

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

1996

The Enforceability of Norms and the Employment Relationship

Edward B. Rock

University of Pennsylvania Carey Law School

Michael L. Wachter

University of Pennsylvania Carey Law School

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Contracts Commons](#), [Dispute Resolution and Arbitration Commons](#), [Labor and Employment Law Commons](#), [Labor Economics Commons](#), [Law and Economics Commons](#), and the [Political Economy Commons](#)

Repository Citation

Rock, Edward B. and Wachter, Michael L., "The Enforceability of Norms and the Employment Relationship" (1996). *Faculty Scholarship at Penn Law*. 1424.

https://scholarship.law.upenn.edu/faculty_scholarship/1424

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

**THE ENFORCEABILITY OF NORMS AND THE
EMPLOYMENT RELATIONSHIP**

EDWARD B. ROCK† & MICHAEL L. WACHTER††

TABLE OF CONTENTS

INTRODUCTION	1914
I. THE THEORY OF INCENTIVE-COMPATIBLE CONTRACTS, ILMS, AND SHASTA COUNTY	1919
A. <i>The Physical Reality of Shasta County</i>	1919
B. <i>The Physical Realities of the Internal Labor Market</i> ...	1920
C. <i>The Norms of the Internal Labor Market</i>	1921
D. <i>The Union Internal Labor Market</i>	1927
E. <i>Relational Commercial Contracts</i>	1928
II. THE RELATIONSHIP BETWEEN NORMS AND LAW: DISCHARGE FOR CAUSE AND THE EMPLOYMENT- AT-WILL DOCTRINE	1929
A. <i>"No Discharge Without Cause" As an Incentive-Compatible Term</i>	1930
B. <i>Why Norms Should Not Be Enforced By Courts</i>	1932
C. <i>Statement of the Thesis</i>	1938
III. TOWARDS A GENERAL THEORY OF THE RELATIONSHIP BETWEEN NORMS AND LAW	1940
A. <i>The Other Half of the Analysis: The Choice to Write a Contract</i>	1940
1. What Is the Difference Between Norms and Contract?	1941
2. When Will Parties Choose Third Party Enforcement?	1942
3. But Why a Bipolar Distribution?	1946

† Professor of Law, University of Pennsylvania; Visiting Professor of Law, The Hebrew University of Jerusalem; B.S. Yale University; B.A. Oxford University; J.D. University of Pennsylvania.

†† Professor of Law and Economics, Director of the Institute for Law and Economics, and Deputy Provost, University of Pennsylvania. B.S. Cornell University, 1964; M.A. Economics, Harvard University, 1967; Ph.D. Economics, Harvard University, 1970.

The authors benefitted from helpful comments and criticism from the other participants in this Symposium, from participants in workshops at the University of Haifa Law Faculty and the Hebrew University, and from Walter Kaniat, Omri Ben Shahar, Alon Harel, and Menahem Spiegel. This research was supported by the University of Pennsylvania's Institute for Law and Economics.

B. *The Benefits of Choice Among Legal Regimes* 1947
 CONCLUSION 1950

INTRODUCTION

Norms have attracted much attention.¹ One of Karl Llewellyn's most significant innovations in drafting the Uniform Commercial Code (UCC) was the incorporation of norms, or customary practices, among merchants.² More recently, Robert Ellickson's detailed account of Shasta County has put norms back onto center stage.³

A fundamental question that emerges from this literature involves the relationship between norms and law. To put this question in its sharpest and most interesting form: Under what circumstances should courts enforce the norms of private parties? In Shasta County, where norms, not laws, apparently govern cattle trespass and fencing disputes among ranchers, should the courts enforce these norms? For Ellickson, the answer seems relatively clear: If one concludes that the norms of close-knit groups are generally efficient, then when a close-knit group acts in contexts in which it is unable to impose losses on members, a utilitarian judge should defer, for example, to the customs of merchants or cattle

¹ For these purposes, we will follow one useful definition of norms as "rules or standards enforced solely by private (that is, non-state) actors." In this usage, the term "norm" is largely descriptive: What is it that we observe parties doing? In this sense, we, like others, use the term in its sociological meaning. This must be distinguished from the use of the term in a "normative" sense, for example, "in this society, there is a norm against cheating." While the two uses are often related, they are distinct.

For important recent contributions, see, for example, ROBERT C. ELICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* (1991); Lisa Bernstein, *Opting Out of the Legal System: Extra-Legal Contractual Relations in the Diamond Industry*, 21 J. LEGAL STUD. 115 (1992); Lisa Bernstein, *Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms*, 144 U. PA. L. REV. 1765 (1996); Eric A. Posner, *Law, Economics, and Inefficient Norms*, 144 U. PA. L. REV. 1697 (1996); Eric A. Posner, *The Regulation of Groups: The Influence of Legal and Nonlegal Sanctions on Collective Action*, 63 U. CHI. L. REV. 133 (1996).

The discussion of the relation between norms or custom and law is, of course, one of the ancient debates of the law, stretching back centuries. One can approach the question from an historical perspective, a conceptual perspective, a sociological perspective, or an ideological perspective. In this Article, we approach it from what can best be called either an economic or a functional perspective.

² For a careful historical account, see Zipporah B. Wiseman, *The Limits of Vision: Karl Llewellyn and the Merchant Rules*, 100 HARV. L. REV. 465, 492-538 (1987).

³ See ELICKSON, *supra* note 1, at 15-28.

ranchers engaged in repeat dealings.⁴ Llewellyn arrived at a similar conclusion. In numerous provisions, the UCC incorporates commercial practice into the terms of the contract.⁵

From this perspective, if the norm in Shasta County is that cattle owners are not liable for occasional trespasses or that adjoining landowners split the cost of a common fence in proportion to the number of cattle, then a utilitarian judge should incorporate these norms into the interpretation of the contract between such parties, if a dispute arises and results in litigation. Indeed, these understandings might even suffice to show the existence of a contract between the parties. If one rancher is willing to go to the trouble of suing an adjoining rancher for the damages caused by a trespassing animal or the cost of fencing, perhaps the judge should adopt the norms that actually govern their interactions, rather than refuse to recognize any enforceable agreement or adopt the prevailing and partially inconsistent legal rules.

In this Article, we approach the question from a different perspective, drawing on the theory of incentive-compatible contracting and its application, the theory of internal labor markets (ILM).⁶ In examining the relationship between norms and law, we shift from the somewhat marginal case of cattle ranching to the more central and complex case of the employment relationship.⁷ The employment relationship is particularly interesting because, unlike Shasta

⁴ See *id.* at 255.

⁵ See, e.g., U.C.C. § 1-205(2) ("Usage of trade" means "any practice or method of dealing having such regularity of observance in a place, vocation or trade as to justify an expectation that it will be observed with respect to the transaction in question."); U.C.C. § 1-205(1) ("Course of dealing" means "a sequence of previous conduct between the parties to a particular transaction which is fairly to be regarded as establishing a common basis of understanding for interpreting their expressions and other conduct."); U.C.C. § 2-208(1) ("Where the contract for sale involves repeated occasions for performance by either party with knowledge of the nature of the performance and opportunity for objection to it by the other, any course of performance accepted or acquiesced in without objection shall be relevant to determine the meaning of the agreement."); U.C.C. § 2-314(3) ("Unless excluded or modified . . . other implied warranties may arise from course of dealing or usage of trade.").

⁶ The "internal labor market," or "ILM," refers to the network of arrangements, understandings, and agreements that constitute the employment relationship within firms, as distinguished from relationships in the external employment market.

⁷ We view cattle ranching as marginal for two related reasons. First, the amounts at issue in the disputes are small and relatively infrequent (except for the accidents with third party motorists which inevitably are governed by law); and second, they tend to be reciprocally distributed. These features yield a relatively simple set of solutions: "live and let live;" "fair shares," etc.

County, it involves numerous interactions among repeat players over a long period of time. In Shasta County, as in Ronald Coase's cases,⁸ the parties enter into occasional side-deals to surmount potential harm. In the employment relationship, by contrast, the parties seek an ongoing relationship and generate incentive-compatible contracts that they intend to be self-enforcing.⁹

The similarities and differences between the ILM and Shasta County generate a number of interesting questions. What is the relationship between the norm-governed interactions of Shasta County's ranchers and a firm's internal labor markets? In both cases, the factors that generate Ellickson's results are present. Ellickson's basic claim is that norms arise in close-knit groups, and that the norms controlling the workaday relationship of such groups will enhance overall welfare.¹⁰ In addition, the groups have an intricate information sharing mechanism that facilitates the identification and punishment of those who violate the norms.

Moreover, in studying the employment relationship, we can evaluate the workings of two distinct types of relationships. Specifically, the National Labor Relations Act system ("NLRA system") provides a bifurcated labor market: employment relationships that utilize norms, the non-union sector, and employment relationships that utilize law, the union sector.¹¹ By studying how these two different frameworks operate, we can begin to evaluate the extent to which norms mirror laws, when parties are more likely to choose norms over law, and vice versa, and the extent to which the parties bargain over the same issues.

Drawing on ILM theory, we examine the question of the relationship between norms and law by considering the two central principles of the non-union employment relationship: the use of an informal governance process to run the relationship, coupled with at-will employment to resolve the question as to how the relationship might be terminated. The legal doctrine of at-will employment

⁸ See Ronald Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960).

⁹ To call such relationships "contracts" is slightly but harmlessly misleading: From a legal perspective, to call an arrangement a "contract" is typically to say that it is legally enforceable. By contrast, "incentive-compatible contracts" are typically not enforceable. In the economics literature, the term contracts is used interchangeably with agreements, independent of whether they are intended to be enforceable.

¹⁰ See ELLICKSON, *supra* note 1, at 167-82. The phrase "close-knit groups" refers to groups in which parties repeatedly interact and have the power to administer sanctions in the future. *See id.*

¹¹ We use the term "NLRA system" to mean the National Labor Relations Act, as amended, and the associated case law.

provides a particularly interesting case for examining the relationship between norms and law. In non-union workplaces, a clear norm exists that an employer will not discharge an employee without cause. As we will see, this norm can be understood to arise in response to the distinctive contracting problems that exist: the combination of match-specific investments,¹² information asymmetry, and transaction costs that create both the potential for joint gains as well as the threat of opportunistic behavior. This norm coexists with the traditional (and apparently inconsistent) legal rule that, in the absence of an explicit contract establishing a specific term of employment, an employer generally can discharge an employee for good reason, bad reason, or no reason at all (the doctrine of "at-will employment"). Even where judicial decision or statute has diluted this rule, when parties are free to provide for at-will employment by contract, they generally do so.¹³ The question can thus be rephrased as: Assuming that the general norm in workplaces, at least when there are investments in match, is "no discharge without cause," should the courts enforce that provision as a term of the implied or explicit employment contract between employees and the firm?

