspects of the employment relationship. These studies argue that companies are motivated to respond to ethical issues because they believe that ethical judgments shape people’s reactions and behavior, an argument supported by studies suggesting that companies regarded as ethical by employees, customers, and other constituencies are more profitable.

I focus on the influence of two particular types of ethical values. The first is the belief held by employees that their organization’s rules and authorities are legitimate. Legitimacy refers to the view held by employees that they are responsible for obeying organizational rules—e.g., that the organization is

31 See Raymond Paternoster & Sally S. Simpson, Sanction Threats and Appeals to Morality, 30 LAW & POL’Y 549, 549-84 (1996); Tyler, supra note 8; COOPERATION IN GROUPS, supra note 7, at 72-75; Tyler & Blader, supra note 30.
34 See generally Daniel Kahneman et al., Fairness and the Assumptions of Economics, 59 J. BUS. 5285 (1986).
entitled to have its rules and policies obeyed. Although early discussions of legitimacy, such as the work of Weber, focus on the perceived legitimacy of government and law, it is clear that legitimacy is also an important concept in the context of work organizations. In work settings, legitimacy refers to the judgment that "the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions." If people feel that their organization has legitimacy, they are motivated to defer to its rules and policies.

The second ethical value is the belief held by employees that corporate policies are congruent with their own personal moral values. If employees believe that such value congruence exists, they will be motivated by their own moral values to follow corporate rules because they will see those rules as being consistent with—and developed from—a set of moral values with which they agree. Thus, they may follow rules in their effort to do what they feel is morally right. For example, in legal settings an important motivation that encourages people to bring their behavior into line with the law is their belief that many behaviors that are illegal are also immoral. Similar moral values are found to shape cooperation within experimental games. If people feel that their organization acts in ways consistent with their own moral values, they are more strongly motivated to support their organization.

Conversely, in situations in which employee behaviors are contrary to official policy but viewed by people as not being immoral—such as drug use, some sexual practices, and the illegal use of copyrighted software—it is more difficult to bring people's behavior into conformity with the law. Employee theft may be another behavior that violates corporate policy but that is not viewed by employees as immoral when it is done to restore the equities in the employee/employer relationship.

40 See SELZNICK, supra note 5, at 139-43 (1969); Suchman, supra note 6, at 571-610.
41 See Suchman, supra note 6, at 571-610
42 See generally TYLER, supra note 8.
Similarly, employees in work organizations evaluate the morality of company policies and practices and react to those policies and practices in moral terms.44 Adherence to those policies is more likely when they are viewed as morally appropriate.

III. **ETHICAL VALUES AND WORKPLACE RULE ADHERENCE**

The findings of recent research support the argument that employees’ ethical values shape their behavior and, in particular, their rule-following behavior. One example is provided by Tyler and Blader,45 who reported two studies: one of a sample of corporate bankers and another of a large and diverse sample of American employees.46 Analysis of both samples indicates that employee rule following and policy adherence was strongly influenced by employees’ ethical values.47 This included distinct influences of legitimacy and moral-value congruence.48

These findings suggest that companies benefit by fostering ethical values in their employees that support rule following. Those ethical values serve as a major motivation for employees to comply with company policies and rules and consequently lead to lower levels of rule-breaking behavior on the part of employees. These results suggest that one promising way to bring the behavior of corporate employees into line with corporate codes of conduct is to tap into their ethical values. Because these values are central to the self-regulatory strategy for achieving employee compliance, companies should activate employee values in order to gain acceptance for corporate rules and policies.

Of course, the activation of employee values is not the only way to influence rule-related behavior. Organizational sanctions for rule-breaking may likewise motivate employees to follow organizational policies, as suggested by the command-and-control model. However, in the two studies reported here, the utility of that approach appears to be smaller in magnitude. These findings suggest that companies have a great deal to gain by going beyond instrumental strategies of

44 See Paternoster & Simpson, supra note 31.
45 See Tyler & Blader, supra note 30.
46 Id.
47 Id.
48 Id.
social control and focusing their attention on the activation of employee values that are consistent with a self-regulatory strategy. Overall, studies indicate that such a strategy is viable and, furthermore, that this strategy is superior to the more traditional command-and-control approach.\textsuperscript{49}

The empirical support outlined above suggests the utility of the self-regulatory strategy. Such an approach also has benefits over a command-and-control strategy. For instance, it prevents organizations from expending resources on creating and maintaining credible systems of surveillance to enforce rules. These enforcement problems are typical of any efforts to regulate conduct using incentive or sanction-based strategies. Even worse, such strategies actually encourage people to hide their behavior and thus make it necessary to have especially comprehensive and costly surveillance systems.

Besides their actual costs, an additional problem associated with these strategies is that they undermine employees' commitment to their company and enjoyment of their jobs. Employees' intrinsic motivations and commitment to their company is undermined when their focus is on avoiding sanctions and, as a consequence,\textsuperscript{50} they contribute less to their workplaces. Hence the downside to sanctions and the surveillance associated with them is that these measures hurt company productivity by undermining the ethical values that encourage commitment to work.\textsuperscript{51}

This is not to say that command-and-control systems cannot work. They can, especially if organizations devote sufficient resources to them. For example, some companies engage in extensive monitoring, even putting cameras in restrooms and monitoring telephone and e-mail communication. They may also try to create conditions under which behavior is easily monitored by, for example, requiring employees to time punch in and out of their workplace, to sign out equipment or tools, or to work in publicly-accessible spaces. Clearly, such efforts consume organizational resources. Even if they work, these strategies are costly and inefficient.

The findings of the studies considered point to the potential value of using the self-regulatory approach to motivate employees. By activating employees' own ethical

\textsuperscript{49} Id.

\textsuperscript{50} Frey, supra note 15, at 88-104.

\textsuperscript{51} See Cooperation in Groups, supra note 7, at 55-57.
values, companies can gain willing cooperation from their employees. By having people regulate themselves, such willing cooperation becomes much more efficient and effective. In such a model, employees take on the responsibility to follow rules and undertake this responsibility without being concerned with the likelihood of being caught and punished for wrongdoing.

In recent decades, it has become widely recognized that self-regulation has value. Self-regulation is widely touted as a means of lessening the costs of government regulatory agencies and generally avoiding the problems that occur when government seeks to regulate business. These same arguments can be applied within companies. Companies benefit when they can develop self-regulatory strategies that encourage their employees to take increased responsibility for rule following.

Earlier studies in the area of everyday law-related behavior highlight the important role ethical values play in encouraging citizens to comply with the law. It has been shown that people are more likely to comply with laws when they feel that legal authorities are legitimate and ought to be obeyed. The findings noted support this argument and extend it to a different arena—employees and their relationship to their corporate employers. Recent corporate scandals have highlighted the importance of understanding better how to motivate employee compliance with corporate codes of conduct.

The influence of ethical judgments in these studies is especially striking because the influence of ethical values in the work arena has traditionally been downplayed in favor of alternative instrumental or "rational" approaches. These studies suggest that a model of motivation that only considers rational motivations is incomplete and does not take account of the important role that social motivations can play in shaping employee rule-following behavior.

The current findings also extend previous work by considering the social value of value congruence (i.e., the match between the person’s moral values and those of the organization) in addition to that of legitimacy. In other words, people who experience justice when dealing with their work


53 TYLER, supra note 8.
organization first think that its rules are legitimate and ought to be obeyed. They also feel that the values of their work organization are more congruent with their own, so that their own motivation to behave morally leads them to support their work organization. Overall, these findings support the argument that developing an appropriately ethical organizational culture is central to the effectiveness and viability of corporations.

It is especially striking that voluntary deference is linked to ethical motivations. Organizations recognize that they depend heavily on the good-will of employees who are motivated to go beyond their job descriptions and to defer to rules even when surveillance is weak. Such voluntary behavior is central to organizational effectiveness and is strongly motivated by legitimacy and moral congruence.

IV. WORKPLACE POLICIES AND PRACTICES AND EMPLOYEE ETHICAL VALUES

The self-regulatory model operates via the activation of employees’ ethical values and feelings of responsibility toward their company. The group engagement model hypothesizes that factors such as employees’ ethical values are shaped by employee perceptions of how fairly they are treated by management. As has been noted, the potentially important role of fairness in motivating positive work attitudes and behavior has been recognized by economists as well as by social and organizational psychologists. This approach is based upon a psychological model suggesting that an organizational environment characterized by fair procedures will activate strong employee organizational identification, thus leading employees to engage in desirable workplace behaviors and to hold positive attitudes towards their work organizations.

Various aspects of an organization’s policies, human resource practices, and culture may potentially influence employee rule following and employee’s ethical values regarding their work organizations. One set of management theories argues that the primary organizational factor shaping employees’ reactions to their work organizations is the

distribution of outcomes in the work environment. According to these theories, employee attitudes and behaviors are responsive to judgments about the favorability of the outcomes (i.e., resources) provided to them by corporate rules and policies, as well as to the incentives and sanctions associated with their workplace behavior. These arguments flow from an instrumental model that views workers as motivated to maximize the outcomes they receive from their work organizations.

Psychological models of equity and distributive justice also suggest that employees are instrumentally motivated and focus on outcomes. The difference, though, is that these psychological models focus on issues of distributive fairness. They suggest that employees are sensitive to whether or not they feel that they are receiving a fair level of wages and benefits. These models are premised on the idea that workers, recognizing that people cannot have all they want, subsequently form their judgments of whether they are receiving their fair share of workplace resources according to how they react towards their work organization.

An alternative set of management theories argues that employee reactions to their work organizations may be based on their judgments about the fairness of the procedures used in their workplace. Factors affecting these fairness judgments may include, for example, whether the procedures allow employees to have input into decision-making processes, whether they require that objective information be used in decision making, whether efforts are made to reduce biased treatment, etc. Widespread evidence from all types of organizations attests to the importance of procedural-fairness judgments in shaping the behavior of employees in work settings. Typical of this research is a study by Kim and


58 See J.A. Colquitt et al., Justice at the Millennium, 86 J. Applied Psych. 425, 425-45 (2001); R. Cropanzano, Moral Virtues, Fairness Heuristics, Social Entities, and Other Denizens of Organizational Justice, 58 J. Vocational Psych. 164, 164-209 (2001); Advances in Organizational Justice passim (Jerald Greenberg & Russell Cropanzano eds., 2001); Authority in Groups, supra note 57, at 58; Tom R. Tyler et al.,
Mauborgne that demonstrates that procedural-justice evaluations influence the willingness of subsidiaries to accept corporate strategic policy decisions in multinational work organizations.\textsuperscript{59} Other studies link the fairness of workplace procedures to employees’ willingness to help their work groups voluntarily, to their intention to stay with their company, and to the quality of their job performance.\textsuperscript{60}

The procedural-justice argument is based upon the belief that people’s procedural-justice judgments are distinct from their instrumental concerns. That is, their reactions to their judgments about the fairness of their organization’s procedures is not related to goals they may have regarding the outcomes that they receive from their organization. Instead, they react to procedures because they make inferences about their relational connections and social identities based on the fairness of those procedures.\textsuperscript{61} These social-identity judgments about issues such as their standing in the organization, the status of the organization, and their level of identification with the organization, in turn influence their workplace attitudes and behaviors.\textsuperscript{62} When organizational procedures are regarded as fair, employees feel that they can safely identify with the work organization and thus become engaged in it.\textsuperscript{63} This approach is based on the idea that people are influenced by the nature of the organizational environment in which they work so that the “fit” between the practices of the organization and a person’s impression of themselves (including their ethical values) is important.\textsuperscript{64}

The findings of procedural-justice research lead us to hypothesize that procedural-justice judgments will impact: (1) employees’ views about the legitimacy of corporate rules, policies, and authorities, (2) employee perceptions that their

\textsuperscript{60} See COOPERATION IN GROUPS, supra note 7, at 77-89.
\textsuperscript{61} See Authority in Groups, supra note 57, at 177.
\textsuperscript{62} See COOPERATION IN GROUPS, supra note 7, at 143-68; Tom R. Tyler & Stephen L. Blader, Identity and Cooperative Behavior in Groups, 4 GROUP PROCESSES AND INTERGROUP RELATIONS 207, 207-26 (2001); Tyler & Blader, supra note 30.
\textsuperscript{63} Tyler & Blader, supra note 30.
organization’s values are consistent with their own, and (3) employees’ rule-related behavior. In other words, fair organizational procedures and processes are hypothesized to foster a sense that corporate authorities are legitimate and that the organization itself possesses moral values similar to those of the individual. This activates employees’ own internal motivations, and they follow company rules and policies more voluntarily—i.e., they become self-regulating.

Note that this approach can be contrasted to one in which employees’ ethical values are shaped by their instrumental concerns. That is, the two instrumental judgments discussed earlier—the favorability or fairness of outcomes received from the organization—may shape the extent to which corporate authorities are viewed as legitimate and the organization itself is perceived as possessing moral values similar to those of the individual. This would be the prediction of instrumental models that emphasize the concern employees have over the outcomes they receive.

We can consider the antecedents of employee ethical values by investigating the relative influence of employees’ outcome judgments (such as outcome favorability and outcome fairness) and procedural-justice judgments. The issue is which of these judgments most strongly shape employee perceptions that (1) organizational rules and authorities are legitimate, and (2) that their personal moral values are consistent with those of the organization. To the extent that employee ethical values are linked to their rule-following behavior, this investigation of the organizational antecedents of those judgments is critical for encouraging employee adherence to organizational policies.

The findings of studies conducted in work settings suggest that one way that work organizations can motivate their employees is by exercising authority in ways that will be judged by those employees as fair. Tyler and Blader, for example, find that procedural-justice judgments are the central antecedent of rule following and policy adherence.65 Those employees who feel that they work in a fair work environment are especially willing to undertake personally the responsibility to follow company policies, with the obvious advantage that the company does not then have to compel such behavior. Studies show that procedural-justice judgments have the potential to

---

65 See Cooperation in Groups, supra note 7, at 77-89.
shape rule-related behavior, and that that influence is primarily explained by the impact that procedural justice has on ethical values. These findings support the arguments of the group-engagement model, which suggests that cooperation is linked to procedural-justice judgments.

These findings directly support the argument that fair behavior on the part of management motivates desirable behavior by employees. Hence, it is important for companies to be concerned about acting in ways that employees will judge to be fair. By acting fairly, companies motivate employees both to follow company policies and refrain from engaging in actions that undermine the company—actions ranging from theft to sabotage. These actions are costly to the company, undermine efficiency and effectiveness, and make clear why companies should be motivated to understand and respond to employees’ feelings about what is fair.

Many organizations already recognize this strategy, and act fairly toward their employees. The findings outlined here indicate that these intuitions are correct and support the wisdom of managing through fairness. Further, they support a particular view about what type of fairness to be concerned about. Both employees and researchers distinguish two forms of fairness: distributive and procedural.⁶⁶ Distributive fairness is concerned with the fairness of a person’s outcomes, while procedural justice is concerned about whether the decision is made in a fair manner. In particular, however, these studies indicate that it is primarily a procedurally-just workplace that encourages ethical values and rule-following behavior.

Of course, companies are hierarchical, with rules and policies flowing down from top levels of management. If upper management does not itself support the value of rule following and conformity to ethical codes of conduct, as appears to have been the case in the recent Enron scandal, then the motivation to create a supportive corporate culture may not exist among managers. Knowing how to create an ethical culture will be unimportant in that case since upper management will not be motivated to act toward the objective. Further, employees are likely to become aware that company policies are not aligned with their own moral values and they will become less committed to following company rules and policies.

⁶⁶ See TYLER ET AL., supra note 58.
In a situation of this type, the effectiveness of regulation falls on the ethical values of semiautonomous groups, such as external lawyers or accountants, whose ethical values may have been activated by their own organizations, and/or to government regulators, who again may be motivated by their own ethical concerns. Or it is shaped by the law and legal institutions through the policies they adopt for dealing with businesses and the people within them.

These findings have optimistic implications for the ability of organizational authorities to encourage rule-following behavior among their employees. Authorities are seldom in the position to expend excessive organizational resources on monitoring and punishing employee misbehavior. The procedural-justice perspective suggests that people will comply with and, more strikingly, voluntarily defer to rules when they feel that their organization’s rule-making authorities are following fair procedures when they exercise their authority and make managerial decisions. This strategy similarly promotes the view amongst employees that organizational authorities are legitimate and that the moral values of the organization correspond with their own personal moral values. From an organizational point of view, what makes such a finding optimistic is that the creation and implementation of procedures that all individuals perceive as fair is not restricted in the same way that allocations of resources are. Procedural fairness is not finite, particularly since it is based on ethical criteria.

Interestingly, the procedural-justice perspective is consistent with emerging trends in law and the legal regulation of business. As command-and-control based strategies of regulation have increasingly been questioned, government regulatory agencies have developed a variety of strategies for enlisting businesses and other “stakeholders” in the formulation and implementation of regulatory policy. These include negotiation to reach consensus on administrative regulations, cooperative arrangements for delivering social services, and joint efforts to manage wildlife and wildlands.


These policies decentralize power to “enable citizens and other actors to utilize their local knowledge to fit solutions to their individual circumstances.” All of these efforts involve procedures for decision making that embody the procedural-justice values of voice, participation, neutrality, and acknowledging the rights, needs and concerns of people involved in the decision. This does not mean that they involve wide employee participation, but rather that they reflect the values inherent in procedural-justice perspectives on management.

V. WHAT IS A FAIR PROCEDURE?

From a management perspective, procedural-justice judgments are most useful to managers if employees distinguish them from outcome judgments and rely on distinct procedural-justice assessments when evaluating the actions of management. Based upon research in work settings, I argue that employees’ views about the fairness of corporate procedures are, in fact, heavily influenced by distinct judgments about procedural fairness that are not linked to the favorability or fairness of the outcomes that results from those procedures. These include, for example, whether the procedures allow employees to have input into evaluations, whether they require that objective information be used, whether they try to control the influence of bias, etc. Recent research draws upon the four-component model of procedural justice and tests the importance of four potential procedural-justice criteria.

Understanding the nature of employees’ procedural justice judgments is central to efforts to design a corporate culture that encourages supportive employee values and that enhances employee rule-following behavior. The argument advanced here is that the potential impact of these procedural issues lies in the ability of corporations to design systems of management that are sensitive to employee procedural concerns even when companies cannot or do not provide workers with the outcomes they desire.

---

71 See COOPERATION IN GROUPS, supra note 7, at 77-89.
72 See Authority in Groups, supra note 57, at 175-76.
73 See Group Engagement Model, supra note 54, at 349-61.
The four-component model of procedural justice identifies four procedural components, or evaluations, each of which contributes to overall procedural-justice judgments. Those components are defined by: (1) two distinct aspects of organizational processes and (2) two sources of information about procedures. I will discuss the influence of each of these four components on employee definitions of procedural justice.

One of the aspects of organizational processes considered in the model refers to the organization’s decision-making procedures. Specifically, the model considers employees’ evaluations of the quality of decision making in their organization. Consideration of these evaluations links to the elements of legal procedures and emphasizes issues of decision-maker neutrality, the objectivity and factuality of decision making, and the consistency of rule application.74

There is a distinct, but potentially equally important issue involving the quality of people's treatment by organizational authorities. Quality of interpersonal treatment issues constitute the second aspect of organizational processes. Quality of treatment involves treatment with politeness and dignity, concern for people's rights, and other aspects of procedures that are not directly linked to the decisions being made through the procedure.

Each of these two aspects of procedures (quality of decision making, quality of treatment) can potentially be linked to two sources of procedure. One source of information involves the rules of the organization. The formal rules and structures of the organization, as well as statements of organizational values, communicate information about organizational procedures. For example, organizations vary in terms of whether they have formal grievance procedures that allow people to voice complaints. They also differ in their statements of corporate values (“corporate vision statements”). For example, one common formal organizational statement that concerns relationships among employees is to “[t]reat each other with respect, dignity, and common courtesy” and “express disagreements openly and respectfully.” These are both statements about the type of procedures that the corporation views as reflecting its values.

The other source of information is an employee’s experience with his or her supervisor or supervisors. While they

74 See generally Authority in Groups, supra note 57.
are constrained by formal institutions and procedures, organizational authorities typically have considerable discretion concerning the manner in which they implement decision-making procedures and how they make decisions regarding issues that have no formal procedures associated with them. Further, they have a great deal of flexibility about how they treat those with whom they deal. The same decision-making procedure can be implemented either in a way that emphasizes the dignity of those involved or in a manner that treats employees rudely or dismissively. A similar situation is found within the law. While there are formal laws and rules constraining the conduct of police officers and judges, those authorities typically have considerable latitude when exercising their authority within the framework of those rules.

The four-component model argues that each of the four components defined by these two dimensions has an important role in the definition of the fairness of procedures. While the four-component model provides a guideline for the types of evaluations that compose overall evaluations of an organization’s procedural justice, the essential argument advanced here is that the nature of those evaluations is non-instrumental and non-material. Neither of the aspects of organizational processes emphasized in this model of the antecedents of procedural justice (quality of decision making, quality of treatment) is directly linked to evaluations of the favorability or fairness of the outcomes people receive.

The four-component model highlights a set of procedural criteria that are distinct from judgments about the favorability or fairness of employees’ outcomes. This is, of course, typical of procedures in any type of organization. We can, for example, distinguish the adversary trial procedure from the verdict of the trial and can contrast that procedure with other ways of making decisions, such as the inquisitorial trial procedure.

Four criteria of procedural justice are typically measured in studies of work settings: organizational-level quality of decision making, organizational-level quality of treatment, supervisor-level quality of decision making, and supervisor-level quality of treatment. Procedural criteria linked to supervisors, rather than organizational rules, are viewed more positively. That is, employees viewed their supervisors as using fair procedures when implementing organizational policies that they generally viewed as being unfair.
VI. CONCLUSION

The argument advanced here is in support of a broader view of the employee and of the antecedents of rule-following behavior among employees. We want to articulate and show the importance of a broader and more complete picture of the motivation of employees in work settings. This model looks at the influence of both instrumental and value-based motivations in shaping rule-following behavior. The results presented suggest that such behavior is best explained when both types of motivation are considered together than when either model is taken alone.

The view presented here includes not only the motivations traditionally studied—motivations that are linked to sanctions—but also includes ethical motivations for following group rules. These ethical motivations are linked to concerns about acting in ethical and fair ways in work settings. The case for this broader model rests on the finding that corporate actors are motivated in their rule following by their ethical values concerning legitimacy and morality, their judgments about the procedural fairness of their workplace, and by their assessments of process aspects of procedures. These findings suggest that we would be better able to understand rule-following behavior in work organizations, as well as other settings, if we adopted a broader model of human motivation that added an account of ethical motivations to our models of employee behavior.

The results outlined suggest that one promising approach to stopping employee misbehavior, and thus the recent wave of corporate scandals that have dominated the business press, is to emphasize the ability of appropriate work cultures to motivate employees to act based upon their feelings of responsibility and obligation to both company codes of conduct and to their own personal feelings of morality. Encouraging such motivations leads to an enhanced likelihood that companies can bring their own behavior into line with their internal principles, as well as formal laws and government regulations, even in the absence of government and corporate regulation.
Structural Holes, CEOs, and Informational Monopolies

THE MISSING LINK IN CORPORATE GOVERNANCE

Lawrence E. Mitchell

Where was the board? This is the question that has resounded throughout the business and scholarly communities, as well as the public more broadly, as scandals from Enron to WorldCom and more have come to light over the last several years. Traditional corporate governance scholarship, as well as generally accepted legal principles, tell us that the board is the ultimate corporate monitor, the failsafe for managerial excesses and the circuit breaker in times of corporate crisis. But case after case of corporate scandal, as well as garden-variety stockholder litigation, reveals that boards were...
unaware of internal corporate misbehavior until matters reached the point of crisis.

The board may not be the culprit. Many of the scandals, such as CEO Dennis Kozlowski’s use of Tyco as his personal piggybank, senior executives and the CEO of WorldCom’s fictionalized financial statements, Andy Fastow's enormous profits at Enron's expense under CEO Jeff Skilling, and a host of shareholder suits, suggest that the board may have been ignorant of what was occurring in the corporation beneath them, not necessarily because they weren’t doing their jobs, but because they were unable to do their jobs. The relevant information was hidden from them or falsified. The implication is that often it was senior executives, and especially the CEO, who were at fault—not the board.

2 Obviously, boards in litigation over corporate scandals have a conflict of interest with CEOs and other corporate managers who they are likely to blame, and claim their own ignorance, which poses some problem of their credibility. But the issue of inadequately informed boards has been sufficiently common through modern corporate history that substantial board ignorance of managerial shenanigans seems like a reasonable assumption.

3 Certainly from the criminal perspective of these scandals it is the CEO and senior executives who are being indicted and convicted, not the board. See, e.g., Press Release, United States Dep’t of Justice, Four Former Qwest Communications Executives Indicted for Fraud (Feb. 25, 2003), available at http://www.usdoj.gov/opa/pr/2003/February/03_crm_112.htm (last visited Feb. 23, 2005) (announcing the indictment of four executives at Qwest Communications, including Qwest’s former Chief Financial Officer for Qwest’s Global Business Unit; a senior vice president, another vice president, and Qwest’s assistant controller).


DYNEGY: In January, the U.S. Attorney’s Office for the Southern District of Texas obtained a seven-count indictment charging Michelle Maria Valencia, a former senior natural gas trader with Dynegy Marketing and Trade, with filing bogus reports to the Federal Energy Regulatory Commission that are used to calculate the “index” price of natural gas. Record Home Sales, WASH. POST, Jan. 28, 2003, at E2.

EL PASO: In Houston last December, a federal grand jury returned an indictment against Todd Geiger, an energy trader at El Paso Corporation, on charges of falsely reporting price information to the Federal Energy Regulatory Commission and wire fraud as part of a scheme to manipulate energy prices. Productivity, Services Grow, WASH. POST, Dec. 5, 2002, at E2.

MERCURY FINANCE: In December, the U.S. Attorney’s Office for the Northern District of Illinois obtained an indictment of Bradley Vallem, the former treasurer of the now-defunct Mercury Finance Company, on bank fraud and wire fraud charges in connection with his participation in a scheme to overstate revenue and hide losses of more than $30 million. John Schmeltzer, Mercury Finance ex-CEO indicted; Lawyer says Client to Plead Not Guilty, CHI. TRIB., Feb. 3, 2005, at C1.

COMMERCIAL FINANCIAL SERVICES: In December, the U.S. Attorney’s Office for the Northern District of Oklahoma and the Justice Department’s Criminal Division obtained the indictment of Commercial Financial Services' former CEO William...
This is particularly striking in light of the fact that corporate governance scholarship, at least since the time of Berle and Means, has focused on the board as the corporate constituent best situated to manage or monitor the corporation's affairs and overcome the intrinsic conflict of interest that arises when corporate managers have access to the shareholders' money. That scholarship, as well as a variety of reform efforts and a substantial amount of case law, has viewed board structure as the solution to what has become known, for the last thirty years, as the "agency problem" in corporate governance.4

Drawing on economic sociology, I argue that corporate law reform efforts have focused on the wrong actors: while the law's principal interest is board governance, it should instead focus on the CEO and how the relationship between internal corporate structures and board structures provide opportunities for misconduct.5 While the CEO is important to the initial inquiry, the real problem goes even deeper; the focus of corporate law should be on the CEO and the entire senior

Enron: Four former Enron employees and former Arthur Andersen accountant David Duncan have pleaded guilty to various charges and are cooperating with investigators. Former Chief Financial Officer Andrew Fastow has been indicted on 78 counts, and three former British bankers have been indicted in connection with an Enron deal. 2002 Will Be Remembered as the Year Executives Paid the Price for Cooking Their Books, Wall Street Shame, SEATTLE TIMES, Dec. 29, 2002, at E1.

management structure of the corporation, examining the relationship of that structure to the board's ability to perform. Questions of whether the board should manage or monitor, whether its role is political, whether it should be concerned with the provision of resources as the most efficient way to limit agency costs, or as a mediator of team production, whether the board is subject to structural bias, whether it is reliable because of its members conflicting interests, or whether the board is subject to the perils of groupthink, all take a back seat to this fundamental structural question. Focusing on the board without paying attention to these structural characteristics of the corporation will not change the status quo, no matter how dramatic the reforms.

I have based my conclusion on the combination of two interrelated hypotheses, which, taken together, offer powerful insight into the role of corporate structure. They are:

1. Corporations that have inside boards will have a weak CEO (one who is dominated by the board); and
2. Corporations that have independent boards will have a strong CEO (one who dominates the board).

---

6 See Hamilton, supra note 1.
While I acknowledge that these hypotheses are contrary to virtually all legal wisdom on the subject, the theoretical construct I present suggests their power. Assuming my hypotheses are correct, the traditional scholarly focus on the board without attention to the CEO and senior management is

---

14 Stephen M. Bainbridge, *Director Primacy in Corporate Takeovers: Preliminary Reflections*, 55 STAN. L. REV. 791, 810 (2002) (“[T]he Delaware courts take the board’s distinct role quite seriously, especially with respect to its independent members.”); James D. Cox, *Managing and Monitoring Conflicts of Interest: Empowering the Outside Directors with Independent Counsel*, 48 VILL. L. REV. 1077 (2003) (“The important role that independent directors have in monitoring and managing conflicts of interest reflects our societal commitment to the power of the outside director. . . . Today, in the post-Enron era, the outside director continues to be the focus of corporate governance reforms.”); Lynne L. Dallas, *The Multiple Roles of Boards of Directors*, 40 SAN DIEGO L. REV. 781, 787 (2003) (“Perhaps the most significant trend in board governance in the United States in the last twenty years has been the increase in the number and proportion of outside directors . . . .”). Dallas proposes a two-tiered board model with a mixed board and outside board that, while this paper does not endorse at the moment, does intuit some of the conclusions I reach. Delaware Chancellor William Chandler recently has expressed his view that courts should not “rely reflexively” on a director’s status as inside or independent in according deference or not, but rather should take matters case by case. William B. Chandler III, *On the Instructiveness of Insiders, Independents, and Institutional Investors*, 67 U. CIN. L. REV. 1083 (1999).

The principal area in which legal scholars and, especially, economists, have noticed the increased power of the CEO is in the realm of compensation. Bebchuk et al., *supra* note 5, at 766; Robert A. Lambert et al., *The Structure of Organizational Incentives*, 38 ADMIN. SCI. Q. 438 (1993); Brian G. M. Main et al., *The CEO, The Board of Directors, and Executive Compensation: Economic and Psychological Perspectives*, 4 INDUS. AND CORP. CHANGE 293 (1995); Kevin J. Murphy, *Explaining Executive Compensation: Managerial Power versus the Perceived Cost of Stock Options*, 69 U. CHI. L. REV. 847 (2002); Edward J. Zajac & James D. Westphal, *Accounting for the Explanations of CEO Compensation: Substance and Symbolism*, 40 ADMIN. SCI. Q. 283 (1995). These studies generally focus on market imperfections relating to the failure of outside boards and consulting firms to understand the true cost of options. See, e.g., Murphy, *supra*. They fail to understand the critical importance of structure which, if misaligned, can hamstring even the most sophisticated boards and consultants. But see Ann K. Buchholtz et al., *Are Board Members Pawns or Watchdogs?*, 23 GROUP & ORG. MGMT. 6 (1998) (finding correlations between increased CEO strength and increased board strength).

There have been dissenting voices from time to time over the wisdom of independent boards as a general reform solution, some of which acknowledges the role of the CEO in either manipulating or otherwise disempowering independent boards. See, e.g., Stephen M. Bainbridge, *Independent Directors and the ALI Corporate Governance Project*, 61 GEO. WASH. L. REV. 1034 (1993) (questioning the value—at least the universal value—of independent directors); Victor Brudney, *The Independent Director—Heavenly City or Potemkin Village?*, 95 HARV. L. REV. 597 (1982); Jill E. Fisch, *Taking Boards Seriously*, 19 CARDOZO L. REV. 265 (1997); Daniel R. Fischel, *The Corporate Governance Movement*, 35 VAND. L. REV. 1259, 1282 (1982); Roberta S. Karmel, *The Independent Corporate Board: A Means to What End?*, 52 GEO. WASH. L. REV. 535 (1984) (questioning the monitoring model of the board based on outside directors proposed by the ALI); Langevoort, *supra* note 5. The explanations as to the uncertain benefits of independent boards vary. The virtue of the approach I take is its theoretical coherence. I recognize, at the same time, that theoretical coherence may result in oversimplification; qualifications to my theory are presented *infra* notes 111-17 and accompanying text.
misdirected. At the very least, my hypotheses suggest that advocacy of independent boards,\textsuperscript{15} which has been the trend over the last thirty years,\textsuperscript{16} is simply wrongheaded.\textsuperscript{17}

This focus on the board has led corporate law scholars (with several notable exceptions\textsuperscript{18}) to ignore the extraordinary increase in CEO power.\textsuperscript{19} Focusing on corporate structure leads us to conclude that the increase in CEO power is the result of increasing board independence.\textsuperscript{20} At the same time, evidence

\textsuperscript{15} Standard terminology divides directors into three categories: inside directors, who are officers of the company, affiliated directors (or “gray directors”), who have some business relationship with the company (investment bankers and lawyers are typical examples), and independent directors, who have no relationship with the company other than their service as directors. Sanjai Bhagat & Bernard Black, The Non-Correlation Between Board Independence and Long-Term Firm Performance, 27 J. CORP. L. 231, 239 (2002). My argument focuses principally on the problem of independent directors, although affiliated directors suffer from problems similar to the ones I shall examine if to a lesser degree.

\textsuperscript{16} Bhagat and Black describe the “conventional wisdom” that only independent directors can be effective monitors. Bhagat & Black, supra note 15, at 232. The advocacy of this “conventional wisdom” has led to regulatory and quasi-regulatory requirements of increased outside directors, most recently the Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745, which requires outside directors on the audit committee, and the New York Stock Exchange’s well-publicized rule revisions which require listed corporations to have a majority of outside directors. New York Stock Exchange, Corporate Rule Proposals Reflecting Recommendations from the NYSE Report of the Exchange Corporate Accountability and Listing Standards Committee as Approved by the NYSE Board of Directors, Aug. 1, 2002, § 303, at A1.

\textsuperscript{17} Certainly this is the implication of Bhagat and Black’s work. Bhagat & Black, supra note 15, at 233.

\textsuperscript{18} See Bebchuk et al., supra note 5, at 783-95; Langevoort, supra note 5.

\textsuperscript{19} Westphal, in an empirical study, concludes that CEOs have developed behavioral patterns that counteract the greater potential board control associated with board independence. See James D. Westphal, Board Games: How CEOs Adapt to Increases in Structural Board Independence From Management, 43 ADMIN. SCI. Q. 511, 529-31 (1998). Klein, consistent with other studies, finds little relationship between corporate performance and board structure. However, consistent with the argument I present here, she does note that the presence of insiders on finance and investment committees is correlated with superior corporate performance, consistent with the notion that insiders provide valuable information to the board. April Klein, Firm Performance and Board Committee Structure, 41 J.L. & ECON. 275 (1998). Robert W. Hamilton argues that in the 1950s much of the work of the board was done by the CEO. Hamilton, supra note 1, at 349-50.

