Yale Law School
Yale Law School Legal Scholarship Repository

John M. Olin Center for Studies in Law, Economics, and Public Policy Working Papers

Yale Law School Other Scholarship

4-25-2003

Contract Theory and the Limits of Contract Law

Alan Schwartz
Yale Law School

Robert E. Scott
Yale School of Management

Follow this and additional works at: http://digitalcommons.law.yale.edu/lepp_papers

Part of the Contracts Commons

Recommended Citation
Schwartz, Alan and Scott, Robert E., "Contract Theory and the Limits of Contract Law" (2003). *John M. Olin Center for Studies in Law, Economics, and Public Policy Working Papers.* Paper 275.
http://digitalcommons.law.yale.edu/lepp_papers/275

This Article is brought to you for free and open access by the Yale Law School Other Scholarship at Yale Law School Legal Scholarship Repository. It has been accepted for inclusion in John M. Olin Center for Studies in Law, Economics, and Public Policy Working Papers by an authorized administrator of Yale Law School Legal Scholarship Repository. For more information, please contact julian.aiken@yale.edu.

CONTRACT THEORY AND THE LIMITS OF CONTRACT LAW

Alan Schwartz[*] *& Robert E. Scott*[**]

I.INTRODUCTION.....2
II. JUSTIFYING AN EFFICIENCY THEORY OF CONTRACTS 11
A. What Firms Maximize.....11
B. Why the State Should Help Firms.....16
III. THE ENFORCEMENT FUNCTION 17
A. Enforcement Often is Unnecessary 18
B. Encouraging Relation-Specific Investment.....20
C. Contracting to Avoid Disruption: The Case of Volatile Markets 25
D. Enforcement and Duress 28
IV. THE INTERPRETATION FUNCTION..... 31
A. The Relevant Interpretive Question 31
B. Two Interpretive Issues: Problems of Meaning and of Language.....33
C. The Parties' Preferences Regarding Interpretive Styles 37
1. The Continuous Payoff Case ..... .38
2. The Invariant Payoff Case 42
3. Summary 47
D. Private Languages, Linguistic Defaults and the Parole Evidence Rule 48
1. The Preferred Linguistic Default 48
2. The Parol Evidence Rule 55
3. Course of Performance Evidence 57

---

[*]Sterling Professor of Law, Yale Law School; Professor, Yale School of Management.

[**] Lewis F. Powell, Jr. Professor and William L. Matheson & Robert A. Morgenthau Distinguished Professor, University of Virginia School of Law.

This paper benefitted from comments received at workshops at the Law Faculty, Cambridge, England and at the Pennsylvania, Texas, Toronto, Virginia and Yale Law Schools. We also are grateful to Bruce Ackerman, Jules Coleman, Sam Issacharoff, John Jeffries, Jason Johnston, Paul Mahoney, Tom Nachbar, Paul Stephan, William Stuntz and George Triantis for helpful comments.

V. THE LEGAL DEFAULT PROJECT 59
A. The Case For Defaults 61
B. The Cost Concern..... 63
1. Default Rules 63
2. Default Standards 67
C. The Asymmetric Information Concern 70
D. Summary 73
VI. MANDATORY RULES 75
A. Parties Cannot Ban Modifications.....76
B. Parties Must Accept Substantial Performance 79
C. Parties Cannot Agree to Penalties 81
VII. CONCLUSION 83

I. INTRODUCTION

Contract law has neither a complete descriptive theory, explaining what the law is, nor a complete normative theory, explaining what the law should be. These gaps are unsurprising given the traditional definition of contract as embracing all promises that the law will enforce. Even a theory of contract law that focuses only on the enforcement of bargains must still consider the entire continuum from standard form contracts between firms and consumers to commercial contracts between business firms. No descriptive theory has yet explained a law of contract that comprehends such a broad domain. Normative theories that are grounded in a single norm -- such as autonomy or efficiency -- also have foundered over the heterogeneity of contractual contexts to which the theory is to apply.[1] Pluralist theories attempt to respond to the difficulty that unitary