We argue that in this case the courts should generally not enforce norms, unless the parties go to the trouble of writing them into their contracts. In taking this position, we reject the "traditional labor law" scholars' near unanimous condemnation of the doctrine of at-will employment as a one-sided contract that makes employers better off and employees worse off. The traditional view lacks an appreciation that norms, as well as collective bargaining agreements (CBAs), can govern the employment relationship in a manner that serves the interests of both parties. Our norm-based theory better tracks how the parties themselves structure their relationship in the great majority of cases.

This then raises a further and separate issue: When will parties choose norms over laws, and vice versa? As we will see below, a system in which parties can choose between norms and

¹² A match-specific investment is one which has value when used by the parties to the match, but no value otherwise.

¹³ When, for instance, individuals attempt to limit the at-will employment doctrine by interpreting an employee handbook as promising employment for a specified term, proponents of the at-will doctrine can return the default to "at-will" by inserting clear language to the effect that nothing in the handbook should be understood to limit the employer's ability to discharge the employee at any time. See Richard H. Winters, Note, *Employee Handbooks and Employment-at-Will Contracts*, 1985 DUKE L.J. 196, 218.

contracts potentially disciplines and improves each. Ellickson does not address this point in his depiction of Shasta County, which suggests that his claim that norms—apart from and uninfluenced by the law—can structure social life may not be generalizable beyond the low stakes, low transaction cost context of Shasta County.

In Part I, we review the ILM literature and argue that norms are an example of the more general phenomenon of self-enforcing, incentive-compatible contracts, that is, contracts that self-enforce, without the need for third party intervention. These contracts encourage parties to maximize their joint welfare. In reviewing this literature, we argue that norms, not laws, govern the internal labor markets of non-union firms, even more so than cattle ranching in Shasta County.

In Part II, we draw on the ILM literature to clarify the relationship between norms and law. Based on this literature, we argue that making the incentive-compatible contracts among close-knit groups legally enforceable often does not, and would not, further the interests of the group. On the contrary, because of the information and transaction costs that give rise to the particular structures in the first place, legal intervention is likely to be, and to be perceived to be, inconsistent with the (*ex ante*) desires of the parties. Under this analysis, the incorporation of custom into law should only occur when the parties themselves choose to write contracts and incorporate norms by reference.¹⁴

Finally, in Part III, we tentatively propose a more general theory of the relationship between the domains of norms and contracts. We argue that the principle difference between norms and contracts

¹⁴ Some argue that courts should not enforce norms because to do so favors the rich and powerful over the poor and weak. According to this argument, the courts should impose standards that protect the weak to a greater extent than do the prevailing norms. In the area of labor law, for example, traditional labor law scholars view employment norms as unilaterally imposed by employers and inadequate to protect employees. *See, e.g.*, PAUL WEILER, GOVERNING THE WORKPLACE: THE FUTURE OF LABOR AND EMPLOYMENT LAW 48-104 (1990). While we likewise reject at least some court enforcement of norms, our alternative—namely, leaving the parties to their own governance structures—is quite different. In essence, we argue that when the courts do not enforce norms, they should yield to alternative governance structures—that is, play a passive role while those who reject norms as insufficiently protective of the weak would argue that rather than refusing to enforce norms at all, courts should enforce standards more favorable to the poor—that is, play a more active role. The difference between these two views revolves around the presence of effects on third parties, the voluntariness of participation in the norm, and the normative implications of consent.

lies in the greater binding power provided by state enforcement. Based on this, we argue that parties are more likely to choose state enforcement when the potential for opportunistic behavior is too great to be constrained by norms. The bipolar distribution between norms and detailed contracts that characterizes both ILMs and commercial relational contracting reflects a consequence of the bipolar choice between self-enforcing norms and third party enforcement. We then turn to the benefits of preserving the choice between norms and law and argue that the legal option likely plays a significant role in protecting norm-governed transactions against opportunistic behavior.

I. THE THEORY OF INCENTIVE-COMPATIBLE CONTRACTS, ILMS, AND SHASTA COUNTY

A. *The Physical Reality of Shasta County*

Shasta County presents neighbors dealing with potential problems of negative externalities, such as roaming cattle. Most of the disputes involve adjacent landowners so that the transaction costs in each occurrence are small, as are the damages from each harmful event. Given the recurring nature of the interactions, coupled with the ability of each party to benefit or harm the other, the residents settle their accounts through reciprocal acts rather than through monetary payments.

In Ellickson's account of these disputes, several features stand out. First, norms—not legal rules—are the fundamental sources of entitlements and the fundamental normative structures that govern the relationships among the members of the community. The parties utilize norms, in part, out of ignorance of the law, but primarily because the individuals in the community view their norms as satisfying their needs. Second, in pursuing remedial actions, residents first attempt to resolve the dispute on their own. They may, for example, take action to harm the reputation of the wrongdoer. If that proves insufficient, they resort to physical acts of self-help, including the seizing of livestock, even if such acts are unlawful.¹⁵ If that also fails, they may complain to public officials. These complaints rarely trigger official sanctions, only the possible threat of such actions.¹⁶ Only as a last resort, and this appears to

¹⁵ See ELLICKSON, *supra* note 1, at 218.

¹⁶ See *id.* at 57-58.

happen rarely, do ranchers seek "attorney assisted claims."¹⁷ Indeed, one of the strongest norms in Shasta County is the norm against going to court to seek monetary or other relief. Finally, Ellickson claims that these norms differ significantly from the rules that a court would apply.¹⁸

B. *The Physical Realities of the Internal Labor Market*

The nature of the ongoing employment relationship within firms, the ILM, is much more complex than is the ongoing Shasta County relationship. In the ILM, many employee-participants operate in a hierarchical organizational structure. In contrast to the occasional and typically unwanted nuisances that characterize the Shasta County scenario, the ILM parties engage in a continuous, beneficial working relationship.

A distinctive feature of the ILM is that it is a structured arrangement with its own set of informal rules governing wages, hours of work, promotion opportunities, grievance procedures, and other terms and conditions of employment. Most hiring takes place in a limited number of entry-level jobs. Based on performance, employees are then promoted, retained in position, or discharged.

There are two very different arrangements for the ILM. The non-union sector employs approximately ninety percent of private sector workers.¹⁹ The non-union ILM can be described as a close-knit group, and the governance arrangement can be described as a norm in the Ellickson sense of the term. Because of the repeat-play

¹⁷ *Id.* at 62.

¹⁸ Ellickson fails to clarify whether Shasta County's norms are ex ante Pareto superior or Kaldor-Hicks superior. In the former case, all parties are better off, in the latter, the overall community is better off, although some individuals may not be. There is no evidence to support the stronger conclusion of Pareto superiority. The weaker claim, that the norms are ex ante Kaldor-Hicks superior, is, however, compelling.

If membership in Ellickson's close-knit groups were entirely voluntary, the norms would be Pareto superior. But this is not the case. Shasta County's norms work because of the disgrace of a bad reputation and nasty self-help moves by aggrieved parties. Residents may be negatively affected even if they have not made a voluntary commitment to observe the norms.

Following Hoffman and Spitzer, Ellickson appears to argue that the norms are Lockean contracts. See *id.* at 143; see also Elizabeth Hoffman & Matthew L. Spitzer, *Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice*, 14 J. LEGAL STUD. 259, 264-66 (1985) (discussing Locke's theory of distributive justice).

¹⁹ See RONALD G. EHRENBERG & ROBERT S. SMITH, *MODERN LABOR ECONOMICS: THEORY & PUBLIC POLICY* 452 tbl. 13.1 (5th ed. 1994).

nature of the ILM contacts, the parties have broad power to administer sanctions, and information sharing takes place to allow the parties to understand the governance arrangements. Moreover, unwritten and unenforceable agreements govern the relationship, and the parties eschew any third party enforcement. Finally, at least the procedural rules that govern the relationship differ substantially from the rules that would apply if the parties unionized.

For these reasons, we can tell the same type of story for the non-union sector that Ellickson tells for Shasta County. Moreover, although Shasta County is but one example, the non-union sector story, if told in terms of stylized empirical regularities, explains behavior across the great bulk of non-union ILMs.

C. *The Norms of the Internal Labor Market*

In the non-union ILM literature, commentators generally agree that an array of stylized regularities exists.²⁰ Generally: wages increase with seniority; a business downturn results in employee layoffs rather than wage reductions; if layoffs occur, junior workers lose their jobs before senior workers; discharges are for cause; if an employer catches an employee shirking, the employer will discharge the employee rather than reduce his wages; and if firms discharge older workers before younger workers, they do so through voluntary retirement mechanisms in which the firm buys out the "contract" of the older worker. Prior to laws against mandatory retirement, firms would allow older workers with below satisfactory productivity to reach the mandatory retirement age and would not dismiss these employees.²¹

These practices are puzzling from a "legal centrist" perspective.²² In each of these cases, the firm could, legally, fire the workers or reduce wages unilaterally and would seem to have an incentive to do so; yet they usually follow a different course of action. Why does the firm provide incentive payments to induce a

²⁰ The ILM literature takes these stylized facts as given, and seeks an explanation for their persistence. There is general agreement that these stylized facts describe the general practice in non-union firms, even if there are departures in various firms at various times. See Michael H. Riordan & Michael L. Wachter, *What Do Implicit Contracts Do?* 3 (Dec. 1982) (unpublished discussion paper for the Center for the Study of Organizational Innovation, on file with author); Sherwin Rosen, *Implicit Contracts: A Survey*, 23 J. ECON. LITERATURE 1144-75 (1985).

²¹ See Edward P. Lazear, *Why Is There Mandatory Retirement?*, 87 J. POL. ECON. 1261, 1261-84 (1979).

²² See ELLICKSON, *supra* note 1, at 137-47.

senior employee to retire when it could simply discharge the employee? Why do firms pay severance pay when they could simply fire a worker? Why do wages increase with age, even when one might have reason to believe that older workers are less productive than younger workers? Why do firms typically lay off junior workers before senior workers in the event of a business downturn when senior workers cost more and may be less productive? Why do firms incur the greater loss of discharging or laying off workers when a business downturn occurs (and run the risk of losing those workers' skills and experience), rather than responding by reducing wages? Why, when workers are, to a large degree, legally unprotected, do they nonetheless receive these protections?