To a limited extent, the recently enacted Sarbanes-Oxley Act seems to intuit the importance of the CEO, requiring his certification of the corporation’s financial statements and thereby increasing his incentives for integrity and careful monitoring through the device of federal sanctions for his failure. This seems more sensitive to the identity of the real culprits in the corporate scandals that led to the Act’s passage, but it doesn’t really follow through on the implications of this intuition. Moreover, following traditional reform approaches, it also places great importance on the board, and particularly the importance of an independent audit committee.

\textsuperscript{20} Bebchuk, Fried and Walker acknowledge the increase in recent years in the number and power of independent directors. Bebchuk et al., supra note 5, at 773. That the level of the CEO’s power can be seen as along a continuum should not be surprising. “Organizations are information processors,” and the critical variable is the
shows that the increasing independence of boards has, predictably, decreased trust between CEOs and boards.\textsuperscript{21}

The wisdom of board independence has come under serious question. Bhagat and Black, in an important empirical study, examine a thirty year trend toward greater board independence during which “the composition of public company boards of directors has changed radically . . . .”\textsuperscript{22} This study provides striking evidence that independent boards not only fail to improve corporate performance, they may in fact make it worse.\textsuperscript{23} While Bhagat and Black speculate as to the causes of this phenomenon, by their own admission they are unable to fully explain it, and continue to suggest remedies that focus on board reform.

The theory set forth here does explain the phenomenon observed by Bhagat and Black. In Part I of this article, I will explain the significance of my theory to corporate law reform efforts. Parts II and III will explicate the underlying theory. Part IV will demonstrate how my hypotheses explain the direct...
relationship between increased board independence and CEO power, and Part V will discuss the extent to which managers a level or two below the CEO might also be in a structural position to manipulate both the board and the CEO.

Following this discussion, Part VI will address a second independent variable that might have an important magnifying effect on CEO power: the bureaucratic organization of the corporation itself. While corporate reformers have been advocating independent boards, important changes in corporate structure have taken place. Large public corporations, once rigidly hierarchical, have, at least in some industries (and sometimes within industries), substantially shifted to more horizontal management systems. This horizontal management structure magnifies the strength of the CEO, whether the board is an inside board or an independent one. Because this conclusion is still tentative, however, I reserve this discussion for the end of the article. Part VII will conclude with some possible directions for further research and reform.

I. THE POWER OF THE THEORY

My hypotheses derive from a subgenre of economic sociology and specifically a subgenre of network theory, known as the theory of structural holes. I will reserve a more detailed explanation of structural hole theory for the next section. For now, I will briefly define and illustrate structural holes and suggest why structural hole theory holds such great promise for corporate scholarship, even as it reveals the importance of focusing on internal corporate structures.

Network analysis is a genre of sociology that, contrary to typical sociological analysis which begins from studies of individuals and classifies them into social structures by grouping their characteristics, instead begins with the structure as the unit of analysis, representing social structures as networks and actors as nodes within those networks, in order to identify the constraints on individual behavior arising from the social structure. BARRY WELLMAN & S.D. BERKOWITZ, SOCIAL STRUCTURES: A NETWORK APPROACH 3, 4 (1988).

Let me make it clear at the outset that the theory I am applying is an adaptation of the essential aspects of structural hole theory. Structural hole theory, as presented by Burt, while an intuitively apparent idea, has enormously complexity and applicability, and Burt ranges from explanations of market behavior to a theory of the firm to a theory of personality. The adaptation of the theory I apply takes the basic defining aspects of structural holes and implicitly combines it with Granovetter’s weak ties theory (to which it is intimately related) for the purpose of creating a hybrid theory that allows us to take a broad view of corporate structure and its potential deformities.
To understand structural holes, one must begin with the proposition that people are socially organized into distinct networks. Sometimes networks overlap through common members. Sometimes they are completely distinct, with no ties to each other. When two networks are distinct and lack ties to each other, the gap between them is a structural hole. The structural hole provides an opportunity for a person to establish contact with each of the two networks, bridging the structural hole and giving him or herself informational and control advantages.

Think, for example, of a university’s anthropology department and its sociology department. Assume that the departments are relatively small, and that the members of the respective departments know each other reasonably well, at least as colleagues. Each department is a single network. Its members have significant professional and, perhaps, social ties to one another. It is likely, however, given the nature of the two disciplines, that at least several members of each department will know each other well. The two networks overlap through these associations, and while this does not destroy the integrity of each department as a network, it does connect the networks so that no structural holes exist. This connection provides members of each department who are unacquainted with members of the other department with some substantial information (gossip or scholarly) about what is going on in the other department, through colleagues who bridge the networks.

Now consider the same university’s physics department and its law school. As with the preceding example, each division of the university forms a network. It is unlikely (with the possible exception of chance acquaintance on university committees or relationships off campus) that any of the members of the physics department and any of the members of the law faculty know each other. The gap between these networks is a structural hole.

There may be no reason for physics professors and law professors to bridge this gap. On the other hand, there might. Assume that women form a relatively small minority of each of the physics and law faculties. Next assume that the university neither provides on-site childcare nor provides child care subsidies. Further assume that most of the men on each faculty leave childcare principally to their wives and therefore don’t
view it as of particular concern. But some of the women do. However, these women do not form a sufficiently large constituency in either department to compel their respective faculties to act, or to challenge the university’s parsimony. A woman who is a member of the law faculty, forced to deal with this problem, can consciously make the effort to become acquainted with at least one woman on the physics faculty. She is now bridging the structural hole. The utility of such a bridge is obvious. Through her contact with the physics professor, the law professor unites the two networks (or in this case sub-networks) of women faculty. By so doing, she is able to gauge the strength of the inchoate demand for childcare and, by joining the networks (particularly if the law professor repeats this effort in every department and school of the university), she may be able to compel the administration to act. This is one value of bridging structural holes.

There is a negative side to structural holes too. Assume that no woman on the university faculty sees this structural hole opportunity. The only bridge among the different departments and schools is the university administration (and for sake of simplicity, let us identify the administration as the university president). If the women in the physics department approach the president and ask for child care relief without themselves bridging the structural hole, the president is in a position to play the faculties against each other to his or her advantage. He or she might, for example, threaten to cut a portion of the law school’s budget to pay for on-campus childcare because the physics department, which already has budgetary constraints, is demanding the service. The law school will object to having its budget cut to subsidize the physicists, even if the women law faculty might benefit. The president can then cite the law school’s protest to the physics department as a reason to deny child care benefits. As the sole bridge of the structural hole, the president can set the two sides off against each other.

The president can also manipulate the situation by creating a structural hole that remains unfilled—in other words, by intentionally splitting two connected networks. For example, assume that the president sets up a management structure in which department heads and deans report only to

---

26 I realize this example is gendered but plead current social realities as a justification.
the provost, and that all university-wide faculty committees, whether ad hoc or formal, can gain audience only with the provost or one of his subordinates. The president delegates authority to the provost to make all decisions that come out of these reports, ensuring that the provost has a good sense of the president’s interests. By cutting off reporting from the provost, the president has created a structural hole between herself and the provost that is unfilled and, given her control and the university structure, unfillable. The provost can deny the request and there can be no appeal to the president.

Of course every faculty member knows that such a situation, if enacted in the world of academia, would create an uproar that would make the president’s life miserable. I offer this example because it is simple and familiar. In the corporate context, to which I will later turn, such manipulation is much easier, for power is more clearly defined, less democratically wielded, and job termination is always a threat. For the moment, however, the foregoing should clarify the definition of structural holes and provide some insight into their utility. Figure One illustrates a structural hole bridged by you.

**Figure 1:** Structural Hole

![Figure 1: Structural Hole](image)

Structural hole theory was developed by Ronald Burt in the 1980s and early 1990s, culminating in *Structural Holes*, its most comprehensive examination. Burt explains competition as a function of social structure by looking at the ways in which competitors can maximize their opportunities by

---

manipulating to their advantage the social structures in which they operate. This contrasts with neoclassical economic analysis, which focuses on individual transactions and the wealth maximizing motivations of individuals. Structural hole theory is closely related to transaction cost economics,\(^28\) which explains the origin of organizations in, among other things, the desire to restrain opportunistic behavior arising from the neoclassical goal of diminished transaction costs.\(^29\) But it is richer than these theories because structural hole theory allows us to dispense with the unrealistic essentialized motivations of actors that characterize the competing theories as well as to broaden our perspective beyond the dyadic transaction.\(^30\) Instead, structural hole theory leads us to see competition as a process occurring within preexisting structures,\(^31\) obviating the need for unrealistic assumptions and situating the theory in the complexities of the world in which competition takes place.\(^32\) Structural hole theory treats the

\(^{28}\) Transaction cost economics, or “the new institutional economics,” finds its origin in Ronald Coase’s 1937 theory of the firm, Ronald Coase, *The Nature of the Firm*, 4 ECONOMICA 385 (1937), and was developed most thoroughly by Oliver Williamson. OLIVER WILLIAMSON, MARKETS AND HIERARCHIES: ANALYSIS AND ANTITRUST IMPLICATIONS: A STUDY IN THE ECONOMICS OF INTERNAL ORGANIZATION (1975).

\(^{29}\) Williamson’s theory is considerably more complex than this, describing the origins of organization not only in terms of restraining opportunism but also as a result of conditions of uncertainty, asset specificity, bounded rationality, and the like. See WILLIAMSON, supra note 28, at 7. The analysis does focus on transaction cost reduction, however, and I focus on opportunism in the text because it is the aspect of Williamson’s concern most related to this paper.


Structural hole theory also has close relationship to resource dependence theory which is not especially relevant to my analysis. JEFFREY PFEFFER & GERALD R. SALANCIT, *The External Control of Organizations* (1978) is the ur-text on resource dependence theory.

To be sure, Burt takes great pains to show the consistency of his theory with transaction cost theory and resource dependency theory (and even traces some of his intellectual roots back to Coase) in his development of his theory of the firm. BURT, supra note 27, at 238-45. My characterization of the differences in the text that lead me to argue that structural hole theory is an improvement over these theories is my own, not Burt’s.

\(^{30}\) Williamson proclaims his unrelenting reliance on the transaction as the unit of analysis. WILLIAMSON, supra note 28, at 1-2.

\(^{31}\) To put it differently, for all of its sophistication in trying to come to grips with organizational structure and its sources and consequences, transaction cost theory remains mired in the unrealistic assumptions of neoclassical economics.

\(^{32}\) Burt himself, especially in his argument in Chapter Seven, sometimes seems to lapse into the same motivational assumptions (although he doesn’t describe them this way) as Coase and Williamson. Nevertheless, from a purely structural perspective, one can easily read his basic theory as dispensing with these assumptions.
process of competition as a function of relations that are visible only by their absence—an absence we might refer to as gaps in the social structure. These gaps—these structural holes—allow actors within a social network the freedom to behave entrepreneurially (or opportunistically). As a general matter, the kinds of social networks most relevant to this discussion are networks of managers within the corporation.

While there are many ways in which an actor in this context can manipulate the social structure, the kind of opportunism most relevant here is that of a manager placing herself in a position where she will maximize the likelihood of receiving information and the opportunity to disseminate it as she desires.33

The freedom to behave opportunistically within social networks—including the social networks within corporations—arises because they create circumstances of imperfect competition by placing given actors in advantaged positions (or creating opportunities for them to seize advantaged positions). Further, social networks structure these positions and the relations between other actors in such a way that movement for those who are not in advantaged positions is relatively difficult. (In economic terms, social structure introduces friction into the “market” so that competition is imperfect.) Circumstances of imperfect competition create opportunities for advantage. In practical terms, imperfect competition in this context means that some actors are “stuck” in place in the corporate hierarchy or are positionally situated to be constrained by other actors, while others, who have the ability to identify gaps in the structure, are free to move in and fill the gaps.34 This is most likely to be the case when a given actor performs the same (or largely the same) function as another actor. As I will later explain, in network terms, such actors are redundant—they have the same contacts—and so only one, if either, will be of use to an actor in an advantaged position and

33 See Burt, supra note 27.
34 The use of the economic concept of imperfect competition here may be jarring to those who are used to seeing it used only in the context of markets. But competition can occur in other spheres of life as well and it is at least metaphorically, if not literally, useful to adopt the phrase to describe the circumstances under which social behavior within social networks takes place. I ask the reader to be careful, however, because I am using the term (and applying the theory) in the corporate context, which itself is an economic realm and could easily lead readers to think of the competition I describe in economic terms. In my application, the competition is social, not economic (even if in some cases it may have economic consequences).
only one will therefore have the opportunity to move into the advantaged position.

The principal benefits accruing to the advantaged actors—the actors who can see and are free to occupy structural holes—are access to information and the ability to control others. The actor who best understands how to exploit the informational and control advantages of the structural holes (and indeed knows best how to recognize and occupy those structural holes) is the actor who will emerge as the most successful: he is a network entrepreneur. These opportunities are resources existing in every social network and organization and they are waiting to be exploited. Those who can identify and fill the structural holes will do better—whatever the rewards of the network—than those who lack this ability.

I should note that Burt presents his theory as a positive one; that is, a theory of enhancing value that explains how competitors can improve their positions and attain the advantages that go with this improvement. But there is a dark side to structural hole theory as well. Though it is a theory of value, it can also be seen as a theory of manipulation, opportunism, and inefficiency. This article will focus on these aspects of the theory in the context of corporate governance.

This very brief introduction to structural hole theory demonstrates why it is such a powerful analytical tool. Corporate governance scholarship has traditionally centered on what has come to be known as the “agency problem.” The question the agency problem presents is how to restrain corporate managers from shirking responsibility or stealing corporate assets. In the bulk of corporate scholarship over the last thirty years, the problem has taken the form of finding ways to reduce agency costs, i.e., the costs that arise from monitoring and preventing shirking and stealing (deadweight economic losses) or, to put it differently, of finding the most efficient ways to restrain shirking and stealing. While agency cost theory has made substantial contributions to our


36 Despite the predominance of neoclassical economic analysis since that time, as I noted earlier, the focus has existed at least since the time of Berle and Means. ADOLF A. BERLE & GARDINER MEANS, THE MODERN CORPORATION AND PRIVATE PROPERTY (1933). See also Adolf A. Berle, Corporate Powers as Powers in Trust, 44 Harv. L. Rev. 1049 (1931).
understanding of corporate governance, its flaw is that it takes the traditional analytical approach of the subject as a given—it begins with the received structure of corporate law and treats the internal workings of the corporation largely as a black box. Of course this is a flaw in the traditional approach itself.

Structural hole theory refocuses our inquiry by taking us into the box. Its power derives from several aspects of the sociological approach in which it originates. First, because the theory focuses on structure, it need not resort to the simplified assumption used by agency cost and transaction cost theorists that corporate actors seek only to maximize their wealth. While this assumption does have utility in economic modeling, it places the corporate reformer in a difficult position. If we start with greed as an immutable motivation for behavior in the corporate context, our goal of reforming corporate law is hamstrung by the fact that we need to devise tools to restrain corporate actors’ pursuit of their own wealth even to the point of transgressing corporate norms. But history has demonstrated that restraining greed is an enormously difficult task. Even the attempts to align managerial and stockholder interests by encouraging executive compensation in the form of stock options has encouraged greed, which has led to substantial abuse, and is thus a partial cause of much corporate misbehavior.

This is actually not true of the work of Fama and Jensen themselves, who build the corporate structure as evolving to provide the most efficient solutions to the agency problem. Fama & Jensen, supra note 11. See also Eugene Fama, Agency Problems and the Theory of the Firm, 88 J. POL. ECON. 288 (1980). Bainbridge also provides an exception, looking within corporate hierarchies to evaluate the quality of information flows throughout the corporation. Stephen M. Bainbridge, Participatory Management Within a Theory of the Firm, 21 J. CORP. L. 657 (1996).

Williamson extends the self-interest model to opportunism. See WILLIAMSON, supra note 28, at 26-30. I recognize that the pursuit of maximum self-interest assumed by neoclassicists can extend beyond monetary greed and apply to other preferences as well. In the corporate context, however, it is almost invariably assumed that the issue is money. Thus the reference to greed seems perfectly appropriate and accurate.

See LAWRENCE E. MITCHELL, CORPORATE IRRESPONSIBILITY: AMERICA’S NEWEST EXPORT (2001); Bebchuk et al., supra note 5; Murphy, supra note 14. For an excellent example of the relationship between managerial opportunism and corporate bureaucracy (which nonetheless focuses more on inefficient information production rather than structure per se), see Walter Novaes & Luigi Zingales, Bureaucracy as a Mechanism to Generate Information, CENTRE FOR ECON. POL’Y RES., June 2003, which argues that the information (managerial performance) creating potential of bureaucracies, coupled with managerial incentives to extract undue rent from the
A focus on structure as creating the conditions for opportunistic behavior allows us not only to consider broader motivations, but also enables us to see the way in which structural freedoms and constraints, in contrast to monitoring and mistrust, can alleviate problems that are the concern of agency theorists and other corporate scholars. In other words, rather than focusing on the restraint of greed, structural hole theory identifies places in the corporate structure that provide room for opportunistic behavior and allows us to concentrate on eliminating these opportunities for corporate actors to behave in self-serving ways. For the same reason, it enhances transaction cost analysis (which begins with the same motivational assumptions) by allowing us to begin with given structures rather than individually modeled behavior that leads to institutional structures. In this way, it reveals the power of position within structures, and enables us to see ways of restructuring hierarchies to reduce the circumstances in which opportunistic behavior can flourish. In addition to these benefits, structural hole theory is well conceived for rigorous empirical testing, as Burt demonstrates. Instead of relying upon reductionist accounts of motivation or attitudinal or psychological self-reporting, all structural hole theory requires is that survey subjects disclose those people with whom they have contacts, as well as the regularity and intensity of those contacts. These results can be verified by independently obtaining the same information from those with whom the subject claims to be in contact. Thus the structure emerges as empirical fact, with relatively little room for distortion.

As a normative matter, structural hole theory can also show how a corporation can both eliminate structural blockages and enhance the efficiency of information flows. It does this by ensuring the proliferation of structural holes within its networks. As I noted earlier, Burt describes this

corporation, lead boards to choose a bureaucratic structure that produces more information about executive performance instead of what perhaps might be a more optimal structure for the maximization of profits. Id.

42 This structural approach bridges the micro approach of transaction cost economics with the macro approach of network theory. Burt, supra note 27, at 181. The search for ways to bridge micro and macro models has been a significant project for sociologists. The bridge provided by network theory is its recognition that structure provides the resources and individuals attempt to benefit from the resources structure provides. This is especially noted by social capital theorists. Kenneth A. Frank & Jeffrey Y. Yasumoto, Linking Action to Social Structure Within a System: Social Capital Within and Between Subgroups, 104 AM. J. SOC. 642, 645 (1998).
positively, in an efficiency-enhancing way.\textsuperscript{43} Efficiency comes from the placement of trustworthy and loyal actors in positions where they can bridge structural holes, facilitating the transfer of information without the fear of manipulation. The theory can also be used positively for self-advancement, as a network entrepreneur sees the opportunity to fill a structural hole within the corporate bureaucracy.

Finally, and most practically, the theory has the power to explain why simple governance reforms such as creating independent audit committees, nominating committees, and compensation committees, or composing boards principally of independent directors, may not be capable of resolving problems of inadequate monitoring. All of the board reforms currently underway will fail unless the structural opportunities for CEOs and other senior managers to control and manipulate information are reduced or eliminated. The failure of corporate law scholars to focus on the special role of the CEO and his subordinates has limited the set of possible solutions to corporate governance problems.\textsuperscript{44} The structural approach, however, presents new solutions.

II. THE INTELLECTUAL PROVENANCE OF STRUCTURAL HOLE THEORY

Structural hole theory is a theory of social capital. While social capital is a concept that recently has garnered scholarly

\textsuperscript{43} I will address the efficiency—or positive—aspect of structural holes infra note 72 and accompanying text. My purpose in this paper is to use the theory in a way that it has not been used, to examine the dark side of structural holes in the manipulation of information within the corporation.

\textsuperscript{44} The CEO may not always be the appropriate focus. Wayne Baker, studying a commercial real estate development firm that appears to have been a partnership, explicitly designed to maximize intrafirm networks, but with a designated CEO, concludes that the absence of the CEO would have little or no effect on firm performance. See Wayne E. Baker, The Network Organization in Theory and Practice, in NETWORKS AND ORGANIZATIONS: STRUCTURE, FORM, AND ACTION 397 (Nitin Nohria & Robert G. Eccles, eds., 1992). The reason for his conclusion, however, appears to me to lie both in the partnership nature of the form and the conscious design of the firm to maximize networks (an approach to firm organization that he believes is highly unusual, since networks tend to develop spontaneously rather than consciously and wind up less complete and neat). Moreover, given the partnership nature of the firm, there does not appear to have been a board of directors, and so the essential structural hole I am exploring could not have existed. The likelihood of such a situation developing in a large public corporation, even one that is relatively flat in terms of structure, is not high, so the focus on the CEO in that context seems perfectly appropriate. See id.
and public attention," its definition is both elusive and debated. For purposes of this article, social capital, defined in its minimal formulation, is the set of resources available to a person, organization, or community that inheres in its social structure. Social structure includes the webs—or networks—of people and institutions that collectively constitute families, friendships, organizations, communities, and societies.

Although all treat social capital as a resource, the particular understanding of how the resource works and where it is found (in addition to whether it is more in the nature of a public good or more in the nature of an individual asset) differs among the various social capital theorists.

One of the earliest approaches to social capital was the work of French sociologist Pierre Bourdieu in the 1960s and 1970s. Bourdieu began from a distinctly anthropological perspective, understanding culture as "a dynamic and creative, but also a structured phenomenon." His theory established a triad of economic, cultural, and social capital, with economic capital the most dominant factor.

James Coleman's theory of social capital is a frank attempt to draw together economics and sociology, originating in his own theories of rational choice. Coleman asked how human capital, which is educational attainment and skills, affected the equality or inequality of persons in society. Coleman theorized that social capital consisted of two parts: first, the social structures themselves, and second, the way they affected the actions of given actors within the structure.

---


47 Network theory itself has been recently popularized in DUNCAN E. WATTS, SIX DEGREES: THE SCIENCE OF A CONNECTED AGE (2003) (laying out a new way to understand the way networks grow, work, and how they drive collective behavior).


49 See generally James S. Coleman, Social Capital in the Creation of Human Capital, 94 AM. J. SOC. 95 (1988) for an explication of his functionalist view of human capital at a late stage in his career.

50 COLEMAN, supra note 46, at 302.
Coleman’s argument focused on the way in which powerful people remained powerful through their social networks with other powerful people. Coleman also saw social capital itself as a resource that could be manipulated through the creation of social relations with other actors, including trust, obligations of reciprocity, and specific social expectations. Given the common understanding of human capital as the educational attainment and skills of an actor, Coleman believed that social capital and human capital were interrelated.¹¹ The principal flaw in Coleman’s work is the circularity of its conclusion—powerful people tend to remain powerful because they are powerful people, without a clear explanation of how they become powerful in the first place.¹²

Robert Putnam, whose famous early work on social capital derived from his study of regional governments in Italy, noted a distinct difference in the performance of governments in the north and south, leading him to focus on the extent to which civic engagement made a difference in explaining the greater effectiveness of northern governments.¹³ The particular variables he examined were associational life, newspaper reading, voter turnout, and voter preferences. He expanded his examination of these variables in his book, Bowling Alone, in which he looked at the decline in associations from bowling leagues, coffee time with neighbors, sewing circles, and similar activities, concluding that the level of civic engagement in the United States had seriously declined. He hypothesized that this decline was largely due to a dramatic increase in television watching.¹⁴ The core of social capital, in Putnam’s definition, lies in three factors: social networks, social norms, and trust. Unlike Bourdieu and Coleman, who focused on the benefits of social capital to the individual actor, Putnam took a more global view. Thus, he treated social capital as a true public good, essential to the maintenance of civic society, as is foreshadowed by the title of his seminal work on Italy, Making Democracy Work.¹⁵ This, of course, is not surprising given Putnam’s training as a political scientist. In fact, the different

¹¹ See Schuller, supra note 46, at 6.
¹² See Coleman, supra note 46.
¹³ See id.
¹⁴ It’s probably worth noting that criticism of Putnam’s empirical base and social capital categories has become something of a cottage industry. It may be that this has pushed him to a more instrumental view of social capital.
¹⁵ See Coleman, supra note 46
professional backgrounds of these three major theorists of social capital help to explain their understanding of the importance of the concept, as illustrated by Bourdieu's interest in culture, Coleman's in rational behavior, and Putnam's in democratic governance.\textsuperscript{56}

Structural hole theory grows out of the confluence of social capital theory with economic sociology. Social capital theory and economic sociology—and thus a greater interest in how social structure contributes to private goods—developed along a somewhat parallel historical track. In the mid-1970s, Mark Granovetter, working with the sociological tool of network theory, which seeks to explain the ways in which particular social structures affect relations among people,\textsuperscript{57} developed his seminal theory of “the strength of weak ties.”\textsuperscript{58} The paradox, Granovetter explained, was that network theorists treated close relationships among people as the key variable affecting their behavior, relationships that he described as “strong ties.” But it was not these strong ties that were essentially important in the economic realm.\textsuperscript{59} Rather, he found, most people surveyed found their jobs not through close friends and family but through more casual acquaintances, relationships he described as “weak ties.” The more weak ties a person had, the broader and more far-reaching was her social network, and thus the more likely it was that she could maximize her own opportunities by exploiting those ties in order to advance her career. Granovetter later retested this hypothesis with somewhat mixed results,\textsuperscript{60} but the theory itself has become one of the principal building blocks of economic sociology. It should be clear that weak ties are a form of social capital, a resource deriving from the social structure that actors can use to better their positions and which actors can

\textsuperscript{56} As Putnam’s work has developed, his focus has shifted from pure participation in associational life to the ways in which such participation develops norms of reciprocity that form the core of social obligation. In other words, Putnam seems to have moved from an understanding of social capital as more like public good to social capital as more like a private resource, albeit necessarily sustained by the social structure.

\textsuperscript{57} This is in contrast, for example, to the ways in which the structures of particular institutions might affect behavior.

\textsuperscript{58} Mark S. Granovetter, \textit{The Strength of Weak Ties}, 78 Am. J. Soc. 1360 (1973).

\textsuperscript{59} Granovetter’s particular interest at the time was how people found work. Mark S. Granovetter, \textit{Getting a Job: A Study of Contacts and Careers} (1974).

\textsuperscript{60} Mark S. Granovetter, \textit{The Strength of Weak Ties: A Network Theory Revisited}, in \textit{Sociological Theory} 201 (Randall Collins ed., 1983).
consciously seek to accumulate. As I will discuss, the theory of structural holes begins in part with Granovetter’s theory.

Structural hole theory is a theory of social capital because as it develops it takes the two component words of the term quite literally. First, it is “social” in that the exploitive opportunities of an actor derive from the social networks in which he is embedded. Second, it is “capital” in that the actor’s particular placement in that network—in the position described as a structural hole—is, like weak ties, a resource that can be used to maximize other resources by allowing the actor to affect the terms of his own relationships with others.

While structural hole theory is a theory of social capital, it is different from the work that I described earlier because of its relentless instrumentality. Moreover, unlike much sociological work, structural hole theory is not simply descriptive, but is also predictive: based on extensive empirical testing, Burt has identified the specific network structures and hierarchies that allow us to predict things like which managers in a given corporation will advance at the fastest rate. The ideal network structure for this purpose varies with the age, gender, and duration of a particular manager’s employment.

This predictive aspect of the theory can also help us determine the best structures of governance for public corporations by allowing us to see how given social structures are defective. That is, structure reveals informational and control opportunities which network (in our case, corporate) actors could use either to the advantage or disadvantage of the corporation while serving themselves. The theory further provides insight into how we might alter structures to prevent the proliferation of these locii of self-interest or, to the extent they are desirable, increase them. It is for this reason that structural hole theory holds great promise as a tool for focusing

---


62 Burt’s empirical data were derived from his study of “one of America’s largest high-technology firms.” Burt, supra note 27, at 118. He later used what appears from his description to be the same firm in his refinement of the theory although he reports data from other studies in this article. Ronald S. Burt, The Contingent Value of Social Capital, 42 ADMIN. SCI. Q. 339 (1997). He based his study of market competition and structural holes on seventy-seven product markets as classified by the U.S. Department of Commerce. Burt, supra note 27, at 85.
the debates over the most effective methods of corporate governance.

Burt’s theory of structural holes developed from the groundwork of Granovetter’s theory of weak ties. Recall that Granovetter saw the weakness of social ties as a correlate of information flows that enables actors with weak ties to have access to information that they otherwise would not. Burt’s departure from Granovetter’s theory rests on two aspects of that theory. First is the issue of causation. What is causative of the effectiveness of weak ties, he argues, is not the strength or weakness of the particular ties themselves, but of the structural holes they span (or the opportunity presented by the structural hole for an entrepreneur to place himself in the middle of potential information flows among networks). The second is that Burt finds that the strength or weakness of the ties does not matter—information flows over both. The relevant question for Burt is whether or not those ties are redundant (which, as I will later explain, means whether they efficiently and effectively reach the same people or whether they reach new networks).

Burt criticizes Granovetter’s theory by arguing that “the weak tie argument obscures the control benefits of structural holes.” The theory of weak ties is about bridges—about the paths along which information flows. The structural hole argument is about “chasms”—the interruptions of those flows. The person who can see the chasm and bridge it is in a position not only to receive information, but to control the information flows himself, regardless of whether the tie is weak or strong.

III. WHAT IS A “STRUCTURAL HOLE”?

With this brief introduction to the development and basic insight of structural hole theory, it is time now to explain

---

63 BURT, supra note 27, at 26-28.
64 Id. at 26-27.
65 Id. at 28. Burt is not always clear about whether structural holes are spaces in the social structure or intersections of networks that occur within individuals. My best reading is that he means they are both. Compare id. at 190, 192 (describing structural holes as relations that intersect). I don’t believe that the ambiguity is at all important. Structural holes, it appears, can be both gaps unfilled between or among networks, or the gap once filled. The important point is that the possibility of filling the gap creates the competitive opportunities that are the essence of structural holes.
66 Id. at 28.
some of the technical aspects of the theory.\footnote{The theory itself is rather complex, although it depends upon one central, fairly simple, insight. The complexity of the theory lies more in its various implications than in the concept itself. I have thus drawn on the centrally important aspects of the theory as being those that are relevant to an initial explication of the problems of corporate governance in an attempt to avoid complicating the story and confusing the reader with details that, at this point, are only of subsidiary importance.} A structural hole is a bridge between nonredundant contacts.\footnote{Burt, supra note 27, at 17-18.} Nonredundancy is characterized by the presence of two criteria: lack of cohesion and lack of structural equivalence. Cohesion refers to the strength of the relationship in question. Strong relationships signal that no structural hole exists. Burt provides examples of cohesive ties as between father and son or husband and wife. If either one of the actors in that relationship is a strong contact of yours, you effectively have access to the other person. Think of a group of close friends who live in the same neighborhood, travel together, and meet socially on a regular basis. A strong tie with one of these friends, in effect, gives you access to the information possessed by the entire network.\footnote{Limited, although Burt disregards this, by norms of confidentiality and perhaps somewhat tempered by in-group psychology. For a recent application of in-group psychological theory see generally O'Connor, supra note 12. One of the gaps in Burt's work, as I read it, is his failure to account for the role of group norms in regulating behavior within networks. However, this appears to be quite intentional and characteristic of the structural approach taken by Burt, because it also provides one of the virtues of the structural approach, in its relentless empiricism based on observable facts (which norms, beliefs, and values are not). Sociologists working in the general area of structural equivalence (of which structural hole theory is an example) have been criticized for failing to account for actors' "beliefs, values, and normative commitments." Mustafa Emirbayer & Jeff Goodwin, Network Analysis, Culture, and the Problem of Agency, 99 Am. J. Soc. 1411, 1425 (1994). See generally Frank & Yasumoto, supra note 42, for a more norms based social capital approach to competition and cooperation within competitive networks.} Leaving aside information that is understood by the group to remain confidential within it, contact with one is contact with all.\footnote{The likelihood of information moving from one member of the group to another is a direct function of the strength of their ties. Burt, supra note 27, at 18-19.} There is no gap in the network for you to occupy. As a result, you derive no informational benefits (although you might derive social pleasure) from establishing ties to other members of the group. The relationship is characterized by cohesion and is therefore redundant. Figure Two illustrates this idea.
Figure 2: Cohesion

Structural equivalence refers to people who have the same contacts. Take the group of friends mentioned earlier. Let us assume you have contacts with Ivan and Alyosha. If Ivan and Alyosha each have contacts with Fyodor and Dmitri, Ivan and Alyosha are structurally equivalent—your connections with Fyodor and Dmitri are indirect, but redundant through Ivan and Alyosha. You can magnify the structural equivalence even further if Fyodor and Dmitri each have contacts with Adelaida and Sofya. Your contacts with Ivan and Alyosha do nothing to increase your network advantages. Each provides you with exactly the same network benefits as your own tie to the other. Figure Three is an illustration of this point.

Figure 3: Structural Equivalence

Once we have determined that your contacts are neither cohesive nor redundant, we turn to the question of the efficiency and effectiveness of the network ties provided by your contacts. An efficient network is one in which you
maximize your nonredundant ties in a particular network (that is, ties that lack coherence and structural equivalence). It does you no good to maintain contact with members of the same network, since this produces precisely the redundancy you are trying to avoid. Maximizing your nonredundant ties maximizes the number of structural holes you get for each contact. In other words, you can maintain contact with the same number of people (that is, the network), at significantly lower cost by focusing your energies on only one primary contact in the network. As is implied by the concept of economic efficiency upon which Burt draws, such an approach allows you to maintain your contacts at the lowest possible cost.

For example, assume that you know Ivan and Alyosha, who also know each other. You also know Fyodor and Dmitri, who know each other but neither of whom know Ivan or Alyosha. You are wasting your time maintaining all four contacts. You're far better off forgetting about Alyosha and Dmitri. Your contact with Ivan gives you access to information from Alyosha, and your contact with Fyodor gives you access to information from Dmitri. By transferring your energies from all four contacts only to Ivan and Fyodor, you have increased the efficiency of your networks by maintaining four contacts for the cost of two. The time saved by dropping your contact with Alyosha and Dmitri can be used to expand your network by making other contacts. Figures Four and Five illustrate an inefficient and efficient network.

**Figure 4: Network Inefficiency**

71 There is an additional element necessary to create the efficient network as Burt discusses, and that is to select the right person in each network with whom to make contact. This largely turns on the question of who you think you can most trust in the network to provide you with timely and accurate information. Burt also notes that network efficiencies imply that you should choose as your contact not only the most trustworthy member of the group but the one with whom contact is easiest to maintain and who is most likely to reciprocate personal favors. *Ibid.* at 20-21.
Figure 5: Network Efficiency

This concept of nonredundancy explains why the existence of structural holes might increase corporate efficiency. The greater the extent to which managers are connected only to primary contacts in a network, the less time they waste by spending it with people who will provide redundant information. In order to see this, simply take the preceding example and imagine that Ivan and Alyosha, and Fyodor and Dmitri, respectively work in different divisions of the corporation. By dropping Alyosha and Dmitri, you get the same informational benefits with less work. To the extent that bureaucracies are information-producing organizations, using structure to achieve this result is a relatively low-cost way of improving the efficiency of this informational function. There is, however, a tradeoff for this increased efficiency, as I will later demonstrate. The corporation that undertakes to create structural holes may do so at its peril.