[1] See Michael Trebilcock, THE LIMITS OF FREEDOM OF CONTRACT (1993). Autonomy theories thus require elastic notions of consent in order to regulate the full scope of contracting behavior with one norm. Peter Benson, *Abstract Right and the Possibility of a Nondistributive Conception of Contract: Hegel and Contemporary Contract Theory*, 10 Cardozo L. Rev. 1077 (1989); Peter Benson, *Contract* in *A Companion to Philosophy of Law and Legal Theory, (*Dennis Patterson ed., 1996); Peter Benson, *The Idea of a Public Basis of Justification for Contract,* 33 Osgoode Hall L. J. 273 (1995); Randy Barnett, *A Consent Theory of Contract*, 86 Colum. Rev. 269 (1986); Randy Barnett, *The Sounds of Silence: Default Rules and Contractual Consent*, 78 Va. L.

normative theories pose by urging courts to pursue efficiency, fairness, good faith and the protection of individual autonomy. Such theories need, but so far lack, a meta principle that tells which of these goals should be decisive when they conflict.[2] We attempt to make progress here with a more modest approach -- to set out and defend a normative theory to guide decisionmakers in the regulation of business contracts.[3]

The theory's affirmative claim, in brief, is that contract law should facilitate the efforts of contracting parties to maximize the joint gains (the "contractual surplus") from transactions. The theory's negative claim is that contract law should do nothing else. Both claims follow from the premise that the state should choose the rules that regulate commercial transactions according to

---

Rev. 821 (1992). Efficiency theories tend to have a more limited scope. Positive articles analyze broad doctrinal patterns in the attempt to find fundamental consistency between these patterns and the efficiency norm, but the authors do not purport to provide a fully descriptive theory of contract law. See, e.g., Charles J. Goetz & Robert E. Scott, *Enforcing Promises: An Examination of the Basis of Contract*, 89 Yale L. J. 1261 (1981); Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 Yale L. J. 729 (1989). Normative economic theories, on the other hand, typically evaluate discrete doctrines by the efficiency norm. See e.g., Charles J. Goetz and Robert E. Scott, *Liquidated Damages, Penalties and the Just Compensation Principle*, 77 Colum. L. Rev. 554 (1977); Alan Schwartz, *The Case for Specific Performance*, 89 Yale. L. J. 271 (1979); Christine Jolls, *Contracts as Bilateral Commitments: A New Perspective on Contract Modification*, 26 J. Leg. Stud. 203 (1997); Robert E. Scott, *The Case for Market Damages: Revisiting the Lost Profits Puzzle*, 57 U. Chi. L. Rev. 455 (1990).

[2] The problems that pluralist theories without meta norms pose are nicely illustrated in Melvin Eisenberg's effort, that purports to solve the "broad scope of contract problem" by proposing overlapping sets of norms. See Melvin A. Eisenberg,, *The Bargain Principle and its Limits,* 95 Harv. L. Rev. 741 (1982); Melvin A. Eisenberg, *The Theory of Contracts* in THE THEORY OF CONTRACT LAW: NEW ESSAYS (Peter Benson ed. 2000). For example, Eisenberg's schema restricts the domain of freedom of contract by norms of reciprocity, trust, and fairness. He recognizes that this "multi-value" approach can generate conflicting social propositions. When conflicts actually occur, "the lawmaker must make a legal rule that gives a proper weight and role to each of the conflicting values or goals in the context at hand." Further, "when social propositions conflict, the legislature must exercise good judgment concerning the weight or role to give to each proposition in the issue at hand." Eisenberg, *The Theory of Contracts* at 243-44. Eisenberg recognizes that his theory lacks a metric that would tell the lawmaker just how to give the proper "weight and role" to each social proposition or value when conflicts occur. Since courts or legislatures are likely to be involved when the relevant social propositions or values arguably favor more than one party type or interest group, pluralist theories such as Eisenberg's tend to be least helpful when they are most needed.

[3]In a thoughtful critique of autonomy and efficiency theories of contract, Michael Trebilcock concludes that both theory types are "valid in their own right", but without a "meta-theory that weights and ranks these various values", both values should be pursued in various social contexts according to the relative competency of different legal institutions to perform effectively. See Trebilcock, supra note 1, at 248. This article takes up Trebilcock's invitation and proposes a normative theory that fits business contracts, the subsidiary category of contractual relationships that the law most affects.

the criterion of welfare maximization.