The ILM literature shows that these initially puzzling but persistent features of the employment relationship—features that are common to both the union and non-union sectors, but do not occur in the ever present external labor market—can be understood as terms adopted by the parties to allow them to solve a set of strategic problems that threaten joint-profit maximization. Labor economists distinguish between two fundamentally different labor markets. The external labor market (ELM) is the market in which firms seek to fill vacancies and workers search for new jobs. The textbook ELM breaks down according to the general skills of the workers, covers broad geographical regions, and contains large numbers of firms and workers. Hence, both the supply and demand sides of the market have numerous parties. Since employers and employees have no investments that are specific to the relationship, the relationship can be terminated at low cost. As a result, ELMs are typically competitive markets, with little potential for successful rent-seeking.²³

But, as long recognized, the textbook ELM fails to describe the employment relationship observed within both union and non-union firms.²⁴ In the observed ILM, firms and workers both make invest-

²³ "Rent-seeking," for our purposes, means the expenditure of resources or efforts by one party in order to transfer resources from the other party to itself. This investment by the rent-payers wastes the joint profits of the parties because it creates no new wealth. Moreover, rent-seeking by one party typically causes the prospective rent-payer to expend resources in order to protect its share of the joint investment. In ELMs, because there are no joint investments, there is no possibility of rent-seeking.

²⁴ See PETER B. DORINGER & MICHAEL J. PIORE, *INTERNAL LABOR MARKETS AND MANPOWER ANALYSIS 2* (1971) (positing that jobs within the internal market are shielded from direct influences of competition in the external market).

ments in their match, investments that are lost if the relationship terminates. These investments encompass the identification and training of employees, as well as joint investments in the organization of work in a firm. Once tied together in the ILM, relevant information is asymmetrically distributed. Firms have private information as to product market conditions and available technologies, while workers have information advantages with respect to their own work efforts and opportunity wages in alternative jobs. As a result, the ILM, unlike the ELM, is not a competitive market but, rather, can better be modeled as a bilateral monopoly with considerable potential for rent-seeking.

Why might firms and workers voluntarily eschew the competitive ELM, with the protections that come from competition, for the bilateral monopoly of the ILM? The ILM literature offers two explanations for this phenomenon. First, ILMs generate surpluses over the returns available in the ELM, surpluses sufficiently large that they leave both parties better off even after the costs imposed by rent-seeking behavior. Second, the ILM structures constrain rent-seeking behavior. Another question—central to this Article—emerges here: Under what circumstances will the parties choose to constrain such behavior through norms, and under what circumstances will they opt for written contracts?

ILM theorists have identified three central economic factors that affect an ongoing employment relationship and that are necessary to explain the above summarized ILM patterns: (1) match-specific investments, (2) asymmetric information, and (3) transaction costs.²⁵ Relying on a combination of these factors, ILM theorists have been able to analyze the otherwise puzzling features described above as incentive-compatible contracts designed to solve the twin problems endemic to ILMs, namely, to encourage the optimal match-specific investments while deterring the parties from using their asymmetric information in a rent-seeking manner.²⁶

From this perspective, one can understand why both the firm and the worker typically invest in the match: It encourages the maintenance of the relationship. If only one party made such a commitment, the noninvesting party could “hold up” the other party, threatening to terminate the relationship unless that party

²⁵ See Michael L. Wachter & Randall D. Wright, *The Economics of Internal Labor Markets*, 29 *INDUS. REL.* 240, 244 (1990).

²⁶ Incentive-compatible contract terms are defined as terms that provide incentives to both parties to maximize the joint surplus available to them.

paid a higher return. Such a threat would be credible because the investing party (but not the threatening party) would lose the return on its match-specific investment. By contrast, joint investments deter such behavior: because the threatening party would absorb a loss in carrying out its threat to terminate the relationship, the opportunity to use such a threat in a rent-seeking manner diminishes.²⁷

Similarly, the ILM literature provides a general analysis of the termination of the employment relationship: discharge for cause and permanent reductions in force. In each of these cases, the observed practices manifest a preference for permanent reductions in employment over less severe adjustments. While this fact may initially appear to conflict with the presumed interest of the parties in maintaining the employment relationship, the employment reduction of firms is indeed incentive-compatible for joint-profit maximization, while less severe adjustments would not be.

For example, firms rarely make minor reductions in wages when they discover worker dereliction. Instead, they typically discharge the worker, thereby terminating the relationship. The explanation for this practice rests on an irreducible asymmetry of information: workers know their work effort; firms do not. Firms can learn by monitoring, but constant monitoring is very costly. To save on costs, firms infrequently monitor workers. If most shirking goes undetected because of high monitoring costs, firms must penalize employees an amount greater than the expected loss of any specific incident.

A low detection rate thus accounts for the seeming paradox of a harsh penalty, discharge, for relatively minor incidences of shirking. The paradox manifests the classical law enforcement problem originally investigated by Becker and Stigler.²⁸ Employees will not shirk as long as the expected penalty for shirking equals or

²⁷ Other examples of incentive-compatible terms include seniority clauses, wage profiles that increase with experience, reductions in employment rather than wages during declines in economic activity, and unilateral implementation by firms of decisions affecting levels of output and product prices. For a discussion of these features, see Riordan & Wachter, *supra* note 20, at 6-8.

²⁸ See Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169, 170 (1968) (studying the optimal amount of law enforcement and its relationship to the costs of catching and convicting offenders); George J. Stigler, *The Optimum Enforcement of Laws*, 78 J. POL. ECON. 526, 526 (1970) (posing a rational enforcement theory of law enforcement).

exceeds the benefit from shirking. The ILM parties have two parameters to control the expected penalty—the probability of detection and the actual (as distinct from expected) penalty. Because any expenditure on detection reduces the joint surplus, risk-neutral parties would minimize the probability of detection and offset its impact on the expected penalty by raising the actual penalty, that is, setting the penalty for detection equal to the maximum amount that the shirker could pay. Specifically, the firm must set the penalty so that the expected cost of shirking exceeds the benefit to the employee. When detection rates are low, firms set the penalty very high to raise the expected value to required levels.

Conversely, to encourage workers to consummate work effort, firms must reward employees when the firm learns of superior effort. Promotions serve this purpose. Rewards, as well as penalties, appear to exceed the benefits conferred on the firm of specific proof of consummate work. The reason is the same. Since most consummate work goes undetected, the reward must be high enough so that the expected reward exceeds the cost to the employee of consummate work effort.²⁹

But why not penalize the shirking employee through an equally harsh large wage reduction? If the employer could simply declare that an employee had shirked and cut her wage, the employer would have an incentive to overstate the degree of shirking, thereby reducing costs and increasing profits. Forcing the employer to discharge workers eliminates the firm's incentive to overstate the degree of shirking because it forces firms to lose valued workers in the process.

A parallel explanation applies to the observed behavior of firms in response to cyclical changes in output: firms typically reduce employment rather than wage rates. The firm knows about product market conditions; the worker does not. The firm informs its work force about product market conditions indirectly through changes in output. If a firm could lower wages in response to a decline in its product market, it would have an incentive to misstate the condition of its product market in order to lower wage rates. The incentive-compatible rule is for the firm to lay off workers. Because layoffs result in a reduction in output, and hence a reduction in the

²⁹ See Edward P. Lazear & Sherwin Rosen, *Rank-Order Tournaments As Optimum Labor Contracts*, 89 J. POL. ECON. 841 (1981).

firm's revenues and profits, such a rule eliminates the firm's incentive to misstate information.

The rules concerning firms exiting an industry parallel those governing temporary layoffs. A firm retains the right to decide unilaterally to go out of business, just as the worker enjoys the unilateral right to quit the firm. The process of going out of business typically occurs in several stages. Stylized rules governing the "effects" on workers of the decision to close have evolved: firms typically lay off workers over time using a seniority schedule; workers typically receive some severance pay and always receive their pension rights; and if the firm discharges older workers before younger workers, it typically does so through voluntary retirement mechanisms where it buys out the "contracts" of the older workers. These process rules, which characterize the non-union ILM, are almost always included in union contracts as well.

The rule that firms can unilaterally implement the multi-steps of going out of business has strong incentive-compatible properties. Once out of business, the firm has lost whatever value is in the ILM, and hence will only close when the ILM is indeed unprofitable. The rules governing the process control the firm's incentives to profit from the process, by imposing direct costs (severance pay or voluntary retirement programs) and indirect costs (laying off less expensive junior workers before more expensive senior workers) on such a firm.

The old rule that firms would allow employees to retire when they approached the mandatory retirement age, rather than discharge such employees for bad performance, also had strong incentive-compatible properties. Older workers, at least in this prior regime of mandatory retirement, tended to be in the stage of their career when they were recouping the benefits of earlier investments in match. In this stage, employers would have an incentive to discharge the employees for small lapses or no lapses at all. The norm against discharging such employees prevented the employer from capturing the employee's return on investment.

Each of these practices is a "norm" in precisely the sense in which the practices of ranchers in Shasta County are norms: they are practices that parties adopt and follow in an ongoing relationship, practices that are largely not legally enforceable. Parties invent these practices or they evolve in order to constrain opportunistic behavior that stands in the way of joint-profit maximization when match-specific investments, asymmetric information, and transaction costs characterize the relationship. Finally, and

most importantly from the perspective of norms, the practices are incentive-compatible and largely self-enforcing; that is, the parties have incentives to abide by the norms, without third party enforcement.

D. *The Union Internal Labor Market*

The union ILM operates very differently from the non-union ILM. In the union sector, a detailed collective bargaining agreement governs the parties' relationship, with the agreement enforceable in contract law or by an arbitrator who settles disputes by filling in the gaps in the existing contract.³⁰

In the Ellickson terminology, the union ILM, like the non-union ILM, is a close-knit group. As is true in the non-union sector, the union ILM involves repeat relationships, the parties have broad power to administer sanctions, and information sharing takes place to allow the parties to understand the governance arrangements. Contrary to Ellickson's prediction, however, the union ILM functions through a complex set of statutory obligations enforced through the NLRA system and contract law, with the parties making frequent appeals to the NLRB, to courts, and to arbitration. In the union ILM, the parties do not just bargain in the shadow of the law, they frequently argue their positions before administrative law judges and arbitrators in time consuming and costly litigation.

Note, however, that the parties to the collective bargaining agreements write explicit terms that enforce the stylized behavior of the non-union ILM parties. Generally, wages increase with seniority, firms lay off employees in a downturn following a seniority schedule, firms discharge employees rather than reduce their wages if the firms prove that employees are shirking, and permanently laid-off workers receive some severance pay.