Network effectiveness is about increasing the total size of your network. The concept mandates that you deploy your resources efficiently by maintaining your primary contact, thus enabling you to use the remainder of your resources to establish and maintain primary contacts in different networks. Doing so, you will maximize the total size of your network by connecting to people who themselves have connections to other networks. In effect, you ideally choose one primary contact, and you leave to this person the job of developing and maintaining the cluster of contacts that form his network.\[^{72}\] In a sense, you

\[^{72}\] Burt is not entirely clear as to whether you should choose primary contacts who have the largest networks or choose primary contacts and maintain them while they build their networks by including more people (which increases your number of secondary contacts, and thus total contacts). The ambiguity arises from his assertion in the text that the primary contact maintains the network. *Id.* at 21. His diagram shows you maintaining the same number of primary contacts, but with their networks growing, which he also describes as the goal of network effectiveness. *Id.* at 20-23. I believe the second reading is more consistent with his theory and it is the reading I
are using your primary contact to diversify your network portfolio, while at the same time she increases it. Because you are the bridge among various networks, you receive information from one network that might be desirable to members of other networks. You are thus in a position to use that information as you see fit, or to use it (as we will later see) to manipulate the various networks to which you are connected. Network effectiveness contrasts with network efficiency. Network efficiency diminishes the cost of maintaining networks for each contact. Network effectiveness increases the overall size of your network.

Choosing effective contacts means choosing contacts in each network who are the most likely to make contacts with others. If they are not, they are probably not good choices for helping you to establish an effective network. Similarly, you should choose the contact in each network who is likely to be the most trustworthy and loyal to you. It is this combination of ability to build their own networks, trust, and loyalty in your primary contacts that enables you to build an effective network. As an example, assume that you know Ivan and Alyosha, neither of whom knows the other. Ivan and Alyosha become acquainted with Fyodor and Dmitri, respectively. Fyodor and Dmitri respectively become acquainted with Adelaida and Sofya. And Adelaida and Sofya become acquainted, respectively, with Pyotr and Vorokhov. Simply by maintaining your primary contact with Ivan and Alyosha, your network has expanded from two contacts to eight contacts. Assuming they are trustworthy and loyal, Ivan and Alyosha are effective primary contacts. This is an effective network.

While Ivan and Alyosha grow and maintain their networks, you are free to seek out other primary contacts. Now assume that you meet Ahab. From your conversation with him, you learn that the only other contact he has is Moby and, given your assessment of his personality, you don’t believe he is likely to expand his network beyond Moby. This is reason enough for you not to establish primary contact with him (not to mention your suspicions about his trustworthiness and loyalty). You are far better off using your energy to make a more promising additional primary contact. Figure Six illustrates network effectiveness.
Figure 6:

![Diagram showing network structure with nodes such as Ivan, Alyosha, Akab, Fyodor, Dmitri, Moby, Adelaida, Sofya, Pyotr, Vorokhov, and You.]

Effective Network

There are several situations in which network efficiency is not optimal. One, specifically mentioned by Burt, is a dense cluster of resources, like a board of directors and a CEO. The other is friendship networks because, in Burt’s felicitous phrasing, “Judging friends on the basis of efficiency is an interpersonal flatulence from which friends will flee.” As he says with respect to networks like the board:

> These clusters are so important to the vitality of the rest of the network that it is worth treating each person in them as a primary contact, regardless of redundancy. Saturation [the phrase he uses to describe such networks] minimizes the risk of losing effective contact with the cluster and minimizes the risk of missing an important opportunity anywhere in the cluster.

As I will later discuss, while treating the board as a saturated cluster (a single network in which the actor has ties to each member) may be important to the CEO’s success, it has more nefarious implications as well.

Thus far we have principally focused on the informational benefits of structural holes. There is, however, a second benefit of structural hole theory—control. Effectively

---

73 Id. at 24-25.
74 Id. at 25.
75 As with the efficiency-increasing properties of structural holes for information flows within organizations, the control benefits of structural holes may
exploiting a structural hole allows you to manipulate various networks to which you are connected (or individual people who are nonredundant contacts) by playing them off against one another in two different ways. The first way is exemplified by a simple economic transaction. Assume that you want to sell your house and you have two bidders. Clearly you can increase the price you will receive if you encourage them to bid against one another. Fostering this type of bidding war ties the control advantage of structural holes to the informational advantage previously discussed.

The second kind of control benefit occurs when one is situated between two players who have conflicting desires. This can occur, for example, when an associate in a law firm is working on major deals with two different partners, both of which heat up simultaneously, leaving each partner to compete for the associate’s time. The associate is in a position to play one off the other to his advantage. Robert Merton describes an actor in the associate’s position as “a more or less influential bystander” whose function is to make the partners’ conflicting demands their problem rather than the associate’s problem. (This is similar to the role of a child in a two parent family). Both strategies have implications for the structure of corporate governance.

All of this sounds terribly manipulative, and one might reasonably ask again what using structural hole theory to analyze corporate governance adds to the already rich tool box of neoclassical and behavioral economics. After all, these approaches assume manipulative or opportunistic actors. And the theory resembles one that describes ways in which actors can maximize their wealth by becoming social capital entrepreneurs.

Also increase efficiency depending upon the way they are exploited by the actor who identifies them.

76 All the better if neither knows the other’s bid—this often can get you a higher price than you would otherwise receive because their lack of information may lead one to bid more than is necessary to beat the other. Real estate brokers in hot markets have caught on to this idea by including escalation clauses in their clients’ contracts which set up an auction among multiple bidders. Such contracts specify incremental increases in bidding price a potential buyer is willing to offer in a contest to obtain a house and is usually capped at some amount. Such clauses have become de rigueur in parts of the Washington, D.C. metropolitan area.

77 ROBERT K. MERTON, SOCIAL THEORY AND SOCIAL STRUCTURE 431 (1968). The technical term for such a person is a tertius gaudens from an Italian proverb meaning “one who benefits.” I avoid the use of the term in the text to avoid introducing jargon unnecessary to the paper.
But the theory does add significantly to these other methodologies. First, it eliminates the neoclassical presumption that actors are autonomous and that the dyadic transaction is the appropriate focus of inquiry. Actors are not autonomous—they are embedded in social networks, and for all of its occasional reversion to neoclassicism, structural hole theory advances considerably beyond asocial neoclassical approaches by explaining the ways in which networks are themselves the sources of competition. Gaps in the network facilitate competition by creating opportunities to obtain positional advantages.

In addition, structural hole theory provides a powerful alternative to neoclassical analysis by eliding the need to identify wealth (or anything else for that matter) as a maximand. People behave entrepreneurially for a variety of reasons: for instance, the simple psychological desire to succeed, the fun of the game, and the maximization of wealth, among others. Burt believes that clarifying opportunities is motivation enough in and of itself. That is to say, given two opportunities an actor will take the clearer path. Thus, Burt treats motive and opportunity as equivalents. This may appear to be a cop-out; after all, the manipulative behavior observed and, to some extent, prescribed by, structural hole theory is unpleasant to many of us. Moreover, the opportunities described are frequently, if coincidentally, wealth maximizing opportunities.

78 In this respect it also advances beyond transaction cost economics which remains mired in neoclassical assumptions about motivations and goals.

79 BURT, supra note 27, at 35.

80 Id. at 35-36.

81 Burt effectively admits that this stance is a cop-out. “I am begging the question of how opportunity and motivation are connected. I emphasize the causal priority of opportunity. The opposite emphasis is traditional in sociology. . . . Here I emphasize opportunity because I can analyze it in a rigorous way with network concepts and describe a great variety of empirical events.” Id. at 275 n.13. His dispensation with concerns of motivation (although they do appear in Chapter Seven) isn’t especially troubling for my purposes because the point of this project is to develop a purely structural analysis that can identify particular points of weakness or strength within the corporate structure.

82 Others, however, believe that structural instrumentalists like Burt actually rely upon utility maximization. See Emirbayer & Goodwin, supra note 69, at 1428.

82 Some of this may derive from Burt’s use of American institutions as his empirical base. This practice embeds his theory in a society which, at least in the economic realm, has come to embody many of the characteristics described by the neoclassical model. See Nicole Woolsey Biggart & Gary G. Hamilton, On the Limits of a Firm-Based Theory to Explain Business Networks: The Western Bias of Neoclassical
For purposes of this article and its focus on the CEO (who has already demonstrated network entrepreneurship by climbing the corporate ladder), it may be enough to leave the question of motivation open, since control is central to structural hole theory and control is, after all, the job and goal of the CEO.\textsuperscript{83} But for those who demand a maximand in order to appreciate a theory, perhaps we can make do for purposes of this discussion with maximization of control as the behavioral motivation. For the broader uses of structural hole theory (perhaps even broader uses within the context of analyzing corporate behavior at lower levels, or at the board level), it is enough to say that—at least in its descriptive and, in my adaptation of it, diagnostic aspects—structural hole theory needs no particular motivational assumption in order to enable us to examine organizational structure. Thus, our goal is to identify, as points of weakness, holes that can be occupied by actors who are then able to manipulate corporate behavior or to control information flows in ways that may be disadvantageous to the corporation. Whether or not corporate actors take advantage of the opportunities provided by structural holes, the important observation is the manner in which structure creates these opportunities. In this way, structural holes can be seen as a source of freedom obtained by opening up networks, where desirable, to either permit a freer flow of information or to destroy the possibility of control monopolies. Whatever the motivation of particular actors might be, the opportunities that are created by network structure can be used for good or ill (or not used at all). Knowing where they are and their potential

\textit{Economics, in Networks and Organizations, supra note 44, at 471, 488. See also Granovetter, supra note 59, at 447 (noting that rational or instrumental behavior aims not only at economic goals but also “at sociability, approval, status, and power”). On the other hand, the structural equivalence approach to sociology of which Burt’s work is an example has been criticized for conceptualizing actors in “narrowly utility-maximizing and instrumental ways.” Emirbayer & Goodwin, supra note 69, at 1425. I do not find this especially troubling; despite my personal distaste for neoclassical analysis, because these motivations, while perhaps implicit (especially in economic sociological work) are not necessary aspects of the theory which itself relies solely upon structure. So while narrow instrumentalism might be implicit in a structuralist account, it is tangential and irrelevant to the explanatory power of the theory. For a motivational account analyzing the interrelationship between structure and psychological motivation, see James N. Baron & Jeffrey Pfeffer, The Social Psychology of Organizations and Inequality, 57 Soc. Psych. Q. 190 (1994).}

\textsuperscript{83} Harrison C. White, Agency as Control in Formal Networks, in Networks and Organizations, supra note 44, at 92. (“To manage is to make use of ties. To gain and maintain control requires attending to networks of ties.”) The classic work on the subject of the role of the CEO is Chester Barnard, The Functions of the Executive (1938).
effects allows us greater insight into appropriate organizational design.

IV. THE APPLICATION OF STRUCTURAL HOLE THEORY TO PROBLEMS OF CORPORATE GOVERNANCE

Corporations are webs of social networks. The power of an actor in a social network depends more on his connections between networks than on his position within a given network. Simply put, the corporate actor with the most nonredundant network contacts (who has bridged the greatest number of structural holes) is better positioned to monopolize information, engender competition between networks or other actors (i.e., the house auction), and control the behavior of other actors (i.e., the university president and the law firm associate), which may translate into the ability to determine his own power and profit. These opportunities are available to him because, by taking advantage of structural holes, he effectively becomes the sole vector of a variety of separate networks, a position that gives him substantial autonomy in managing intracorporate relations and, as a consequence, his own position in the corporation.

The basic aspects of structural hole theory and their applicability to the problem of corporate boards should now be clear. Prior to the 1980s, most public corporation boards were comprised of a majority of non-independent directors. The tradition of inside directors (or at least directors with substantial ties to the corporation and corporate management below the level of CEO) began in the early stages of bureaucratic corporate growth and persisted into the 1980s. These directors were, generally, high level executives of the corporation or so-called “gray” directors: the corporation’s bankers or lawyers who regularly worked with upper management. Whatever the problems of these inside directors,

---

84 This is a bit of an oversimplification, but suffices for purposes of discussion. For a more complex depiction, see Burt, supra note 27, at 22 fig. 1.4, which illustrates the process of optimizing your structural holes.
85 See id. at 45-49; Burt, supra note 62. See also Mark S. Granovetter, The Strength of Weak Ties, 78 Am. J. Soc. 1360 (1973)
86 Bhagat & Black, supra note 15, at 238.
87 These are commonly referred to as “affiliated directors” and are typically professionals like lawyers and investment bankers, whose day to day work with the corporation typically involves substantial and sustained contact with executives other than the CEO.
88 Gray directors are typically classified as a form of outside director. I depart
real or perceived, the very nature of their relationships with
the corporation ensured that they had ties to corporate
managers below the level of CEO. Put in terms of network
theory, they had ties to networks that were independent of the
CEO. Thus the number of nonredundant corporate contacts
they had, and therefore their access to corporate information,
was not dependent solely upon the CEO. These ties to networks
below the CEO gave them independent access to corporate
information, putting them in a position to verify information
provided by the CEO (whether or not they actually took
advantage of it by challenging the CEO).

Beginning in the 1980s, the number of inside directors
on corporate boards began to diminish as corporate boards
moved toward independence. This shift was spurred, in part, by
two developments. First, the American Law Institute adopted
its Principles of Corporate Governance, which initially
advocated independent directors (directors without personal or
financial ties to the corporation) in a variety of contexts.
Although the ALI took a more moderate position in its final
draft, the Principles do envision a “monitoring” board with a
substantial number of independent directors, a concept which
itself implies substantial directorial independence. Second, the
Delaware Supreme Court, in fighting its way through a whirl
of takeovers, tried to develop doctrine appropriate to evaluate
boards’ fulfillment of their fiduciary obligations. The Court
noted repeatedly that boards dominated by independent
directors were likely to receive more deference than boards
with a substantial number of inside directors. As finance came

from this usage because, with respect to the problem I’m discussing, they are situated
more like inside directors than independent directors who are true outside directors.

Westphal and Zajac have noted the substantial increase in boards
-dominated by outside directors. James D. Westphal & Edward J. Zajac, Defections from
the Inner Circle: Social Exchange, Reciprocity, and the Diffusion of Board Independence

The moderation largely was a result of political pressure from large firm
lawyers protecting their perception of their clients’ interests.

be a sociological dimension to the proliferation of outside boards. Gerald Davis and
Henrich Greve, examining the spread of golden parachutes and poison pills among
corporations, look to network theory to explain how the adoption of these devices may
well be attributable to information flows through interlocking directorates. Gerald F.
Davis & Henrich R. Greve, Corporate Elite Networks and Governance Changes in the
1980s, 103 AM. J. SOC. 1 (1997). The same mechanism might at least in part have been
to dominate management and the intercorporate transaction became an important tool of continued corporate development, the need for corporations to be free of fiduciary taint increased. Boards became increasingly independent. 92 Finally, the corporate scandals of 2002 led not only to the passage of the Sarbanes-Oxley Act, with its emphasis on directorial independence, but also to New York Stock Exchange reforms requiring listed corporations to have a majority of independent directors. Other business associations joined in advocating largely independent boards as well. 93 With directorial independence, and the absence of corporate insiders other than the CEO and perhaps one other executive on the corporate board, directors began to lose their network ties into the corporate structure. Ties with managers and other corporate insiders were eliminated, and replaced by a single tie into the network, the CEO. 94 As a result, directors became more dependent on the CEO for their information. Directors were not left completely isolated. Executives themselves belong to networks of other executives that tend to produce institutional conformity among corporations. 95 The relatively small number of directors of large American corporations compared to the total number of corporations suggests some degree of director overlap. In addition, a substantial subset of directors are CEOs of their own corporations, which provides them with significant network ties outside of the particular corporations they serve. While this may produce a level of conformity among directors, it does not itself suggest that the majority of independent directors have network ties within a given corporation beyond the CEO and any other insider who happens to sit on the board.

---

92 See the survey data in Bhagat & Black, supra note 15, at 239.
93 Other organizations, like the Business Roundtable, have jumped on the bandwagon. BUSINESS ROUNDTABLE, STATEMENT ON CORPORATE GOVERNANCE (1997). The Business Roundtable is composed of leading CEOs. Their support for independent boards should not be surprising once one understands that this is entirely in their self-interest.
94 But see James D. Westphal, Collaboration in the Boardroom: Behavioral and Performance Consequences of CEO-Board Social Ties, 42 ACAD. MGMT. J. 7, 16 (1999) (finding empirical evidence that corporations with CEO social ties to board members improve collaboration and advice seeking and that such corporations show improved performance).
These observations, in light of the insights provided by structural hole theory, suggest two hypotheses:

1. Corporations that have inside boards\(^{96}\) will have a weak CEO; and
2. Corporations that have independent boards will have a strong CEO.

These hypotheses are counterintuitive and contrary to most of the accepted wisdom of corporate scholarship.\(^ {97}\) It has long been an article of faith, as well as accepted by corporate scholarship\(^ {98}\) and in caselaw,\(^ {99}\) that corporations with inside boards have the strongest CEOs.\(^ {100}\) Insiders, dependent for their jobs on the good will of the CEO, are unlikely to oppose him. This argument is fundamentally flawed. It assumes that the work of the board actually takes place at the board level. As a number of studies have demonstrated,\(^ {101}\) this simply is not true. To the extent that board members seek to challenge CEOs, they tend to do so in discussions outside of the board meeting. And, if confrontation with the CEO is to occur, it does so either informally or is choreographed in advance by the dissidents. While I lack evidence to substantiate this point, it follows logically that challenges to the CEO would be posed by independent directors, if at all. But the clear implication of network theory is that the independent directors will have received the information to challenge the CEO from their relationships, established on the board, with inside directors whose networks extend deep into the corporate hierarchy. By virtue of these relationships, the independent directors have sources of information independent of the CEO. They need not rely upon the kind or quality of the information the CEO

\(^{96}\) By “inside board” I do not mean to suggest that the board is exclusively or, even necessarily mostly, composed of insiders, only that a significant number of insiders other than the CEO sit on the board.

\(^{97}\) But see note 14 and accompanying text, for recent questioning of the desirability of independent boards and attention to CEO power in relation to the board. Although none of these studies provides a direct connection between independent boards and CEO power (although Westphal, supra note 19, comes close), they clearly intuit the problem. The hypotheses not only clearly identify the problem but provide an explanation for these intuitions.

\(^{98}\) See supra note 14 and accompanying text.


\(^{100}\) But see the work done on the relationship between independent boards and executive compensation cited supra note 14.

\(^{101}\) The classic, if dated, studies are Jay Lorsch & Evelyn MacIver, Pawns or Potentates: The Reality of America’s Corporate Boards (1989) and Myles L. Mace, Directors: Myth and Reality (1971).
presents. Given this greater access, the directors on an inside board, whether inside or independent, will be in a better position to challenge the CEO (even if the challenge is fronted by the independent directors to shield the inside directors from termination or retaliation for perceived disloyalty to the CEO). Hypothesis one is illustrated by Figure Seven.

Figure 7: Corporations that have inside boards will have weak CEOs

Hypothesis two suggests that in a corporation with an independent board, the CEO will be strong. The reason should by now be obvious. Unlike the corporation with an inside board, containing members with substantial network ties into the corporate hierarchy—the corporation with an independent board has only one structural hole between the subordinate managers and the board. In other words, when the board is completely independent, the nonredundant contacts bridging structural holes between the corporation and the members of the board do not exist. The single structural hole is bridged by the CEO, who, as the only contact between managers and the board, is the board’s sole source of information. Unless a

102 Unless of course the CEO chooses to present insiders at board meetings or facilitate their contact with independent board members. It would, it seems, be irrational for a CEO to do this because in so doing he relinquishes his informational monopoly and the full power of his control position.

Of course auditors may provide a source of financial information to the board independent of the CEO. But the work of auditors is typically done in conjunction with management and, following Sarbanes-Oxley, we can expect the CEO to have an active role in the process of preparing audited financial statements. Moreover, the corporate scandals of 2002 demonstrated that independent auditors have greater incentives to please their employer than to provide truly independent audits. While this is likely to change, it only mitigates the problem at issue—it does not eliminate it.
subordinate manager has independent network ties to one or more of the independent directors, only the CEO is in a position to control and manipulate information flows to the board. This leaves the CEO in an enormously powerful position, with every incentive to present information to the board in a light that is most favorable to him. The board is constrained by the information the CEO chooses to present and how he chooses to present it. Consequently, the CEO in a corporation with an independent board should be strong. The following figure illustrates hypothesis two.

**Figure 8:** Corporations that have independent boards will have strong CEOs

It should therefore come as no surprise that there is some (albeit disputed) evidence that CEOs with independent boards tend to receive the highest compensation. But the reason may not be, as some suggest, the board’s lack of sophistication with respect to option pricing and a consequent

---

103 Langevoort, *supra* note 5, at 812 discusses the CEO’s incentive to manipulate information to outside directors. See also Westphal, *supra* note 19 (discussing and analyzing CEO ingratiatiion and persuasion to offset the power of outside directors).

104 Westphal, *supra* note 19, provides empirical support for what he calls greater CEO ingratiatiation and persuasion behavior where the corporation has an independent board. Ingratiatiation and persuasion may not rise to the level of manipulation, but they’re not far removed from it.

105 This helps to explain the superior bargaining position of the CEO in compensation matters observed by Bebchuk et al., *supra* note 5, as well as suggest opportunities for the CEO to engage in what they interestingly describe as “camouflaging” his compensation package. *Id.* at 789.

106 The empirical evidence is mixed on whether this is accurate. See *supra* note 23 and accompanying text.
tendency to be overly generous with options. Instead, the reason may be that the independent board is reliant upon the CEO for information with respect to his own performance, information that can easily be manipulated or suppressed by the CEO because of his position as the sole source of information. This also may explain Bhagat and Black’s evidence that independent boards do not necessarily improve corporate performance and may in fact make it worse.

Finally, and perhaps most powerfully, it may account for one of the principal ways in which opportunities for executive misbehavior—and resulting corporate scandals—are created. As the sole bridge between corporate management and the board the CEO is put in an enormously powerful position. He has a monopoly over the information delivered to the body ultimately responsible for the integrity of corporate management and information.

At this juncture, it is important to note some qualifications. Corporations of any size, and especially public corporations, are highly complex. As they face a variety of different challenges, there will be exceptions to the hypotheses. One clear case, presented in Joy v. North, is where a CEO has an insider dominated board that he has, by force of personality, largely subjugated to his control. Or, a corporation with an independent board may have directors who themselves possess sufficiently strong personalities as to insist upon verifying the information presented by the CEO. There are also certain

---

107 See Murphy, supra note 14.
108 Bebchuk, Fried, and Walker recognize this problem, supra note 5. See also Novaes & Zingales, supra note 41 (discussing the relationship between bureaucratic structure and compensation arrangements).
109 Bhagat & Black, supra note 15, at 263.
110 This does not mean that the CEO is unconstrained. Obviously the CEO will be dependent for his information on the information supplied to him by his subordinates. This can create its own informational problems, as I discuss infra Part VI.
111 692 F.2d 880 (2d Cir. 1982). This is likely to be a relatively rare situation. Persons—even insiders—who sit on corporate boards tend to be powerful and highly successful in their own right, and while psychological theories, as well as studies of director selection, suggest that these directors are or become friendly to the CEO, the situation in which directors are complete pushovers is unlikely to be frequent.
112 This may well be increasingly the case, strong personalities or not, in light of the recent corporate scandals in which various boards were accused of failing to pay proper attention to internal corporate transactions, like the conflict of interest transactions approved by the board in the Enron case. William C. Powers, Jr., 107th Congress, Report of Investigation by the Special Investigative Committee of the Board of Directors of Enron Corp. 1-17 (Feb. 1, 2002) [hereinafter Powers Report].
types of transactions, like board-authorized compensation, for which an inside board may be beneficial to the CEO (particularly if the compensation plan simultaneously benefits the senior managers who sit on the board). Other types of transactions, like takeover situations, may focus the board (at least in part because of potential legal sanctions for failing to fulfill their duties in highly visible transactions), such that they become more demanding of verified information (as, for example, in the employment of investment bankers to give fairness opinions). And a situation such as a corporation in reorganization may result in a grant of extraordinary power to the CEO (or result in a new CEO with such power), notwithstanding board composition or corporate structure. Nonetheless, for ordinary monitoring situations of the type that evidently failed in the recent corporate scandals, the hypotheses hold as a theoretical matter.

Finally, I recognize that in basing the hypotheses on a wholly independent board (except of course for the CEO), I am relying on an ideal type that may not exist in practice. But using the ideal type allows us to clarify the problems an independent board faces with respect to the CEO, and formulate better questions for analysis.

V. EXTENDING THE IMPLICATIONS—SENIOR MANAGERS

It should be apparent that what holds true for the CEO has the potential, under the right circumstances, to hold true for senior managers. For purposes of this paper, I define senior managers as those executives who are the CEO’s immediate subordinates and, depending upon the corporation’s internal structure, the managers directly below them.

In order for senior managers to seize the opportunities presented by structural holes at the top of the corporate structure, at least one of several conditions must exist. One potential condition is that the CEO, despite his rank, is a poor network player, either because of his inability to recognize structural holes or because he has allowed them to proliferate.

---

113 The evidence on whether independent or inside boards overcompensate CEOs is disputed. Even in these circumstances, the CEO’s ability to control information about his own performance with respect to an independent board may give him an edge.

114 I know this is the typical law professor’s cop-out but I am truly unequipped to do this kind of empirical work. I do hope, in the not too distant future, to find a way to conduct adequate empirical testing of the hypotheses.
at high levels. Another is that one or more senior executives might collude with the CEO in order to control and manipulate information with respect to the board. A third possibility is that the CEO intentionally creates structural holes in order to separate himself from the manipulation of information and control that would then be possible at the senior manager level. A final condition would be that the senior executives are good network players, with efficient and effective networks within the corporate hierarchy, and the ability to identify structural hole opportunities. Several examples will illustrate the existence of these conditions and the resulting possibilities for senior management misconduct.

The poster child of corporate scandals, the Enron fiasco, presents the first example. The facts of the decline and fall of Enron are well known. Enron engaged in a number of off-balance sheet transactions which, contrary to generally accepted accounting principles, it failed to disclose, resulting in highly overstated earnings and highly understated assets. Moreover, some of the transactions engaged in were shams. These transactions provided certain executives (most notably Andrew Fastow, the corporation's chief financial officer, and some of his subordinates), with opportunities for rich rewards. The collapse of Enron resulted in an internal investigation and well-publicized congressional hearings which involved the testimony of a number of the Enron participants.

Most striking in both the reports and the hearings were CEO Jeff Skilling's continuous denials of knowledge of the transactions, their financial structures, and the extent of Fastow's conflicts of interest and profit opportunities. (Some of the transactions took place under the watch of Skilling's predecessor CEO, Kenneth Lay, who refused to testify. The Powers report suggests the extent to which Lay may have lacked knowledge.) Particularly striking was Skilling's congressional testimony, under oath, that he was not aware of the mayhem transpiring beneath him in the managerial ranks.

115 POWERS REPORT, supra note 112.
117 The Powers report is critical of the board, which evidently approved certain transactions on the basis of incomplete knowledge. POWERS REPORT, supra note 112, at 1-17. This of course is consistent with the notion that boards lacking strong network ties into the corporate hierarchy are more easily manipulated.
The Powers report also suggests that Skilling may have been uninformed at some level.\textsuperscript{118}

The common wisdom, I believe, was that Skilling and perhaps Lay were lying.\textsuperscript{119} But it may well be that Skilling was telling the truth. How could a chief operating officer (later CEO) be so ignorant of what his immediate subordinates were doing? One answer is that Skilling, a highly accomplished network player, created a structural hole between himself and Fastow, which he intentionally left unbridged. This break established chains of command and reporting systems designed to stop with Fastow. Information would only reach Skilling selectively, if at all. If this was the case, it is a superb example of how a CEO can create a structural hole in order to protect himself from potential liability.

Why would a CEO do this? In Skilling's case, the answer may be because he needed Fastow's manipulations in order to maintain Enron's appearance as a highly valuable and continually growing corporation. That, after all, is what the off-balance sheet transactions were designed to do. The price for Fastow's cooperation may well have been the enormous compensation he received for his role in managing these entities. Assuming Skilling was aware both of Fastow's skill (as he undoubtedly was) and his lack of character, it would have been to Skilling's advantage to allow Fastow to operate as a free agent, while Skilling created plausible deniability through the creation of a structural hole—severing the link between COO (later CEO) and CFO.\textsuperscript{120}

The financial fraud of MCI/Worldcom presents another possibility: senior managers colluding with the CEO. This case also presents a nice example of board ignorance as a result of managerial manipulation of information. Worldcom (as it is

\textsuperscript{118} Id. at 169-72.

\textsuperscript{119} Of course this is not in itself necessarily probative of their lack of knowledge; only that the government has yet to make a case.

\textsuperscript{120} This is where Burt's description of structural holes both as chasms and as interlocks creates possibilities for expanding the utility of the theory. The use of structural holes in the context described in the text envisions the creation of a chasm which remains unfilled despite the fact that presumably both Skilling and Fastow perceived the opportunity to fill it. This defensive use of structural holes is not part of Burt's explanation of the theory. It is consistent with the theory I develop here based upon Burt's work.

I also do not mean to suggest that as a moral or ethical matter we ought to accept Skilling's denial of knowledge as a legal defense if indeed this is what happened. If I am right about the structure, however, it could make it difficult for prosecutors to establish a strong case against him.
now known), through chief financial officer Scott Sullivan and several of his subordinates, simply made up numbers that showed higher profits and revenues and lower costs than were actually the case. CEO Bernie Ebbers was convicted in 2005 for his role in the fraud. Yet it remains unclear how much he knew. A jury found that, despite Ebbers’ denial of knowledge, he in fact was aware of the fraud and may have participated in it.\footnote{Robert Frank et. al., Executives On Trial: Scandal Scorecard, WALL ST. J., Oct. 3, 2003, at B1, B4.} As in the case of Enron, the clear purpose of the financial fraud was to support and increase the company’s stock price despite underlying business conditions that would have in fact damaged the company (or at least its stock price) had they been properly disclosed.

To the extent that Ebbers was involved, Worldcom presents a paradigmatic example of collusion among the CEO and senior managers to take advantage of a structural hole. In contrast to the Enron case in which, on my reading, Skilling intentionally broke the network tie between himself and Fastow, in this case Ebbers and the senior managers involved themselves formed a network. This network stopped at the CEO and allowed the board to remain in the dark.

\section*{VI. Extending the Theory—The Relationship Between Board Structure and Internal Corporate Structure}

Thus far my analysis has proceeded as if there were no significant differences among managerial systems. In fact there are two basic paradigms of internal corporate structure. The first is the traditional, hierarchical corporate bureaucracy, the development of which is described thoroughly and compellingly by Alfred Chandler.\footnote{See generally Alfred D. Chandler, The Visible Hand: The Managerial Revolution in American Business (1977).} By this account, American big business grew through the first half of the twentieth century as multi-unit enterprises that, of necessity, employed legions of corporate managers at various levels in huge bureaucracies. Such bureaucracies persisted well into the end of the century.\footnote{See David M. Gordon, Fat and Mean: The Corporate Squeeze of Working Americans and the Myth of Managerial “Downsizing” (1996).}

The second kind of corporate structure is more condensed. Less hierarchical corporations rose in prominence
as the large bureaucracies began to come apart. These “flatter” corporations are organizations in which layer upon layer of middle management has been stripped out, leaving in its place a CEO, a number of high level executives on a relatively equal level, and several middle level managers above the worker. Raghuram Rajan and Julie Wulf, after surveying relevant information with respect to more than 300 large American corporations over fourteen years, concluded that the number of managers reporting directly to the CEO has increased, and the levels of managers between the CEO and the lowest manager with profit responsibility has decreased. The number of managers reporting directly to the CEO increased by 61% between 1986 and 1999, while the number of positions between the CEO and the lowest manager with profit responsibility decreased by 25%. This is clear evidence that American corporate hierarchies are flattening.

Each organizational type is designed to facilitate the flow of information. Organizational theorists argue that the particular internal structures, or information paths, vary with the circumstances of the corporation. Bureaucratic structure tends to be found in stable industries where such factors as technology, supply, demand, and production are relatively predictable and unvarying. In this situation, the bureaucracy works precisely because it is designed for control, and control is possible and likely efficient because of the predictability of the business environment. Control is ensured by the creation of rigid hierarchies and separated divisions in which information flows up the pyramid to the responsible senior executive. By contrast, horizontal structures work best in industrial environments, where information and technology are rapidly changing, and the need for control is supplanted by the need

---

124 This is not to say that there are not a significant number of American corporations that still operate as hierarchical bureaucracies—only that they have become less dominant in the corporate population and tend to be concentrated in certain industries.

125 Raghuram Rajan & Julie Wulf, The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies, NAT'L BUREAU ECON. RES., Apr. 2003, available at http://papers.nber.org/papers/w9633.pdf. Rajan and Wulf are interested primarily in the effects this trend has had on managerial pay as well as the causes of the flattening (which they conclude is due to exogenous pressure in the form of technological and environmental changes.) These insights, while obviously important, are not especially relevant to the central argument of this paper.

126 Nohria & Berkley, supra note 20, at 118 (“Organizations control on the basis of knowledge; They are information-processors, or what could be seen in a metaphorical sense as a kind of human-based computer.”)
for flexibility. Such flexibility is best obtained by loosening the bureaucratic hierarchy and allowing information to flow freely through the various intracorporate networks. Building upon structural hole theory as applied to CEO-board relationships, these observations regarding internal corporate structure lead me to formulate two additional, although tentative, hypotheses:

1. Corporations that are structured hierarchically and have independent boards will have a moderately weak CEO; and

2. Corporations that are structured horizontally and have inside boards will have a moderately strong CEO.

Before examining the two hypotheses individually, I propose four expanded hypotheses that result from the conflation of the two theories:

1. Corporations that are structured hierarchically and have inside boards will have a weak CEO;

2. Corporations that are structured hierarchically and have independent boards will have a moderately weak CEO;

3. Corporations that are structured horizontally and have inside boards will have a moderately strong CEO; and

4. Corporations that are structured horizontally and have independent boards will have a strong CEO.

The hypotheses in this broader iteration can also be presented in the form of a matrix:
I previously explained the two principal hypotheses without reference to bureaucratic structure. Let me now elaborate the more complex hypothesis. The information flows to independent directors from inside directors are likely to be substantially richer in bureaucratically-structured corporations than in horizontal corporations. In a bureaucratic corporation, almost by definition the CEO will have relatively few contacts in the corporate hierarchy. That is, he will have relatively few subordinates and bridge relatively few structural holes (as many structural holes as he has immediate subordinates with different responsibilities and therefore different networks). Recall that the ideal situation for an actor to exploit his information and control opportunities is to maximize the number of structural holes he bridges, or to maximize the number of his nonredundant network contacts. Bureaucratic efficiency implies a chain of command, narrowing to only a few immediate subordinates below the CEO. Thus the bureaucratic structure provides, at least in an ideal state, few structural hole opportunities—few nonredundant contacts—for the CEO.