A simple categorization of the universe of bargaining transactions will clarify the domain of our theory. A transaction involves a seller (whether of goods or services) and a buyer. Parties to transactions can be partitioned into individuals and firms. This yields four transactional categories: (1) A firm sells to another firm; (2) An individual sells to another individual; (3) A firm sells to an individual; and (4) An individual sells to a firm. Category 2 contracts, between individuals, are primarily regulated by family law (antenuptial agreements and divorce settlements) and real property law (home sales and some leases). Few litigated contracts between individuals are regulated by the rules of contract law. Category 3 contracts, between a firm as seller and an individual as buyer, are primarily regulated by consumer protection law, real property law (most leases) and the securities laws. Category 4 contracts, between an individual as seller and a firm as buyer, commonly involve the sale of a person's labor, and are regulated by laws governing the employment relation. The rules in Article 2 of the Uniform Commercial Code and the provisions of the Restatement (Second) of Contracts constitute what commonly is meant by contract law. These provisions are primarily invoked to resolve disputes arising under Category 1 contracts, between firms. Our theory applies only to these contracts, and thus has important implications for the content of the UCC and the common law of contracts.

Category 1 contracts, however, can be partitioned into two subcategories. Some parties obviously are sophisticated economic actors (i.e., the General Electric Corporation). Other parties function in commercial contexts but have many of the characteristics of ordinary persons(i.e., a gift shop owned and run by a retired teacher). Any effort to analyze contracts between "firms" thus confronts a boundary issue – to define a firm for purposes of the theory. We draw this boundary here by defining a Category 1 firm as (a) an entity that is organized in the corporate form and that has five or more employees; (b) a limited partnership; and (c) a professional partnership such as a law or accounting firm. These economic entities apparently understand how to make business contracts, and the theory we develop here applies only to contracts between two such firms. The extent to which this Article's conclusions hold when one or both of the parties to a commercial contract falls on the other side of our boundary awaits further work.

Firms that maximize profits face the canonical "contracting problem" of ensuring both efficient ex post trade and efficient ex ante investment in the subject matter of the contract.[4] Parties trade efficiently when, and only when, the value of the exchanged performance to the buyer exceeds the cost of performance to the seller. Parties invest efficiently when their actions maximize a deal's expected surplus. Many observers would agree that contract law should attempt to facilitate efficient trade and investment. The novelty of our theory lies in its systematic development of the implications of this goal and in its claim that contract law should restrict itself to the pursuit of efficiency alone (for Category 1 contracts).

Four objections may be made to the claim that contract law should restrict itself to encouraging efficient trade and investment. First, one can argue that firms sometimes do not maximize profits and, owing to the systematic cognitive errors made by the people who run them, are incapable of doing so should they try. Thus, a law that presupposes profit maximization will be misguided. Second, firms that maximize profits sometimes do bad things – pollute the environment , for example -- that the law should attempt to deter. Third, the state should promote fairness in contracting in addition to efficiency. And, finally, the state should pursue distributional goals although they sometimes conflict with efficiency.

These objections should trouble a unitary efficiency approach to the regulation of all contract types, but we will argue that the objections have little force when Category 1 contracts *alone* are considered. Thus, we will argue, firms and markets are structured so as to minimize the likelihood of systematic cognitive error by important decisionmakers within the firm. Cognitive error, that is, is more likely to afflict Category 2 and 3 contracts than Category 1 contracts. Further, the bad things that firms do commonly entail imposing costs on third parties, such as creating environmental harms or erecting barriers to entry. These behaviors – the creation of negative externalities – are regulated by the environmental and antitrust laws. An analysis of contract law as such therefore can assume the absence of externalities. Finally, it usually is futile

[4]Legal scholars commonly refer to investment in the contract's subject matter as "reliance". We use reliance and the economist's term "investment" interchangeably.

to pursue contractual fairness when firms are permitted a large measure of contractual freedom. This is because firms will contract away from fair legal rules that do not maximize joint surplus. In sum, efficiency is the only institutionally feasible and normatively attractive goal for a contract law that regulates deals between firms.[5]