In addition, the scope of the collective bargaining agreement is similar to the scope of the non-union agreement. The union contract, like the non-union norms, covers wages, hours, and other terms and conditions of employment. It rarely includes contract terms or norms that significantly limit the employer's ability to direct the firm. Although many traditional labor law scholars have

³⁰ The NLRA system governs the procedures for unionizing and drafting a collective bargaining agreement.

advocated the characterization of employees as quasi-joint owners of the firm, that has not occurred.³¹

The difference is thus limited to what might be described, according to the Ellickson framework, as the procedural and constitutive norms. The NLRA contains considerable rules governing the certification of unions. Both the employer and the union must act in a manner that does not violate a series of rules governing unfair labor practices. The parties may resort, as noted above, to the NLRB for violations of the NLRA and to arbitration and the courts for violations of the collective bargaining agreement. Thus, while the non-union norms are intended to be self-enforcing, the union contracts are enforced by third parties.

E. *Relational Commercial Contracts*

Relational commercial contracts are similar to ILMs and include match-specific labor investments that are not brought inside the firm. Subcontracting arrangements and personal service contracts are two examples of relational labor contracts. Non-labor examples include merchants in continuing commercial relationships³² and long-term relationships between suppliers and customers.³³

Relational contracts, unlike other commercial contracts, are ongoing. As is true for the ILM, ongoing relationships are efficient when there are opportunities for investments in the match that will generate a surplus beyond what sequential contracting across different parties can achieve. As is true in the labor case, the sunk investments in match give rise to quasi-rents and the potential for opportunistic behavior. Hence, neither side can go forward without protection.

The generic relational contract is a mix of contract terms and norms. In the case of merchants interacting repeatedly in relatively small-stakes transactions, Macaulay showed that the norms that

³¹ For excellent statements of the traditional labor law approach, see JAMES B. ATLESON, *VALUES AND ASSUMPTIONS IN AMERICAN LABOR LAW* 173 (1983); WILLIAM B. GOULD IV, *AGENDA FOR REFORM: THE FUTURE OF EMPLOYMENT RELATIONSHIPS AND THE LAW* 128 (1993).

³² See Stewart Macaulay, *Non-Contractual Relations in Business: A Preliminary Study*, 28 AM. SOC. REV. 55 (1963) (explaining how what merchants do differs from what contract law prescribes).

³³ See OLIVER E. WILLIAMSON, *MARKETS AND HIERARCHIES: ANALYSIS AND ANTITRUST IMPLICATIONS* 91-95 (1975) (discussing incentive compatibility and its relationship to incomplete long-term contracts).

governed their interactions differed significantly from the doctrines of contract law that a court would apply. In place of the legal rules, one observes a set of workaday norms that, at least from the parties' perspective, serves to maximize their joint profits. But, as a feature of relational contracts, the norms are insufficiently self-enforcing and the parties thus choose to include explicit contract terms as well.

The paradigm relational contract in the law and economics literature is the coal mine next to the power plant.³⁴ It is in the parties' joint interest, *ex ante*, to enter into a long-term contract; yet the number of contingencies that can arise over a twenty year contract are such that writing a fully specified contract is impossible. Parties thus draft a long-term contract with open terms that they later fill by negotiation and, when negotiation fails, by third party adjudication (either by a court or an arbitrator). In addition, the contract includes the norms of the relationship by reference.

II. THE RELATIONSHIP BETWEEN NORMS AND LAW: DISCHARGE FOR CAUSE AND THE EMPLOYMENT-AT-WILL DOCTRINE

Ellickson's work and the earlier work on norms establish that norms exist and govern behavior. Ellickson's work suggests that norms, at least in business contexts, are generally joint-profit maximizing. When you combine these two hypotheses, the next step seems almost inescapable: if the norms exist and maximize the joint profits of the participants, then of course the law should incorporate and enforce them. After all, the parties, by their own actions, have demonstrated that they wish to live by these rules. In the absence of third party effects (the constant assumption here), what business do the courts have enforcing any other rule? As we show in this Part—through an analysis of the coexistence of a norm of “discharge only for cause” with a legal rule of “employment-at-will”—this deeply held intuition is wrong.

³⁴ See *id.* at 91-95; Robert A. Hillman, *Court Adjustment of Long-Term Contracts: An Analysis Under Modern Contract Law*, 1987 DUKE L.J. 1, 23. There is an extensive literature on relational contracting beginning with the pioneering contributions of Goetz and Scott. Much of this literature discusses how to fill in the inevitable gaps. Generally, the literature concludes that the joint-profit maximizing norms of the contracting parties should fill in the gaps. See Charles J. Goetz & Robert E. Scott, *Principles of Relational Contracts*, 67 VA. L. REV. 1089, 1099 (1981).

A. *"No Discharge Without Cause" As an
Incentive-Compatible Contract Term*

We know that generally, even in non-union firms, the norm governing termination is "no discharge without cause." One can understand the persistence of this norm, even in a legal world in which "employment-at-will" predominates, on the same basis as described above.³⁵ A norm of "discharge for cause," in contrast to a reduction in wage rates, is joint-profit maximizing for the parties because it induces optimal investment in match. If the firm could discharge without cause, it could use such a threat to appropriate an additional share of the joint surplus ex post. The threat of such ex post appropriation would, as in the earlier analyses, stand in the way of optimal investments in match ex ante.

In general, and in most cases, such a provision can be expected to be self-enforcing, for reasons introduced in Part I. Firms will not discharge employees without cause because if they did so they would lose the benefit of that employee. By contrast, if the norm provided for a wage reduction when firms detected shirking, the firm would have an incentive for claiming shirking even though it never occurred. In such a case, the firm could retain the valued employee, while having higher profits from the wage reduction.

The role of discharge for cause plays an important role in allowing the ILM parties to use norms rather than laws. For norms to work, the parties must have power to administer sanctions. Discharge serves that purpose in the ILM.

For norms to be self-sustaining, however, the parties must have the power to administer sanctions. Ellickson's account of Shasta County demonstrates the importance of the availability of adequate sanctions. There, the neighbors have a limited set of options,

³⁵ There is an enormous literature on "employment-at-will" including prominent contributions attacking the doctrine on various grounds. See, e.g., Lawrence E. Blades, *Employment at Will vs. Individual Freedom: On Limiting the Abusive Exercise of Employer Power*, 67 COLUM. L. REV. 1404, 1405 (1967) (arguing that the at-will doctrine does little to protect an economically dependent employee from employer power); Clyde Summers, *Individual Protection Against Unjust Discharge: Time for a Statute*, 62 VA. L. REV. 481, 483 (1976) (criticizing the at-will doctrine for leaving a majority of workers without protection against arbitrary discipline). A number of articles analyze the doctrine and the arguments against it from a law and economics perspective. See Richard A. Epstein, *In Defense of the Contract at Will*, 51 U. CHI. L. REV. 947 (1984); Mayer G. Freed & Daniel D. Polsby, *Just Cause for Termination Rules and Economic Efficiency*, 38 EMORY L.J. 1097 (1989); Stewart J. Schwab, *Life-Cycle Justice: Accommodating Just Cause and Employment at Will*, 92 MICH. L. REV. 8 (1993).

including negative gossip and social ostracism.³⁶ In addition, they sometimes employ physical, sometimes unlawful, self-help. The difficulties they had with Frank Ellis illustrate the weakness of the available sanctions. The good players had difficulties dealing with him because of limited ability to apply sanctions that were meaningful to him. He was unfazed by negative gossip. Unlike the ILM scenario, where a firm can discharge a bad player, bad neighbors can stay put. In the end, as in a fairy tale, bad things happened to Ellis, and he eventually went bankrupt and moved. But absent such "acts of God or nature," the good players faced a difficult situation.

In the ILM, employees likewise have sanctions available to discipline the errant employer. First, when firms are continuing in business, reputational effects will push them to avoid opportunistic behavior. For the repeat player, departure from the norm will impose costs on the employer: in hiring new employees, in discouraging workers from making investments in match, in "loss of morale" among existing employees (for example, in a self-determined and relatively unsanctionable reduction in output), and in higher wages. For a large employer, the benefit of acting opportunistically with respect to a particular employee will thus be offset by the increased costs of all the other employees, aggregated.³⁷ Second, employees can engage in hard to detect, but often effective, work slowdowns (what in England is called "work-to-rule"). Third, as we discuss further below, employees can discipline the bad norm player (employer) by forming a union and replacing the system of norms with a collective bargaining agreement.

³⁶ Gossip on the job plays the same critical role as does gossip among neighbors. Gossip among employees spreads the required information as to whether the employer follows the norms, and is thus a good employer, or does not, and is thus a bad employer. In doing so, gossip lowers the cost of information asymmetries among employees. Employees can differentiate the good employees from the bad, at lower cost than can the employer. For discharge-for-cause to function as a norm, employees must be able to validate that a discharged employee violated the norm and was indeed shirking.

³⁷ H. Lorne Carmichael emphasizes the importance of reputational effects in labor markets. See H. Lorne Carmichael, *Reputation in the Labor Market*, 74 AMER. ECON. REV. 713, 713-25 (1984). The significance of these costs and the extent to which they adequately restrain opportunistic firing of employees during the "payback" period are empirical questions. The persistence of "at-will employment" in unregulated markets provides some evidence (albeit perhaps not conclusive evidence) that these restraints sufficiently constrain such behavior. For a fuller discussion of this question, see Epstein, *supra* note 35, at 967-68; Freed & Polsby, *supra* note 35, at 1127-30; Schwab, *supra* note 35, at 26-28.

Norms work least well when the power of future sanctions is eliminated or reduced. This scenario most often occurs when an employer has a reasonable prospect of going out of business. This raises the "end-game" problem and converts a relationship that is near termination from a repeat game into a nonrepeat one. Because most firms go out of business in steps, employees can protect themselves by reducing their investments in match, thereby reducing their losses if forced to switch employers. Alternatively, employees can protect their interests through contract. We discussed this issue in the context of the relational commercial contract above. We will discuss the general issue of choosing between contracts and norms in the ILM context in Part III below.

B. *Why Norms Should Not Be Enforced By Courts*

If non-union ILMs have a norm of "no discharge without cause," why are parties who have committed themselves to a norm of "no discharge without cause" in non-union ILMs, unwilling to back that commitment with legal enforceability? Why rely on norms to self-enforce, when the option of third party enforcement is available? There are two primary explanations: the difficulty in accurately verifying that "cause" exists and the inability of third parties to enforce highly incomplete contracts.

Suppose that the employer needed to prove "just cause." As a starting point, proving cause to a court would require that the employer engage in additional detection costs. It is always more difficult to prove a case to a third party than to learn the facts independently. This alone reduces the value of the match.