Rajan and Wulf's study found that, indeed, the number of subordinates reporting directly to the CEO was significantly smaller in a more bureaucratic era. As recently as 1986, the median CEO in their study group had 4.4 executives reporting directly to him.\textsuperscript{127} A web of hierarchy extended below these executives, networks upon networks. This web of networks goes much deeper in a bureaucratic corporation than in a horizontal corporation and it broadens considerably as one proceeds down the hierarchy from the CEO and his handful of subordinates. One implication is that there are a greater number of points within the pyramidal hierarchy at which network entrepreneurs can exploit structural holes that are likely unavailable to the CEO. The opportunities for informational and control advantage are pushed further down the chain of command to lower level managers. These managers, if they are good network entrepreneurs, can exploit these structural hole opportunities with the possible result that they can manipulate or stop information before it reaches the level of the CEO or his immediate subordinates.

While this argument suggests that top level executives who would serve as inside directors also have a limited number of direct reporting contacts in the bureaucratic corporation, the

\textsuperscript{127} Rajan & Wulf, supra note 125, at 1.
issue is more complex. The theory does not necessarily imply that the executives’ only contacts in the hierarchy are those reporting directly to them (in a way that is probably more characteristic of the CEO), and thus that they face informational constraints similar to the CEO. On the contrary, these managers are likely to have more subordinates (and thus more network contacts) that reach deeper into the hierarchy than the CEO. Glenn Carroll and Albert Teo found that higher level managers tend on average to have wider social networks than do lower level managers, both within and independent of the corporation. These social networks include other high level managers not directly involved in that particular manager’s area of expertise, suggesting more horizontal networks across the corporation that can serve as crucial sources of information. Moreover, Burt’s work shows that not only do managers’ network sizes increase with rank, but that in establishing their networks, managers typically choose a “strategic partner[,]” and that the most effective strategic partner tends to be one of their bosses’ boss. The implication is that higher level managers also have networks that go deep into the bureaucratic structure, as lower managers leapfrog the middleman and align themselves with higher level managers. The higher the level of manager, therefore, the more likely they are to have both broad and deep networks. This line of reasoning may also suggest that some second-tier executives

---

128 Carroll & Teo, supra note 95.
129 The specific position of the CEO is not addressed. But it seems likely, given the nature of bureaucratic corporations, that the CEO will be more isolated in the executive suite and less likely to have network ties to lower level managers. Moreover, given the presumed desire of CEO subordinates to succeed the CEO, one should expect these executives to adopt the CEO as their strategic partner (which of course puts the CEO in a position to manipulate them in the manner of the law firm associate I earlier described).
130 BURT, supra note 27, at 125.
131 Id. at 146, 150. It is not at all clear, though, that Burt would agree with my analysis. Although he doesn’t address the point directly, he does point out that a manager with a disorganized work force (which may correlate with my description of some horizontal corporations) may be more able to exploit his workers to his advantage than the manager of a well-organized work force (which may correlate to my description of some bureaucratic corporations). Id. at 189.
132 It is, of course, possible that the CEO serves as strategic partner to executives below the level of his immediate subordinates and thus establishes network contacts for him deeper in the corporate structure than I have suggested in the text. But even if there is some depth to CEO contacts in hierarchies, the flattened corporation provides more subordinates of the CEO’s subordinates with whom he can make contact, thus suggesting that the bureaucratic or flattened structure of the corporation should be expected to have an effect on the CEO’s power independent of board composition.
strategically partner with the CEO, but given the CEO’s ultimate responsibility for the corporation and his relatively limited time, it may well be more likely that the CEO remains relatively isolated in the executive suite. If the CEO’s subordinates maximize their nonredundant contacts, they can take advantage of the structural hole opportunities to gain greater information and control than is available to the CEO.

All of this suggests that the managers on a level directly below the CEO (the managers reporting directly to the CEO) are likely to have significant nonredundant network contacts that provide them with important information flows. How and what they choose to report to the CEO who, theoretically, is more constrained, is within their control. The fewer the number of executives reporting directly to the CEO, the more limited the CEO’s sources of information, and the more control maintained by the senior executives. Thus it would appear that the CEO’s position in the bureaucratic corporation is highly dependent upon the extent to which his immediate subordinates refrain from opportunistic use of their informational positions, gained through their multiple networks which provide them with more structural hole opportunities than the CEO. The CEO, in choosing his subordinates, must choose effectively—subordinates who are trustworthy and reliable—in order to maximize his power within the bureaucratic structure and prevent his own informational isolation.

The likely conclusion is that inside directors in a bureaucratic corporation are structurally more wired than the CEO, and have independent relationships (by virtue of their board seats) with independent directors. The insiders’ nonredundant contacts—their structural hole opportunities—are more numerous and likely deeper than those of the CEO. Now, the opportunity for critical information to bypass the CEO is structurally apparent. Thus, network theory suggests that the bureaucratic corporation with insiders on the board is likely to have a relatively weak CEO. Figure Nine illustrates this hypothesis.
Figure 9: Corporations that are structured hierarchically and have inside boards will have weak CEOs

The second hypothesis logically follows from the first. Bureaucratic corporations with an independent board are likely to have moderately weak CEOs (CEOs that are stronger than in the bureaucratic/inside board corporation and weaker than CEOs in each configuration of the horizontal corporation). The same structural conditions I discussed above with respect to the hierarchy continue to exist. This time, however, the CEO's subordinates lack direct ties to the independent directors. Information flows through the CEO. The CEO is not in as strong a position of informational control as he is in a horizontal configuration because his immediate subordinates (and therefore network contacts) are still few in number and sufficiently connected within the corporation to maintain their ability to control and manipulate the flow of information to the CEO. But unless they have relationships with at least one independent director beyond the context of the particular corporation (which is possible if, for example, a chief financial officer sits on the board of a corporation on which one of the independent directors also sits), they have no informational bypass route around the CEO. Thus, whatever the quality and completeness of the information the CEO receives, he fills the single structural hole between the board and his subordinates and is in a position to control and manipulate the information received by the board. The structure of the corporation suggests that the quality and quantity of information he receives will vary with the integrity and loyalty of his subordinates, but his position with respect to the board is strong. Figure Ten illustrates the hypothesis.
Figure 10: Corporations that are structured hierarchically and have independent boards will have moderately weak CEOs

Some empirical evidence suggests that this indeed is the case. Ranjay Gulati and James Westphal demonstrated (through the empirical examination of joint venture formation) that “[i]ndependent board control over management may actually produce a negative relationship between the CEO and the board characterized by a lack of mutual understanding and distrust.”\(^{133}\) (This, of course, is consistent with Bhagat and Black’s work and the studies cited earlier suggesting that independent boards actually impede corporate performance.\(^{134}\)) Gulati and Westphal attribute this largely to the dissatisfaction felt by a CEO who believes he is under significant scrutiny by an independent monitoring board. While their conclusion with regard to mistrust seems compelling, their explanation relies upon psychological motivations that are difficult to verify. While the two are not necessarily mutually exclusive, I present a theory that does not rely on speculation, self-reporting, or psychological analysis. The board is right to distrust the CEO because the CEO is in a structural position to block or manipulate the flow of information to the board. This conclusion, as will be clear, has its strongest implications for hypothesis four.

The third hypothesis introduces the horizontal corporation, a structure that has continued to proliferate over the last several decades. The bureaucratic corporation derives

\(^{133}\) Gulati & Westphal, supra note 21, at 477, 498.  
\(^{134}\) See Bhagat & Black, supra note 15, and supra text accompanying notes 15, 17, 23, and 105.
from the traditional, Taylorist-inspired,\textsuperscript{135} pyramidal structures, consisting of line workers to legions of middle managers to Janissary corps of junior executives to the chosen few in the executive suite and, ultimately, the CEO. This structure has, in many industries (and to some extent within industries) significantly flattened. Instead of pyramidal hierarchies, we are now more likely to see corporations that have stripped out much of the middle managerial ranks and spread the executive power over a larger horizontal range of senior managers, who report not to the CEO’s subordinates, but to the CEO directly. Rajan and Wulf found that the average number of senior managers reporting directly to the CEO had increased by 61\%, from 4.4 to 7.2, between 1986 and 1999.\textsuperscript{136}

In the horizontal corporation, the managerial levels below the top executives are dramatically reduced and the number of executives reporting directly to the CEO is significantly increased. From the standpoint of structural holes, the consequences are immediately apparent. While the CEO’s subordinates may still retain some informational control based upon their networks, two things change. First, those networks are considerably reduced in size. The manager has fewer subordinates, and the manager’s subordinates have fewer subordinates. Even if, as a consequence, the manager’s network becomes flatter (i.e. managers at the same level will interact in networks),\textsuperscript{137} their ability to control information flows to the CEO is considerably diminished. This is because the number of contacts the CEO has into the various networks within the corporation corresponds with the number of executives reporting to him. In other words, he bridges more structural holes than he does in the hierarchical corporation. Unless one assumes that the executives will collude to restrict the information flow to the CEO,\textsuperscript{138} the CEO will have


\textsuperscript{136}Rajan & Wulf, supra note 125, at 1. But see Gordon, supra note 123, for an argument that bureaucracies which flatten as the result of significant layoffs in the middle management ranks relatively quickly re-establish their hierarchies through rehirings.

\textsuperscript{137}This situation, by the way, has been found by Burt to be most advantageous to managers in their ability to take advantage of structural holes. Burt, supra note 62.

\textsuperscript{138}This is possible, but unlikely, given the normal ambitions of second-level executives eventually to ascend to a CEO position either in their own or another
materially greater informational opportunities. Moreover, when applying this theory to the ideal model of the horizontal corporation, each of these senior executives performs different functions, and has access to (and reports to the CEO) different types of information. As the number of CEO subordinates multiplies, the diffusion of job descriptions is likely to increase. The result is that the CEO is increasingly able to receive, convey, and control information—not to mention manipulate his subordinates against one another. Thus we should expect that the CEO in the horizontal structure is, counterintuitively, more powerful than the CEO in a bureaucratic organization with either an inside or an independent board. Hypothesis three is illustrated in Figure Eleven.

Burt also notes that this kind of structure, in which managers have few or no peers, increases the value of social capital to them in terms of their chances of promotion. This is true for bureaucratic organizations, but in network organizations (or horizontal corporations) it is also valuable to managers with multiple peers. Burt, supra note 62. This observation is not especially significant for my argument because in the stylized firm I've created each CEO subordinate can only rise within the firm by becoming CEO. But there is a highly important implication of this analysis which is most relevant to hypothesis four. The structure puts them in the position of monopolist over the kinds of work they do. The CEO is, in this analysis, the ultimate monopolist, since nobody in the corporation does the work of the CEO, and it is the ability to use this monopoly position (which would be constrained by inside directors) that gives him the greatest power.
Figure 11: Corporations that are structured horizontally and have inside boards will have moderately strong CEOs

A qualification is in order. If bureaucratic organizations are built for control, horizontal organizations are built for freedom. Information flows are significantly less constrained in a horizontal organization, as are relationships defined by position. Freedom is the opportunity provided by structural holes. The implication is that in the horizontal organization, managers are freer to establish their own networks across the corporation, and thus skillful network entrepreneurs can achieve a level of informational control that is more widespread than is likely in the more constrained bureaucracy. In such a case, the network entrepreneur could be in a better position to control broader information flows to the CEO and his immediate subordinates, and thus the CEO would be weakened. While I believe that the CEO’s greater access into the network structures of the horizontal corporation is likely to give him greater power, this is not certain, and the hypothesis remains tentative.

The key variable now becomes the board. If the board has insiders as well as independent directors, the CEO’s position is likely to be weakened because, as in hypothesis one, the presence of insiders on the board gives network contacts to the independents into the corporate networks by virtue of their contacts with the insiders. In light of the CEO’s strong structural position in the horizontal corporation (with his own multiple nonredundant contacts into the corporation’s networks), the independent directors’ ability to circumvent the CEO will be reduced. It will not be entirely eliminated, because opportunistic senior managers hypothetically working in different aspects of the business might take advantage of their board positions to manipulate the independents. The risk of
detection by the CEO, however, is increased: the senior managers' corporate responsibilities are more distinct, therefore the CEO's ability to discern the source of information that bypasses him will be improved. Thus, in hypothesis three, we can expect the CEO in the horizontal corporation with an inside board to be stronger than the CEO in a bureaucratic corporation, but not as strong as he might be were the board composed solely of independent directors as in hypothesis four.

My conclusions are tentative with respect to the four restated hypotheses that address both board structure and bureaucratic structure, and require further study. As I noted earlier, corporations are highly varied in their internal organizations despite the existence of distinct and identifiable patterns. It may be that in a particular bureaucracy, CEO control is so tight that defection in the form of information blockage or distortion at lower levels is unlikely. But if true, it is most likely to be true in a corporation that does business in a single location. Burt found that in a corporation with multiple and geographically dispersed business operations, managers whose jobs put them at the “frontiers,” i.e., required them to be in contact with managers in other locations, had greater structural hole opportunities than managers whose responsibilities were confined to a single location. Thus the tightly controlled bureaucracy of the Taylorist model may be increasingly rare.

At the same time, as I noted earlier, horizontal corporations are designed to permit freer and less structured information flows. In terms of network theory, the networks are not as constrained by bureaucratic structure. As a result, it may well be that greater structural hole opportunities exist in these corporations because the diffusion of job responsibility implied by the horizontal structure creates internal frontiers which enable even managers working in the same location to identify more numerous structural holes. Since structural holes are opportunities to receive, control, and manipulate information, it may be the case that in the horizontal corporation the CEO will be less well informed than in the bureaucratic corporation. On the other hand, the broader and shallower management structure of the horizontal corporation suggests that the likelihood of detecting informational blockage and manipulation is greater than in the hierarchical structure, with potentially dire consequences for those acting opportunistically to the disadvantage of the corporation or the CEO.
Because of the uncertain affect of the internal corporate structure on information flows, these hypotheses are tentative. For the time being, I suggest that bureaucratic structure be seen as an independent variable that magnifies the effects of the presence of an inside or independent board. While this variable is not likely to have as substantial an explanatory impact as do the simpler hypotheses stated at the outset, it merits additional consideration and empirical research.

VII. MOVING BEYOND THE BOARD IN CORPORATE LAW REFORM

Almost since the advent of the modern corporation, corporate scholars have focused on the board of directors as the body most capable of protecting the corporation and its stockholders from managerial depredation. Beginning in the 1980s, the reform effort increasingly focused on the importance of independent directors as the solution to this monitoring problem. Recent evidence, however, suggests that the resulting changes in corporate boards have not been entirely beneficial. For example, the CEO of a corporation with an independent board has the ability to distort his compensation and the corporation’s broader compensation structure in ways that are disadvantageous to the corporation. Structural hole theory provides an explanation for these phenomena, as well as for the recent corporate scandals that occurred beneath the boards’ radar screens. Independent boards magnify CEO power. Since independent boards have been the consensus solution after almost a century of attempts at board reform, this traditional focus on the board of directors is misplaced. Instead, governance scholars must begin to study the internal workings of the corporation, and how these networks link to the CEO, senior executives, and the board, if they are to devise effective solutions to problems of corporate governance.

One specific application relates directly to the current state of the law on board monitoring. In the widely discussed opinion in In re Caremark International, the Chancellor represents the boards’ obligation as the duty to create information flows in both directions (from the board as to corporate policy and from the bureaucracy as to compliance

---

140 698 A.2d 959 (Del. Ch. 1996).
My argument suggests that in order to fulfill this duty, it cannot be enough for the board simply to establish lines of reporting. Structural holes, which provide opportunities to block or distort information throughout the hierarchy up to and including the CEO, must be identified in order for the board effectively to monitor the flow and accuracy of corporate information. Monitoring information flows is only effective when one accurately identifies the network nodes that are the corporate actors who bridge structural holes and thus control the flow of information. In the absence of attention to internal structure, monitoring will have limited utility: the more limited the more independent the board.

The point of this paper is not, at this early stage, to prescribe specific corporate reforms. While the theory is powerful, reform must rest upon empirical evidence. Instead, I suggest that certain paths of research must be investigated in order to achieve effective corporate reform. The debate over board function has been one of manager versus monitor, with the latter clearly victorious. Although I do not suggest that the board as an institution is irrelevant, my argument raises the question, starkly, of the board’s utility. While others have proposed alternative conceptions of the board, my theory indicates the difficulty in evaluating, much less prescribing, the board’s function and composition without an inquiry into the relationship between the board and the internal structure of the corporation. I have presented the theory as a general one, as it would be useful first to study it that way, although individual corporate evaluation may ultimately be necessary. It will likely be found that one size doesn’t fit all.

At least as important, my hypotheses suggest that we take a hard look at the role of the CEO and senior managers. In recent years in particular, CEOs of major corporations have developed something of a “rock star” quality, followed by the popular press, lionized in success while derided in defeat. No doubt this has contributed to CEO self-image, if not actual power. But board reform without attention to the role and power of the CEO and the corporation’s internal managerial

---

141 I have elsewhere argued that the opinion is disingenuous, see Mitchell, supra note 38, but it is the articulated state of the law.

142 See, e.g., Dallas, supra note 14 (suggesting a dual board); Fisch, supra note 14 (suggesting individually tailored boards).
structure is a fruitless endeavor. Corporate law scholars have, with rare exceptions, ignored this critical aspect of corporate governance. Attention should turn to this subject.
Corporate Misbehavior and Collective Values

Margaret Gilbert

“Our merely social intolerance kills no one, roots out no opinions, but induces men to disguise them or to abstain from any active effort at their diffusion.”

I. INTRODUCTION

In recent years there have been many scandals in which highly paid corporate executives have apparently acted in morally unacceptable ways. Why is this? Is it a matter of a few “bad apples” or is there some other, or some additional, explanation? In order to answer such questions, one needs to know what the possibilities are: what, generally, goes on in the life of a corporation and its members? What factors might influence the behavior of a given executive or other member of the corporation?

It is plausible to suggest that collective value judgments or, for short, collective values, are important components of corporate life and must be considered in examining immoral behavior within corporations. This article carefully articulates a particular interpretation of this idea.

It is argued, first, that considerations of our everyday ascriptions of beliefs and value judgments to groups of people point away from an account of collective values that might initially be proposed. A different account is then offered. The

© 2005 Margaret Gilbert. All Rights Reserved.

† Professor of Philosophy, University of Connecticut, Storrs. I thank Professor Lawrence Solan for inviting me to the stimulating interdisciplinary conference on corporate misbehavior held at Brooklyn Law School on November 12, 2004, and to contribute a paper on relevant ideas of my own to the associated volume of the Brooklyn Law Review.

likely influence of collective values according to this account, and phenomena akin to them, is discussed. It is argued that they are indeed likely to be influential. To echo Mill in the above quotation, collective values are apt to induce people to abstain from any active effort to counter them. Some practical consequences for corporations and those concerned with them are then sketched.

II. GROUP BELIEF STATEMENTS

Among the possible sources of malfeasance that one finds informally proposed are a climate of opinion or corporate culture in which “anything goes” as long as the corporate bottom line—the maximization of profit—is served. Most people have a rough idea of what is at issue when a climate of opinion or a culture is mentioned. It is harder to say exactly what phenomena are in question. If we are to make more of this explanation of behavior, we need to go further.

If pressed, many would most likely propose the following: the judgment that “anything goes” is part of the climate of opinion or corporate culture in a given corporation if and only if most of the people working for that corporation personally endorse that judgment. This accords with philosopher Anthony Quinton’s statement that “[i]n some cases, which may be called summative, statements about social objects are equivalent to statements otherwise the same that refer explicitly, if at some level of generality, to individual people. To say that the French middle class is thrifty is to say that most French middle class people are.”

It can be argued, however, that everyday statements about the beliefs and attitudes of social objects, to use Quinton’s phrase, are not always summative in his sense. That is, they are not always equivalent to statements “otherwise the same” that refer to individual people generally described. Consider some further statements about the beliefs and attitudes of social objects: “The union believes management is being unreasonable,” “In the opinion of the court, this law is

---

unconstitutional,” “Corporations value nothing but profit,”
“Our family favored Bush,” “We”—said by an unmarried
couple—“want to get married.” If it is sometimes appropriate to
take these as summative statements, it is surely not always so.

After all, in many cases, a formal voting procedure
determines what the social object or group believes. What
counts as far as any individual group member goes, then, is his
or her vote—not what he or she personally believes. Evidently,
one’s voting in favor of the proposition that, say, Arthur was
the best candidate for the position does not logically entail that
one personally takes this proposition to be true.

Less formal processes are also taken to determine what
a group believes. These can be argued to be analogous to the
case of voting insofar as public expressions as opposed to
private thoughts are what matter. Thus, after some discussion,
a literary discussion group may reach a point of quiescence
after which no one would dispute that “We thought Plath’s
poem a very strong one.”

It appears, then, that there is an important sense in
which for a group to believe something (or value it in a certain
way or want it, and the like) it is not necessary that all or most
of the members believe it. Were group belief statements always
summative statements in Quinton’s sense this would not be so.

Given this sense of group belief (and so on), is it
sufficient for a group to believe something that all or most of
the members do? Apparently not. Consider a court. A certain
matter may not yet have come before it. It would then seem
right to say that, as yet, the court has no opinion on the matter.
The individual justices may, at the same time, have definite
personal opinions about it. What they think, however, does not
determine what the court now thinks. This is true even if the
court is in session and the matter in question is before it. All
may be of the same personal opinion but before a vote is taken,
the court itself has none.

The same goes for less formal groups as well. Asked
what her discussion group thought of a particular poem,
someone might respond, with some emphasis, “The group has
no opinion: we’ve not discussed it!” At the same time, she would
acknowledge it to be perfectly possible that all of the members
have the same personal view of the poem.

In sum, everyday group belief (value, goal, and other)
statements are not always interpretable as summative
statements. In other words, when we talk of a group’s belief, we
may well be talking about something other than what all or
most of the members believe. What, then, are we talking about? In what follows I refer to the phenomenon in question simply as group belief.

III. OBSERVATION: THE STANDING TO REBUKE

One clue as to what defines a group belief has to do with informal rebukes or reproof. A rebuke is a form—albeit a mild form—of punishment.\(^4\) I take it that although one can cause pain to someone without any special standing, one cannot punish them without such standing.\(^4\) Similarly, one can speak harshly to a person without any special authority, but one cannot rebuke them without such authority. This is supported by the fact that people sometimes respond to purported rebukes in such terms as, “What’s that to you?” or “What business is that of yours?”

In contexts where group members believe their group to have a particular belief, they understand that they have the standing to rebuke any member who bluntly expresses the opposite belief. Opposed to a “blunt” expression is one that makes it clear that the speaker is “speaking personally.” That the other parties have the standing to rebuke the member in question appears to be a function of the collective belief itself.\(^5\)

This clue is, perhaps, little more than a provocation or a question. What is it about group belief that gives the group members the standing to rebuke each other for blunt expressions of a contrary opinion? An adequate account of group belief should give a plausible answer.

IV. THE PLURAL SUBJECT ACCOUNT OF A GROUP’S BELIEF

The foregoing discussion suggests three criteria of adequacy for an account of a group’s belief. Such an account should explain how the existence of such a belief gives the group members the standing to rebuke each other for bluntly expressing a view contrary to that belief. It should neither logically entail that all or most of the parties personally have the belief in question nor should it suppose that if all or most of


\(^5\) Perhaps it should be said that “one cannot punish someone in the strict sense of ‘punish’ without a special standing,” since at this point the term “punish” would appear to have both a broader and a narrower (the so-called “strict”) meaning.

\(^6\) For extended discussion on this point see Gilbert, Living Together, supra note 3, at 200-03.
the members have a given belief then the group believes it. For the former is not necessary for group belief and the latter is not sufficient for group belief.

The following account of group belief—whose terms will be explained—meets all of these criteria.\(^7\) Here the letter \(p\) stands for any particular proposition.

A group, \(G\), believes that \(p\) if and only if its members are jointly committed to believe as a body that \(p\).

Two aspects of this account need to be explained: what is a joint commitment, and what is it to be jointly committed to believe something as a body? These will now be discussed in turn.

What is joint commitment, as understood here? The answer to that question can usefully be broken down into two parts, one concerning the commitment side of things, the other relating to joint-ness.

The relevant general concept of commitment is illustrated, for example, in the following judgment: if Sandra decides to read the newspaper this evening, then she is committed to doing so. In the case of a personal decision such as Sandra’s, the commitment is personal. That is, the one whose commitment it is creates it unilaterally and is in a position unilaterally to do away with it. I take it that, once committed, Sandra has reason to read the newspaper and will continue to do so unless and until she rescinds her decision. The concept of a joint commitment is the concept of a commitment of two or more people. It is not a conjunction of the personal commitment of one party with the personal commitment(s) of the other(s).\(^8\)

Joint commitments can be created in various ways. One such way is to informally agree that one or more of the parties to the agreement is to act in a certain way. Less explicit means are also possible. Absent special background understandings, what is needed generally speaking is an expression of readiness by all parties to be jointly committed in the relevant way, in

---

\(^7\) This account was introduced in Margaret Gilbert, *Modeling Collective Belief*, 84 SYNTHESE 185, 185-204, reprinted in GILBERT, LIVING TOGETHER, supra note 3, at 195-214, and has been further elaborated in subsequent publications.

conditions of “common knowledge.” Common knowledge is intended here in roughly the sense introduced by the philosopher David Lewis. Rather than going into the details of that here, suffice it to say that the expressions in question must be “out in the open” as far as the parties are concerned. In parallel with the conditions of its creation, the—more or less explicit—concurrence of all is required for the dissolution of a given joint commitment.

There may be special background understandings that impact upon how joint commitments are formed. For example, people may jointly commit to believe as a body whatever a certain person says about the group. Thus the members of the board of a corporation may be jointly committed to believe as a body whatever the chairman of the board says about the corporation. As a consequence of this background commitment, if the chairman says that the corporation is doing fine, then the other members of the board are jointly committed to believe as a body that the corporation is doing fine. Given this background, the chairman need not heed to what anyone else does or thinks: the board will believe whatever he says about the corporation.

A given joint commitment can always be described in a sentence of the following form: the parties are jointly committed to $x$ as a body. Acceptable substitutions for $x$ are psychological verbs such as “believe,” “value,” “intend,” and so on. What is it to be jointly committed to $x$ as a body? This can be spelled out further as follows: it is to be jointly committed to constitute, as far as is possible, a single body that $x$s. The guiding idea of a single body that $x$s includes nothing about the intrinsic nature of the single body in question. In particular, it does not imply that it is in some way made up of two or more distinct bodies who are its members. Thus there is no circularity in the proposed account of group belief. It does not say that a group believes that $p$ if and only if its members are jointly committed to constitute, as far as possible, a group that believes that $p$. The point is, rather, that if and only if they are jointly committed to constitute, as far as is possible, a single body that believes that $p$, they will constitute a group that believes it.

---

9 David K. Lewis, Convention: A Philosophical Study 52-60 (1969). See also Gilbert, Social Facts, supra note 3, at 188.
V. JOINT COMMITMENT AND THE STANDING TO REBUKE

It can be argued that by virtue of their participation in a joint commitment, the parties gain a special standing in relation to one another's actions. To put it briefly, each can call on the other in the name of the joint commitment. If one violates it, for instance, he has not violated a commitment that is his alone, but a commitment to which others can lay claim. Each can say: “You violated our commitment.”

Important aspects of the special standing of the parties in relation to one another are as follows. The parties to a joint commitment are answerable to one another with respect to their conformity. Further, one who violates a joint commitment has offended against all of the parties to the joint commitment, as such. The offense in question can plausibly be characterized in terms of a violation of right. In other words, when I am subject to a joint commitment requiring me to do certain things, all of the parties to the commitment have a right against me to the relevant actions. Correlatively, I am under an obligation to all of them to perform these actions. These obligations and rights are derived directly from the joint commitment. They are not a matter of moral principle.

Once it exists, they exist also, irrespective of the surrounding circumstances.

In consequence of the existence of these rights and obligations, failing special background circumstances, those who are party to a joint commitment have the standing to demand that others conform to it, if non-conformity is threatened. They also have the standing to rebuke one another for defaults that have taken place.

\[10\] For amplification see MARGARET GILBERT, SOCIALITY AND RESPONSIBILITY: NEW ESSAYS IN PLURAL SUBJECT THEORY 50-70 (2000) (discussing obligation in relation to joint commitment).

\[11\] I have in mind a general moral principle such as philosopher Thomas Scanlon's Principle of Fidelity. See Thomas M. Scanlon, Promises and Contracts, in THE THEORY OF CONTRACT LAW 86, 95 (Peter Benson ed., 2001) (discussing the Principle as “Principle F”). For a critique of Scanlon's account of promissory obligation, see Margaret Gilbert, Scanlon on Promissory Obligation: The Problem of Promisees' Rights, 83, 94 (Feb. 2004) (arguing that one cannot account for the rights of a promise by reference to a moral principle such as Scanlon's).

\[12\] See GILBERT, LIVING TOGETHER, supra note 3, at 305 (stating that coercive circumstances do not affect the obligating quality of joint commitments). See also MARGARET GILBERT, A THEORY OF POLITICAL OBLIGATION (forthcoming) (stating that joint commitments with immoral content obligate in the usual way; it may well be that one ought not to fulfill such an obligation, all things considered).
The present author uses the phrase “plural subject” as a technical term to refer to those who are jointly committed to \(x\) as a body, for some \(x\).\(^{13}\) Those who are jointly committed to \(x\) as a body constitute, by her definition, the plural subject of \(x\)-ing. Accordingly, one can label the above account of group belief the plural subject account.

There are some important aspects of that account that have not yet been discussed. A joint commitment to believe as a body that \(p\) does not require each participant personally to believe anything. The requirement at issue is precisely to constitute, as far as is possible, a single body that believes that \(p\). It does not concern any other bodies that may bear some relation, however close, to the body in question.

More positively, the joint commitment will be fulfilled, to some extent at least, if those concerned say that \(p\) in appropriate contexts, with an appropriate degree of confidence, and do not call \(p\) or obvious corollaries into question. Their behavior generally should be expressive of the belief that \(p\), in the appropriate contexts. That does not mean that they must personally have that belief. In other words, this expressive behavior need not connote or be the expression of a personal belief that \(p\).

Certain contextual conditions are likely to be understood. Thus, for example, members of a seminar on rights may form a joint commitment to believe as a body that the notion of a group right is a viable one. This would involve a requirement to express that belief when acting as a member of the seminar. Presumably the parties are not always so acting. If a friend who is not a member of the seminar engages one of them on the topic in the middle of a picnic, it would presumably be appropriate for each to speak in propria persona, without preamble.

Suppose, however, that one is in a context where it is appropriate to act in accordance with a given joint commitment. It is then open to one to use such qualifiers as “Personally speaking,” to preface the expression of a belief contrary to the collective one. This makes it clear that one is indeed now speaking for oneself and not as a member of the relevant group. One makes it clear that one’s utterance is not a violation of the joint commitment in question. This is one way in which the plural subject account of group belief accords with

\(^{13}\) See generally GILBERT, ON SOCIAL FACTS, supra note 3.
the logic of “Our group believes that $p$” and so on, as this is understood in everyday life. It allows for the possibility that a party to a supposed group belief aver without fault—though not necessarily without danger—that he personally does not believe that $p$.

What danger might there be in making such an avowal? Other members may subsequently regard one with suspicion, thinking one more liable to default on the joint commitment, either inadvertently or deliberately. If one does default, they have the standing to rebuke one for doing so. They may also begin to think of one as an “outsider.” To be regarded with suspicion, to be thought of as “not one of us,” to risk inadvertently incurring rebukes, are things that most people would prefer to avoid. It is clear then that group beliefs according to the plural subject account are likely to suppress the development of contrary ideas at both the individual and the collective level.

VI. COLLECTIVE BELIEFS, VALUES, GOALS

The plural subject account of a group’s belief that $p$ may well articulate a central everyday conception. Among other things, it meets the criteria noted earlier: it explains the standing of the parties to rebuke each other for bluntly speaking as if $p$ were false, and it is neither necessary nor sufficient for group belief, on this account, that most members of the group personally believe that $p$. It is necessary and sufficient only that they are jointly committed to believe that $p$ as a body. Whatever its relationship to everyday conceptions, if there are group beliefs in the sense of the plural subject account, they will be an important aspect of the lives of the parties to them.

In what follows group beliefs as these are understood on the plural subject account will be referred to as collective beliefs. Members of a given population—members of the board of a certain corporation, for instance—will be said collectively to believe that $p$, by definition, if they are jointly committed to believe as a body that $p$.

Analogous plural subject accounts can be given collective values, collective goals and intentions, and so on. As to collective values, the relevant account would run along the following lines: “Members of population P collectively value
item $I$ in a certain way if and only if they are jointly committed to value item $I$ in that way as a body.”

Evidently, in relation to the ideas about corporations mentioned at the outset of this discussion, one might interpret the idea of a climate of opinion or corporate culture in such terms. If members of a given corporation were jointly committed to value as a body the maximization of the corporation’s profits above everything else, one could reasonably say that such valuation of the maximization of profit was part of that corporation’s culture. Important parts or accompaniments of this culture would be collective beliefs of a factual nature, such as beliefs about the capabilities and temperament of various corporate executives, and collective goals and plans.

It is plausible to hypothesize that there are collective values and beliefs according to the definitions given. That is, it is plausible to hypothesize that in the world as it is, there are a variety of such phenomena. This hypothesis, like the plural subject account of group belief, is based on observation of the judgments people make in the context of their ascriptions of beliefs to groups and so on. On the assumption that such phenomena abound among those living and working together, in corporations and elsewhere, the concluding sections of this discussion briefly discuss their nature as rational motivators, and some practical implications for those who are concerned about morally unacceptable behavior in the corporate realm.

VII. COLLECTIVE BELIEFS, VALUES, AND GOALS AS RATIONAL MOTIVATORS

Something may be considered a rational motivator if the behavior of a perfectly rational agent would be influenced by it, all else being equal. A perfectly rational agent, for purposes of this idea, is one who always acts as reason dictates he should act. Thus, he does whatever he has reason to do, all else being equal.

---

14 For a different version, still in plural subject terms, see generally Margaret Gilbert, *Shared Values, Social Unity, and Liberty*, 19 PUB. AFFAIRS Q. 25, 33 (2005). For present purposes the precise details of the account are not important.

15 I take it that one may have reason to do something irrespective of one’s own preference ranking of the various possible outcomes of one’s action. The conception of the dictates of rationality at issue here, then, is different from that employed in the mathematical theory of games.
I take it that perfectly rational participants will be motivated by their joint commitments. That is, all else being equal, a perfectly rational agent who is party to a joint commitment will conform to that commitment. If he is jointly committed with other members of the firm to assigning as a body the highest possible value to the maximization of his firm's profits, then all else being equal, he will act in ways expressive of such valuing. Insofar as human beings are at least imperfectly rational agents, their joint commitments are liable to motivate them in a similar way. This does not mean that, all things considered, it would be rational to pursue the corporate bottom line come what may. All else being equal, one is rationally required to conform to a commitment, but all else may not be equal. Morality, many would say, is an external consideration—one not founded in the corporation, at least—that may change the picture.

Of course, if you violate a standing joint commitment for moral reasons, you will be liable to the rebukes of at least your less morally aware colleagues. Your preference to avoid such rebukes, or worse, may lead you not to violate the commitment. Depending on one's colleagues and one's circumstances, it may be little short of heroic for one to violate the commitment. It may then be that even morally speaking it is permissible for you not to take a moral stand. Evidently, much depends on the case.

VIII. PRACTICAL IMPLICATIONS

Collective values and factual beliefs make a difference. If one wishes to understand the pressures on corporate executives and other members of corporations to behave in a manner contrary to central moral principles, one must be aware of the collective values, beliefs, and goals to which they are party.