An efficiency theory restricted to contracts between firms (as firms are defined above) has four major implications for contract law. To understand the first implication, realize that contracts often would be performed even if there were no legal sanction for breach. These contracts are "self-enforcing" in two senses. First, when parties contemplate making a series of contracts, neither party would breach an early contract if the gains from one breach are lower than the expected profit stream from future contracts that breach would cause to vanish. Second, neither party will breach if the gains are exceeded by the reputational sanction the market will exact. When contracts fall outside the self-enforcing range, legal enforcement is necessary to ensure performance in two principal cases: in volatile markets, when a party's failure to perform could threaten its contract partner's survival; and when contractual surplus would be maximized if one or both parties made relation-specific investments.[6] "Enforcement" includes more than simply requiring parties to perform, however. The enforcement function comprises policing contracts for fraud and duress, and the congeries of rules that encourage or facilitate performance, such as the damage rules. Perhaps a third of the sections in UCC Article 2 are enforcement rules under the definition here. The initial implication of our theory is that enforcement, when needed, is much the most important thing the state does. Put more starkly, a modern commercial economy

---

[5]As another example of the criticism that we sidestep here, Professor Melvin Eisenberg has criticized theories holding that contract law should maximize welfare alone on the ground that "these theories are impoverished ... because they exclude other important policy values, such as the value of keeping intimate and affective relationships free from the intrusion of state power". Melvin A. Eisenberg,*The Theory of Contracts*, in THE THEORY OF CONTRACT LAW: NEW ESSAYS 206, 238 (Peter Benson, Ed. 2000). This objection may have force as applied to Category 2 contracts, between persons, but seems irrelevant to the Category 1 contracts we analyze. Thus, contracts between General Electric and General Motors seem not to involve "intimate and affective relationships."

[6]A relation-specific investment is not fully "redeployable". As an example, assume that a seller purchases standard steel tubes to make a machine for the buyer. The seller's investment would be "general" if breach occurred before the seller began work on the tubes because the tubes could be resold on the market. The investment would become "relation specific" if breach occurred after the tubes had been fabricated into shapes that only the buyer could use; for then the transmuted tubes could only be resold as scrap, probably for less than their cost.

can function well with little more than honest courts and a set of enforcement rules. The rest is of second-order importance.

A court cannot enforce contracts, however, without a theory of interpretation that "maps" from the syntactic content of the parties' writing to the writing's legal implications. An interpretive theory that is grounded in efficiency holds, in contrast to the UCC and much modern scholarship, that textualist interpretations best suit parties to Category 1 contracts. Business firms, that is, prefer courts to adhere as closely as is possible to the ordinary meanings of the words the parties used, apply a "hard" parol evidence rule,[7] and honor "merger clauses" reciting that the parties intended their writing to be interpreted as if it were complete. A textualist theory of interpretation, however, will not suit all parties all of the time. Therefore, our second implication actually holds that textualist interpretation should be the default theory for Category 1 contracts. Courts should use narrow evidentiary bases when interpreting agreements between firms, but also should comply with party requests to broaden the base that is applicable to them. This implication is at variance with current law, which holds that interpretation is an issue for courts to decide and is conducted according to rules that parties cannot vary.

Contract law has more rules than are needed to perform the enforcement and interpretation functions. These rules, that regulate various aspects of the contracting relationship, commonly are defaults, controlling only when parties do not contract out. Creating good defaults is widely believed to be the principal function of a law of contracts. This belief is misguided because the state could create defaults that business firms would want only under quite stringent conditions. A good default *rule*[8] must apply in very few possible states of the world, be relatively simple in form, be efficient in a highly heterogenous set of circumstances, and not rely on information that courts cannot conveniently recover. A default *standard* should be written when parties do not

---

[7]A "hard" parol evidence rule treats writings that appear to be complete contracts *as* complete contracts. See TAN 81-89, infra.