But what set of facts would constitute cause? Flagrant bad play, as in the case of Frank Ellis, might be easy to prove, but provides little guidance. Since the ILM parties want to minimize monitoring or detection costs, they want the penalty to be large relative to the damages from any violation. This means that the courts would have to determine whether a relatively minor violation of a norm was meaningful enough to constitute "cause" for discharge. Doing so would be a daunting task. The court would need to learn about the relevant norm,³⁸ the damages caused by the violation of the norm, the probability of detection and punishment, and the effect of the sanction on providing optimal incentives for other employees.

³⁸ For example, does one or several cases of shirking constitute just cause according to the norm?

The difficulty of valuing match-specific investments increases the likelihood of error when third parties enforce norms. In valuing assets, third parties typically rely on market prices. Valuations of match-specific investments, by contrast, rely on the parties' own valuations. From the employer's perspective, just cause exists when the continuation of the match with the particular employee has a negative net present value, including the reputational cost of taking too tough or too easy a stance in the face of the perceived violation. The problem is that a third party could not verify the firm's own net present value calculation at low cost and perhaps not at all. Courts would therefore be prone to reaching erroneous decisions.³⁹

Proof for the assertion that courts would find it difficult to resolve non-union ILM disputes manifests itself in the widespread use of arbitration clauses in similar types of relationships. For example, in ongoing relational commercial contracts and union ILMs, where third party enforcement is common, the parties often rely on arbitrators who specialize in resolving problems for that particular match. This behavior illustrates the difficulty of relying on the courts. Although arbitrators may generate fewer costs than courts, including the costs of verifying asset specific valuations and incorrect decisions, the costs would still be high.

Generally, when the facts needed to resolve disputes are difficult to verify, the parties themselves are in a better position than a third party to evaluate them. In such cases, norms built around self-enforcing rules are superior to third party enforcement. Ellickson also supports this conclusion, which plays a large role in the incentive-compatible economics literature.

The second explanation for why norm based relationships are not amenable to third party enforcement involves the issue of incomplete contracts. When agreements are built primarily around norms, any contract will be highly incomplete. In the extreme case, there may be no contract at all. The more incomplete the contract, the more limited the court's ability to infer the parties' intentions.

Schwartz has pointed out that courts take one of two very different approaches when confronted with an alleged contract breach.⁴⁰ In cases where the courts find a detailed, albeit incomplete contract, they take an activist stance, attempting to fill in the gaps in the contract. Courts fill in the gaps based on their interpre-

³⁹ See Alan Schwartz, *Relational Contracts in the Courts: An Analysis of Incomplete Contracts and Judicial Strategies*, 21 J. LEGAL STUD. 271, 313-15 (1992).

⁴⁰ See *id.* at 271-74.

tation of what the parties would have agreed to had they known that the contingency would arise. In doing so, the courts help the parties maximize the value of their exchange. Contract writing is costly, and contracts will always be incomplete. By filling in the gaps in such a manner, contract law serves the joint welfare maximizing interests of the parties.

When confronted by a highly incomplete contract, the courts typically become passive, treating whatever language exists as the complete contract. As Schwartz points out, in such cases, courts deem the contract complete—without gaps for the courts to fill.⁴¹ Courts take this position not because they really believe that the contract is complete—it clearly is not—rather, because they do not know how to fill in the gaps in a manner that assists the parties in maximizing their joint profits. The more incomplete the contract, the more limited the court's ability to infer the parties' intentions.

Information asymmetries pose a related hurdle to an activist court. Information asymmetries may render a contract incomplete because they prevent third parties from observing or verifying relevant information. One type of information asymmetry, mentioned above, involves the parties' valuation of their asset. If the relevant information concerning a variable is unobservable or nonverifiable, the parties will not condition a promise on its outcome, and a third party cannot enforce it. The only solution is to adopt a self-enforcing contract term. These information asymmetries drive the development of the ILM's self-enforcing terms.⁴²

⁴¹ *See id.*

⁴² If the court endeavors to serve the parties, it will hesitate to override their conscious decision to leave the contract incomplete. It will not require more disclosure from the informed party than the contract demands. In such cases, the court views the parties as unwilling to force the information asymmetry into the open, presumably because doing so would entail high costs.

The term "contract gap" describes inadvertent missing terms in the contract. Gaps do not include areas that the parties did not intend the contract to cover. Instead, the court treats the contract as complete and does not consider the missing areas part of the contract. Historically, courts have been unwilling to assume the degree of activism necessary to broaden the scope of a contract beyond its intended limits.

Commentators disagree on this issue. Ian Ayres and Robert Gertner suggest that the courts establish penalty defaults. By "penalty defaults," they mean defaults that the parties would not choose on their own because they thwart rent-seeking ends. *See* Ian Ayres & Robert Gertner, *Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules*, 101 YALE L.J. 729, 735-36 (1992). But the penalty defaults that the courts

In the absence of clear evidence that the governance structure has failed the parties, third party enforcement—an activist judicial approach—will likely reduce the joint-profits of the parties. Suppose that, in spite of the aforementioned problems, the courts developed a rule which required the firm to prove the case of shirking to a third party even in the absence of explicit or implicit contract terms.⁴³ In response to this uninvited policy innovation, the parties would have a number of alternative options. First, they might adopt an explicit term either specifically ruling out a requirement of providing proof to third parties, or indicating under what circumstances to enforce such a requirement.⁴⁴ Absent public policy to the contrary, even an activist court would enforce explicit contractual language. But requiring the parties to complete contracts, when they otherwise would not do so, forces them to incur extra costs.

Second, parties could incorporate the court-imposed rules and standards of proof into their relationship. This incorporation will most likely occur when the cost of the court-imposed inefficiencies is relatively low and when the parties can mitigate those costs by adopting alternative strategies. Finally, the parties might decrease

adopt appear to mirror the self-enforcing contracts that the parties would indeed adopt on their own. Moreover, Jason S. Johnston challenges Ayres's and Gertner's emphasis on information revealing rules. See Jason S. Johnston, *Strategic Bargaining and the Economic Theory of Contract Default Rules*, 100 YALE L.J. 615 (1990). Johnston shows that the choice of rules is consistent with game-theory results which would allow the parties to be silent with respect to the profits they expect from the contract. See *id.* at 627-30.

⁴³ For a discussion of this issue, see Ronald G. Ehrenberg, *Workers' Rights: Rethinking Protective Labor Legislation*, in 8 RESEARCH IN LABOR ECONOMICS 285, 290-97 (Ronald G. Ehrenberg ed., 1986) (discussing proposed labor reforms of the employment-at-will doctrine to provide non-union workers protection against "unjust dismissal").

⁴⁴ Where courts have altered the default rule of employment-at-will by enforcing language in employer handbooks as "implied contracts" or "implied terms," employers have tended to return to the prior default by memorializing the default in explicit language. On the enforceability of employee handbooks, see generally 1 E. ALLAN FARNSWORTH, FARNSWORTH ON CONTRACTS § 3.15a (1990); E. Allan Farnsworth, *Developments in Contract Law During the 1980's: The Top Ten*, 41 CASE W. RES. L. REV. 203, 206-10 (1990); Winters, *supra* note 13.

The ability of employers to avoid "implied terms" by explicit contractual terms is consistent with general principles of contract law. See, e.g., U.C.C. § 2-208(2) ("The express terms of the agreement and any such course of performance, as well as any course of dealing and usage of trade, shall be construed whenever reasonable as consistent with each other; but when such construction is unreasonable, express terms shall control course of performance and course of performance shall control both course of dealing and usage of trade (Section 1-205).").

the intensity of use of the ILM relationship by making greater use of ELM contracts whenever possible and decreasing their joint investments in match-specific investments.⁴⁵

A middle ground would be to establish a default rule that stated that third party proof would be required unless the parties expressly agreed to waive the requirement. The costs of such a requirement would depend on the breadth of the default rule. For example, if the default rule applied broadly, it could mean that third party proof would be required to fire a seasonal, part-time worker, absent waiver. On the other hand, the cost could be relatively small if the requirement of explicit waiver were narrowly construed and if the parties were allowed to adopt broad or "boilerplate" contract language waiving third party enforcement across all issues. Making the middle ground the default would have the effect of forcing the employer to write the waiver into a contract. In doing so, the employer would be notifying the workers of their limited rights, thus allowing workers to adjust their behavior accordingly. If this were the rule, the standard non-union ILM contract would include but the single, employment-at-will term.

Such a middle ground is a negative default in the spirit of Ayres and Gertner.⁴⁶ An interesting question is whether judicial activism should attempt to push the parties further into a broader, middle-ground contract setting. For example, should the courts interpret the terms of the employee handbook as contract terms? Our answer is that the employment contract does not constitute middle-ground contract terms. Rather than being contract terms, the prototypical employee handbook contains the parties' agreed upon norms. Writing the handbook is not akin to writing a contract. Instead, it is the cost efficient mechanism for communicating norms to the ILM community. For reasons discussed below, employment contracts tend to be bipolar, with at most a narrow middle ground.⁴⁷

⁴⁵ See James N. Dertouzos & Lynn A. Karoly, *Employment Effects of Worker Protection: Evidence from the United States*, in EMPLOYMENT SECURITY AND LABOR MARKET BEHAVIOR 215, 215-27 (Christoph F. Buechtemann ed., 1993).

⁴⁶ See Ayres & Gertner, *supra* note 42. For commentators concerned that the ILM is fraught with employer opportunism, the middle ground provides little additional protection to workers. To the traditional labor law scholars, for example, only the polar case of union contract can protect employees.

⁴⁷ Other terms that can fit into middle ground are severance provisions or other provisions that deal with the effects of the termination of the relationship. However, the middle ground would not contain terms that govern the functioning of the ongoing ILM relationship.

Statutory intervention is potentially more troublesome than court intervention because policymakers are likely to be more willing to override the expressed intent of the parties and it is more difficult to contract around statutory terms. For example, a statute making the norm "no discharge without cause" legally enforceable would undermine the norm-based system discussed above which protects all those employees for whom the norm is incentive-compatible, in order to provide protection for those employees subject to employer opportunism and who cannot protect themselves through unionization.

Another interesting example involves the rule against mandatory retirement. One of the most prevalent stylized practices of non-union ILMs prior to the prohibition of such rules was the term or norm specifying a mandatory age of retirement. With the mandatory retirement norm in effect, an associated norm provided that firms would not discharge employees whose performance failed to meet threshold requirements in the years immediately prior to retirement. Employers who discharged older employees violated the norm and acted opportunistically. The mandatory retirement term replaced difficult to observe terms (effort and performance) with an easily observable term (age).