These may work to stifle the expression of moral concerns. To echo my opening quotation from J. S. Mill, collective beliefs, values and goals are apt to induce people to disguise their contrary opinions—however morally perspicacious—and to abstain from any active effort at their diffusion. Indeed, contrary to the suggestion in the quotation, they may tend to root out opposing opinions themselves. Forced regularly to couch their arguments in terms of profit maximization, or other corporate goals, executives and other members of a corporation may simply lose sight of moral
constraints and values they previously held. At first, executives may cease to understand fully the constraining maxims they previously endorsed, and gradually they may cease to entertain them.

How might a situation of this kind be rectified? If in a corporate context adherence to central moral norms is highly valued collectively, then this will encourage the critique of other collective assumptions. If the members of corporation C collectively accept that certain moral constraints must rein in the pursuit of corporate profit, respecting such constraints will be collectively understood as part of being a good corporate citizen. The same goes, of course, for any other type of citizenship.
A Party That Won’t Spoil

MINOR PARTIES, STATE CONSTITUTIONS AND FUSION VOTING

I.  INTRODUCTION

In the 2000 Presidential race 2,882,955 Americans cast their votes for Ralph Nader, the Green Party candidate.¹ When Republican George W. Bush won a narrow victory, many argued that Ralph Nader caused Democrat Al Gore’s defeat.² Nader was deemed a “spoiler.”³ Democratic Party leaders were angry, and they made that anger public. For instance, according to one elected Democrat, Nader “divorced himself from the very ideals that made him a worthwhile political actor. He sold out his constituency.”⁴ The anger over the

¹ © 2005 Elissa Berger. All Rights Reserved.
³ See, e.g., Sheila R. Cherry, Nader Raids the Democrats, INSIGHT, Dec. 4, 2000, at 24; James Dao, Angry Democrats, Fearing Nader Cost Them the Presidential Race, Threaten to Retaliate, N.Y. TIMES, Nov. 8, 2000, at B3; The Spoiler, PITTSBURGH POST-GAZETTE, Dec. 17, 2000, at A9. For the argument that, contrary to the conventional view, Buchanan had a greater impact on the election than Nader, see David Leonhardt, Was Buchanan the Real Nader?, N.Y. TIMES, Dec. 10, 2000, at 44.
⁴ See, for example, the news articles listed supra, note 2. A “spoiler” is a candidate who has no chance of winning, but whose candidacy deprives another of success. See MERRIAM WEBSTER ONLINE DICTIONARY, at www.m-w.com (last visited May 15, 2005).
⁵ See Peter DeMarco, D.C. Dems Gore Nader for Crashing the Party, N.Y. DAILY NEWS, Jan. 30, 2001, at 2 (quoting Rep. Robert Wexler). See also Michael Powell,
obscure candidates, and they must accept that their candidate has no hope of winning.\footnote{See Steven J. Rosenstone et al., Third Parties in America 3 (2d ed. 1996).}

Nonetheless, people sometimes choose to vote on a minor party ballot line to send a message about their frustration with the two major parties and their candidates.\footnote{Id. at 9.} This is probably why 2.8 million people voted for Nader in 2000.\footnote{See, e.g., James Dao, The 2000 Election: The Green Party, N.Y. Times, Nov. 10, 2000, at A29 (quoting voters as saying “I voted for Nader because he was most aligned with my values” and “I voted my conscience”).} The risk of “spoiling” or “wasting votes,” however, makes it hard for minor parties and independent candidates to consistently secure voters’ support at the ballot box, even if voters remain committed to the party and candidate’s ideology.\footnote{See Rosenstone et al., supra note 11, at 81, 174-75 (explaining modern minor parties fielding their own candidates rarely last more than two election cycles).} This is probably why 2.4 million fewer voters chose to vote for Nader in 2004 than in 2000.\footnote{Television host Bill Maher, who voted for Nader in 2000, explained his switch in 2004 this way: “We all got a little reality slapped into us by George W. Bush . . . I see [voting for Nader] as a bratty thing to have done.” See Tom Shales, Bill Maher: Back for More, WASH. POST, Aug. 2, 2004, at C1. Former Nader supporter Ronnie Dugger changed his position after the 2000 election and described Nader’s second presidential candidacy as a mistake in strategy and harmful to the progressive values he and Nader share. See Ronnie Dugger, Ralph, Don’t Run, NATION MAG., Dec. 2, 2002, at 14.}

If minor parties are allowed to endorse major party candidates, they could be effectively released from their political cage. Imagine a closely contested race between a Democrat (candidate X) and a Republican (candidate Y). A minor party, the Purple Party, endorses candidate X. The ballot has a column for each political party and in each column the name of that party’s nominee is printed. Candidate X, then, is listed twice. (See Appendix for a sample ballot.) Those who prefer candidate X to candidate Y can vote on either the Democratic or the Purple party ballot line for candidate X.

Let’s say the Purple Party platform strongly supports the right of same-sex couples to marry. A voter, who similarly supports gay marriage, hopes candidate X will win the election, but is frustrated with the way candidate X and the Democratic Party avoid the issue of same-sex marriage. Our hypothetical voter would be able to cast her vote for both the candidate she prefers and the political party she feels best represents her
views. This practice—voting for candidates that are endorsed by more than one political party—is known as fusion voting.

In a fusion voting system, a single candidate can be nominated to run on more than one party’s ballot line. All votes for that candidate are added together to determine whether the candidate wins a majority of votes, regardless of the ballot line on which the vote was cast. But each party’s votes are also tallied separately, so the election results reflect each party’s contribution to the candidate’s electoral success. Minor parties are able to influence the outcome of elections even when they do not field a viable candidate of their own. They can endorse a candidate who has a reasonable chance of victory, while also demonstrating that voters support their platform. Likewise, voters are able to express their support for the party’s principles, while avoiding the danger that their vote has been wasted in symbolic protest. As one pro-fusion party has boasted, fusion voting makes one vote count twice—first it sends a message about the issues the voter cares about and then it helps elect a candidate.

16 The broadest definition of fusion voting would include ballot rules that allow more than one party to be listed as endorsing a single candidate, but only list a candidate once. Returning to our previous hypothetical, this would mean that candidate X would be listed under a column labeled Democrat and Purple Party, and candidate Y would be listed under a column labeled Republican. For the sake of this note, I use fusion voting to mean a system where each party would have its own ballot line. For an argument that any other kind of fusion voting scheme is unconstitutional, see Note, Fusion Candidacies, Disaggregation, and Freedom of Association, 109 HARV. L. REV. 1302 (1996).

17 Of course, in a fusion voting system, minor parties could choose to run their own candidate if they felt so motivated. When fusion parties proliferated in America, minor parties often debated whether they should endorse a major party candidate or run an independent candidate. See DISCH, supra note 10, at 53-54. Generally, fielding a minor party candidate was a move of last resort. See ROSENSTONE ET AL., supra note 11, at 79. However, just the threat of not receiving a minor party endorsement kept major parties on their toes. Major party candidates would adopt parts of minor party platforms in order to secure the votes of minor party adherents. See DISCH, supra note 10, at 41.

18 A party’s strong showing at the polls could translate into influence on policy. Returning to our hypothetical candidate after Election Day illustrates this point. Imagine Candidate Y receives 48% of the vote, more votes than Candidate X receives on the Democrat ballot line. However, when all votes are tallied, Candidate X is elected with 52% of the vote—44% from the votes on the Democratic ballot line, and 8% of the votes from the Purple Party ballot line. (See Appendix for hypothetical election results.) Candidate X assumes her elected office knowing that she would not have won without the Purple Party’s support. Since she is constantly thinking about the future of her political career, she looks to the Purple Party’s platform and incorporates it into her policy agenda. For real world examples of this, see infra notes 79-82 and accompanying text.

19 Working Families Party Campaign Literature (on file with the author).
Unfortunately for minor parties, fusion voting is illegal in most states. And in spite of the burden anti-fusion laws place on minor parties’ freedom of association, the U.S. Supreme Court has upheld state bans on fusion voting. But when the Supreme Court closes a door, state constitutions may provide an open window.

This Note argues that state anti-fusion laws violate the rights granted to political parties and voters under state constitutions. Part II of this Note describes the role of minor parties in American politics and briefly sketches the history of fusion voting in America. Part III summarizes the Supreme Court’s approach to the rights of political parties and discusses Twin Cities Area New Party v. Timmons, in which the Supreme Court upheld Minnesota’s fusion ban. Part IV argues that state courts can strike down anti-fusion laws based on state constitutional rights.

II. MINOR PARTIES AND FUSION VOTING

Increasingly, voters identify themselves as “independent” and voter registration information suggests Americans want more than the Democrats and Republicans have to offer. Third parties would provide more variety in the political landscape if allowed to thrive. Minor parties broaden
the debate; they raise issues that the major parties refuse to address. The abolition of slavery and women’s suffrage, for instance, first found homes in the platforms of minor parties. With more choices at the ballot box, voter participation would likely increase. And with more viable parties, resulting competition might make major parties more responsive and accountable to voters.

Contemporary politics in the United States, however, is a game with only two teams: the Democrats and the Republicans. Minor parties watch from the sidelines rather than play on the field. Candidates running solely on minor party ballot lines barely make a blip on election return charts unless they have extreme wealth or celebrity status. Moreover, when non-major party candidates have the opportunity to impact an election, that impact is often considered destructive. On learning of Ralph Nader’s intent to run in 2004, for example, progressive commentators predicted “mind-boggling irrelevance—but with a potential for catastrophic mischief.”

(explaining that when there are only two political parties, their platforms tend to reflect the center of the political spectrum).

See ROSENSTONE ET AL., supra note 11, at 221-23.

See id. at 8; SIFRY, supra note 24, at 8.

See SIFRY, supra note 24, at 53. Cf. Arend Lijiphart, Unequal Participation: Democracy’s Unresolved Dilemma, 91 AM. POL. SCI. REV. 1, 7 (1996) (suggesting multiple parties in a proportional representation system would increase voter turnout because it would give “voters more choices and . . . [eliminate] the problem of wasted votes”).

See Richard L. Hasen, Entrenching the Duopoly: Why the Supreme Court Should Not Allow The States to Protect the Democrats and Republicans From Political Competition, 1997 SUP. CT. REV. 331, 344 (“Without third parties to challenge the positions of the two major parties and their candidates, the major parties are likely to become (some would say, remain) complacent and unresponsive to social pressures and movements.”).

Disch notes Americans would never accept only two options as consumers, but they seem to accept such limited choices as voters. See DISCH, supra note 10, at 7. Hasen similarly suggests the absence of marketplace competition results in poor representation by the party duopoly. See Hasen, supra note 29, at 344. And Sifry quips, “For a country that prides itself as the heartland of free market capitalism, this lack of competition in the political arena is not just perverse. It is positively unhealthy.” See SIFRY, supra note 24, at 7.

Reform Party candidate Jesse Ventura became Minnesota Governor in 1998 because of his celebrity status and Minnesota’s public financing laws. See DISCH, supra note 10, at 1-4; SIFRY, supra note 24, at 42. Similarly, presidential candidate Ross Perot was able to garner 19% of the popular vote because he was independently wealthy. See SIFRY, supra note 24, at 3. On the high cost of even getting an independent candidate on the ballot, let alone garnering votes, see Samuel Issacharoff, Politics as Markets: Partisan Lockups of the Democratic Process, 50 STAN. L. REV. 643, 687 (1998).

Harold Meyerson et al., He’s Back: Nader is running for president again.
Minor parties have not always been in this predicament. They used to play a major role in American elections. Up until the early 1900s, minor parties could endorse major party candidates and the same candidate could appear more than once on the ballot. Voters could vote for a candidate on the ballot line of whichever party appealed to them most. When votes were tallied, minor parties simultaneously helped elect candidates to office and demonstrated the parties’ own popularity at the polls. Votes translated into power over elected officials who wanted to run for reelection with minor parties’ endorsements. Major parties, too, watched closely the votes that minor parties garnered. Key issues of vote-getting minor parties would be absorbed into major party platforms. The ability of minor parties to “fuse” their endorsements with major parties’ endorsements “[guaranteed] that dissenters’ votes could be more than symbolic protest, that their leaders could gain office, and that their demands might be heard.”

The decline of fusion voting began when state governments took charge of elections. In the 19th century, political parties controlled the electoral process. Parties themselves used to be responsible for printing ballots listing their slate of candidates. As voters went to the polls, party activists would pass out party tickets. Casting a vote was as simple as dropping the ticket into the ballot box. The simplicity of this process created opportunities for corruption. For example, without government regulation, voters could be tricked into casting their vote on what they thought was a
Republican ballot, but was actually a listing of Democratic candidates. Moreover, different parties' ballots were different sizes and colors. Onlookers could see by the ballot in voters' hands for whom they were voting. The 1888 presidential campaign seemed particularly crooked. It proved a catalyst for states to adopt the Australian system of secret ballot.

Under the Australian ballot system, state governments started printing ballots. State printed ballots protected the privacy of voters' choices—every ballot looked the same. In addition, the new voting system eliminated the distribution of ballots that looked like the slate of one party but actually listed the candidates of another party.

The new system brought the need for new rules. Procedures were required to decide which candidates' names would be printed on the ballot. As part of the new laws, legislators delivered a near fatal blow to minor parties, enacting anti-fusion rules that prohibited multiple parties from endorsing the same candidate.

While most of the Australian ballot laws were enacted to rid the electoral process of corruption, anti-fusion laws had a less noble motivation. The electoral successes of fusion tickets threatened some lawmakers. Majority Republican legislatures were first to realize they could use the trend of ballot reform to remove a tool that often benefited their rivals. In 1893, South

---

42 See, e.g., Daniel v. Simms, 39 S.E. 690, 694 (W. Va. 1901) (decrying the vulnerability of the old system in which “[a] voter, coming upon the ground and desiring to vote the Democratic ticket, might have one of these fraudulent tickets placed in his hands, and, without examining it closely, deposit it, and thus be defrauded out of his vote as to that particular office in which he felt most deeply interested”).

43 See ROSENSTONE ET AL., supra note 11, at 19-20.

44 See id. at 19-20.

45 See id.

46 See id.


48 See Argersinger, supra note 33, at 290-91.

49 See id.

50 See ROSENSTONE ET AL., supra note 11, at 20.

51 See Argersinger, supra note 33, at 291.

52 See id. at 292 (“[T]he [anti-fusion] law . . . was intended to promote the dissolution of party ties while giving Republicans the residual benefits of them.”). In addition to anti-fusion laws, other mean-spirited laws were included in states' reform packages. It was during this time that legislators instituted poll taxes and literacy tests with the goal of disenfranchising African Americans. See Paul R. Petterson, Partisan Autonomy or State Regulatory Authority? The Court as Mediator, in THE U.S. SUPREME COURT AND THE ELECTORAL PROCESS 113-14 (David K. Reyden ed., 2d ed. 2002).

53 See Brief for the Respondent at 7, Timmons, 520 U.S. 351 (No. 95-1608);
Dakota lawmakers enacted the first anti-fusion law, preventing a candidate from being listed more than once on a ballot. By 1895, Oregon, Wisconsin, Michigan, and Ohio passed analogous laws. By 1899, eight more states had passed anti-fusion laws. All were passed by majority Republican legislatures, whose members wanted to prevent the cooperation between Democrats and minor parties. A Michigan lawmaker forthrightly declared, “We don’t propose to allow the Democrats to make allies of the Populists, Prohibitionists, or any other party, and get up combination tickets against us. We can whip them single-handed, but don’t intend to fight all creation.”

Legislators were rarely so blunt. Most defended anti-fusion laws as good-government reform. But the actual motivation for these laws was not lost on one journalist, who renamed the anti-fusion law “the law providing for the extinction and effacement of all parties but the Democratic and Republican.” Nor was the partisan motivation lost on Democratic or minor party members. A Populist Party member declared that the anti-fusion law “practically disfranchises every citizen who does not happen to be a member of the party in power . . . . They are thus compelled to either lose their vote . . . or else to unite in one organization. It would mean that there could only be two parties at one time.” Without fusion, what once was an effective way to express a voter’s ideology now became a wasted gesture—a throwaway vote. Unsurprisingly, voters stopped voting for minor parties.

ROSENSTONE ET AL., supra note 11, at 48-80; Argersinger, supra note 33, at 289-90. Fusion voting was most common in the Midwest and West where Republicans more often were in control of the state legislatures. See Argersinger, supra note 33, at 289-90. Several eastern states did not immediately pass anti-fusion laws because the major parties were strong enough to prevent support for minor parties without legislating against them. See ROSENSTONE ET AL., supra note 11, at 20 n.5 (noting the following states’ history of fusion: Maryland, Massachusetts, Pennsylvania, Rhode Island, and Connecticut). Except for Connecticut, each of these states has since enacted anti-fusion laws. See COBBLE & SISKIND, supra note 20, at 8.

54 See Argersinger, supra note 33, at 297.
55 See id. at 298-301.
56 See id. at 302 (listing the following states as having enacted anti-fusion laws: California, Indiana, Illinois, Iowa, Nebraska, North Dakota, Pennsylvania, Wisconsin and Wyoming).
57 See id. at 302-03.
58 See Argersinger, supra note 33, at 296 (quoting DETROIT FREE PRESS, Feb. 1, Jan. 5, 1893).
59 See id. at 292.
60 See id. at 304.
61 See id. at 302; DISCH, supra note 10, at 52.
62 See Argersinger, supra note 33, at 304 (quoting KALAMAZOO WEEKLY TELEGRAPH, Mar. 20, 1895).

116
and the parties were forced to the sidelines of American politics.\(^63\)

Today, anti-fusion laws exist in all but eight states.\(^64\) In the states that allow fusion voting, seven have laws or party rules that make it difficult, if not impossible to establish a statewide fusion party.\(^65\) For instance, Connecticut laws allow a candidate to be endorsed by more than one official political party in a given election,\(^66\) but to become an official political party, with the ability to endorse candidates in all Connecticut elections, the party must run an independent candidate in a gubernatorial race and win 20% of the vote.\(^67\) New parties, without official status, cannot endorse candidates already nominated by existing parties.\(^68\) Winning 20% of the vote in a statewide election with an independent candidate is difficult. More importantly, it is unlikely a pro-fusion party would want

\(^{63}\) See Brief of Amici Curiae of Twelve University Professors and Center For A New Democracy In Support of Respondent Twin Cities Area New Party, Timmons, 520 U.S. 351 (No. 95-1608); ROSENSTONE ET AL., supra note 11, at 149.

\(^{64}\) The eight states are Connecticut, Delaware, Idaho, Mississippi, New York, South Carolina, South Dakota, and Vermont. See supra note 20 and accompanying text.

\(^{65}\) See COBBLE & SISKIND, supra note 20, at 10-45 (describing major party rules and election laws that may make fusion voting difficult in states that do not have anti-fusion laws).

\(^{66}\) See CONN. GEN. STAT. § 9-453t (2003) (“Nothing in this section shall be construed to prohibit any candidate from appearing on the ballot as the nominee of two or more major or minor parties for the same office.”).

\(^{67}\) Parties retain official status across the entire state when they win 20% of the vote for governor or have registered 20% of voters. See CONN. GEN. STAT. § 9-372/6 (2003) (“Major party’ means (A) a political party or organization whose candidate for Governor at the last-preceding election for Governor received, under the designation of that political party or organization, at least twenty per cent of the whole number of votes cast for all candidates for Governor, or (B) a political party having, at the last-preceding election for Governor, a number of enrolled members on the active registry list equal to at least twenty per cent of the total number of enrolled members of all political parties on the active registry list in the state’); Parties who win 1% of the vote in a given election only have party status in that district, for that office. See CONN. GEN. STAT. § 9-372/6 (“Minor party’ means a political party or organization which is not a major party and whose candidate for the office in question received at the last-preceding regular election for such office, under the designation of that political party or organization, at least one per cent of the whole number of votes cast for all candidates for such office at such election.”).

\(^{68}\) See CONN. GEN. STAT. § 9-453t (2003) (“[T]he nomination of a candidate by a major or minor party under this chapter, for any office shall disqualify such candidate from appearing on the ballot by nominating petition for the same office.”). If a party does not have official status, it must nominate its candidate through the petitioning process, thereby precluding the nomination of a candidate supported by an official political party. See CONN. GEN. STAT. § 9-379 (2003) (“No name of any candidate shall be printed on any official ballot at any election except the name of a candidate nominated by a major or minor party unless a nominating petition for such candidate is approved by the Secretary of the State as provided in sections 9-453a to 9-453p, inclusive.”).
to risk spoiling in its election debut. Therefore, establishing a fusion party through a statewide election in Connecticut is almost impossible.

New York is the only state in the nation where fusion voting has remained a common practice. As a result, minor parties have thrived. Currently, three minor parties have official party status in New York state. In one recent election, minor parties captured more than 20% of the total vote. In New York’s local, statewide and federal elections, minor parties’ vote totals have tipped major party candidates to victory. Rudolph Giuliani, for example, became mayor of New York City only because the votes he won on the Liberal Party ballot line were added to the votes he won on the Republican Party ballot line. George Pataki was able to secure the governorship only by adding the votes cast on the Conservative Party line to the votes cast on the Republican Party line. Similarly, New York’s Electoral College votes have been determined by minor party votes: neither Franklin D. Roosevelt, John F. Kennedy nor Ronald Reagan would have

---

69 See Lee Foster, 3rd Party Planting Roots in State, HARTFORD COURANT, Oct. 6, 2004, at B1 (reporting that one Working Families Party candidate dropped out “to avoid stealing votes” from the Democratic candidate in a closely contested state senate race).

70 However, a new party that has won 1% of the vote for an independent candidate in a non-statewide election is able to “fuse” endorsements with other parties the next time that local office is up for election. Winning 1% of the vote in a local election is more achievable and strategically less problematic than winning 20% of the vote in a statewide election. Activists in Connecticut are running independent candidates in local races, and have achieved official party status in at least 66 districts. See SIFFY, supra note 24, at 297; Gail Ellen Daly, Working Families Happy With Results, CHRONICLE (Willimantic, Conn.) (Nov. 5 2004), available at http://www.ct-workingfamilies.org/WCPleased.html (last visited May 15, 2005).

71 See SIFFY, supra note 24, at 228-89.


74 The minor party vote in three recent statewide elections are as follows: in 2004 U.S. Presidential race, 5%; 2004 U.S. Senate race, 9%; 2002 Gubernatorial race, 22%. Elections results are available from the New York State Board of Elections at www.elections.state.ny.us.

75 See SIFFY, supra note 24, at 228-29.

76 See id.
carried New York by the votes cast on major party lines alone. Each of them needed the votes that were cast on the minor party lines to win New York.

New York’s minor parties do not only influence the outcome of specific elections; their success at the polls also leads to influence over politicians in office. The Conservative Party, for example, pressures Republican legislators to oppose abortion and gay rights with threats of running independent challengers against Republican incumbents. The Working Families Party has also linked its possession of a ballot line to policy gains. For instance, a few months after the Working Families Party helped elect a Democrat to the Suffolk County legislature, the county enacted a living wage law, a legislative priority for the Working Families Party. The Working Families Party believes that members of the county legislature saw the decisive part the minor party played in the election and thought of their own upcoming reelections when passing the living wage bill. Highlighting the importance of fusion voting in this legislative victory, Daniel Cantor, executive director of the Working Families Party had this to say:

The ability to clearly demonstrate a minor party's electoral strength via the fusion vote was absolutely essential to winning the living wage in Suffolk. In fact, it's no overstatement to suggest that the 210 votes we got on our line that proved the "margin of victory" in one legislative race resulted in 4,000 low-wage workers getting an increase in salary of nearly $2,000 per year. That's the power of fusion.

77 See id. at 229.
78 See DANIEL A. MAZMANIAN, THIRD PARTIES IN PRESIDENTIAL ELECTIONS 130-32 (1974) (describing the impact fusion parties have had on the direction of New York policy).
81 See WORKING FAMILIES PARTY, Fusion Voting—Our (Not So) Secret Weapon, at http://www.workingfamiliesparty.org/fusion.html (last visited Apr. 14, 2005) (“Every member of the Republican controlled County Legislature noticed the WFP’s role in the Lindsay victory. They were soon up for reelection and realized the importance of appealing to our voters. So, they decided to pass the [living wage] bill.”); See also Michael Tomasky, Inside Agitators, N.Y. MAG., Mar. 4, 2002 (“The 210 votes William Lindsay got on the WFP line provided his margin of victory. For a small party, that means leverage, which the WFP converted into the passage of living-wage legislation in Suffolk.”).
82 E-mail from Daniel Cantor, Executive Director, Working Families Party, to author (Dec. 15, 2004) (on file with author).
Inspired by the successes of minor parties in New York, a few pragmatic idealists began thinking about how to export the New York model to other states in the late 1980s. Joel Rogers, a political science professor at the University of Wisconsin, and Daniel Cantor, then a political organizer in New York, believed that American democracy would be improved if minor parties had a stronger voice in politics. They recognized, however, the political irrelevancy of modern minor parties. The solution was simple: fusion. The solution to state anti-fusion laws was also simple: sue.

Joel Rogers and Daniel Cantor spent the next few years building the “New Party,” a progressive political party that they hoped would win a ballot line by challenging the constitutionality of anti-fusion laws. The New Party’s strategy was to build a grassroots, membership base while planning litigation. The party’s motto was fitting: “Start Small. Think Big.” They were confident that their day in court would result in victory. They were wrong.

III. THE FEDERAL CONSTITUTION AND ANTI-FUSION LAWS

The New Party sought to challenge anti-fusion laws on First Amendment grounds. The First Amendment of the federal Constitution can be construed to protect fusion voting in two ways. First, voting might be seen as expression of political views, and thus protected by the right to freedom of expression. Fusion voting is a means by which voters and parties can critique the major parties’ position on issues and

---

83 See SIFRY, supra note 24, at 228-31.
84 See id. at 229-30. According to a 1999 memorandum, Cantor and Rogers believed that the major parties offered progressives “a ‘devil’s bargain’ . . . in which support is generally exchanged for frustration” but “that the history of third party alternatives seems even more grim.” Creating a new political party with the power to endorse major party candidates would “give voice to [progressives]’ political aspiration in ways that matter in conventional electoral arenas.” See Memorandum from Dan Cantor & Joel Rogers, Party Time, 7, 8, 12 (May 1990) (on file with author).
85 See Cantor & Rogers, Party Time, supra note 84, at 10-11.
86 See Memorandum from Dan Cantor & Joel Rogers, Sue!, 1-2 (May 1990) (on file with author).
87 See SIFRY, supra note 24, at 231-32.
88 See id. at 230-31; Cantor & Rogers, Party Time, supra note 84.
90 They were at least part wrong. They had some initial success, but the Supreme Court ruled against their challenge to anti-fusion laws in Timmons. See infra notes 123-35 and accompanying text.
anti-fusion laws limit this expression. Second, activities of political parties are activities of individuals associating with each other for a common purpose, and as such might be protected by the right to freedom of association to achieve expressive goals. Preventing a minor party from endorsing a major party candidate could be seen as interfering with the party’s core functions. This section will summarize the federal jurisprudence on these two claims, and then discuss Timmons v. Twin Cities Area New Party, in which the U.S. Supreme Court ruled against the New Party’s constitutional challenge to anti-fusion laws.

Checking a box, pulling a lever, punching out a chad, or touching a screen in the ballot booth is a statement of belief as well as a declaration of preference. The U.S. Constitution protects the right to participate in elections as part of the Equal Protection clause of the Fourteenth Amendment, but voting may also be understood to be protected by the right to freedom of expression as guaranteed in the First Amendment. An election marks the temporary end of a political debate and casting a vote is the “official expression of [a voter’s] judgment on issues of public policy.” Denying the ability of multiple political parties to endorse a single candidate denies voters the opportunity of using the ballot to communicate their opinions effectively on their government’s course of action.

Although this argument may be philosophically compelling, the Supreme Court rejected the concept of ballot based expression in Burdick v. Takushi. In that case, the Court considered whether a state could prohibit voters from writing in names of preferred candidates who did not appear on the printed ballot. Several lower federal courts had previously

---

92 See generally Adam Winkler, Note, Expressive Voting, 68 N.Y.U. L. REV. 330 (1993) (arguing that the right to vote should be protected because of the expressive function voting serves).
95 Id.
97 Id.
concluded that write-in votes were a form of political expression and therefore protected by the First Amendment. In *Burdick*, the Supreme Court dismissed this view, declaring that “the function of the election process is ‘to winnow out and finally reject all but the chosen candidates,’ not to provide a means of giving vent to ‘short-range political goals, pique, or personal quarrel[s].’” The Court concluded that treating voting as an act of expression would “undermine the ability of States to operate elections fairly and efficiently.” Given Supreme Court precedent, the New Party rested its argument against anti-fusion laws on parties’ freedom of association rights, rather than on individuals’ expressive rights.

The Supreme Court first formally announced that the constitution protected associational rights in 1958. In *NAACP v. Alabama ex rel. Patterson*, a unanimous Court held that Alabama could not require the NAACP to provide its membership list to the state Attorney General because to do so would offend the NAACP’s right of association. The Court based the right of association in the right of expression, declaring that protecting effective advocacy requires protecting the right of individuals to act collectively. In *Roberts v. United States Jaycees*, the Court further explained that the right to associate was an extension of other First Amendment freedoms. The Court said the “freedom to speak, to worship, and to petition the government for the redress of grievances could not be vigorously protected from interference by the State unless a correlative freedom to engage in group effort toward those ends were not also guaranteed.” Associations formed

---

99 *Burdick*, 504 U.S. at 438 (quoting Storer v. Brown, 415 U.S. 724, 735 (1972)).
100 *Id.*
101 The New Party did argue that the ballot serves an expressive function, but it made the argument from the perspective of a party, not an individual voter. See Brief for the Respondent at 25, *Timmons*, 520 U.S. 351 (No. 95-1608) (“[T]he fusion ban interferes with the message sent to voters by the party, in the voting booth, that it has nominated a particular candidate, and it does so despite the fact that the State otherwise uses its ballot system for precisely this purpose.”).
103 See id. at 462-63.
104 See id. at 460-61.
106 *Id.* at 622.
with the purpose of engaging in activity protected by the First Amendment have been termed "expressive associations."\textsuperscript{107}

The rights of political parties as expressive associations can run headlong into state regulations of elections.\textsuperscript{108} The U.S. Constitution charges state governments with regulating federal elections,\textsuperscript{109} and the Court has implied that states have a duty to ensure all elections are fair and honest.\textsuperscript{110} Therefore, when faced with a law that infringes on the rights of political parties, the Court will balance the interest of the state in regulating elections against the burden on the party’s rights of association.\textsuperscript{111} The “rigorousness of [the Court’s] inquiry into the propriety of a state election law depends upon the extent to which a challenged regulation burdens First and Fourteenth Amendment rights.”\textsuperscript{112} Only regulations that severely burden those rights must be narrowly tailored to further a compelling state interest. A state can justify less serious infringements on associational freedom by showing a regulation furthered important interests.\textsuperscript{113}

Prior to Timmons, the Supreme Court had applied this balancing test and struck down several state laws involving major parties’ associational rights. The Court had held that states cannot require parties to use a closed primary system\textsuperscript{114}


\textsuperscript{108}For a discussion of how political parties fit within the expressive association framework, see id. The First Amendment applies to state action by way of the Due Process clause of the Fourteenth Amendment. The first case to recognize that First Amendment rights are incorporated into the Due Process clause of the Fourteenth Amendment was Gitlow v. New York, 268 U.S. 652 (1925).

\textsuperscript{109}U.S. CONST. art. I, § 2, cl. 1; U.S. CONST. art. I, § 4, cl. 1.

\textsuperscript{110}See Storer v. Brown, 415 U.S. 724, 730 (1974) (“[T]here must be a substantial regulation of elections if they are to be fair and honest and if some sort of order, rather than chaos, is to accompany the democratic processes.”); Bullock v. Carter, 405 U.S. 134, 145 (1972) (“[A] State has an interest, if not a duty, to protect the integrity of its political processes from frivolous or fraudulent candidacies.”).

\textsuperscript{111}The balancing test was articulated in Anderson v. Celebrezze: [The Court] must first consider the character and magnitude of the asserted injury to the rights protected by the First and Fourteenth Amendments that the plaintiff seeks to vindicate. It then must identify and evaluate the precise interests put forward by the State as justifications for the burden imposed by its rule. In passing judgment, the Court must not only determine the legitimacy and strength of each of those interests, it also must consider the extent to which those interests make it necessary to burden the plaintiff’s rights. Only after weighing all these factors is the reviewing court in a position to decide whether the challenged provision is unconstitutional. Anderson v. Celebrezze, 460 U.S. 780, 789 (1983).

\textsuperscript{112}Burdick, 504 U.S. at 434.

\textsuperscript{113}Id.

A PARTY THAT WON’T SPOIL

prohibit parties from endorsing candidates in primary elections or compel parties to accept state delegates to a national convention who were not selected according to party rules. In these decisions, the Court established that the right of association meant “not only that an individual voter has the right to associate with the political party of her choice, but also that a political party has a right to identify the people who constitute the association, and to select a standard bearer who best represents the party’s ideologies and preferences.”

The cases decided in the decades before Timmons also suggested that limiting ballot access for independent candidates might be especially hard for states to justify. One of the earliest cases about a minor party candidate proved to contain the strongest language. In Williams v. Rhodes, the Court considered an Ohio law that required a party to gather signatures from 10% of Ohio voters in order to secure a space on the ballot. The Court struck the law because “[n]ew parties struggling for their place must have the time and opportunity to organize in order to meet reasonable requirements for ballot position, just as the old parties have had in the past.” The Court dismissed the state’s argument that this law was justified out of protection for the two-party system:

"[T]he Ohio system does not merely favor a “two-party system”; it favors two particular parties—the Republicans and the Democrats—and in effect tends to give them a complete monopoly. There is, of course, no reason why two parties should retain a permanent monopoly on the right to have people vote for or against them. Competition in ideas and governmental policies is at the core of our electoral process and of the First Amendment freedoms."

New Party members’ analysis of this precedent left them feeling optimistic. Insofar as anti-fusion laws prevented
parties from endorsing the candidate of the party’s choosing simply because another party had already nominated that candidate, they seemed vulnerable to First Amendment challenges. Given the particular burden anti-fusion laws placed on minor parties, the laws would seem especially difficult for a state to justify. In the words of Professor Theodore Lowi, the case “look[ed] like a constitutional no-brainer.”

The first federal court challenge to anti-fusion laws took place in the Western District of Wisconsin. There, the district court upheld Wisconsin’s ban on fusion voting, and the Seventh Circuit affirmed. Several years later, a district court in Minnesota also upheld an anti-fusion voting law, but this decision was reversed by the Eighth Circuit. The Supreme Court then reversed the Eighth Circuit in Timmons v. Twin Cities Area New Party.

Timmons considered the right of the New Party to nominate a candidate for Minnesota State Representative previously nominated by the Democratic-Farmer-Labor Party. The candidate, Andy Dawkins, wanted to run with the endorsements of both the Democratic-Farmer-Labor Party and the New Party. The Democratic-Farmer-Labor Party raised no objection to the New Party’s endorsement. When the New Party attempted to file the petition to nominate Dawkins, county officials refused to accept the nomination because Minnesota’s anti-fusion laws prevent a candidate from being twice nominated.