[8]The decisionmaker specifies the content of a rule in advance. Thus, drivers cannot exceed a 55 mile per hour speed limit. The decisionmaker specifies the content of a standard ex post. Thus, parties must drive "reasonably" in the circumstances.

need, or it is too costly to provide them with, concrete guidance regarding the performance obligation. Because standards permit parties much latitude (the seller must deliver in a "reasonable" time), a good standard will confer discretion only when a party's likely actions under it will maximize joint rather than individual gains. Statutory drafters and courts, we will argue, often adopt default rules and standards that fail to satisfy these stringent conditions. This is itself inefficient because parties respond to bad rules or standards by contracting out of them. The creation of inefficient defaults thus raises business parties' contracting costs but does not otherwise affect their behavior. Our theory's third implication holds, in consequence, that the *effective domain* of business contract law is much smaller than is commonly thought. Another way to put this point is that the difficulty of creating good defaults makes much of what today is called contract law irrelevant to commercial life.

In addition to the many defaults, contract law contains a number of mandatory rules that are applied to contracts between firms as well as to contracts between firms and persons. The fourth implication of our efficiency theory is that many of the rules regulating business contracts should not be mandatory. We discuss a number of mandatory rules, including the interpretation rules, the modification rules and the rules relating to liquidated damage clauses. The justification for these rules apparently is a form of paternalism. The contract terms the rules override do not create externalities and are not unconscionable. Rather, contract law overrides terms that appear to decision makers to conflict with the parties true substantive intentions. We argue, to the contrary, that business firms have good reasons to adopt the terms that today are prohibited, and that a commitment to party sovereignty requires those reasons to be respected.

There are several reasons why an attempt to develop a general efficiency theory of business contracts is particularly salient now. First, the specification of a good contract law has become an important priority in many countries as they have made a commitment to markets. It is a consensus that a good contract law is a necessary condition for a modern commercial economy. It is less well understood just how such a law is supposed to function. Our article thus attempts to address concerns that have global implications.

A further reason motivating our decision to develop a theory of business contracts is that the building blocks for such a theory are only now becoming available. Contract theory has become one of the most significant fields in modern micro and industrial organization economics. Three recent Nobel prizes, to George Akerloff, Michael Spence and Joseph Stiglitz, were awarded largely for work in contract theory, but the field itself is less than thirty years old.[9] Moreover, much of the work in the field takes a mathematical form, and thus has not been easily accessible to nonspecialists. We draw heavily on contract theory to construct our normative theory of contracts.[10]

Finally, as we suggested earlier, the current state of contract law scholarship suffers from the absence of a successful theory of contract. Thirty years ago, Grant Gilmore described what he called the classic Willistonian model of contract law, a model grounded in formalist notions of the centrality of written agreements voluntarily exchanged between contracting parties, and that emphasized the limited role of the law in enforcing and interpreting these agreements.[11] According to Gilmore, this classical model owed more to Holmes' imagination than to a careful reading of the case law.[12] But whether this was so or not, Gilmore believed that the modern case law repudiated the model. The disjunction between the dominant scholarly view and the lived

---

[9]The work of these scholars is concisely summarized in Karl-Gustaf Lofgren, Torsten Perssons and Jorgen Weibull, *Markets with Asymmetric Information: The Contributions of George Akerlof, Michael Spence and Joseph Stiglitz*, 104 Scan. J. Econ. 195 (2002).

[10]Law and economics scholars such as Aaron Edlin, Ian Ayres and Jason Johnston have used contract theory in illuminating fashion when discussing particular legal rules. See, e.g., Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules,* 99 Yale L. J. 729 (1989); Jason Johnston, *Strategic Bargaining and the Economic Theory of Contract Default Rules,* 100 Yale L. J. 615 (1990). Aaron Edlin, *Cadillac Contracts and Upfront Payments: Efficient Investment Under Expectation Damages*, 12 J. L. Econ. & Org. 98 (1996). The genre of model that we and these scholars use has performed well in empirical tests. See P. A. Chiappori and B. Salanie, *Testing Contract Theory: A Survey of Some Recent Empirical Work*, Working Paper #2002-11, Institut National De La Statistique Et Des Etudes Economiques (2002).

[11]GRANT GILMORE, THE DEATH OF CONTRACT (1974).