How are the parties likely to respond to this statutory intervention? One possibility, and perhaps the one that would be predicted by Ellickson, is that older employees would continue to honor the norm governing retirement age. Preliminary evidence suggests that this is not the case, at least not at age sixty-five. Of the three strategies noted above, overwriting the rule is obviously not possible. Reducing the intensity of use of ILMs is possible and would add to other statutory interventions that appear to have made certain ILMs less profitable. The second strategy—live with the rule but mitigate its effects—is most likely. In particular, discharge for cause of older employees is likely to become more common. Such an increase would not be, and should not be viewed as, discriminatory in that the incidence of termination for cause would increase toward the rate for younger employees.

This theory thus suggests that activist judicial intervention in ILM governance structures is unlikely to improve the joint profits of the parties.⁴⁸ In the face of incomplete contracts and informa-

⁴⁸ To the extent that the norm against discharge without cause is designed to protect deferred compensation, the employer will have an increased incentive to discharge as the employee moves into the "payback" period. Stewart Schwab has

tion asymmetries, individual courts would find it difficult to write contract terms that would maximize joint profits. Even if one assumes that the parties have failed to negotiate the optimal contract, the ability of courts to know which terms would improve the contract, rather than impair it, is suspect. Thus, in these contexts, if the courts or the legislature make norms enforceable when doing so is not joint-profit maximizing for the parties, the unanticipated and undesired effect will be to cause the parties either to contract around the term, if permitted, or to abandon the norm-governed relationship entirely.

C. *Statement of the Thesis*

If the foregoing analysis is generally correct, it points towards a specific and perhaps counterintuitive thesis about the relationship between norms and law: when a system of norms is self-sustaining and does not impose costs on third parties, the courts should not enforce those norms. Rather, the courts should do nothing, adopting a rule which, although different in substance from the pattern of behavior that the parties themselves follow, forces the parties to rely on the self-enforcing properties of their arrangement. When a norm is self-enforcing, judicial enforcement only provides opportunities for abuse without any off-setting benefits. Thus, in the ILM context, to the extent that incentive-compatible norms evolve to solve the contracting problems that interfere with the parties' abilities to protect match-specific investments, those norms should not generally be enforceable at law. On this analysis, the fact that elaborate norms protecting workers' investments in match against ex post appropriation by firms (and vice versa) coexist with a regime of "employment-at-will" makes perfect sense, and does not reflect the sort of "gap" that the courts should fill with norms. Likewise, the fact that firms tend to contract around encroachments on the doctrine is likewise consistent and unsurprising. From this perspective, even if the courts of Shasta County concluded that the

argued that many of the departures from employment-at-will reflect attempts to protect against such "opportunistic" behavior. See Schwab, *supra* note 35. Schwab's argument does not address the question of whether the norms are sufficiently strong to protect against such behavior (and thus whether judicial intervention is necessary in such circumstances). Rather, he makes the positive claim that the cases represent an attempt to protect against opportunistic behavior by the employer. But this reasoning immediately raises a red flag: if the parties could not solve the problem of opportunism by themselves, as Schwab supposes that the courts hypothesize, why is it that the parties did not address the problem contractually?

ranchers' norm governing the allocation of the costs of fencing was joint-profit maximizing, they should not enforce it.

Note that our defense of employment-at-will differs substantially from that offered by Richard Epstein.⁴⁹ In essence, Epstein argues that employment-at-will is the parties' equilibrium choice.⁵⁰ This means that, to contract for greater protection than offered by employment-at-will, the employees would have to trade off more in wages and benefits than they would be willing to sacrifice. We argue, on the other hand, that most employees work under a norm that offers considerably more protection than employment-at-will.

Our argument will not, however, assuage critics of employment at-will, like Paul Weiler, in the least.⁵¹ They assert that all employees should receive protection from wrongful discharge, either through union CBAs or stringent court protection.⁵² We argue

⁴⁹ See Epstein, *supra* note 35, at 982 (arguing that employment-at-will affords flexibility for the "ceaseless marginal adjustments that are necessary in any on-going productive activity" conducted under conditions of change). Because we emphasize the costs of litigating "just cause" and related cases, our analysis does not deal with the class of statutory interventions such as the Employment Retirement Income Security Act that provide for minimum terms in retirement plans. In addition, as a conceptual matter, our defense of employment-at-will would not apply to special public policy goals such as Title VII. Of course, as more groups receive additional protection, less remains of the doctrine.

⁵⁰ See RICHARD A. EPSTEIN, *FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS 147-58* (1992).

⁵¹ See WEILER, *supra* note 14, at 48-104. Some will also argue that the foregoing discussion is unconvincing because it puts the rabbit in the hat: if one assumes that employment markets are reasonably efficient and that workers have choices and can contract freely, then reasonably efficient outcomes will result from the freedom of contract. The problem, according to such critics, is precisely that workers cannot contract freely: they lack choices, information, and contractual capacity. The essence of the attack on the employment-at-will doctrine is that workers, because they are weak in comparison to employers, cannot protect themselves.

A full response to this argument transcends the scope of this Article. For a thorough reply to the argument, see Epstein, *supra* note 35; Freed & Polsby, *supra* note 35.

Because incentive-compatible contractual structures characterize a variety of "incomplete" contracts, many of which involve parties of relatively equal bargaining power, the claimed incapacity of workers cannot completely explain the persistence of such structures in the non-union sector.

⁵² This is the position taken by Walter Kamiat in his Comment on our Article, when he states that the lack of contractual protection from discharge "may reflect serious contracting problems in the non-union internal labor market."

Professor Kamiat suggests that employees do not request contract language stating "no discharge without cause" lest they be signalling that they are "lemons," in the spirit of Akerlof's famous article. See George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488 (1970). But the problem is not an "employee-as-lemon" problem at all. In a just-cause regime, low

that most employees already enjoy that protection via norms. If we are correct, then employees working in efficient, norm-governed ILMs have a better overall package than would Weiler's employees working with court protection.

III. TOWARDS A GENERAL THEORY OF THE RELATIONSHIP BETWEEN NORMS AND LAW

In this Part, we try to sketch out a more general theory of the relation between norms and law. To do so, we first discuss the alternative to relying on norms, that is, contract, and some of the factors that push parties to choose one or the other. We then turn to the conditions necessary to maintain competition between these two alternatives and the implications of that competition.

A. *The Other Half of the Analysis: The Choice to Write a Contract*

In Shasta County, in commercial interactions, and in labor markets, one observes the use of norms as well as the use of contracts. What factors explain the choice between the two options?

productivity employees are no more protected than they are in a norm-governed relationship. In either case, they are appropriately discharged for cause. On the other hand, the employee who demands contractual protection against arbitrary dismissal may do so for fear that the employer is a lemon. We discuss this below in noting that perceived employer untrustworthiness is one reason why employees choose the union ILM.

For Professor Kamiat's model to work, the just-cause provision, whether incorporated in an individual or collectively bargained contract must better protect the shirking employees than does the norm regime. This does not provide an attractive defense for the existence of just-cause contractual protection, and we doubt that it is correct. The ILM, whether union or non-union, is a relationship of frequent repeat-plays. The low productivity employee is inevitably discharged.

More generally, Professor Kamiat, like others who take a traditional approach to labor law, fails to appreciate the connection between the existence and viability of this system of norms and the doctrine of employment-at-will which, as we argue in the text, is what makes the norms self-regulating. For example, Kamiat argues that the widespread adoption of just-cause protection in union contracts reflects employees' interest in such protection and that its absence from non-union contracts thus reflects market failure. But, where norms are adequately self-enforcing, non-union employees enjoy all of the protections without writing them into the contract. The gulf between the union contract and the non-union governance mechanism is not indicative of market failure, but of the ability of parties working as a close-knit group to enjoy all of the protection of contract without the cost of writing and enforcing those contracts.

1. What Is the Difference Between Norms and Contract?

A striking feature of labor markets is that contracting is polarized: either firms opt for almost entirely unwritten (and unenforceable) agreements or firms adopt intricate, detailed, written contracts, along with governance mechanisms for adjudicating disputes and filling in gaps. The non-union sector relies on unwritten and unenforceable contracts while the union sector uses detailed collective bargaining agreements.

In union workplaces, parties typically negotiate detailed terms and, when disputes arise, resort to legally mandated grievance procedures, with arbitration as a final step. In a grievance, the parties will first try to negotiate a resolution, in effect, renegotiating the terms of the contract, or, perhaps more precisely, negotiating to fill in gaps that were (often intentionally) left in the contract at the time of signing. When the parties fail to negotiate a solution, and an arbitrator adjudicates the dispute, the arbitrator will look to the practices of the ILM in order to fill in the gaps. Here, one observes incorporation of norms in contract interpretation. Indeed, one of the perceived advantages of labor arbitration over court enforcement involves the greater familiarity that an arbitrator is thought to have with firm and industry norms.⁵³

But note a significant difference with the non-union sector: in incorporating norms into their contract by means of the arbitration provision, the union sector, paradoxically, relies on norms less than the non-union sector, where the relationship hinges almost entirely upon norms. In the non-union sector, parties precommit not to use third party enforcement.

⁵³ As stated in *United Steelworkers v. Warrior & Gulf Navigation Co.*, 363 U.S. 574 (1960):

The collective bargaining agreement states the rights and duties of the parties. It is more than a contract; it is a generalized code to govern a myriad of cases which the draftsmen cannot wholly anticipate. . . . [Moreover, a]rbitration is the means of solving the unforeseeable by molding a system of private law for all the problems which may arise and to provide for their solution in a way which will generally accord with the variant needs and desires of the parties.

Id. at 578, 581; *see also* *United Steelworkers v. American Mfg. Co.*, 363 U.S. 564, 568 (1960) ("Whether the moving party is right or wrong is a question of contract interpretation for the arbitrator. . . . The courts, therefore, have no business weighing the merits of the grievance."); *United Steelworkers v. Enterprise Wheel & Car Corp.*, 363 U.S. 593, 596 (1960) ("[Arbitrators] sit to settle disputes at the plant level—disputes that require for their solution knowledge of the custom and practices of a particular factory or of a particular industry as reflected in particular agreements.").

This commitment not to use third party enforcement distinguishes norm-dominated interactions. The norm against suing other ranchers is an intrinsic feature of the norm-structured resolution of cattle trespass and fencing disputes in Shasta County. The employment-at-will doctrine which protects non-union ILMs from third party enforcement is intrinsic to the self-enforcing character of the interactions. From this perspective, the link between explicit contracting and third party enforcement becomes more clear: the choice to write contracts is fundamentally a choice to move from a relationship governed by norms to a relationship with enforcement by state actors.