---

122 Theodore J. Lowi, Editorial, Supreme Court Should Ban Fusion Tickets, PLAIN DEALER, Dec. 31, 1996, at 9B.
123 Swamp v. Kennedy, 950 F.2d 383 (7th Cir. 1991).
124 See id.
125 See Twin Cities Area New Party v. McKenna, 73 F.3d 196, 200 (8th Cir. 1996).
126 Timmons, 520 U.S. at 370.
127 See id. at 354. In Minnesota the two major parties are the Democratic-Farmer-Labor Party and the Republican Party. Id. at 354 n.2.
128 Id. at 354.
129 Id.
131 Minnesota statute provides:
(1) Major party candidates. No individual shall be named on any ballot as the candidate of more than one major political party. No individual who has been certified by a canvassing board as the nominee of any major political party shall be named on any ballot as the candidate of any other major political party at the next ensuing general election.
(2) Candidates seeking nomination by primary. No individual who seeks nomination for any partisan or nonpartisan office at a primary shall be nominated for the same office by nominating petition, except as otherwise
The New Party filed a complaint in district court alleging a violation of the party's associational rights as guaranteed by the First and Fourteenth Amendments. The court granted summary judgment in favor of Minnesota, rejecting the minor party's claim that the state's anti-fusion law was unconstitutional. The Court of Appeals reversed that decision, finding that the fusion ban created a severe burden on minor parties' associational rights and that the state could have enacted a more narrowly tailored law to achieve its goals. The Supreme Court granted certiorari and, in a 6-3 decision, reversed the Court of Appeals.

The New Party argued that the anti-fusion law severely burdened its associational rights because it prohibited the party from nominating its preferred candidate. It claimed that "nothing is more fundamental to a party than the choice of candidates to represent it in electoral competition. Nothing is more important to a party's ability to mobilize its supporters around candidates than its ability to identify those candidates, on the ballot, as its own."

The Supreme Court's majority opinion, authored by Chief Justice Rehnquist, recognized that political parties are guaranteed associational rights under the First Amendment. Those rights, however, could be limited by state laws reasonably regulating parties, elections and ballots. To determine if the New Party's rights were violated, the Court engaged in a balancing test, "weigh[ing] the 'character and magnitude' of the burden the State's rule imposes on those rights against the interests the State contends justify that burden, and consider[ing] the extent to which the State's concerns make the burden necessary."

provided for partisan offices in section 204D.10, subdivision 2, and for nonpartisan offices in section 204B.13, subdivision 4.


See Twin Cities Area New Party, 863 F. Supp. at 988. The First Amendment applies to state action through the Fourteenth Amendment Due Process clause. See, e.g., NAACP v. Alabama ex rel. Patterson, 357 U.S. 449, 460 (1958). ("It is beyond debate that freedom to engage in association for the advancement of beliefs and ideas is an inseparable aspect of the 'liberty' assured by the Due Process Clause of the Fourteenth Amendment, which embraces freedom of speech.").

See id. at 994.

See Twin Cities Area New Party, 73 F.3d at 200.

Timmons, 520 U.S. at 356.

Brief for the Respondent at 12, Timmons, 520 U.S. 351 (No. 95-1608).

Id.

See Timmons, 520 U.S. at 357-58.

See id. at 358.
The Supreme Court conceded that fusion bans interfere with minor parties’ associational rights, but the Court did not find the burden to be severe.\(^{140}\) To be sure, the Minnesota statute would prevent the New Party from having its preferred candidate listed on the ballot, but it would not prohibit the party from campaigning and supporting a candidate.\(^{141}\) Even if the ballot restriction limited the party’s ability to send a message to voters and to its preferred candidate, the Court was not convinced that a party had a right to use the ballot to send a message to voters.\(^{142}\) The Court explained, “[b]allots serve primarily to elect candidates, not as forums for political expression.”\(^{143}\) Adhering to *Burdick*, the Court declined to acknowledge any constitutional protection for the expressive value of voting.\(^{144}\)

Finding that the anti-fusion law did not severely burden minor parties, the Court held that the state did not need to survive strict scrutiny analysis in order be valid. The state articulated four reasons to justify the law: avoiding voter confusion, promoting candidate competition, preventing electoral distortions and ballot manipulations, and discouraging party splintering and unrestrained factionalism.\(^{145}\) After declaring these justifications sufficient to support Minnesota’s law, the Court introduced an additional reason to justify the ban on fusion, one that had not been raised by the state. For the first time, the Court declared that a state’s interest in the stability of its political structure allowed it to enact legislation promoting the two-party system.\(^{146}\)

Building on previous cases that acknowledged a state’s interest in the stability of its government,\(^{147}\) the majority said that to achieve that goal, state laws could favor the two-party system.\(^{148}\) Although this was a new approach to election law for the Court, the opinion devoted relatively little space to the exploration of how fusion threatened the two-party system or how the two-party system encouraged stability.\(^{149}\) With

---

\(^{140}\) See *id.* at 363.

\(^{141}\) See *id.* at 358-59.

\(^{142}\) *Id.*

\(^{143}\) *Timmons*, 520 U.S. at 363 (citing *Burdick*, 504 U.S. at 438).

\(^{144}\) See *id.* See also *supra* notes 92-100 and accompanying text.

\(^{145}\) See *Timmons*, 520 U.S. at 364.

\(^{146}\) See *id.* at 367.

\(^{147}\) *Eu*, 489 U.S. at 226; *Storer*, 415 U.S. at 736.

\(^{148}\) *Timmons*, 520 U.S. at 367-68.

\(^{149}\) See *id.*
references to James Madison’s fear of factions, the decision declared that states are permitted to enact laws that favor the two-party system in order to “temper the destabilizing effects of party-splintering and excessive factionalism.”150

The majority’s unsolicited defense of the two-party system was contrary to the Court’s past political party jurisprudence. In Williams, the Court had implied that protecting the major parties would not sufficiently justify infringement on a minor party’s associational rights.151 At oral arguments for Timmons, it is no wonder that counsel and courtroom observers were surprised by Justice Scalia’s questioning on the protection of the two-party system.152 When it looked like counsel for Minnesota was not willing to admit anti-fusion laws were intended to protect the major parties, Justice Scalia interjected: “Well, you wouldn’t concede the major point, would you, that there is something wrong about the state establishing its electoral machinery . . . to facilitate and encourage a two-party system . . . ?”153 Justice Scalia proceeded to guide counsel to argue that states should be able to choose whether and in what way they will protect the major parties, a point which, admittedly, counsel had not planned to assert.154

In dissent, Justices Stevens, Ginsburg, and Souter argued that the Court should not have considered this justification, since the state had never raised it.155 Justice Stevens, joined by Justice Ginsburg, suggested that the protection of the two-party system was the “true basis” for the majority’s decision against the New Party and argued that even if the state had properly raised this justification, it would have been insufficient.156 In their view, the risks of government instability resulting from fusion voting were speculative, and

150 See id. at 368 (citing THE FEDERALIST NO. 10 (James Madison)). Associations have great power and if not checked, Madison thought, associations could destroy popular government by implementing policies in opposition to the will and benefit of the majority. Indirect elections of the senate, the Electoral College and the tripartite nature of the federal government were designed to shield against the dangers of factions. See THE FEDERALIST NO. 10 (James Madison). See also infra note 229 and accompanying text.
151 Williams, 393 U.S. at 23.
152 See SIFRY, supra note 24, at 297.
154 See id. at 26-29.
155 Timmons, 520 U.S. at 377-78 (Stevens, J., dissenting) (Justice Ginsburg joining, Justice Souter joining in part).
156 See id.
the burden created by anti-fusion laws demanded a greater demonstration of threat. Unlike Justice Stevens, Justice Souter believed that anti-fusion laws might be justified based on a state interest of protecting the two-party system. To satisfy constitutional scrutiny, however, Justice Souter would have had the state demonstrate that fusion voting would, in fact, threaten the two-party system and that the disintegration of the two-party system would risk state instability. As the state had failed to do this, Justice Souter would not have upheld the law.

The New Party’s day in court came and went, and a revival of fusion voting now seemed permanently buried in America’s electoral graveyard. State constitutions, however, have the ability to revive constitutional issues that the Supreme Court has killed.

IV. CHALLENGING ANTI-FUSION LAWS BASED ON STATE CONSTITUTIONS

A. Protecting Rights through State Constitutions

State constitutions are wholly independent documents; they are not drafted to echo the federal Constitution. In fact, many state constitutions were written and ratified prior to the federal Constitution. Framers of the federal Constitution decided to have the Bill of Rights apply only to the federal government because they believed that state constitutions sufficiently protected citizens from state governments. Even those state constitutions that were written after the adoption of the federal Bill of Rights borrowed from the language of the existing state constitutions more than from the federal Constitution. This history leads many scholars and jurists to agree with Justice William Brennan’s conclusion that “the decisions of the [Supreme] Court are not, and should not be,

---

157 See id.
158 See id. at 383 (Souter, J., dissenting).
159 Timmons, 520 U.S. at 384 (Souter, J., dissenting).
160 See JENNIFER FRIESEN, STATE CONSTITUTIONAL LAW: LITIGATING INDIVIDUAL RIGHTS, CLAIMS AND DEFENSES § 1-3(a) (3d ed. 2000).
161 See id.
163 See FRIESEN, supra note 160, § 5-2.
dispositive of questions regarding rights guaranteed by counterpart provisions of state law.”

Although state courts are free to interpret their constitutions without reference to the Supreme Court’s interpretation of the federal Constitution, many courts look to the Supreme Court for guidance. Especially in the area of civil rights, Supreme Court decisions have influenced the scope of state constitutional protection for individual rights. There are exceptions. Some provisions of state constitutions are unique to the states and do not have federal counterparts. State courts are left to understand the meaning of constitutional guarantees to public education, for instance, without direction from the Supreme Court. Furthermore, for the first hundred and fifty years of Supreme Court jurisprudence, state governments were not limited by the federal Bill of Rights. Courts intent on curbing abusive state action relied on their state, not federal, constitution.

---


566 See id.

567 See FRIESEN, supra note 160, § 1-3(b).


569 The Supreme Court did not apply the First Amendment to the action of state governments until 1925. See supra note 108.

570 For example, in 1859, Wisconsin declared that government appointed counsel for indigent defendants was part of the right to fair trial guaranteed by the state constitution. See Carpenter v. Dane County, 9 Wis. 274 (1859). More than a century later, the U.S. Supreme Court took a similar position. See Gideon v. Wainwright, 372 U.S. 335, 344 (1963). See generally Shirley S. Abrahamson, Reincarnation of State Courts, 36 Sw. L.J. 951 (1982).
Over the last several decades, state courts have revived a practice of independent interpretations of their constitutions, finding greater protections for individual rights than those provided by the federal Constitution.\textsuperscript{171} Even state constitutional provisions that mirror language in the federal Constitution have been interpreted as more expansive than the federal Constitution.\textsuperscript{172} For instance, after the U.S. Supreme Court upheld a state’s sodomy law in \textit{Bowers v. Hardwick} against a right to privacy challenge,\textsuperscript{173} the Georgia court struck down its state’s sodomy law based on the Georgia constitution’s right to privacy.\textsuperscript{174}

This Note explores state constitutional law generally as it applies to fusion voting, but each state’s constitution deserves its own analysis. That being said, there are shared characteristics of state constitutional law that make anti-fusion laws vulnerable to state constitutional challenges even though a federal challenge failed. First, there is an absence of federalism concerns when state courts are interpreting state constitutions.\textsuperscript{175} This means state courts may adopt a less deferential approach in analyzing state legislatures’ justifications for anti-fusion laws. Second, the history of state constitutional development reflects dedication to broad and diverse political participation.\textsuperscript{176} This conception of politics may mean minor political parties receive more protection under state constitutions than under the federal Constitution. Finally, several state courts have articulated a broader interpretation of freedom of expression than the Supreme Court has found in the federal Constitution.\textsuperscript{177} This means state

\begin{footnotesize}
\begin{enumerate}
\item The Maryland courts, for example, have “emphasized that, simply because a Maryland constitutional provision is \textit{in pari materia} with a federal one or has a federal counterpart, \textit{does not} mean that the provision \textit{will always} be interpreted or applied in the same manner as its federal counterpart.” Dua v. Comcast Cable of Maryland, Inc., 805 A.2d 1061, 1071 (Md. 2002).
\item Powell v. State, 510 S.E.2d 18 (Ga. 1998). The Georgia Court relied on language of the state constitution that is almost identical to the Due Process clause of the Fourteenth Amendment in the federal Constitution. \textit{Id.} at 21. \textit{Compare} U.S. CONSTIT. amend. XIV (“No State shall . . . deprive any person of life, liberty, or property, without due process of law.”) \textit{with} Ga. CONSTIT. art. I, § 1, para. 1 (“No person shall be deprived of life, liberty, or property except by due process of law.”).
\item See infra Part IV.B.
\item See infra Part IV.C.
\item See infra Part IV.D.
\end{enumerate}
\end{footnotesize}
courts may adopt a more protective approach to rights of associations and the value of voting.

B. The Strength of State Court Judicial Review

In *Timmons*, as in other election law cases, the Supreme Court balanced a state’s role as regulator with a party’s rights of association. Acceptance of state justifications in this balancing act are in keeping with a general reluctance of federal courts to interfere with the way a state “defines itself as a sovereign.” State courts, for obvious reasons, need not be concerned about disrespecting the sovereignty of their own state. Without federalism concerns, deference to the political branches need not be as extreme as it is in the federal courts. This is particularly true when state courts face claims from a minority of the population, who by their very numbers will never have control of the legislature.

Some state constitutions explicitly authorize judicial review of state legislation. While federal courts rely on precedent to support their powers of judicial review, they are cautious in exercising that power, especially when asked to invalidate legislative actions. State courts, however, have

---


180 See Burt Neuborne, *State Constitutions and Evolution of Positive Rights*, 20 RUTGERS L.J. 881, 900 (1988). For example, the Georgia Constitution provides “Legislative acts in violation of this Constitution or the Constitution of the United States are void, and the judiciary shall so declare them.” GA. CONST. art. I, § 2, para. 5. The judiciary articles of other states’ constitutions similarly declare the power of judicial review, although often less explicitly or with limitations. See, e.g., ARIZ. CONST. art. 6, § 2 (“The Supreme Court . . . shall not declare any law unconstitutional except when sitting in banc.”); LA. CONST. art. V, § 5, para. D (“In addition to other appeals provided by this constitution, a case shall be appealable to the supreme court if a law or ordinance has been declared unconstitutional.”); NEB. CONST. art. V, § 2 (“The judges of the Supreme Court, sitting without division, shall hear and determine all cases involving the constitutionality of a statute.”); UTAH CONST. art. VIII, § 2 (“The court shall not declare any law unconstitutional under this constitution or the Constitution of the United States, except on the concurrence of a majority of all justices of the Supreme Court.”); VA. CONST. § 1, para. 2 (“[T]he Supreme Court shall, by virtue of this Constitution, have appellate jurisdiction in cases involving the constitutionality of a law under this Constitution or the Constitution of the United States and in cases involving the life or liberty of any person.”).

enhanced legitimacy in reviewing legislation because they act according to explicit state constitutional provisions.\textsuperscript{182} In addition, many state court judges are elected, so their role in reviewing legislative action is less subject to charges of anti-majoritarianism.\textsuperscript{183} In fact, when states amended their constitutions to allow for the popular election of judges, they simultaneously limited the power granted to the legislature, intending that judges should help “make public policy.”\textsuperscript{184}

Therefore, when reviewing anti-fusion laws, state courts carry with them more than just substantive law to find the laws invalid. They are also, perhaps more importantly, draped with a cloak of legitimacy.

The \textit{Timmons} opinion allows states broad power to enact laws that protect the major parties. State courts interpreting state constitutions in response to a challenge to anti-fusion laws would be more critical of the state’s justification than the Supreme Court was in \textit{Timmons}. If state courts accept that there is a constitutionally permissible interest in protecting the two-party system, a more rigorous analysis would likely strike down anti-fusion laws because there is a striking lack of evidence to support the claim that fusion destroys the two-party system. Rather, New York’s experience suggests fusion voting creates a “modified two-party system,” where minor parties play an important role but do not replace the major parties.\textsuperscript{185}

Perhaps more importantly, there is a lack of evidence to show that stable democracy requires limiting the number of major parties to two. State constitutions protect broad participation in electoral government,\textsuperscript{186} and courts scrutinizing the justification of anti-fusion laws should find the state’s protection of major parties suspect and inconsistent with state constitutional conceptions of popular sovereignty.


\textsuperscript{182} Neuborne, \textit{supra} note 180, at 900.

\textsuperscript{183} See \textit{TARR}, \textit{supra} note 165, at 174-75 (suggesting that the election of state judges may explain why the legitimacy of state courts is less questioned than that of the U.S. Supreme Court).

\textsuperscript{184} See \textit{id.} at 122.

\textsuperscript{185} MAZMANIAN, \textit{supra} note 78, at 115 (“New York State has a highly competitive party system with two major contenders and third party contestants that are able to sustain themselves over time.”).

\textsuperscript{186} See discussion \textit{infra} Part IV.C.
C. Political Participation Protected in State Constitutions

Compared to the federal Constitution, state constitutions are extremely specific regarding their dedication to political participation. The first article of many state constitutions is a declaration of commitment to popular sovereignty.\(^{187}\) For instance, Article I, section I of the Washington constitution announces, “All political power is inherent in the people, and governments derive their just powers from the consent of the governed, and are established to protect and maintain individual rights.”\(^{187}\)

In addition, many state constitutions explicitly provide for the right to vote.\(^{189}\) This is dramatically different than the federal Constitution, which may prohibit discriminatory denial of the right to vote, but “does not confer the right of suffrage upon any one.”\(^{190}\) State constitutions typically have language similar to the Pennsylvania constitution: “Elections shall be free and equal; and no power, civil or military, shall at any time interfere to prevent the free exercise of the right of suffrage.”\(^{191}\) The right to vote has been one of the state constitutional rights most frequently expanded by amendment.\(^{192}\) States have similarly demonstrated their

---

\(^{187}\) See Tarr, supra note 165, at 11-12. See, e.g., N.H. Const. art. I (“All men are born equally free and independent; therefore, all government of right originates from the people, is founded in consent, and instituted for the general good.”); N.J. CONST. art. I (“All persons are by nature free and independent, and have certain natural and unalienable rights, among which are those of enjoying and defending life and liberty, of acquiring, possessing, and protecting property, and of pursuing and obtaining safety and happiness.”); WIS. CONST. art. I, § 1 (“All people are born equally free and independent, and have certain inherent rights; among these are life, liberty and the pursuit of happiness; to secure these rights, governments are instituted, deriving their just powers from the consent of the governed.”).

\(^{189}\) WASH. CONST. art. I, § 1.


\(^{190}\) See Minor v. Happersett, 88 U.S. 162, 178 (1874).

\(^{191}\) PENN. CONST. art I, § 5. See also, e.g., IND. CONST. art. II, § 1 (“All elections shall be free and equal.”); S.D. CONST. art. VI, sec. 19 (“Elections shall be free and equal, and no power, civil or military, shall at any time interfere to prevent the free exercise of the right of suffrage.”); TENN. CONST. art. I, § 5 (“That elections shall be free and equal, and the right of suffrage, as hereinafter declared, shall never be denied to any person entitled thereto, except upon a conviction by a jury of some infamous crime, previously ascertained and declared by law, and judgment thereon by court of competent jurisdiction.”).

\(^{192}\) See Tarr, supra note 165, at 105-08. There is a major exception to this trend of broadening the right to vote through constitutional revisions: In many southern states constitutional amendments were used to disenfranchise African Americans during the end of the Nineteenth Century. See id. at 107.
commitment to popular sovereignty by amending their constitutions to create procedures for referendum, initiative, and recall elections. 193

Constitutional amendments expanding the voting population and permitting direct democracy were adopted in response to fears of government corruption. 194 Drafters believed that the “main threats to rights, both collective and individual, were despotic officials and those seeking special privileges, rather than the people as a whole.” 195 Thus, state constitutions reflect cynicism of the motives of government officials and they aim to prevent manipulation of the electoral system that protects the power of the few. 196 Like Madison’s fear of factions, drafters of state constitutions worried that minorities could impede the will of the majority. 197 In contrast to Madison, however, the minorities the state constitutional drafters worried about were the ones elected to positions of power, not the ones advocating for political change. 198

When parties first began challenging election regulations, state courts were concerned with the rights of voters, not parties. 199 Decisions from the late 1800s and early 1900s expressed distress over corruption by party leaders and party bosses. 200 Judges believed manipulative political parties hampered political participation. 201 For some judges, distrust of political parties was an argument in favor of fusion voting, since “[p]olitical fusions among minority parties often serve as a check upon arrogant majority parties, or rather political

194 See TARR, supra note 165, at 170 (“[T]he initiative does provide a mechanism for circumventing the power of political elites within state government, just as its early proponents had expected.”); Gardner, supra note 189, at 649 (“Progressives . . . sought to reform state and local government by creating institutions of direct democracy, such as the initiative, referendum, and recall election, which would allow ordinary voters to thwart plans by incumbent power-holders to serve their own interests and to assure their own continuation in office.”).
195 TARR, supra note 165, at 78. More recently, this logic has motivated constitutional amendments providing for term limits for elected offices. See id. at 170, 172.
196 Cf. Gardner, supra note 189, at 649-50 (arguing that state constitutions’ focus on electoral responsiveness suggests partisan gerrymandering claims may be advanced under state constitutions).
197 See supra note 150.
198 See supra note 165, at 178, 100, 150-51.
200 See id. at 890.
201 See id. at 875.
parties whose thorough organization has enabled them to repeatedly elect officers that are dishonest and corrupt.”

Most judges facing early cases on fusion voting, however, did not consider the impact anti-fusion laws had on minor parties. Professor Adam Winkler notes that in an era of genuine party competition it is “easy to understand how the courts overlooked the duopoly-enhancing nature of many turn-of-the-century reforms.”

History now shows that anti-fusion laws were enacted to protect the political parties in power and they succeeded far better than they could have hoped. In *Timmons*, the Supreme Court was willing to allow states to purposefully favor two major parties, but state constitutions would not provide that leeway. State constitutions are drafted to prevent laws that are enacted to protect the privileges of the elected. Unless current legislators can defend anti-fusion laws with less partisan motives than those of the past, this manipulation of the electoral scheme should not only fail to justify such laws, it should result in their invalidity.

D. Freedom of Expression Protected in State Constitutions

Like the federal Constitution, state constitutions contain specific provisions protecting freedom of speech and assembly. In the resurgence of state constitutional law of the last twenty-five years, freedom of expression has been one of the most watched areas. Several courts have held that their

---

202 State ex rel. Dunn v. Coburn, 168 S.W. 956, 964 (Mo. 1914) (Brown, J., dissenting).

203 Early court challenges to anti-fusion voting laws were generally brought as claims under state constitutional rights to vote by ballot, and rights of free and equal elections. See, e.g., Dunn, 168 S.W. at 964; State ex rel. Runge v. Anderson, 76 N.W. 482 (Wis. 1898); State ex rel. Bateman v. Bode, 45 N.E. 195 (Ohio 1896); State ex rel. Sturdevant v. Allen, 62 N.W. 35 (Neb. 1895). This note explores modern challenges to anti-fusion laws based on freedom of expression and association, but even future challenges based on the right to vote are not precluded by these cases. The value of competition in the electoral arena may play a different role in challenges to anti-fusion laws brought today, as opposed to ones brought a century ago since the context of elections has changed dramatically. Cf. Winkler, *Voters’ Rights and Parties’ Wrongs*, supra note 199, at 892-95 (describing party competition at the turn of the century).

204 See Winkler, *Voters’ Rights and Parties’ Wrongs*, supra note 199, at 892.

205 See Argersinger, supra note 33, at 288.


state constitutions provide a broader right to free expression than the federal Constitution. This suggests state protections for expression may cover a broader range of activities than the First Amendment, specifically, the activities of voting and association.

The earliest state constitutions were drafted during the American Revolutionary war. Almost all included guarantees of freedom of speech. This is hardly surprising given the resentment towards British attempts to limit expression in the colonies. Freedom of expression was seen as having “a direct relationship to freedom from government oppression.” State courts have relied on this history in finding state constitutional law protects a wide range of expressive activities. Of course, some states drafted constitutions after the adoption of the federal Bill of Rights. But those states borrowed the broad language of older state constitutions in protecting free speech, instead of copying the federal Constitution.

Forty-one state constitutions protect the right of expression with affirmative avowals of the right to speak.


See, e.g., People v. Ford, 773 P.2d 1059, 1066 (Colo. 1989) (“[O]ur constitution extends broader protection to freedom of expression than does the first amendment to the United States Constitution.”). See also infra notes 213, 226 and accompanying text.

See TARR, supra note 165, at 61 (providing a table of states constitutions and the date of their adoption).


See id. at 21.

See id.

For example, in Pennsylvania, the highest court has paid special attention to the history of the Pennsylvania’s founder, William Penn, in analyzing the text of its constitution:

[Since] William Penn[] was prosecuted in England for the “crime” of preaching to an unlawful assembly and persecuted by the court for daring to proclaim his right to a trial by an uncoerced jury . . . [i]t is small wonder . . . the rights of freedom of speech, assembly, and petition have been guaranteed since the first Pennsylvania Constitution, not simply as restrictions on the powers of government, as found in the Federal Constitution, but as inherent and “invaluable” rights of man. Commonwealth v. Tate, 432 A.2d 1382, 1388 (Pa. 1981).

See TARR, supra note 165, at 61.

See FRIESEN, supra note 160, § 5-2. Today, every state provides for the rights of speech in their constitution. See Force, supra note 206, at 125.

For a listing of state free speech and press provisions, see FRIESEN, supra note 160, at app. 5 and Note, Private Abridgement of Speech and the State Constitutions, 90 YALE L.J. 165, 180-81 n.79 (1980).
Kansas’ constitution provides, “The liberty of the press shall be inviolate; and all persons may freely speak, write or publish their sentiments on all subjects.” Similarly, the Michigan constitution says, “Every person may freely speak, write, express and publish his views on all subjects.” These provisions are typical.

Forty-six states have provisions guaranteeing the right to assembly and these provisions are often expressed as positive declarations as well. Using Kansas and Michigan as examples once again, Kansas’ constitution provides, “The people have the right to assemble, in a peaceable manner, to consult for their common good, to instruct their representatives, and to petition the government, or any department thereof, for the redress of grievances.”

Although these provisions of state constitutions protect a right similar to the First Amendment of the federal Constitution, their distinct language implies they deserve a distinct analysis. Most state constitutional provisions are in sharp contrast to the federal Constitution, which simply declares that “Congress shall make no law” restraining expressive rights, and does not provide a positive guarantee. The affirmative nature of the state provisions illustrates the spirit in which they were enacted, celebrating the fundamental rights of state citizens of which freedom of speech was a priority. Relying on this, some states have found that under

217 KAN. CONST. Bill of Rights § 11.
218 MICH. CONST. art. I § 5.
219 See FRIESEN, supra note 160, at app. 5.
220 See Force, supra note 206, at 139.
221 KAN. CONST. Bill of Rights § 3.
222 See MICH. CONST. art. I § 5 (“The people have a right to peaceably to assemble, to consult of the common good, to instruct their representatives and petition the government for redress of grievances.”).
223 See Kevin Francis O’Neill, The Road Not Taken: State Constitutions as an Alternative Source of Protection for Reproductive Rights, 11 N.Y.L. SCH. J. HUM. RTS. 1, 31 (1993) (“If a court were interpreting contractual terms, would it conclude, as readily as some courts have, that these clauses are coextensive?”).
224 The First Amendment in the federal Constitution reads, “Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.” U.S. CONST. amend. I.
225 See Todd F. Simon, Independent but Inadequate: State Constitutions and Protections of Freedom of Expression, 33 U. KAN. L. REV. 305, 310 (1985) (“Freedom of the press was considered the right of greatest importance, at least initially, and assuring freedom of expression was a primary concern of settlers in new states.”).
their constitutions infringements of the rights of speech can occur when there is no state action. In New Jersey, for instance, state courts have relied on the affirmative nature of the free speech provision to hold that the state constitution provides a right to distribute political leaflets in shopping centers even though the federal Constitution would not provide that right.

A broad right to freedom of expression is valuable in challenging anti-fusion laws for two reasons. First, freedom of association is an offshoot of freedom of expression, so the scope of protection for speech is indicative of the protection associations will be given. Second, voting is an expressive act. Freedom of expression in the federal Constitution does not protect the act of voting, but freedom of expression in state constitutions should.

1. Broad Right of Expression Protects Associations

The rights of political parties are in essence the rights of voters who have collectivized in order to engage in more efficient expression. Since state constitutional language and history suggest broader protection for expression than the federal Constitution, state constitutions should be construed to provide greater protection for associational rights of political parties.

Although political parties did not exist at the time the federal Constitution was drafted, its framers sought to guard against the danger of “factions,” which James Madison defined as groups of citizens “united and actuated by some common impulse of passion, or of interest.” The Supreme Court, therefore, was understandably slow in developing a freedom of association doctrine to protect the very group activity Madison

226 N.J. Coalition Against War in the Middle East v. J.M.B. Realty Corp., 650 A.2d 757, 771 (N.J. 1993) (“[T]he State right of free speech is protected not only from abridgement by government, but also from unreasonably restrictive and oppressive conduct by private entities.”).

227 See id. at 770-71.

228 See supra notes 92-100 and accompanying text.

229 THE FEDERALIST NO. 10 (James Madison). Similarly, of the Democratic-Republican societies forming in 1794, George Washington said, “All combinations and associations, under whatever plausible character, with the real design to direct, control, counteract or awe the regular deliberation and action of the constituted authorities are . . . of fatal tendency.” ROBERT J. BRESLER, FREEDOM OF ASSOCIATION 23 (2004) (quoting President George Washington, Farewell Address to the People of the United States (Sept. 17, 1796), in INDEP. CHRON., Sept. 26, 1796).
State courts, however, were quicker to recognize the democratic value and necessity of associations.231 Half a century before the Supreme Court said there was constitutional protection for associations, state courts had recognized that political parties are protected under fundamental rights of speech and assembly.232 The highest court of California wrote in 1900,

> No one, it would seem, can be so thoughtless as not to realize that government by the people is a progressive institution, which seeks to give expression and effect to the wisest and best ideas of its members. . . . [E]lectors . . . may freely assemble, organize themselves into a political party, and use all legitimate means to carry their principles of government into active operation through the suffrages of their fellows. Such a right is fundamental.233

Similarly the Wisconsin Supreme court declared in 1910 that “[t]he right of members of a political party to freely assemble, deliberate and act, to promote the interest of such party, is a right guaranteed by the Constitution, state and national. Freedom to do those things, reasonably appropriate to the effective maintenance of party organization, cannot be abridged.”234 While acknowledging constitutional protections for political parties, state courts also allowed state legislatures to regulate them. Political parties, these courts recognized, are more than private associations. They are part of the machinery of democracy.235 State courts upheld Australian ballot laws and other reforms, not because they rejected the constitutional rights of parties, but rather because they believed electoral regulations would increase voter choice and opportunity.236 These courts sustained regulations that they thought would protect the rights of voters to participate effectively in party organizations.237

---

230 See BRESLER, supra note 229, at 25, 32.
231 The fact that state courts considered the rights of political parties before the U.S. Supreme Court is likely a result of the historical development of the incorporation doctrine. It was not until 1925 that the Supreme Court applied the First Amendment to state action. See supra note 108.
232 See Winkler, Voters’ Rights and Parties’ Wrongs, supra note 199, at 874.
233 See Britton v. Bd. of Election Comm’rs. of San Francisco, 61 P. 1115, 1117 (Ca. 1900).
234 See, e.g., State ex rel. Van Alstine v. Frear, 125 N.W. 961, 976-77 (Wis. 1910).
235 See Winkler, Voters’ Rights and Parties’ Wrongs, supra note 199, at 884.
236 See id. at 884 (“Protecting and preserving the ability of voters to make effective use of electoral opportunities free from the corrupting influence of party leaders led most state courts to uphold laws restricting ballot access.”).
Today, anti-fusion laws limit voter choice rather than ensure it. By preventing parties and their supporters from nominating their selected candidates, anti-fusion laws run afoul of a long tradition of state protection for voter participation as expressed through political parties.

2. Broad Right of Expression Protects the Act of Voting

In addition to association as a derivative right of freedom of speech, freedom of expression on its own might prohibit anti-fusion laws. As discussed above, voting may be considered an expressive act.\textsuperscript{238} Using the facts of \textit{Timmons} as an example, New Party members wanted to vote for Andy Dawkins on the New Party ballot line to express a message that they felt would not be expressed by voting for him on the Democratic-Farmer-Labor ballot line, namely, that the Democratic Party was too centrist.\textsuperscript{239} That the medium for this voter communiqué would be the ballot does not change its essential expressive nature.

The majority of the U.S. Supreme Court in \textit{Timmons} rejected the link between voting and expression. State courts, however, are free to take another approach. No trend of protection for the expressive nature of voting has yet emerged in state courts, but there are promising harbingers. Several states have found state constitutional protection for write-in votes, for instance.\textsuperscript{240} Recently, the Utah Supreme Court described the constitutional right to vote for a ballot initiative as important because it “encourages political dialogue” as well as “allows the general populace to have substantive and meaningful participation in enacting legislation.”\textsuperscript{241} Oregon’s Judge Landau has gone further in acknowledging the ballot as a place of expression. In \textit{Freedom Socialist Party v. Bradbury},

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{238} See supra notes 92-95 and accompanying text.
\item \textsuperscript{239} See \textit{Disch}, supra note 10, at 17-18.
\item \textsuperscript{240} See Littlejohn v. People \textit{ex rel.} Desch, 121 P. 159 (Colo. 1912); Smith v. Smathers, 372 SO. 2d 427 (Fla. 1979); Thompson v. Wilson, 155 S.E.2d 401 (Ga. 1967). Even though these cases have protected write-in votes under the right to vote, and not under freedom of expression, they suggest state constitutions differ in their understanding of the value of voting from the federal Constitution. See supra notes 92-100 and accompanying text.
\item \textsuperscript{241} Gallivan v. Walker, 54 P.3d 1069, 1081 (2002).
\end{itemize}
\end{footnotesize}
the Oregon Court of Appeals considered the constitutionality of a statute preventing the Freedom Socialist Party from using their party name on the ballot because the Socialist party already had been given an exclusive right to the use of its name and the word “socialist.” In a concurring opinion, Judge Landau found the statute limited the ability of a political party to communicate its message to the public, and this was a violation the Oregon constitutional right to free speech.

State constitutions value voting more than the federal Constitution. Moreover, they offer more protection for expressive activities. Therefore, state courts should understand voting as an act of expression. Fusion voting is especially motivated by an urge to express one’s political views. As recognized by Justice Stevens, fusion allows voters to indicate views they feel are not sufficiently represented by the major parties, while still allowing them to vote for the candidate they hope will win the election. Fusion voting, then, should receive constitutional protection as part of states’ protection of expression.