[12]"The theory of contract, as formulated by Holmes and Williston, seems to have gone into its protracted period of breakdown almost from the moment of its birth" Id. at 57. But see, Richard E. Speidel, *An Essay on the Reported Death and Continued Vitality of Contract,* 27 Stan. L. Rev. 1161 (1975).

doctrine, he thought, produced incoherence.[13] Modern scholars commonly share Gilmore's rejection of Williston but have yet to disprove his incoherence thesis.[14] We, too, lack a theory of everything. Rather, the theory we develop here is Willistonian in spirit, but applies in a limited domain -- to contracts between firms that do not create externalities.

This limited scope permits our normative thesis to develop according to a particular logic. The market's social function is to maximize welfare, subject to distributional and fairness constraints. Firms, we show below, have incentives to choose the contracts and contracting strategies that will maximize the surplus their deals can create. Further, firms are more able than courts or statutory drafters at choosing efficient terms and strategies. It follows that, when externalities are absent, a contract law that regulates firms should be the contract law that firms prefer. The state, that is, should let the preferences of firms control because firms can better pursue the objective that both the state and firms share. The central organizing question this Article asks thus is: What contract law would commercial parties want the state to provide?

We proceed as follows. Part II defends the welfare maximization norm as applied to the contracts of sophisticated actors. In Part III, we describe commercial parties' first order preference: To have the state enforce contracts in order to protect relation-specific investments or to guard against especially disruptive market movements. Part IV argues that firms want the state to supply a theory of interpretation, but not the theory currently advanced by the UCC and the Restatement of Contracts. Rather, we defend a textualist theory of interpretation as the optimal default approach for business contracts. In Part V, we develop the restrictive conditions under which the state can create defaults terms that satisfy typical party preferences. Part VI analyzes a set of unjustifiable mandatory rules -- rules that rest on a misplaced view of the parties' interests. We conclude, in Part VII, that today's contract law is a series of category mistakes. Rules that are

---

[13]"Consider the Restatement's definition of consideration [in §75] taken in connection with its most celebrated section, § 90 (promissory estoppel)...One thing that is clear is that these two contradictory propositions cannot live comfortably together: in the end one must swallow the other up." Id. at 60-61.

[14]Eric A. Posner, *Economic Analysis of Contract Law after Three Decades: Success or Failure?* Yale L.J.(2002).

appropriate for contracts involving individuals (Categories 2 through 4 above) are too frequently applied today within the domain in which sophisticated parties function. Another way to put this conclusion is to remark that commercial law for centuries has drawn a distinction between mercantile contracts and others.[15] Modern scholars have not systematically pursued the normative implications of this ancient distinction, however. We attempt to cure this neglect by setting out the theoretical foundations of a law merchant for our time

II. JUSTIFYING AN EFFICIENCY THEORY OF CONTRACT

A. What Firms Maximize

It had been traditional to assume that firms attempt to maximize expected profits. The accuracy of this assumption recently has been challenged. An economic actor may not maximize wealth[16] for two reasons: (a) She is maximizing something other than her own wealth, perhaps because she is concerned with fairness; (b) She cannot maximize wealth in the context under study, perhaps because she is prone to cognitive error. These reasons apply much less to firms than to persons. A firm is directed by its owners, who often are shareholders. Shareholders prefer their firms to maximize profits that the shareholders then can consume or save. Firms thus will choose to maximize profits unless the managers who run them cannot be controlled by the shareholders who own them. In these cases, the managers may be maximizing their own earnings or perks at the expense of profit maximization.

No one doubts that managers sometimes successfully sabotage owners, but for two reasons we will assume here that managers obey shareholder instructions. First, managers

---

[15]Vestiges of this distinction exist in the few Uniform Commercial Code sections that regulate deals "between merchants" differently than deals between a merchant and a person. See, e.g., UCC §§ 2-104(3), 2-201(2), 2-207(2), 2-209(2), 2-603, and 2-609(2).

[16]Individuals are assumed to be risk averse while firms are assumed to be risk neutral. The utility function of a risk neutral party is linear in money: that is, the party values each additional dollar of wealth it may receive as much as it valued all previous dollars. Because monetary gains are coextensive with utility gains for risk neutral parties, we assume that firms maximize profits, a monetary measure.

sabotage shareholders either by diverting corporate wealth to themselves or by failing to take appropriate risks on behalf of the firm. Managers, however, have no incentive to degrade the quality of the contracts that they do write; after all, these contracts create the wealth that the managers later can divert. Second, the legal rules that attempt to deter bad manager behavior fall in the domains of the criminal, corporate and securities laws. Contract law should exploit this specialization by assuming that the agreements it regulates reflect the parties' maximizing choices.