2. When Will Parties Choose Third Party Enforcement?

When will parties choose third party enforcement (contracts) over norms? The answer must lie in the ability to construct self-enforcing contracts. If the parties can construct incentive-compatible contracts that encourage joint-profit maximization and are self-enforcing, then they will use norms rather than contracts. This answer, however, is somewhat definitional. What predicts when self-enforcing contract terms are likely to be available? The answer seems to be that the choice is not driven by either match-specific investments or asymmetric information directly, but rather, by the difference in the binding power of norms versus contracts.

To understand when norms will work, one must understand why norms dominate in both Shasta County and the non-union ILM, two very different circumstances, but fail in commercial power plant/coal mine type relational contracts and union ILMs. We will focus first on the norm side and then shift to the contract side.

Shasta County is a Coasian-type world of low transaction costs, where interactions are sporadic, unconnected, low stake, and simple. The high transaction cost ILM interactions, on the other hand, are high frequency, connected, and complex. How can such very different circumstances share a reliance on norms?

Shasta County and the non-union ILM, although differing in significant ways, are fundamentally similar in salient respects. Both are repeat-play situations (occasional in Shasta County, but continuous in the ILM) in which the parties can enforce sanctions against each other. In Ellickson's terminology, both are close-knit groups.⁵⁴ In both, the stakes for individual plays are sufficiently low

⁵⁴ See ELLICKSON, *supra* note 1.

so that the benefits of defecting will not likely outweigh the ongoing value of the relationship to the parties. In other words, the agreements will self-enforce.

The fundamental difference between Shasta County and the non-union ILM—the fact that Shasta County involves market transactions governed by norms while the non-union ILM exists within a firm—tracks a very different question than that of norms versus contracts, and does not undermine the analysis of when norms will likely work. Rather, the difference tracks the distinction between organizing economic transactions through the market versus organizing transactions within a firm.⁵⁵ The dominance of norm-governed transactions within firms should not be surprising: Coase himself cited bringing activities within the firm as a way of internalizing externalities in a high transaction cost world.⁵⁶

⁵⁵ In terms of the extent of economic activity, within-firm norms are clearly the most important home of norm-directed activity. Thus, we think that Shasta County is the marginal case and the ILM the central case for the use of norms.

Moreover, the industrial organization literature dealing with the theory of the firm points to exactly these factors—high frequency, connectedness, and complexity—as key components of an economic transaction that is best coordinated by bringing the activity inside the firm. More generally, firms organize activity when resorting to the intrafirm authority relationship costs less than writing contracts and using the market. In the ILM, it would be prohibitively costly to write individual contracts to organize tasks.

As Milgrom and Roberts have argued, intrafirm organization is superior (1) the greater the specificity of the investments involved in the transaction; (2) the higher the frequency and the longer the duration of the transaction; (3) the more uncertain and more complex the needed performance; (4) the more difficult the measurement of performance; and (5) the more connected the tasks. See PAUL R. MILGROM & JOHN ROBERTS, *ECONOMICS, ORGANIZATION, AND MANAGEMENT* (1992).

⁵⁶ See Ronald Coase, *The Nature of the Firm*, 4 *ECONOMICA* 386, 394 (1937). Indeed, as we argue elsewhere, the advantages of norms provide the best answer to Coase's question on the difference between firms and markets. Coase argued that firms and markets differ in that firms use contracts and authority relationships, with the choice driven by the relative transaction costs. See *id.* at 391-92. In their important contribution to the debate, Alchian and Demsetz pointed out that, contrary to Coase, there are often contracts within the firm and authority relationships within the market. See Armen A. Alchian & Harold Demsetz, *Production, Information Costs, and Economic Organization*, 62 *AM. ECON. REV.* 777, 794 (1972) (describing the classical firm as a contractual structure in which no authoritarian control is involved); see also Steven N.S. Cheung, *The Contractual Nature of the Firm*, 26 *J.L. & ECON.* 1, 3 (1983) (arguing that the word "firm" is used to denote the organization of activities under contractual arrangements that are unlike the arrangements in ordinary product markets); Oliver Hart, *An Economist's Perspective on the Theory of the Firm*, 89 *COLUM. L. REV.* 1757, 1761 (1989) (noting the conceptual weakness in the dichotomy between the role of authority within the firm and the role of consensual trade within the market); Benjamin Klein et al., *Vertical Integration, Appropriable Rents, and the Competitive Contracting Process*, 21 *J.L. & ECON.* 297, 297 (1978) (exploring the

If what unites Shasta County and the non-union ILM is that, in both cases, self-enforcing contracts are available, how, then, should we understand the union ILM? On the one hand, the union ILM is like the non-union ILM in that it is brought inside the firm and is characterized by the same types of transactions. On the other hand, it relies more on contracts and law than on norms.

To answer this question, we turn first to the simpler case of commercial relational contracts and norms. As discussed above, the paradigm case for analysis in the law and economics literature on relational contracting is the case of the coal mine next to the power plant. For these parties, it is in their joint interest, *ex ante*, to enter into a long-term contract; yet the many contingencies that can arise over a twenty-year contract render it impossible to write a fully specified contract. In addition, because the sunk investments in the mine and power plant give rise to quasi-rents and the potential for opportunistic behavior, neither side can go forward without contract protection.

The relational contract and non-union ILMs are thus similar in that match-specific investments motivate both of them. In the prototypical relational contract, however, individual transactions are often single shot and usually high stakes. In such circumstances, even though the relationship is long term, parties cannot rely on norms—the chances of opportunistic behavior are too great. Instead, the parties write detailed contracts. Contracts, even incomplete ones like those written between the coal mine and the power plant, have a greater capacity to bind parties than do norms. When the bonds of norms are too weak to neutralize the threat of opportunistic behavior, the expense of contracts is justified.

To realize the benefits of the long-term nature of the match, the parties draft long-term contracts with open terms that they later fill by renegotiation and, when that fails, third party adjudication (either by a court or by an arbitrator). Since the relationship involves a long-term commitment, norms will likely develop. Indeed, the parties will typically incorporate these norms into the

possibility of post-contractual opportunistic behavior in the market system). The analysis in the text makes clear that the difference between firms and markets is that norms, working within the hierarchical structure of the firm, are—as a descriptive matter—the typical mode of organization within firms, while explicit contracting (even if sometimes incomplete as in the case of long-term relational contracts) is the typical mode of organization throughout markets. According to this analysis, the choice between organizing production in firms or through markets amounts to a choice between norms and contract.

contract by reference. The norms, however, need not self-enforce. Instead, the arbitrator or judge resolves conflicts by interpreting the contract and filling the inevitable gaps.⁵⁷

The critical factors driving the choice between norms and contracts cannot be match-specific investments, asymmetric information, or long duration. All are present in the non-union ILM and the relational contract and largely absent in Shasta County. Rather, the combination of large stakes and infrequent or single play drives the parties towards a contractual solution. In these cases, norms do not have sufficient binding power to constrain opportunistic behavior, so the parties must resort to contract. In short, the need to constrain defection drives the choice.

How should this analysis affect our understanding of union CBAs? From a contractual perspective, union CBAs share many similarities with commercial relational contracts. Both are long term and use third party adjudication intensively, most often through arbitration where the neutral party attempts to fill the gaps in the contract with the norms of the parties. But does the choice between the norm-dominated non-union sector and the incomplete contract/CBA solution of the union sector parallel the choice between norms and relational contracts more generally? In other words, are CBAs relational contracts of the sort that we see between the power plant and the coal mine?

From one perspective, the answer seems to be negative. If what distinguishes the domain of commercial relational contracts from norms is the existence of high stakes, single-shot transactions, then CBAs fall on the norms side of the line: in the union, as in the non-union sectors, there are numerous small transactions among the same parties.

But, from a larger perspective, union ILMs more closely resemble relational contracts than do non-union ILMs. There are two potential explanations for the choice of the union ILM over the non-union ILM. First, if norms imperfectly self-enforce, there will be a class of players for whom norms do not provide sufficient limitations to protect the parties' match-specific investments. In the ILM context, if an employer proves untrustworthy, the parties may opt for the contractual solution over the cheaper but less powerful normative solution. Hence, the union ILM model fits well if the employer is untrustworthy—a view widely held by labor unions. This

⁵⁷ See Goetz & Scott, *supra* note 34, at 1117.

analysis would include, but would not be limited to, the end-game problem that occurs when the ILM lacks sufficient repeat interaction.

The second explanation rests on the notion that the union ILM is chosen in order to achieve wage premia. This is a complex argument because it interacts with the statutory protection provided by the NLRA. Here, again, the more powerful constraints imposed by contract drive the choice of contracts over norms. Employers are unlikely to be good and faithful norm players when the outcome disfavors them in a redistributive sense. The contract also prevents individual employees from defecting from the union norm by offering to undercut the wage premia. According to either explanation, the greater binding power of contracts over norms and the greater need for constraint drive the choice.

3. But Why a Bipolar Distribution?

Assuming that contract's greater ability to bind motivates the choice between norms and contract, why does there seem to be a bipolar distribution of solutions, in both the ILM and commercial contexts? Why are commercial relational contracts so detailed? Why is there such a sharp distinction between the unwritten and unenforceable non-union ILM and the detailed CBA of the union ILM?

The best explanation for the bipolar choice between norms and contracts is the similarly bipolar choice between self-enforcement and third party enforcement. Once it is worthwhile to the parties to write a contract, for whatever reason (to protect sunk investments in the power plant/coal mine context, to bind the untrustworthy employer, or to pursue union wage premia in an ILM), they must try to write contracts that are as explicit as possible. By opting for third party enforcement and thereby foregoing the benefits of a self-enforcing norm-governed interaction, the parties expose themselves to third party error and opportunistic behavior. The two available mechanisms for minimizing this vulnerability are, first, the choice of the third party decisionmaker (that is, providing for expert arbitration), and, second, detailed terms to guide the decisionmaker. Not surprisingly, one finds these provisions in both long-term commercial relational contracts and union CBAs.⁵⁸

⁵⁸ An additional factor in the labor context that discourages contractual forms lying between the union and non-union models may be the NLRA's restrictions on

B. *The Benefits of Choice Among Legal Regimes*

In non-union ILMs, norms dominate. In union ILMs, detailed contingent-state contracts with specified governance structures to fill in gaps prevail. One cannot know, a priori, which approach is more efficient. Perhaps one model is more efficient for some contexts while the other model works better elsewhere. In Shasta County, although not much emphasized, norms and law likewise coexist happily. Norms dominate with regard to trespass and fencing matters. One can be certain, however, that when a rancher buys or sells a ranch or property, he relies on enforceable contracts.