V. CONCLUSION

Legislators enacted anti-fusion laws in order to ensure their reelections, not as part of a noble defense of government stability. In Timmons, the Supreme Court declared states have the right to enact such laws to protect the two-party system. State courts interpreting state constitutions should treat challenges to anti-fusion laws differently. Drafters of state constitutions were dedicated to expansive political participation and were cynical of elected power. Sustaining laws that have the purpose of limiting the viability of minor parties reduces voter choice and shields established politicians from challenges. Anti-fusion laws, then, are incompatible with the goals of state constitutions. Moreover, protection of the two-party system is an especially weak defense for these laws in

242 Freedom Socialist Party v. Bradbury, 48 P.3d 199, 200 (Or. Ct. App. 2002). The majority found the statute unconstitutional under the federal Constitution’s First Amendment and never addressed whether there was a state constitutional violation, noting that parties did not raise a state constitutional issue on appeal. See id. at 201 n.2.

243 See id. at 208 (Landau, J., concurring) (“[T]he statute prohibits a political party from using specified words in communicating a message to members of the voting public.”).

244 See supra Part IV.C.

245 Timmons, 520 U.S. at 381 (Stevens, J., dissenting).
states that value free expression. Anti-fusion laws infringe upon the rights of voters to express their political beliefs and the rights of parties and their adherents to associate.

Many voters are unhappy with their choices on Election Day but anti-fusion laws allow them no satisfying options. They can “hold their nose” and vote for the candidate they believe is the lesser of two evils or they can cast a vote that is unlikely to translate into actual political power. A revival of fusion voting would solve this dilemma, but after the Timmons opinion was issued, a revival of fusion voting appeared unlikely. Examining state constitutions reveals a different future—anti-fusion laws are not as unassailable as they may seem. State courts have the ability, authority and obligation to invalidate anti-fusion laws and thereby liberate voters and parties alike.

Elissa Berger

---


247 B.A., Macalester College; J.D. Candidate 2006, Brooklyn Law School. Thanks to Dean Joel Gora, Professor Ed Cheng, and my family, especially Bethany Berger for her guidance. For my father.
Appendix:

HYPOTHETICAL BALLOT IN A FUSION VOTING SYSTEM

<table>
<thead>
<tr>
<th>Party</th>
<th>Democrat</th>
<th>Republican</th>
<th>Purple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate for Office</td>
<td>Candidate X</td>
<td>Candidate Y</td>
<td>Candidate X</td>
</tr>
</tbody>
</table>

**BALLOT**

*Make your selection by filling in one of the circles.*

**HYPOTHETICAL ELECTIONS RESULTS**

Votes for Candidate Y as Republican ...........................................48%
Votes for Candidate X as Democrat ...........................................44%
Votes for Candidate X as Purple ...........................................8%

*Candidate X is declared the winner with 52% of the vote.*
In Defense of Market Self-Regulation

AN ANALYSIS OF THE HISTORY OF FUTURES REGULATION AND THE TREND TOWARD DEMUTUALIZATION

I. INTRODUCTION

The debate over the efficacy of the U.S. system of market self-regulation—where the securities and futures industries regulate themselves with oversight from the federal government—has been ongoing since its inception some seventy years ago. Populist politicians have long compared the system, primarily in times of regulatory failures or market crises, to the proverbial fox guarding the henhouse. The controversy surrounding recent market scandals has led to new scrutiny of whether exchanges can properly police themselves.

1 In 1963 the efficacy of the self-regulatory system was called into question by stock market abuses reported in a Securities and Exchange Commission study, but the SEC concluded that self-regulation should be maintained. Roberta S. Karmel, Turning Seats Into Shares: Causes and Implications of Demutualization of Stock and Futures Exchanges, 53 HASTINGS L.J. 367, 401 (2002) (citing SEC, REPORT OF THE SPECIAL STUDY OF SECURITIES MARKET, H.R. DOC. NO. 95, at 502, 83, 414-15 (1st Sess. 1963) [hereinafter SPECIAL STUDY]); see generally JERRY W. MARKHAM, THE HISTORY OF COMMODITY FUTURES TRADING AND ITS REGULATION (1987) [hereinafter HISTORY OF COMMODITY FUTURES] (illustrating periodic strife between Congress and futures industry since late 1800s regarding self-regulation). Between 1880 and 1920 there were some 200 bills introduced in Congress to regulate futures and options trading. Id. at 10.

2 See, e.g., Laurie P. Cohen & Kate Kelly, NYSE Turmoil Poses Question: Can Wall Street Regulate Itself?, WALL ST. J., Dec. 31, 2003, available at http://online.wsj.com/article_print/0,,SB107282097396518500,00.html (while the current system of self-regulation with government oversight has survived numerous crisis, the scandal surrounding a $188 million pay package awarded to former New York Stock Exchange Chairman Richard Grasso has forced Wall Street to confront whether the system must be “drastically reformed or even replaced” in order to restore the confidence of investors). Some suggest that a for-profit exchange will not vigorously
The catalyzing events include a price-fixing investigation at the Nasdaq in the late 1990s, an ongoing investigation of floor trading at the New York Stock Exchange (NYSE), and governance failures at the NYSE that resulted in the ousting of former chairman Richard Grasso. Some have called for reforms to overhaul the system, while others say it should be scrapped altogether.

The current debate over self-regulation was actually well under way before these regulatory failures pushed it onto the business pages. Industry and government officials have been actively debating how self-regulation can be adapted to address new conflicts of interest caused by a much more secular change: the ongoing trend of demutualization, where securities and futures exchanges convert to for-profit entities from not-for-profits. The latest example is the NYSE's announced plan to demutualize as it acquires electronic trading system Archipelago Holdings.

This Note argues that the self-regulatory model, while in need of some type of reform, will survive the latest round of scrutiny because time has shown that it is the most efficient and practical alternative. The debate over self-regulation must be made in the context of the alternatives: the government as the sole securities regulator or no government oversight at all. It is important to remember that direct governmental regulation raises some of the same concerns that self-regulation does, as well as different concerns that are particular to large government bureaucracies.


See, e.g., Jenny Anderson, Big Changes at Exchanges Bring Their Self-Regulation Into Question, N.Y. TIMES, April 28, 2005 (At issue is whether self-regulation works . . . virtually every regional exchange has been cited for turning a blind eye to improper or illegal behavior of one sort or another.)

Cohen & Kelly, supra note 2.

Id.


See, e.g., Andre Postelnicu et al., NYSE to Merge with Archipelago Exchange, FIN. TIMES, Apr. 21, 2005.


See, e.g., Raymond Urban & Richard Menkel, Federal Regulation of
advantages for self-regulation as an adjunct to government oversight are to minimize intrusion into the marketplace, take advantage of the expertise of professionals working within the exchanges, and defray the costs of monitoring and policing trading practices to the private sector.10

The self-regulatory model has been in continuous evolution since it was created under the Commodity Exchange Act in 1936, first operating under the oversight of the Commodity Exchange Authority (CEA), and later the Commodities Futures Trading Commission.11 In the case of the securities industry, modern self-regulation began under the Securities Exchange Act of 1934, when the Securities and Exchange Commission was created. For decades the government sought to increase its authority over exchanges, focusing on rooting out fraud and bolstering the financial strength of market participants.12 With regards to the futures industry, in recent years the government has changed tacks by relaxing the regulatory framework and giving self-regulatory organizations more authority, not less, as a means to improve efficiency.13

Whiskey Labeling: From the Repeal of Prohibition to the Present, 15 J. LAW & ECON. 411 (1972); William A. Jordan, Producer Protection, Prior Market Structure and the Effects of Government Regulation, 15 J. L. & ECON. 151 (1972), as cited in Miller, supra note 8, at 861. Miller argues that in some cases no regulation at all (other than that provided by privately enforced rights) might be a preferable alternative. Id. New York Attorney General Eliot Spitzer said that the Securities and Exchange Commission has become too close with the securities industry, as evidenced by its early rejection of efforts to force Wall Street banks to reform the way they provide stock research. Interview by Reuters News with Eliot Spitzer, New York State Attorney General, New York, NY (Nov. 12, 2003). Spitzer went on to say that the SEC is slow to uncover abuses, is too soft on violators, and is too slow in crafting new rules that address abuses. Id. Throughout history, no single government authority has ever been entrusted with regulatory authority over all American banks. See Jerry W. Markham, Super Regulator: A Comparative Analysis of Securities and Derivatives Regulation in the United States, the United Kingdom, and Japan, 28 BROOK. J. INT’L L. 319, 405 n.435 (2003) (citing BLUEPRINT FOR REFORM: THE REPORT OF THE TASK GROUP ON REGULATION OF FINANCIAL SERVICES 8 (1984)).

10 See WHITE PAPER, supra note 6.
11 See generally HISTORY OF COMMODITY FUTURES, supra note 1.
12 Id.
13 See Hearing Before the Subcomm. on Research, Nutrition and General Legislature, Subcomm. on Agric., Nutrition and Forestry, 107th Cong. 2 (2000) (statement by C. Robert Paul, General Counsel, Commodity Futures Trading Commission, detailing the content and purpose of the Commodity Futures Modernization Act of 2000, at the time it was being proposed to Congress).

[T]he proposed RFE [(Recognized Futures Exchange)] offers significant regulatory relief compared to the current requirements applicable to designated contract markets. . . . The second category, the derivatives transaction facility [DTF], would be subject to a lesser degree of Commission oversight. . . . Finally, the third category, the exempt multilateral transaction
It is natural that self-regulatory bodies be forced to undergo further reforms as new governance concerns arise. A current concern is the creation of new potential conflicts of interest that stem from the transformation into for-profit entities. But this paper also argues that the underlying forces fueling the move to demutualize—for instance, the advent of electronic exchanges, internationalization of markets and increased competition both domestically and abroad—will create new incentives for exchanges to better police themselves by increasing more competition among exchanges.

This analysis will begin by looking at the historic evolution of self-regulation in the U.S. futures industry, including how power has shifted back and forth between the federal government and the exchanges. The Note then reviews how different exchanges have reacted to recent allegations of conflicts of interest and failures in corporate governance, and finally, it reviews the most widely considered options being proposed to change the system.

II. THE EVOLUTION OF SELF-REGULATION

A. Prior to the Commodities Exchange Act of 1936

Organized futures trading dates back to the founding of the Chicago Board of Trade (CBOT) in 1848, where execution facility [MTEF], or exempt MTEF, would operate on an unregulated basis.

See White Paper, supra note 6, at 1.

The regulatory regimes of both futures and securities exchanges are similar. This Note addresses issues relevant to both, though unless otherwise noted, it will focus on futures exchanges. The regulation of futures trading is less politicized than that of securities—at least by some measures—because futures and other derivatives contracts are mainly traded by large financial institutions and other sophisticated investors. By contrast, the shares in publicly held companies traded on securities exchange are to a large degree held, both directly and indirectly, by retail investors and pensioners. As a result, the tendency of elected officials to pressure for stricter regulation in the securities industry is more tightly linked to the ups and downs of the market than it is for the futures industry (though regulation of the futures is certainly not immune from political lobbying by market participants). See generally Markham, supra note 9, at 399-403 (describing the hysteria within Congress and by various regulatory offices to quickly react to scandals surrounding publicly traded companies like energy trader Enron Corp. and investment bank Merrill Lynch & Co.)

Demutualization is the process of reorganizing a mutualized, or member owned, entity into a for-profit corporation with shareholders. Organized securities and futures exchanges where traditionally operated as non-profit membership organizations, but advances in technology, increased competition and other market forces have led numerous exchanges to demutualize. Karmel, supra note 1, at 368-69.
commodities ranging from grain to coal to alcohol were traded. At the outset, U.S. futures exchanges regulated themselves, with some oversight from state regulators, though none from the federal government. The main incentive for self-regulation was to assure high standards of conduct and decorum on the trading floor. It became clear early on that a more rigid form of government oversight would be needed.

As trading volumes on futures exchanges grew during the late 19th century, there developed a steady flow of allegations that the market was vulnerable to manipulation. The failures of the system became apparent in the 1880s with the rise of so-called bucket shops, which were poorly financed, off-exchange establishments where speculators bet on commodities prices. Many banks would not lend money to brokerages that were not members of the most prominent exchanges, so smaller brokerages would often “bucket” their clients’ money in order to get capital with which to trade. By 1891, the practice had become so prevalent that one member of the CBOT, one of the most respected commodities exchanges, wrote a pamphlet actually defending bucket shops. The Consolidated Exchange, a securities exchange which at the time was a powerful rival to the New York Stock Exchange,

17 See History of Commodity Futures, supra note 1, at 4.
18 Id.
19 See id. (citing Jonathan Lurie, The Chicago Board of Trade, 1859-1905, at 25 (1979)).
20 See id. at 10.
21 Id. at 4.
22 Thomas Hieronymus, Economics of Futures Trading 88 (1971).
24 H. S. Irwin, Legal Status of Trading in Futures, 32 Ill. L. Rev. 155, 155 n.5 (1938). Bucket shops accepted customers' orders and funds but did not execute the orders on any exchange. Rather, they simply bet the customer would lose and kept the customer's money when they did. If the customer won too much, the bucket shop would simply fold its operations and move to a new location. Comparative Analysis, supra note 15, at 339 n.92 (citing John Hill, Jr., Gold Bricks of Speculation 37-39 (1904)).

The term “bucketshop” as now applied in the United States, was first used in the late [18]70’s, but it is very evident that it was coined in London as many as 50 years ago, when it had absolutely no reference to any species of speculation or gambling. It appears that beer swillers from the East Side (London) went from street to street with a bucket, draining every keg they came across and picking up cast off cigar butts. Arriving at a den, they gathered for social amusement around a table and passed the bucket as a loving cup . . . the den soon became called a bucket shop. Later on the term was applied, both in England and the United States, as a byword for reproach, to small places where grain and stock deals were counterfeited.

Id.
came to be regarded as a "den of bucketeers." The Chicago Open Board of Trade, or the "Little Board," was said to be captured by bucketeers (The Chicago Open Board of Trade later became the Mid-America Commodity Exchange, which eventually joined with the CBOT in 1985).25

As the bucket shops became more influential, the exchanges found themselves in a conflicted position with state regulators. Those that were losing business began to push for more assistance from state regulators, but at the same time, the exchanges tried to defend their right to regulate themselves against mounting criticism that futures trading had become corrupted. Some of the more prominent exchanges tried to put the bucket shops out of business by cutting off access to market quotations.26 In a major victory, the CBOT established the legal right to prevent bucket shops from obtaining market quotations in the 1904 case, Chicago Board of Trade v. Christie Grain & Stock Co.27 The CBOT used undercover detectives and eventually prosecuted a number of its members for engaging in bucket shop activities.28 Expulsions resulted, with 281 people in Illinois indicted for violating the state’s anti-bucket shop laws.29

Political pressures to create federal oversight of the exchanges mounted, but the industry still had some powerful supporters. President Herbert Hoover was quoted as saying that the CBOT is "the most economical and efficient agency of the marketing of foodstuffs anywhere in the world."30 Other politicians stated that it would be a mistake to supplant a system that is dictated by the market with an inefficient government bureaucracy.31

26 See HISTORY OF COMMODITY FUTURES, supra note 1, at 9.
27 Bd. of Trade of Chicago v. Christie Grain & Stock Co., 198 U.S. 236, 253 (1905) (upholding an injunction which cut off quotations to the operations of C.C. Christie). “[T]he plaintiff’s collection of quotations is entitled to the protection of law. It stands like a trade secret . . . that others might do similar work, if they might, does not authorize them to steal the plaintiff’s.” Id. at 250.
28 See HISTORY OF COMMODITY FUTURES, supra note 1, at 10.
29 See id.
30 Hearing on Futures Trading Before the House Comm. on Agric., 66th Cong., 583 (1921).
31 Id. at 125. Representative Thaddeus Caraway:
I have never believed that someone who sits here in the basement of some Government building with his hair parted in the middle can run this country better than all the people can run their own private business. I have no patience with that. I never went to a department in my life I did not come away thoroughly angry and half ashamed of my government . . . .

Id.
The problems with bucket shops continued, however, giving way to a series of congressional investigations and unsuccessful bills that sought to put heavy regulations on futures trading, or prohibit it altogether. The industry drew on the Supreme Court’s decision in *Board of Trade v. Christie* to defend against accusations that futures trading was similar to gambling, a charge that was grounded in a general failure to understand how trading activity could be legitimate when the majority of commodities contracts were never actually delivered. Justice Oliver Wendell Holmes validated the concept of futures trading in *Board of Trade v. Christie*. Using the CBOT’s grain pit as an example, he acknowledged that in three-fourths of the transactions no grain actually exchanged hands. Still, Holmes determined that the sales were settled legitimately, proclaiming that the fact that contracts were satisfied in this way detracts in no way from the good faith of the parties and is consistent with a serious business purpose.

In 1921, Congress approved the Futures Trading Act (FTA), the first legislation to create federal oversight of futures trading. But the FTA was quickly struck down. In *Hill v. Wallace*, the Supreme Court ruled that the Act was an unconstitutional exercise of Congress’s taxation authority. The Act had sought to give the Secretary of Agriculture authority over exchanges by giving it the power to designate exchanges as “contract markets.” Options and grain futures contracts not traded on government approved exchanges were to be subject to a prohibitive 20-cent per bushel tax.

---

32 See id.
33 See id.

Speculation of this kind by competent men is the self-adjustment of society to the probable. Its value is well known as a means of avoiding or mitigating catastrophes, equalizing prices and providing for periods of want. It is true that the success of the strong induces imitation by the weak, and that incompetent persons bring themselves to ruin by undertaking to speculate in their turn. But legislatures and courts generally have recognized that the natural evolutions of a complex society are to be touched only with a very cautious hand, and that such course attempts at a remedy for the waste incident . . . are harmful and vain.

Id.

35 Id. at 246-47.
36 Id. at 251.
39 Id. at 63-64
40 Id. “To give such magic to the word ‘tax’ would be to break down all
The ruling in *Hill* helps to illustrate the complex nature of the relationship between futures exchanges and the government at a time when it was clear that federal intervention had become inevitable. While members of the futures industry were widely opposed to government oversight, representatives of the CBOT itself were not the ones to challenge the constitutionality of the FTA. A number of members of the CBOT brought the suit, and charged that the exchange’s board of directors refused to challenge the statute itself because it did not want to offend the Secretary of Agriculture.\(^{41}\) In his opinion, Justice William Howard Taft concluded that in not challenging the FTA itself the CBOT’s board of directors had failed in their duty to represent the interests of its members.\(^{42}\)

However, just a few days after the Supreme Court’s decision in *Hill* there was a manipulation of grain prices, which reinforced the belief within Congress that regulation was needed immediately.\(^{43}\) Legislatures quickly passed the Grain Futures Act of 1922,\(^{44}\) this time resting authority on its commerce powers rather than taxation powers. The Supreme Court subsequently upheld the Grain Futures Act based on Congress’s contention that market volatility on the exchanges was a burden on interstate commerce.\(^{45}\) Section 5 of the act permitted the Secretary of Agriculture to regulate futures trading by requiring that such transactions be conducted on a “contract” market that must be licensed by the federal government.\(^{46}\) That provision still forms the core of the current regulatory system today. It also required that the exchanges prevent such conduct as price manipulation,\(^{47}\) marking the onset of the exchanges’ role as self-regulator under oversight of the federal government.

---

\(^{41}\) Id. at 61, 72. “The averments of the bill are that the Board of Directors refused the request to bring suit because they feared to antagonize the public officials whose duty it was to construe and enforce the act, and not because they thought the act was unconstitutional.” Id. at 68.

\(^{42}\) Id. at 61. In a concurring opinion, Justice Louis Brandeis asserted that the Chicago Board of Trade should not be required to contest every statute that its members believe to be invalid. Id. at 74.

\(^{43}\) Markham, *supra* note 9, at 338 (citing H.R. REP. NO. 67-1095, at 2 (1922)).


\(^{45}\) Chicago Bd. of Trade v. Olsen, 262 U.S. 1, 56 (1923).


\(^{47}\) Id. at § 5(b).
The Grain Futures Administration (GFA), the Secretary of Agriculture office that carried out the Act, had the role of investigating practices at the exchanges, while actual regulation of trading was conducted by the exchanges. The arrangement had mixed results, with the government in its new role often just as responsible for regulation failures as the exchanges were. For instance, in an effort to boost surveillance the GFA required the clearing members of each exchange to provide daily reports that included the market positions of its customers. But while members of the CBOT were willing to make their records available, the GFA had only one internal auditor and was therefore unable to monitor the records in a meaningful way. The GFA suspected fraud and market speculation in many of the cases it reviewed, but it had limited success in prosecuting the cases. It also began to supervise the dissemination of news reports, an effort to stop unsubstantiated reports that were moving the market.

A series of trading scandals and the onset of the Great Depression led to a further loss of faith in the exchanges’ ability to regulate themselves. On the political front, a populist movement against futures trading began to gather steam. One Senator in favor of heavier government oversight called the CBOT the “world’s greatest gambling house.”

B. The Commodities Exchange Act

In 1936, Congress approved the Commodity Exchange Act, the result of efforts by President Franklin D. Roosevelt, who had pushed for new regulation of both the securities and

---

48 HISTORY OF COMMODITY FUTURES, supra note 1, at 15.
49 Id.
50 Id.
51 In 1927, however, the Grain Futures Administration suspended its requirements of daily reports for large traders, because of continuing charges that its reports were keeping large bullish speculators from operating in the wheat market and thereby depressing prices. It suspended reporting requirements from February 26, 1927, until November 1, 1927. It then determined that its reports did not have the effect of discouraging bullish spectators.
53 See HISTORY OF COMMODITY FUTURES, supra note 1, at 22-24.
54 See id. at 26-27.
futures industries. Regulation of the two industries was separated because the banking committees controlled securities matters and the agriculture committees controlled commodity exchanges, and neither was willing to cede power. The CEA was created to replace the Grain Futures Act as the authority over day-to-day regulation. Drafters of the Act included statutes that prohibited market price manipulation, created “position limits” in a bid to curtail the taking of big speculative positions, and established registration requirements for futures brokers, known as futures commission merchants (FCMs). The Act also initiated the financial requirement that customers place margins in trust with the FCMs.

The success of the regulatory effort continued to be mixed. The CEA began to regularly audit the FCMs, and found that oftentimes their clients’ investments were not being properly protected. At dozens of firms it was found that the positions held by clients did not match the funds that FCMs held in segregated accounts (where the accounts of each client was segregated from other FCM accounts). In one investigation the CEA looked at approximately 4500 discretionary accounts by financial advisors, finding that many were not properly executing their clients’ trades on time or were leaving unprofitable trades open so that clients would have a false impression of the value of their trading portfolios.

Even though the members of the futures exchanges had resisted passage of the Act, there is evidence they played an important role in furthering the government’s regulatory efforts once it was passed. For example, in August 1938 the heads of the leading commodity exchanges were asked to meet in Washington to consider what rules were needed to meet the discretionary account problems uncovered by the CEA. Thereafter, every contract market represented at the

---

55 See Markham, supra note 9, at 339-40.
56 Id.
58 Commodity Exchange Act § 4(c), 6(b).
60 Commodity Exchange Act § 4d(1).
61 Commodity Exchange Act § 4d(2).
62 See HISTORY OF COMMODITY FUTURES, supra note 1, at 30.
63 See id.
64 See id. at 31.
65 Id.
conference adopted amendments to their rules to prevent such practices.\textsuperscript{66}

In one of the better examples of futures exchanges making an earnest effort to police themselves, the CBOT pursued one of the country’s biggest and most powerful grain traders, Cargill Grain Co., for alleged manipulation of the market.\textsuperscript{67} In 1939 the CBOT required Cargill to liquidate part of a large position in 1937 September corn futures, and later expelled the company from membership on the exchange.\textsuperscript{68} Cargill fought the expulsion with the Commodity Exchange Commission, an overseeing body made up of representatives from the Department of Agriculture, the U.S. Attorney General and the Department of Commerce, on the basis that the board of trade had acted outside of its authority as defined in the Commodity Exchange Act.\textsuperscript{69} In 1940, the CEA dismissed the action, finding that the CBOT had sufficient reason to believe that Cargill’s was attempting to manipulate the market.\textsuperscript{70}

The CBOT’s actions led the Secretary of Agriculture to conduct its own probes of Cargill’s trading,\textsuperscript{71} which would lead to a series of charges against Cargill over the next several decades. In 1940 the Secretary of Agriculture brought charges against Cargill for engaging in “wash trades”\textsuperscript{72} and manipulating the price of corn, and issued a temporary ban on its trading privileges.\textsuperscript{73} Cargill was charged with manipulating

\textsuperscript{66} Id.
\textsuperscript{67} U.S. DEP’T OF AGRIC., REPORT OF THE CHIEF OF THE COMMODITY EXCHANGE ADMINISTRATION 39 (Sept. 25, 1939). At one time the Cargill Grain Co. held 80% of the total long open contracts. Id.
\textsuperscript{68} Id.
\textsuperscript{69} Id.
\textsuperscript{70} U.S. DEP’T OF AGRIC., REPORT OF THE CHIEF OF THE COMMODITY EXCHANGE ADMINISTRATION 44 (Aug. 31, 1940).
\textsuperscript{71} Id. at 44-45.
\textsuperscript{72} Wash trading, as defined by the Commodity Futures Trading Commission, is “entering into, or purporting to enter to, transactions to give the appearance that purchases and sales have been made, without incurring market risk or changing the trader's market position." Wash trading involving futures contracts is prohibited by the Commodity Exchange Act, \textit{available at} http://www.cftc.gov/opa/glossary/opaglossary_wxyz.htm. Wash trades are pre-arranged simultaneous trades entered into for the purpose of artificially inflating volumes or revenues or for the purpose of manipulating prices. C. Bryson Hull, \textit{Suit Says El Paso Engaged in Wash Trades}, \textit{Reuters Bus. News}, Nov. 21, 2002.

[T]he respondents entered into a stipulation with the complainant, admitting that one of the respondents, J.H. MacMillan, Jr., president of Cargill, Inc., directed and was responsible for the trading of the Cargill Grain Co. of Illinois, which executed transactions in grain futures as alleged in the
prices of oats futures in 1951 and 1952, and banned from trading in oats futures contracts. In 1971 a federal appeals court upheld a ruling that Cargill had manipulated wheat contracts on the CBOT eight years earlier, at one point holding 62% of the long interest in all contracts to be delivered in May of 1963.

Still, despite a few notable instances, such as the Cargill case, there was evidence that self-regulators were not effectively policing exchange floors. The CEA conducted infrequent investigations of trading floor practices, and one investigation showed that an estimated 10% of the trading volume came from wash trades (profitless round trades that can be used to inflate trading volume or falsely inflate revenues). In another instance, regulators conducted a probe in 1968 into job lot trading (splitting commodities trades into denominations of less than 5,000 bushels) on the CBOT. The investigation showed a lack of competition in trade execution, and as a result customers were paying a higher premium on their purchases and were being forced to sell at a greater discount. Following the investigation job lot trading on the CBOT was discontinued and complaints were issued against seven floor traders. Such discoveries put pressure on the CEA to perform more investigations, despite operating with limited resources.

Even so, these investigations did not substantiate one of the government's biggest concerns, that floor traders on the exchanges exercised their special advantages to profit at the expense of the trading public. In 1968 the General Accounting Office (GAO) released a study stating that floor traders had such advantages. In response, the Commodity Exchange

---

Id.

Id.

Id.
Administration (CEA) conducted a test study of potato futures contracts traded on the New York Mercantile Exchange (NYMEX). It found that day trading by floor traders represented only a small, relatively stable percentage of trading, and that short-term intra-day price movements resulted principally from trading by the general public, not trading by floor traders.\textsuperscript{81} It also found that two times as often as not, floor traders were trading against price movements, which indicated to the CEA that day traders actually helped to limit volatile price movements.\textsuperscript{82}

Congress approved amendments to the Commodity Exchange Act in 1968 to step up the CEA’s regulatory authority.\textsuperscript{83} The Agriculture Department was a driving force behind implementation of the changes, which helps to show an animosity that had developed between the government and the futures industry.\textsuperscript{84} The CEA’s power was broadened to include the establishment of minimum financial requirements for FCMs,\textsuperscript{85} livestock and concentrated orange juice futures contracts were brought under its jurisdiction,\textsuperscript{86} and criminal penalties (from misdemeanor to felony) for market violations were made more severe.\textsuperscript{87} However, one major amendment also had an unintended effect of weakening self-regulation. The CEA was given the power to disapprove rules implemented at the exchanges.\textsuperscript{88} This government involvement led courts to begin to view the exchanges’ rules as carrying a guarantee that members of the exchanges would not violate them.\textsuperscript{89} As a result, the amendments had the unintended effect of weakening self-regulation because exchanges began to reduce their regulatory schemes in order to minimize the potential liability from private litigation.\textsuperscript{90}

\textsuperscript{81} HISTORY OF COMMODITY FUTURES, supra note 1, at 51.
\textsuperscript{82} Id.
\textsuperscript{84} HISTORY OF COMMODITY FUTURES, supra note 1, at 52.
\textsuperscript{85} Commodity Exchange Act Sec. 4d.
\textsuperscript{87} 7 U.S.C. § 9(a) (2000).
\textsuperscript{89} See HISTORY OF COMMODITY FUTURES, supra note 1, at 62.
\textsuperscript{90} See id.
C. The Commodity Futures Trading Commission Act of 1974

By the early 1970s there was a growing consensus that futures markets were as important to the general public as securities markets were.

A series of market manipulation scandals and a sharp increase in trading volumes for both regulated and non-regulated commodities led to criticism that the CEA was not properly protecting small traders and the consuming public. At the time, the country was in the midst of the Cold War, and there was concern within Congress that the futures markets were susceptible to manipulation by companies from the USSR and other foreign nations. There was sentiment within Congress that regulation within the exchanges was lax, but of equal concern was that the CEA was not properly overseeing the exchanges. An internal report prepared for the inspector general of the Department of Agriculture found that the CEA relied on the exchanges too heavily to enforce its rules, and that self-regulation was insufficient. In 1973 a damaging article in the Des Moines Register stated that the CEA had turned the task of regulation over to the exchanges, and that the exchanges were run in a “club like atmosphere.”

Congress concluded it needed to establish a regulatory authority similar to the Securities and Exchange Commission, which was created by the 1934 Securities and Exchange Act. The Commodity Futures Trading Commission Act (CFTCA) of 1974 dramatically increased the government’s authority over futures exchanges. The most important provision of the Act

---

91 See id. at 56-57.
92 “In the early 1970s, Congress became concerned about the ‘Great Grain Robbery,’ that the Soviets were using the commodity futures markets as a means to manipulate prices and obtain large profits at the expense of American consumers.” Id. at 56. “The ‘Grain Robbery’ of 1972 was one of those economic events that, like the OPEC oil embargo . . . can truly be said to have changed the world.” Id. at 262 n.10 (quoting DAN MORGAN, MERCHANTS OF GRAIN 120-21 (1979)).
94 CEA relied on exchanges to enforce their rules and to insure that all trades were executed “competitively. Insufficient effort was made to determine whether trading rules were enforced. . . . [W]e also found other suspected violations of trading rules which we believe indicate a lack of control, detection and enforcement of rules governing the execution of customer orders.” Id.
96 Markham, supra note 9, at 341.
was the creation of the Commodity Futures Trading Commission (CFTC), an independent five-member regulatory commission. 98 The agency was given exclusive jurisdiction over the trading of futures and options on all commodities. 99 Just as in the securities industry, exchanges would continue to regulate themselves, though the CFTC would be overseer with broader authority and more resources. 100

But while the debate leading up to the CFTCA focused on how to increase the government’s authority, the reality of attaining that goal forced legislatures to concentrate on what is still a major consideration today: cost and efficiency. The result was actually to increase the responsibility of the exchanges, though under broader and stricter guidelines so that the government could be a more efficient overseer. 101 In a hearing before the Senate’s Committee on Agriculture and Forestry in May of 1974, Senator George McGovern from South Dakota argued that registered exchanges should be performing many of the routine checks and investigations that the government was currently responsible for, while the federal regulatory arm should have a heavier hand in making sure the exchanges enforce the government’s rules. 102 In one example of inefficiency, McGovern said that the CEA’s professional staff spent about 25% of its time performing routine audits of the hundreds of FCMs. It made more sense to place the primary responsibility for those audits with the exchanges themselves. 103 The size of the CEA’s staff, he pointed out, remained at 165 between 1970 and 1973, even though the volume of derivatives contracts traded had surged by 73% during the same period. 104 By McGovern’s estimation, the growth in the market was fast outpacing the CEA’s resources.

98 CFTC Act § 201.
99 Id.
100 The CFTC was given increased enforcement powers, the regulatory reach of the Commodity Exchange Act was expanded to include commodity trading advisors, commodity pool operators, and associated persons of futures commission merchants. Markham, supra note 9, at 341.
101 See Hearings Before the Subcomm. on Agric. and Forestry, 93d Cong. 199-200 (1974).
102 Id.
103 Id.
104 Id. at 199. In another example, McGovern said the exchanges should be responsible for regularly reviewing the need for position limits on specific commodities in order to head off attempts to manipulate the market; be given a time limit for implementing and enforcing trading rules; and conduct investigations to seek out abusive practices. As futures trading escalated, the CEA simply could not handle these sorts of responsibilities. Id.
Exchange members who had opposed increased government authority over rulemaking argued that the 1974 Act would destroy the exchanges’ free market traits. In the same senate committee hearing, Carlos Bradley, President of the Board of Trade of Kansas City, Mo., said that “the effect of such a proposal could be to abdicate completely to the government the responsibility of the exchanges, the expertise of their members, their public responsibility etc. There would be no need for exchange governments.”

Those arguments, however, were generally lost in Congress, where there remained strong support to increase government oversight.

Under the CFTCA, the CFTC’s authority was expanded from the statutory list of physical commodities to include futures contracts on all goods, articles, services and rights and interests—thereby defining the term commodity to include anything on which a contract is traded (except onions). The CFTC was given power to grant reparations to any person hurt by a violation of the Act, and commodity trading advisers, commodity pool operators (enterprises who solicited and received funds from others in order to trade in commodity futures) and associated persons (employees of FCMs who solicited or accepted customer orders or supervised such persons) were now required to be registered with the agency.

Also, the CFTC now had a direct involvement in the regulation by the exchanges. Exchanges were required to submit proposals for new rules pertaining to futures contracts or trading requirements to the CFTC for advance approval. The CFTC was given long desired authority to intervene in the trading of contract exchanges when it deemed market disturbances to be emergencies. It was also given injunctive authority, a power that the Department of Agriculture had sought for years. The CFTCA increased the potential

---

105 Id. at 225.
106 CFTC Act § 201 (1974). Onion farmers successfully lobbied to have onions withheld from futures trading after onion futures disastrously plummeted in 1955. The lobbying effort resulted in the so-called Onion Futures Act, passed by Congress in 1958. David S. Jacks, Populists v. Theorists: Futures Markets and the Volatility of Prices, at 14, at http://aghistory.ucdavis.edu/jackspaper04.pdf. Onions are the only commodity to be banned from futures trading. Id. The law was upheld as constitutional in 1959. Id. Subsequent studies have shown that onion prices have been volatile when futures trading was permitted than after the law was passed. Id.
107 CFTC Act § 209.
108 Id. at § 4k.
109 Id. at § 210.
110 Id. at § 215.
111 Id. at §6c.
penalties for market manipulation to $100,000 from $10,000, and provided the framework for the creation of a quasi-public regulator that would, similar to the National Association of Securities Dealers in the securities market, govern the conduct of market participants who are not members of contracts markets.