Firms that attempt to maximize expected profits commonly do as well as their circumstances permit. This is because survivorship pressures tend to induce competence. These pressures take two forms: (i) Firms that systematically make bad economic decisions lose out in competition with profit maximizing firms. Hence, surviving firms generally can do what they set out to do; (ii) Employees who systematically make bad economic decisions are unlikely to be promoted to positions of responsibility. Hence, senior managers generally can do what they set out to do.[17] This is not to say that all firms all the time pursue profit maximizing strategies. But it is to say that owners and the market put systematic pressure on firms to behave optimally; hence, it is a plausible working assumption that firms rationally pursue the objective of maximizing profits.[18]

---

[17]In addition, many corporate executives have attended business school and also attend business school executive programs for working managers. It is a function of business education to teach people to make optimizing (rather than cognitively erroneous) decisions.

[18]Psychologists and economists have shown that persons make systematic cognitive mistakes in laboratory experiments when asked to solve specified individual decision problems. These experiments do not test a general theory of how people make decisions, and thus they raise an issue of external validity: it is an open question whether, or when, real world parties will behave as did the experimental subjects. Two scholars recently noted a consequence of this "lack of theoretical foundations": "the policy implications of BE [behavioral economics] are limited by its inability to predict circumstances in which anomalous behavior will arise (other than in those sorts of circumstances in which it has been observed before) or how it will respond to policy changes". Jessica L. Cohen and William T. Dickens, *A Foundation for Behavioral Economics* 92 Amer. Econ. Rev. 335 (2002). For recent, exhaustive analyses of the psychological literature and a skeptical view of its relevance for the law see Gregory Mitchell, *Taking Behavioralism Too Seriously? The Unwarranted Pessimism of the New Behavorial Analysis of Law*, 43 William and Mary L. Rev. 1907 (2002); Gregory Mitchell, *Why Law and Economics' Perfect Rationality Should Not be Traded for Behavioral Law and Economics Equal Incompetency*, 91 Geo. L. J. ___ (2002).

We provisionally view the individual decision experiments as not relevant to our project for three reasons. First, as we are in a world of speculation, we speculate that individuals in laboratories may perform worse than officers of firms because the experimental subjects had not been trained to make good decisions and were not subject

The assumption that each party to a contract wants to maximize its own profit does not itself imply that parties also want to maximize joint gains. Rather, a party may prefer a larger share of a smaller pie. Thus, parties appear sometimes to have an incentive to behave strategically at the expense of joint welfare maximization. On a deeper view, however, parties at the negotiation stage prefer to write contracts that maximize total benefits.[19] To see why, assume that each party's share of the contractual surplus is set exogenously. This assumption holds that a party cannot affect the size of its share of the parties' bargain by the (nonfraudulent) actions it takes during a negotiation. On this assumption, the parties will want only to maximize the total surplus.[20] To put this point in a contracting context, let parties contemplate making a simple sales contract for goods that the buyer values at $100 and that would cost the seller $80 to produce. Now assume that each party's share in the contracting surplus ($100 - $80) is fixed in advance at one half each. Then the price will be $90, and each party's profit would be $10. Of more importance, assume that the seller could make a $2 investment in the subject matter of the contract that would lower its production cost to $70. The seller would want to make this investment because then its share of the new $28 contractual surplus ($100 - $2 - $70) would be $14, a share that would be realized by a price reduction to $86. Similarly, the buyer has an incentive to make cost justified value increasing investments.