The ability to choose between norms and contract provides substantial benefits. Consider the ILM context. In the preceding analysis, one of the explanations for choosing the union ILM over the non-union ILM was the presence of an untrustworthy employer. As long as norms do not perfectly self-enforce, the possibility of opportunistic behavior will exist. The existence of the union ILM option can therefore be expected to constrain opportunistic behavior in the non-union ILM.

If one views the NLRA system as providing employers and firms with a choice (adopt the NLRA specified union ILM structure or remain non-union),⁵⁹ a competition between the two forms emerges: the non-union ILM form that uses self-enforcing norms versus the union ILM form that uses contract and statutory law to provide third party enforcement of agreements. In such competition, the ILM form that can best maximize joint profits will win labor market share.⁶⁰ Likewise, providing employees with an additional weapon to sanction employers, makes firm rent-seeking less profitable. Hence, although one might exalt the non-union ILM as a triumph of norms over law, some of the credit may go to the law itself: even the parties to the non-union ILM agreement bargain in the shadow of the law. The ability to choose the union ILM

employer involvement in union activity. *See* Electromation, Inc., 309 N.L.R.B. 990 (1992) (stating that a finding of domination may be appropriate if a union has no effective existence independent of the employer's active involvement).

⁵⁹ *See* Wachter & Wright, *supra* note 25, at 256.

⁶⁰ Traditional labor law scholars reject this view, arguing that the NLRA not only created a union ILM form, but favored its acceptance. *See generally* WEILER, *supra* note 14. Whether this view accurately reflects the original legislative intent in the 1930s is debatable, but claiming that it is a positive description of the NLRA today is implausible.

option provided by the NLRA may, itself, have encouraged ILM parties to behave neighborly and not violently.

Against this background, consider the NLRA's protection of union organization activities as an exception to the general doctrine of at-will employment.⁶¹ The untrustworthy employer, that is, the bad actor who flouts the norm of "no discharge without cause" will likely bear a cost for breaching the norm through reputational and other internally administered sanctions. The norm will be less credible in the future, which will affect the employees' willingness to invest in match, in the allocation of the joint surplus, and in the employees' willingness to rely on a norm, as opposed to an enforceable contractual provision, to protect against opportunistic behavior. Nonetheless, because self-enforcing norms may be imperfect in that they do not completely self-enforce, such behavior may still be worthwhile to the untrustworthy employer, both in the short run as well as in the long run. If so, such an employer will also have an incentive to discharge employees who try to organize a union to constrain the employer's behavior with the stronger tools of contract.

The legal rule against discharges for engaging in union organizing activities is important because such behavior threatens the choice between union and non-union workplaces, and thus the competition between the two regimes. Just as, in the corporate context, the restrictions that antitakeover statutes place on reincorporation provide a disincentive for managers to manage well, so, too, would interfering with the ability of non-union ILMs to "reincorporate" as union ILMs (and vice versa) hinder the process that renders the norms of such ILMs joint-profit maximizing.⁶² Moreover, it is clear why such a rule must be a legal rule and not a norm: it only matters in situations where the firm's incentives to breach the norm outweigh any cost it is likely to face, and therefore the normal incentive compatibility disappears. When one confronts a player for whom the norm is not self-enforcing, the benefit of constraining such a player may outweigh the resulting problems of observability and verifiability, even though using law to constrain

⁶¹ It is well settled that firing an employee for union organizing activities is an unfair labor practice. *See, e.g.*, *NLRB v. Gissel Packing Co.*, 395 U.S. 575, 617 (1969) (noting that § 8(a)(1) of the NLRA "prohibits interference, restraint or coercion of employees in the exercise of their right to self-organization").

⁶² For an analogous argument in the corporate context, see ROBERTA ROMANO, *THE GENIUS OF CORPORATE LAW* 67 (1993) (noting that antitakeover statutes entrench management).

the player will run into the problems that gave birth to the norm-based system in the first place, for example, asymmetry of information.

Although the integrity of the union ILM alternative depends on legal intervention, the integrity of the non-union ILM depends on non-intervention. Regulation of properly working self-enforcing norms renders non-union ILMs less efficient. Despite the hopes of the traditional legal scholars, such intervention will not encourage greater unionization. The external labor market—through the use of subcontracting, personal service contracts, and other such relational contracts—is a closer substitute than the union ILM for parties seeking to make match investments with the least amount of legal intervention.

The same benefits of choice between norms and law exist, at least implicitly, in Shasta County. The ranchers are keenly aware of the existence of legally enforceable contracts: they surely encounter such beasts when they buy and sell cattle, ranches, and cars. Their decision to forego the use of contracts in resolving questions of cost allocation of fencing among adjoining ranches does not argue for the irrelevance of law. Rather, their choice only argues for the relative superiority of norms in the particular context.

In this connection, the tale of Frank Ellis becomes important. The informal, norm-based sanctions failed to constrain his behavior. Only his fortuitous bankruptcy and departure solved the problem. Had he remained in Shasta County, his neighbors would have had three options. First, his neighbors could have done nothing—an adequate response in the face of minimal harm. Second, the neighbors might have engaged in more aggressive (and likely illegal) self-help, risking the escalation of violence. Third, the neighbors may have rejected norms in their dealings with Ellis and insisted instead on law and contracts.

For all the interesting and important cases, the first option (do nothing because the stakes are too small to bother with) is unavailable. The choice between the second and the third options implicates the Hobbesian thesis that Ellickson so quickly dismisses—that without a Leviathan to issue and enforce commands, civil strife would ensue.⁶³ The history of the American labor movement and

⁶³ See ELLICKSON, *supra* note 1, at 4-5 ("In so doing, Coase repeated a blunder that dates back at least to Thomas Hobbes. According to Hobbes, without a Leviathan (government) to issue and enforce commands, all would be endless civil strife. The Shasta County evidence shows that Hobbes was much too quick to equate anarchy

the success of the NLRA in ending the cycle of violence between employers and employees provide strong evidence of the importance of law in constraining and improving norm-governed interactions. By contrast, the apparent independence of law and norms in Shasta County may simply be an artifact of the small size and unimportance of the transactions at issue.

If this assertion is right, then the apparent independence of norms from law—whether in Shasta County or in the non-union sector—is more apparent than real. If norms work in Shasta County or the non-union ILM, they most likely function, at least in part, because of the alternative contractual regime that lurks in the background.

CONCLUSION

This brings us to our rather tentative conclusions. Norms are a species of self-enforcing relationship, and the choice between norms and contract is the choice between self-enforcement and third party enforcement. When parties choose to adhere to norms by not writing contracts, the efficiency minded court will stay out, even when the norms are joint-profit maximizing. The reason for this judicial abstention is that the efficiency of a norm-dominated regime lies in its self-enforcing character. Attempting to enforce norms in court is meddlesome and negates the very promise that norms offer the parties.

This is a very important conclusion because it contradicts a deeply felt and widely held intuition that if the patterns of behavior that contracting parties follow differ from the legal rules, the legal rules should change. The extended discussion of the employment-at-will doctrine shows that the existence of a norm and a substantively inconsistent legal rule may, in fact, be optimal. More generally, even if one is unconvinced of the virtues of the employment-at-will doctrine, the analysis shows that an additional step is involved, and an additional argument required, to move from the identification of a difference between a norm and a legal rule to the conclusion that the law should change or that the courts should enforce the norm. This conclusion does not automatically follow.⁶⁴

with chaos. Many entitlements, especially workaday entitlements, can arise spontaneously. People may supplement, and indeed preempt, the state's rules with rules of their own.⁷⁾

⁶⁴ In a context in which norms and law coexist and compete, the old claim that law emerges from norms and custom, and therefore (on conceptual grounds?) cannot or

The coexistence of self-enforcing systems of norms and state-enforced contracts poses the question of when parties will choose one over the other. Parties will choose contracts when the binding power of norms fails to constrain opportunistic behavior, as in the paradigmatic long-term, high stakes contracting of the relational contracts literature. In such circumstances, the courts have no choice but to figure out which norms the parties intended to incorporate and which they did not. This is a separate although related subject and well beyond the scope of this Article.⁶⁵

From this perspective, the legal regimes of norms and contracts, far from existing in isolation from one another, provide interacting alternatives. The competition between activities governed by norms and those governed by contracts must be protected because it improves the quality of each realm and protects the participants from opportunistic behavior.

The coexistence of norms and laws does not diminish the fact that laws trump norms, both in fact and by necessity. To say, as Ellickson appears to in some places, that norms do not form in the shadow of the law is troublesome and we believe incorrect. The shadow of the law is critical to well-functioning norms. Once parties enforce norms by unlawful self-help, they start down a very slippery slope. The history of labor unions in the United States and their attempts at organizing and bargaining with firms provides an example.

Here again, the story of Shasta County's Frank Ellis illuminates the analysis. Definitionally, Ellis was a bad player because he did not play by the norms. He may even have been a bad person, if he did not play by the law, but failure to play by the norms is not proof of that. Once the good players retaliate by engaging in unlawful self-help, however, they become bad persons. The essence of the efficiency of norm-governed interactions lies in their voluntary

perhaps should not conflict with it, can no longer be persuasive. Once law assumes a life of its own, however it got there, it competes with norms and customs as an alternative mechanism to structure productive relationships (and many others as well). The appropriate relation between the two thereby becomes of central importance.

⁶⁵ As Alan Schwartz has argued, the best guide in this task is whether the terms depend on verifiable or nonverifiable information. This asymmetry of information, which is part of what drives the choice to use norms over contracts, likewise constrains the courts in interpreting contracts, notwithstanding the fact that other considerations push the parties to write contracts. See Schwartz, *supra* note 39, at 282-83.

character. If Ellis does not agree to join in the norms, for whatever personal reasons, he becomes indistinguishable from a third party. If the norms injure him, he has a right to ask for legal remedy. This is typically not a problem in the United States today, but only because norms operate so clearly in the shadow of the law. Only Frank Ellis's exit from Shasta County made it unnecessary for the law to enter the fray against the bad persons, whether or not they were bad or good norm players.

These conclusions leave open some important questions. What sort of evidence is necessary before one concludes that the parties are operating according to norms and not law? What about norms that impose costs on third parties while benefiting the participants? What is the cost of incorrectly enforcing norms? And, finally, when should courts read contracts as incorporating norms, and what norms should they incorporate?⁶⁶

⁶⁶ If we correctly conclude that the choice to write a contract amounts to a choice between norm-governed relationships and third party enforcement, then clear demarcation may be useful. Although the question transcends the scope of this Article, contracting formalities, whether they be of the seal type or the statute of frauds type, can serve this purpose.