The CFTCA encouraged exchanges to enforce its rules under the CFTC’s oversight by explicitly authorizing them to discipline their members. The exchanges were required to report any such actions to the CFTC, which could in turn affirm, modify them or set the disciplinary actions aside.

Two years later, in 1976, the CFTC commenced one of its most significant regulatory reforms. The liquidity requirements of FCMs were increased in a move designed to ensure that their financial position could not deteriorate to the point of endangering customer funds. The requirement, while imposing substantial costs on the system, was designed to insure the financial integrity of the system. To defray the cost of the audits from the government, the CFTC required FCMs to be audited by independent financial accountants. Other provisions were included in the Bankruptcy Act of 1978 to further protect owners of the futures contracts from financial failure of a FCM. The capital requirements would set in motion an important trend of consolidating trading into the hands of the most financially sound FCMs, which in turn would allow both government and self-regulators to focus on issues like trading violations and market manipulation.

---

112 CFTC Act § 6B.
113 Id. at §17.
114 Id. at § 8C.
115 Id.
116 See History of Commodity Futures, supra note 1, at 86. For amended statute see 7 U.S.C.A. § 6(b) (2002).
117 See History of Commodity Futures, supra note 1, at 87. The certification requirement, coupled with early warning requirements, is also designed to ensure that the CFTC is on early notice when a firm is in financial trouble, so steps can be undertaken to prevent or limit injury to customers. Id.
118 Id. at 86.
119 Id. at 87.

The primary protection for investors in commodity futures contracts is that their funds are kept in segregated trust accounts. It often happens, however, that a breach of such trust may occur—the broker could convert the funds, either to trade for its own account or to meet the margin calls of another customer. In such instances, the new Bankruptcy Code provisions provide for the equal sharing of all customers in any remaining segregated funds. . . . These customers have a priority over all other creditors in such funds.

Id.
In the ensuing years, regulation of futures exchanges was subject to harsh criticisms, but oftentimes it was directed at failures of the CFTC rather than at the exchanges. In addition to criticism of the CFTC’s handling of a series of options scandals in the late 1970s, a report prepared for Congressman James Whitten of Mississippi by the Surveys and Investigations Staff of the House Committee on Appropriations showed that the CFTC was in many respects incompetent. The report found that the CFTC employed a disproportionately large number of political appointees and that there had been a flagrant misuse of funds, ranging from the use of outside consultants in order to avoid civil service hiring restrictions, to overpayment for parking spaces and driving services for CFTC employees. Abuses were also found in the awarding of government contracts. For example, one former employee in the chairman’s office was given two consulting contracts, which if aggregated, would have exceeded a threshold amount that would have triggered a requirement to open the contract up to a competitive bidding process.

D. The Commodity Futures Modernization Act

After six decades of strengthening its oversight, government regulators began to contemplate changing tacks in the late 1990s. The CFTC had determined that the financial underpinnings of the futures industry had been strengthened enough that it would now be more efficient to loosen its grip on self-regulators. Trading volumes had surged and technological advances were giving way to growth of electronic

---

120 Id. at 94-96. (detailing in particular the attempt by brothers Bunker and Herbert Hunt of Dallas, Texas to corner the soybean market in 1977, at times owning the right to take delivery of over one-third of the U.S. carryover inventory for old crop soybeans covered by the Chicago Board of Trade contracts).

121 HOUSE COMM. ON APPROPRIATIONS, 95th Cong., INVESTIGATIVE STUDY ON THE COMMODITY FUTURES TRADING COMM. (Comm. Print 1978) (by Mr. Whitten for use by the Subcomm. on Agric., Rural Development and Related Agencies).

122 See id. at 51, 64-66, 72. “While the schism between the former CEA employees and the newly hired CFTC group may have contributed to some of the early organizational problems of the Commission, the [Subcommittee’s] Investigative Staff was advised that of even greater significance was the rivalries between the new senior-level staff appointees.” Id. at 23.

123 Id. at 64.

exchanges that could operate across borders. The industry argued that the current regulatory system had bogged it down to the point that it may not be able to compete effectively with foreign exchanges. Under heavy pressure from the industry, the CFTC approved a series of amendments called the Commodity Futures Modernization Act of 2000. The aim of the CFMA was to make the regulatory framework more flexible. Exchanges were now allowed to create a regulatory framework from sets of “core principles” that fit their particular operation, rather than have to adhere to a one size fits all system. The framework also created three regulatory tiers for markets, with a lower level of regulatory oversight where access to the exchange or trading facility is limited to commercial participants and the nature of the underlying commodities provides a low risk of manipulation.

Unlike the fiery debates that preceded prior regulatory legislation, there appeared to be little contention between the industry and government regarding the CFMA. In a March, 2000 hearing before the Senate Subcommittee on Research, Nutrition and General Legislation, of the Committee on Agriculture, Nutrition and Forestry, Senator Peter Fitzgerald, the chairman, opened by saying that “the CFTC has suggested that it is willing to grant broad regulatory relief to futures exchanges.” C. Robert Paul, General Counsel of the CFTC, followed by saying that one of the CFTC’s main policy goals should be “removing any regulatory barriers that hamper these markets from fully exploiting innovations in technology.”

The major futures exchanges chimed in. David Brennan, chairman of the CBOT, told the subcommittee that it “heartily endorses the concept of replacing inflexible, micromanaging

---

125 See Hearings Before the Subcomm. on General Farm Commodities and Risk of the House Comm. on Agriculture, 108th Cong. 16 (2003) (statement of James Newsome, Chairman of the Commodity Futures Trading Commission, in response to question by the subcommittee).
126 Id. Committee member and Representative from Kansas Jerry Moran said in a review of the CFMA two years after it was approved: “[W]e heard continually from the exchanges about the potential threat if we didn’t appropriately deregulate the industry, the threat from competition, foreign sources, from the ability of customers to utilize exchange service offshore.” Id.
129 Id.
130 Id.
131 Hearing Before S. Subcomm. on Research, Nutrition and General Legis., supra note 124, at 1.
132 Id. at 2.
government mandates with core principles. The CFTC is right that exchanges and others are best able to design systems to achieve the desired and shared objectives of market integrity, financial integrity and preventing abuses.132

At a hearing to discuss the CFMA two and a half years after it was approved, congressional members seemed unable to assess whether the new system had been an improvement or not. The CFTC defended the CFMA by saying trade volume in futures has increased by 50% since its passage, and that as it had predicted, the industry has become substantially more competitive.133

But in the interim there had also been a series of market scandals, including the collapse of a major energy trader, Enron Corp., and several congressmen raised concerns about whether the CFTC was properly able to monitor the energy markets.134 CFTC Chairman James Newsome also conceded that foreign exchanges had not followed suit with similar moves to liberalize their regulatory regimes.135

III. DEMUTUALIZATION AND MODERN TRENDS

A. Trend of For-Profit Exchanges

The trend by the world’s largest exchanges to demutualize has pushed the debate about self-regulation back to the forefront.136 Demutualization in this context refers to the conversion of non-profit, membership-owned organizations into for-profit stock corporations. Demutualization among futures and securities exchanges has been driven by forces in the business environment, including advances in technology, globalization of markets, a concentration of investment capital, competitive pricing pressure and government deregulation.137

132 Id. at 64.
133 Hearing Before House Subcomm. on General Farm Commodities, supra note 125, at 16 (statement by James Newsome, Chairman of the Commodity Futures Trading Commission). Newsome said that during the prior one and a half years, the CFTC, which usually handles about 100 investigations concurrently, had conducted roughly 30 investigations in the energy sector. Id. at 27.
134 Id. at 22.
135 Id. at 16-17.
136 See generally Karmel, supra note 1.
137 Id. at 368.

A dramatic shift in the economic and power structure of the securities industry is currently in progress. Although competition to traditional markets from electronic trading markets may be the precipitating cause of this upheaval, more than technology is driving these changes. The worldwide
By demutualizing, management at the exchanges hope to be able to raise larger pools of money, which in turn would allow them to invest more in technology and grow their businesses. Proponents also believe that demutualization will create a fairer marketplace. The members who have traditionally run the exchanges have been driven by the profits earned from their own trading. The demutualized exchanges would in theory be directed by shareholders and experienced management teams who are more focused on the bottom line, which in a competitive environment would mean there would be pressure to provide the best possible services.

There are concerns that demutualization creates new conflicts of interest. One concern is that an inherent conflict exists between the interests of the shareholders and the market users. It has been suggested, for example, that a for-profit exchange would not rigorously undertake self-regulatory obligations if those obligations negatively affect profitability. The issue is whether a commercial entity that is running an exchange and seeking to protect and promote its business can also support the integrity and efficiency of the trading markets by setting and enforcing regulations that are in the public interest.

One must enter this debate with the understanding that the current structure has its own conflicts. Even traditional, not for profit exchanges are run by members interested in

---

rise in stock exchanges trading volume, global integration of the capital markets and competition for trading profits have triggered a disintermediation comparable to the unfixing of commission rates. Decimalization has cut the conventional trading increment, formerly twelve and a half cents, to a penny or less. Futures exchanges similarly have been buffeted by technological change, global competition and resulting cost pressures.

Id.


139 See Erickson, supra note 2, at 5.

140 Id at 5-6.

141 Karmel, supra note 1, at 420. But see Erickson, supra note 2, at 6 n.9: [I]n a briefing of staff for the Senate Agriculture Committee, Robert Colby, Deputy Director of the SEC’s Division of Market Regulation, raised the converse argument. Would a demutualized exchange have a perverse incentive to impose hefty fines for violative activity, thereby creating a profit center for the exchange? While this view might be a short-term profit maximization strategy, an exchange employing such a practice would quickly lose business in a competitive market where similar products are traded on many exchanges.

Id.
making money and enhancing value through trading and seat value.\textsuperscript{142} Furthermore, the current structure of exchange disciplinary programs, where members sanction fellow members, could arguably affect the rigor of an exchange’s self-regulatory program. Therefore, to some extent, questioning the efficacy of a for-profit structure in fulfilling self-regulatory obligations is also questioning the current exchange structure. According to a CFTC report on demutualization that was published by CFTC Commissioner Thomas Erickson in 2000, even if new conflicts arise under the for-profit context, exchanges would continue to have a self-interest in preserving their reputations for providing fair and efficient markets.\textsuperscript{143}

Exchanges like the CME, the New York Stock Exchange and the New York Mercantile Exchange aggressively market their records of regulatory enforcement to attract new business, and as more competitors enter the market place, the reputations of these exchanges will pay a heavier price when their regulatory systems fail.\textsuperscript{144} During a Senate hearing in 2000, Thomas Donovan, the chief executive of the CBOT, told legislators that “the CFTC strictly should be an oversight agency, one that provides the flexibility for us to use our self—the regulatory structure as a marketing tool for people to want to come and trade . . . .”\textsuperscript{145} At the same hearing, James McNulty, the CEO of the CME, concurred, saying that the exchange has built “a highly disciplined self-regulatory body in the CME, and we think that is one of the reasons people come to work with our exchange.”\textsuperscript{146}

B. Initial Reforms: Exchanges and Regulation Bodies React

One issue that exchanges have been forced to address is the independence of boards of directors. A scandal over governance at the New York Stock Exchange led regulators to revisit issues surrounding exchanges’ governance standards.\textsuperscript{147} In September of 2003, NYSE Chairman and CEO Richard

\begin{footnotes}
\footnotetext[142]{Erickson, supra note 2, at 6.}
\footnotetext[143]{Id.}
\footnotetext[144]{Id.}
\footnotetext[145]{Hearing Before S. Subcomm. on Research, Nutrition and General Legis., supra note 124, at 20.}
\footnotetext[146]{Id. In the securities industry, the NYSE has a regulatory staff of 550 people and a budget of $142 million annually. The NASD has a regulatory staff of 2100 people and an annual budget of $500 million. See Cohen & Kelly, supra note 2.}
\footnotetext[147]{See, e.g., Cohen & Kelly, supra note 2.}
\end{footnotes}
Grasso was pressured to resign after it was made public that he was entitled to close to $140 million in compensation and deferred retirement benefits in 2003\(^\text{148}\) (the NYSE as a whole made less than $28.1 million in profits in 2002).\(^\text{149}\) Much of the focus since has centered on the structure of the board, and claims that it did not receive enough information to gain a sufficient understanding of the pay package it was approving.\(^\text{150}\)

A review of the NYSE’s approach is relevant to the study of futures exchanges because the futures industry has historically followed the lead of the securities industry.\(^\text{151}\) Grasso, a 35-year veteran at the NYSE, had been a respected figure in the debate over demutualization for both securities and futures exchanges. In a hearing before the Senate Banking Committee in September of 1999, Grasso said that the NYSE would need to demutualize, and possibly go public, in order to fend off competition from “electronic communications networks,” commercially-owned electronic trading systems known as ECNs.\(^\text{152}\) Grasso argued that ECNs are not subject to cumbersome self-regulatory requirements, and are often owned by wealthy corporations that are willing to invest money to expand and enhance their businesses.\(^\text{153}\) Demutualization would cause the members’ interests to align with the success of the exchange as a whole, as opposed to being skewed toward the success of only the floor trading operations.\(^\text{154}\) Under a for-profit structure the NYSE would also be able to raise money by

\(^{148}\) Jake Keaveny & Brendan Intindola, NYSE Chairman Grasso Resigns Under Pressure, REUTERS NEWS, Sept. 17, 2003. It was later determined that Grasso’s total compensation, including compensation and deferred benefits, was closer to $190 million, when including an additional $50 million he claimed to be owed. Ben White, Former NYSE Chairman Grasso Sued Over Pay, Washington Post, May 25, 2004.


\(^{150}\) Andrew Countryman, NYSE Chief Proposes Independent Board, CHI. TRIB., Nov. 6, 2003, at 1. New York Attorney General Eliot Spitzer, whose office has jurisdiction over not-for-profit corporations like the NYSE, said his office “will probe the role of NYSE board members and others that helped create” Grasso’s compensation package.” Under New York not-for-profit laws directors can only approve compensation packages to executives that are commensurate with the benefits that the executives provide to the corporation. Jake Keaveny and Mark McSherry, Spitzer Says Troubled by Report on Grasso Pay, REUTERS NEWS, Jan. 13, 2004.

\(^{151}\) See Karmel, supra note 1, at 402 (“To a large extent the CFTC is an analogue to the SEC . . . .”).

\(^{152}\) PUBLIC OWNERSHIP OF THE U.S. STOCK MARKETS: HEARING BEFORE THE SENATE COMM. ON BANKING, HOUSING AND URBAN AFFAIRS, 106th Cong. 3-4 (Sept. 28, 1999) (statement of Richard A. Grasso, CEO, New York Stock Exchange) [herinafter Grasso, Hearing].

\(^{153}\) Id.

\(^{154}\) Id.
sellers stock, either publicly or to private investors. Grasso also argued that greater competition in the marketplace would strengthen the NYSE’s commitment to regulation. At the time he made that statement, the exchange had no intentions of altering the compensation of its board—then made up of 50% industry representatives and 50% public directors unaffiliated with its members—as a way to eliminate conflicts.

The public outcry surrounding Grasso’s ouster has since led the 211-year old exchange to dramatically alter its course. John Reed, the former co-CEO of Citigroup Inc. who was brought in as interim NYSE chairman, orchestrated a series of reforms. NYSE members approved a plan that the board be cut down to 8 members, less than a third of its present size, and not include any representatives of the financial firms that are members of the exchange. Under the new structure, the board is responsible for such issues as compensation, independent audits, and self-regulation, while a separate advisory committee that would include member firms would be created to help oversee issues that are strategic to the exchange’s business. The Securities and Exchange Commission has approved the proposal.

Such corporate governance initiatives are implemented on securities markets sooner than on futures markets because of the public nature of the companies that are listed on them. The latest board proposal at the NYSE is an extension to a similar shake-up some 31 years earlier. In 1972, significant changes were made to the NYSE constitution after release of the Martin Report, a congressionally commissioned study that

155 Id.
156 Id. at 5.
158 See Greg Farrell, Reed Proposes Cutting NYSE Board to Eight, USA TODAY, Nov. 6, 2003. Several State treasurers and large public pension officials, some of whom publicly pushed for Grasso’s ouster, felt that Reed’s plan was just more of the status quo. Sean Harrigan, president of the nation’s largest public pension fund, the $145 billion California Public Employee’s Retirement System, said “Investors were expecting a home run proposal to reform the New York Stock Exchange. What we got . . . is not even a base hit.” Arden Dale, Pension Funds See Flaws in NYSE Reform Plan, DOW JONES NEWSWIRES, Nov. 5, 2003.
159 See Plitch, supra note 157.
was critical of the exchange. The report had recommended that the NYSE reduce the number of board seats to twenty-one from thirty-three, and that the number of members representing the public be increased to ten from three. Prior to the reforms half of the NYSE’s board had been composed of public directors.

Traditionally, outside directors representing the public had only a token representation on futures exchanges. In 1989, the Federal Bureau of Investigation ran an undercover sting at the CME and CBOT that resulted in the indictment of forty-eight individuals for various trading practice violations on commodity exchange floors. The controversy surrounding the arrests led Congress to amend the Commodity Futures Act in 1992, including a provision which had previously failed that required at least 20% of the regular voting members of the exchanges’ boards be independent, non-member directors. Other provisions required that a diversity of interests be represented by including the principal groups of the commodities being traded, floor brokers and at least 10% from a group that included farmers, merchants, and exporters.

161 See Karmel, supra note 1, at 405.
162 Id. at 405-06.
These changes occurred in the context of uncertainty about the immunity of stock exchanges form the anti-trust laws, pressures to unfix commission rates and the financial and operational back office crisis of the securities industry. These developments ultimately led to the enactment of the 1975 Act that restructured the regulatory relationship between the SEC and SROs and stripped stock exchanges of some of their former autonomy. The Martin Report was intended to compel the NYSE to discard what vestiges of a private club atmosphere then remained and to become a quasi-public organization.
164 See Karmel, supra note 1, at 408 (citing JERRY W. MARKHAM, COMMODITIES REGULATION: FRAUD, MANIPULATION AND OTHER CLAIMS § 14.10 (1998)).
165 See id.
These criminal indictments were upheld, although the trials had mixed results. In response to the sting operations Congress passed legislation to strengthen trading in the pits . . . audit trails were strengthened, there was increased regulation of floor broker associations, and more outsiders were required to be included on exchange boards and disciplinary committees.
166 Karmel, supra note 1, at 409.
A major futures exchange to come under scrutiny following the NYSE scandal was the CME.\(^{167}\) In 2000 the CME was the first exchange to demutualize\(^{168}\) (preceding plans by the CBOT and the New York Mercantile Exchange),\(^{169}\) and in 2002 it became the only major U.S. exchange to go public.\(^{170}\) In an October, 2003 article, *Business Week* reporter Joseph Weber questioned the independence of the CME’s board, and said that the CFTC is scrutinizing its corporate governance policies.\(^{171}\) At the time only four of the 105-year old exchange’s twenty directors do not have ties to the CME or its trading floor, while fifteen were long time exchange members.\(^{172}\)

In November, 2003 the CME announced that it planned to make a number of changes to its Board that would enhance the independent oversight of key corporate governance issues.\(^{173}\) As part of the plan, the CME would create a new board level committee in 2004, comprised solely of independent, non-member directors.\(^{174}\) The committee would conduct an annual review of issues that include the independence of the CME’s regulatory functions from its business operations; the CME’s compliance with its statutory self-regulatory responsibilities; the funding of the CME’s self-regulatory responsibilities; and

---


\(^{168}\) *Chicago Mercantile Exchange to Dual List on Nasdaq Stock Market*, NASDAQ PRESS RELEASE, April 27, 2005, at http://www.nasdaq.com/newsroom/news/pr2005/ne_section05_047.stm. See Erickson, *supra* note 2, at 4. At the time that the CME demutualized is set in motion a two year plan to reduce its existing 39-member board to 19 members, with the exchange run by a chief executive officer hired by the board. *Id.*


\(^{170}\) *Chicago Merc Nets About $117.8 Million in IPO*, REUTERS NEWS, Dec. 12, 2002.

\(^{171}\) Weber, *supra* note 163 (“The Merc board—including its compensation committee—remains controlled by traders and floor brokers who are regulated by the exchange. Indeed, the Merc [CME] seems rife with the same conflicts of interest that tarnished the Big Board [NYSE] before CEO Richard A. Grasso self destructed.”).

\(^{172}\) *Id.*


Market-structure experts were quick to say the changes . . . are mild compared with the plan put out by Big Board [NYSE] interim Chairman John Reed. . . . Still, the plan will have some immediate results . . . trader William R. Shepard, longtime chairman of the Merc’s compensation committee, will leave that position. Also, the plan, which came out of the governance committee now led by trader Jack Sandner, will require him to give up that position.

*Id.*

\(^{174}\) *Id.*
the compensation of exchange employees involved in regulatory activities.\textsuperscript{175}

By making its boards more independent, the NYSE and the CME hope to preserve its regulatory roles from encroachment by the government. Critics have suggested that self-regulatory bodies should be completely separate from the exchanges’ business operations.\textsuperscript{176} The NYSE has publicly opposed suggestions that it should either spin its regulatory unit off into a subsidiary, or that the industry should move towards a single self-regulatory organization (SRO) for the entire securities industry, such as the National Association of Securities Dealers, a NYSE competitor.\textsuperscript{177} In the futures industry, the National Futures Association (NFA) has aggressively marketed itself as a third party provider of regulatory services to other exchanges, though with limited success.\textsuperscript{178}

Regulatory authorities have been a step slower, but are moving ahead with comprehensive groups of proposals. The SEC, which saw nearly all of the major scandals of the last several years come under its watch, has opened a series of governance and regulatory related proposals for public comment.\textsuperscript{179} Some of the proposals mimic those made by securities industry. The SEC would call for securities exchanges and registered securities associations to require a majority of board members to be independent, and certain board committees would be required to be composed solely of independent directors.\textsuperscript{180} There would also be requirements that a separation be maintained between regulatory functions and

\textsuperscript{175} Id.
\textsuperscript{176} See Karmel, supra note 1, at 424-26.
\textsuperscript{177} Grasso, Hearing, supra at note 152, at 3-4. “John Reed, the interim NYSE chairman brought in after the Grasso uproar, is seeking to preserve self-regulation.” Cohen & Kelly, supra note 2.
\textsuperscript{178} See Hearing Before Senate Subcomm. on Research, Nutrition and General Legislation, supra note 124, 11 (statement of Robert Wilmouth, Chief Executive, NFA). From 1977 to 1999 there were no new futures exchanges formed. In the last six months, at least six different enterprises have stated their interest in creating new electronic futures exchanges. All of them are dedicated to using effective self-regulation . . . but none . . . are really shackled by the past. Everyone is looking for more efficient ways to perform their self-regulatory functions, and everyone has contracted NFA to discuss outsourcing that function to us.
\textsuperscript{179} Id. at 12.
\textsuperscript{180} Fair Administration and Governance of Self-Regulatory Organizations, 70 Fed. Reg. 11 (Proposed Jan. 8, 2005) (to be codified at 17 C.F.R. at pt. 240, 242, and 249) [hereinafter Fair Administration and Governance].
\textsuperscript{180} Id.
Further, the proposals would prohibit members that are brokers or dealers from owning or voting more than 20% of the ownership interest in the exchange. Among other proposals, the SEC would require exchanges to maintain their books and records in the U.S. and add reporting requirements by the exchanges to enhance transparency. Exchanges that go public, and whose shares trade on their own exchange, would have a separate group of requirements to help supervise trading and enforce listing standards.

For its part, CFTC Chairman James Newsome opened a review of self-regulation in May of 2003. The review revolves around an analysis of regulation under the Commodity Futures Modernization Act. As of April, 2003, it had opened a formal request for commentary from futures market participants that includes: board composition, regulatory structure, forms of ownership of the exchanges, the structure of disciplinary committees, and other issues.

C. Potential Alternatives: the NFA model and Market Competition

The nature of the SRO structure makes the potential for conflicts almost inherent. Under the 1934 Commodities Exchange Act, SROs are required to act as quasi-governmental bodies in implementing federal laws as their own. Yet SROs are also membership organizations that represent the economic interests of their partners. In addition, SROs are marketplaces concerned with preserving and enhancing their competitive positions. As competition grows among marketplaces and SROs, it seems that the relationship that SROs have with their members inevitably will strain the SROs’ ability to carry out their regulatory duties impartially.

---

181 Id.
182 Id.
183 Id.
184 Fair Administration and Governance, supra note 179.
186 Id.
187 Id.
188 See White Paper, supra note 6, at 8-9.
189 See id. at 9.
190 See id.
In a recently updated study, the Securities Industry Association (SIA), an industry group, evaluated six potential models for self-regulation.\footnote{See generally id.} In evaluating the models the SIA echoed the original argument that the SRO model puts regulatory decisions in the hands of people familiar with the relevant facts, and that any regulatory change should not abandon the system in favor of a distant, generalist regulator that is not as deeply familiar with the markets it regulates.\footnote{Id. at 6.}

The study outlined a series of possible alternative structures that include the splitting off of SRO functions into subsidiaries of the exchanges; creation of a single SRO to audit and monitor all broker dealers; or putting all of the regulatory responsibility into the hands of the SEC (or CFTC).\footnote{See WHITE PAPER, supra note 6, at 2-12.}

Proponents of a single SRO system argue that the transition could be made relatively easily in the futures industry because the NFA has already been sanctioned by, and works closely with, the CFTC.\footnote{Id.} The NFA, which is similar to the National Association of Securities Dealers in the securities industry, began as a regulator for trading participants that were not registered with exchanges, but has moved into the business of outsourcing its regulation services. The NFA can avoid some of the criticisms arising from the demutualization of SROs because it is a non-profit organization.\footnote{See Board of Directors at the National Futures Association website, at http://www.nfa.futures.org/aboutnfa/board.asp.} Its board is made up of representatives from the industry, though it is not controlled by any one entity.\footnote{See Natasha de Teran, Eurex Signs Up National Futures Association for US Exchange, FIN. NEWS, Nov. 6, 2003. Rudolf Ferscha, Eurex CEO, said of an agreement to farm out is regulatory services  to NFA: “We have agreed on initial plans of a three-year contract, with automatic one-year renewal contracts thereafter. The NFA will ensure that Eurex US's customers are protected at all times, and that business will be fair, orderly and transparent.” Id.}

The economic inefficiencies that come with operating multiple SROs could also help to promote a single SRO system. The NFA has marketed itself as an outsourcing facility for self-regulatory functions.\footnote{See Hearing Before Senate Subcomm. on Research, Nutrition and General Legislature, supra note 124, at 11 (statement of Robert Wilmouth, Chief Executive, NFA).} By regulating numerous exchanges, the NRA (or another comparable outsourcer to enter the market)
would benefit from efficiencies of scale. Several newer exchanges have contracted the NFA in hopes of finding a cost effective means to regulate their trading operations. If pressure on expenditures continues then traditional exchanges like the CBOT and the CME could be pressured to reduce regulatory expenditures, which in turn could diminish the quality of regulation. Traditional exchanges are already being squeezed by the trend towards lower trading fees. However, the NFA’s outsourcing model is not free of potential conflicts either. As the practice becomes more prevalent, issues could be raised regarding the exchange’s continuing responsibility over its contractor and the relationship of both entities to the overseeing government regulator.

The NFA entered into an agreement with Merchants Exchange and BrokerTech, two small U.S. futures exchanges, to perform market surveillance, conduct background checks, investigate and litigate disciplinary matters, and perform audits and financial surveillance. This led to a break through deal that could give the NFA new credibility as a third party regulator. In November of 2003, it signed a three-year contract with Eurex US, the U.S. arm of Eurex, to be its regulator. Frankfurt-based Eurex, which in recent years surpassed the CBOT as the world’s largest futures exchange in terms of volume traded, received approval from the CFTC to set up a U.S. exchange in 2004.

Detractors of the single SRO model say that it would weaken self-regulation. Broker-dealer regulation has its roots in efforts to assure creditworthiness of exchange members. In that regard, the big exchanges have spent hundreds of millions of dollars setting up self-regulatory systems to stand behind assurances that large member firms are financially viable.
single SRO based in a separate location from the exchange and indirectly run by the CFTC or SEC may never be able to attain the same level of intimacy with a particular exchange. Not surprisingly, major exchanges like the NYSE, the CME and the CBOT have sought to preserve their regulatory duties by publicly opposing a single SRO model.\footnote{206}

Ultimately, competition in the market place may dictate the future model. If choice and competition are important in how financial products are offered, perhaps there are analogous benefits in terms of how exchanges are self-regulated.\footnote{207} Just as the NFA has offered its services as third party regulator, the CME or CBOT also may decide to compete by also farming out their regulatory services.\footnote{208} Thomas Erickson, the former CFTC commissioner, indicated there is talk that the CME has studied such an initiative. The NASD has taken similar initiatives in the securities business, and could also work toward offering a similar service to futures exchanges, according to Erickson.\footnote{209} Such a scenario opens up the specter of further conflicts. For instance, the CME could create a structure where it acted as regulator to itself and a competitor. A third party SRO could be acquired by a large financial firm like Citigroup Inc. or J.P. Morgan Chase & Co., which also operate FCMs. Will the quality of self-regulation diminish amid the pressure to cut costs?\footnote{210}

Competitive forces among futures exchanges may also give way to new incentives to uphold regulatory standards.\footnote{211}

\footnote{206} See House Subcomm. Hearing on General Farm Commodities and Risk Management, supra note 125, at 61. (statement of Terrence A. Duffy, Chairman Chicago Mercantile Exchange) (“Rather than detracting from our ability as a self-regulator, the CME’s incentives and capability to maintain an effective program of self-regulation have been enhanced by its reorganization as a for-profit company.”).

\footnote{207} See Erickson, supra note 2, at 8. Moreover, to require exchanges to contract with a super self-regulator would appear to be inconsistent with the CFTC’s role as an oversight agency, especially if regulatory concerns do not outweigh the benefits of the current SRO structure. Id.

\footnote{208} Telephone interview with Thomas Erickson, former CFTC Commissioner and Vice-President, Bunge Ltd., Washington, D.C. (Nov. 7, 2003).

\footnote{209} Id. The NASD has been pushing for a single SRO model and the former Chairman of the SEC, Arthur Levitt, briefly embraced this model. See Jeffrey E. Garten, Manager’s Journal: How to Keep NYSE’s Stock High, WALL ST. J. Sept. 13, 1999, at A44. The main barrier has been that the NYSE opposes the single SRO model. Grasso, Hearing, supra note 152.

\footnote{210} Telephone interview with Thomas Erickson, former CFTC Commissioner and Vice-President, Bunge Ltd., Washington, D.C. (Nov. 7, 2003).

\footnote{211} Karmel, supra note 1, at 370.

An interesting and relevant question is whether current trading technologies and the competition these technologies have engendered should lead to a reduction of SEC market regulation, rather than increase in regulation.
Currently, exchanges like the CBOT and CME say that the integrity of their marketplace is the biggest incentive to uphold such standards. But that argument has weaknesses because for many futures products those exchanges offer the only liquid markets for particular contracts, and thus they enjoy de facto monopolies. If, for instance, the Eurex’s U.S. exchange gains market share, and also offers a more transparent, and fair trading operation, then regulation issues could become more prevalent in traders’ decisions over where they do business. In at least one instance, competitive pressures have already had an impact. In 2003 the SEC issued a report that said the American Stock Exchange had massive shortcomings in its regulation of options trading and that it had attempted to cover up its deficiencies. In October of 2003 board members from the NASD, which holds a majority stake in the American Stock Exchange, and the American Stock Exchange voted to have the NASD take over its self-regulation. In making the decision, the directors at both boards considered the American Stock Exchange’s poor performance in the regulatory arena and the high cost of regulation.

IV. CONCLUSION

Critics have questioned the SROs’ ability to maintain fair and transparent trading, but as this Note has demonstrated, failures in government oversight have also been pervasive. It seems unlikely that the government could operate more efficiently as a sole regulator. If pure government

envisioned by current SEC concept and rulemaking releases, so that competition rather than regulation can determine outcomes.

Id. at 369-70.

Stephen Craig Pirrong, The Self-Regulation of Commodity Exchanges: The Case of Market Manipulation, 38 J.L. & ECON. 141, 154 (1995) (“As the advocates of self-regulation suggest, competition from other exchanges could mitigate, and perhaps eliminate, these problems. Unfortunately, it is by no means clear that competition between exchanges in a particular contract is especially acute.”).


Id.

Markham, supra note 9, at 405. [W]e must be careful of what we wish for in life. A single regulator may also seek to expand its powers after a scandal. A single regulator will also undoubtedly use bad judgment in times of crisis. A single regulator could also stifle competition, over-regulate, and cause a loss of competitive position in
regulation were a preferable alternative, the system of self-regulation would have been scrapped years ago amid the government’s many trials and errors. The SRO system is preferable to a pure government regulatory scheme because it defrays much of the costs onto the market users, and makes efficient use of the expertise at the exchanges.

The role and duties of SROs vis a vis government regulation has steadily evolved since adoption of the Commodity Exchange Act and the Securities Exchange Act, with the government approving legislation throughout the years to increase its oversight authority and impose stricter standards. While the government initially sought to micro-manage regulation at the exchanges, it more recently determined that giving SROs more autonomy is more effective.

Thomas Erickson, who spoke regularly about the future of SROs during his tenure as a CFTC Commissioner, said in an interview that from a political standpoint the SRO system has too much history to ever be scrapped altogether. There will be conflicts of interest in any system, and so the challenge is to create a system with the right oversight and incentives so that as many conflicts as possible are eliminated. In that regard, self-regulation is a preferable system to pure government regulation, and the question becomes one of balance.

The CFTC should take several steps to eliminate the specter of conflicts of interest within exchanges. It should implement a model for independent boards similar to that adopted by the NYSE, or the rules that the SEC has proposed. These rules would ensure that exchange members are separated from regulation related decisions. Secondly, the CFTC should facilitate competition in the marketplace to the extent possible. For example, the Eurex’s entrance into the U.S. could spark the beginning of a period of a competition driven regulation market, where market participants themselves go far in determining what the most effective SROs,

international markets. It could even try to become a Japanese MoF [Ministry of Finance] that seeks to manage the economy by bureaucratic fiat.

Id.  
Karmel, supra note 1, at 401. The 1975 Act sought to preserve and reinforce the concept of industry self-regulation. Id.  
Telephone interview with Thomas Erickson, former CFTC Commissioner and Vice-President, Bunge Ltd., Washington, D.C. (Nov. 7, 2003). Self-regulation is so enshrined in U.S. securities regulation that it is unlikely and probably not in the public interest for it to be supplanted by government regulation. Karmel, supra note 1, at 427.  
Id.
or self-regulatory model, will be. Finally, the CFTC needs to establish a formal structure that facilitates regular dialogue with exchanges and other market participants. The resulting effects of demutualization, electronic trading, and the globalization of marketplaces have yet to be fully realized. In this way, the CFTC would be able to react in a steady and effective manner to conflicts or other issues that arise from demutualization and other forces in the marketplace.

Jake Keaveny†

† J.D. Candidate, 2006, Brooklyn Law School; M.S., Knight-Bagehot Fellow, Columbia University 2005. The author wishes to thank Professor Gary A. DeWaal for his comments and suggestions, Brooklyn Law Review editor Jason Cade, and the rest of the editorial staff, for their attention to detail, and my wife Beth for her endless support.