---

to the pressures to maximize that are described above. Second, recent evidence suggests that behavioral anomalies can be substantially mitigated or made to disappear when individuals are asked to perform as actors in firms -- See Jennifer Arlen, Matthew Spitzer and Eric Talley, *Endownment Effects Within Corporate Agency Relationships*, 31 J. Legal Stud. 1 (2002) –, or when the applicable institutions permit communication within a group of actors and require competition between groups -- see *Do institutions promote rationality? An Experimental Study of the Three-Door Problem*, Discussion Paper #2002-21 (University of St. Gallen, Sept. 2002). The last finding is consistent with experiments showing that when persons are required to function *in markets* rather than to perform individual tasks, the persons reach equilibria that are consistent with individually optimizing behavior. See Vincent P. Crawford, *Introduction to Experimental Game Theory*, 104 J. Econ. Theory 1, note 8 at 3 (2002) ("... repeated play of the same game often converges to equilibrium no matter what subjects are thinking."); Dhananjay K. Gode and Shyam Sunder, *What Makes Markets Allocationally Efficient?*, 112 Q. J. Econ. 603 (1997); Same authors, *Allocative Efficiency of Markets with Zero Intelligence Traders: Markets as a Partial Substitute for Individual Rationality*, 101 J. Pol. Econ. 119 (1993); Vernon L. Smith, *Rational Choice: The Contrast Between Economics and Psychology*, 99 J. Pol. Econ. 877 (1991). Individuals acting on behalf of firms make contracts in markets.

[19]Part III, infra, shows that parties need the law's help to deal with post-contractual opportunism.

[20]Put simply, if a party is to receive a fixed 20% of a joint gain, it would *always* prefer the joint gain to be $200 rather than $100.

It remains for us to show that parties' bargaining shares actually are set exogenously. Bargaining power is a function of two factors. The first factor is the parties' relative patience. The more patient bargainer will reject offers it dislikes to wait for more favorable offers, while the less patient bargainer will accept relatively unfavorable offers just to get a deal.[21] The second factor is each party's disagreement point. In order for parties to agree at all, each party must do at least as well in the deal at hand as it could do elsewhere. Therefore, if party A has many good opportunities should it fail to agree with party B, then party B will have to give party A a good deal in order for the parties to contract. On the other hand, if party B is party A's only hope, A will accept a less favorable offer.[22]

A business party's patience is a function of its ability to finance its projects. Firms that have capital or convenient access to capital can be more patient than firms that need revenue immediately to survive. Parties ordinarily cannot affect the access of prospective contract partners to the capital market. Moreover, one party to a possible contract ordinarily cannot affect the other party's alternative business opportunities (its disagreement point). Thus, each potential contract partner will realize that its share of the maximum surplus the parties could generate jointly has already been fixed before any contract is signed. And this implies a preference by each party to contract so as to maximize the size of the pie.

---

[21]A party's discount rate measures his patience: the higher is a party's discount rate, the more highly the party values current dollars than future dollars. Parties with high discount rates thus are impatient bargainers: they want their share of the surplus now. A party for whom current dollars are relatively less important -- a party with a low discount rate -- suffers less from delay and, as a result, is more willing to reject low current offers. Hence, patient parties do well when bargaining with impatient parties, who will reduce their demands in order to reach agreement quickly.

[22]This paragraph summarizes the Nash bargaining game with disagreement points that function as threat points. In this game, each party receives a payoff that equals the payoff she would receive in her next best alternative plus an exogenously determined share of the surplus from concluding the bargain. The game is widely used in the contract theory literature. See, e.g., OLIVER HART, FIRMS, CONTRACTS, AND FINANCIAL STRUCTURE (1995) and Oliver Hart and John Moore, *Property Rights and the Nature of the Firm*, 98 J. Pol. Econ. 119 (1990). Predictions of bargaining outcomes using this game have received substantial support in the experimental literature. See, e.g., Joep Sonnemans, Hessel Oosterbeek and Randolph Sloof, *The Relation Between Asset Ownership and Specific Investments*, 111 The Econ. J. 791 (2001); Kenneth Binmore, Alexander Shaked and John Sutton, *An Outside Option Experiment*, 104 Q. J. Econ. 753 (1989). An accessible treatment of Nash bargaining with threat points is in JOEL WATSON, STRATEGY: AN INTRODUCTION TO GAME THEORY 172-78 (2002). We use this bargaining game here and in the analyses below.

This result is contrary to the common view. It is widely believed that parties exercise bargaining power by requiring weaker contracting partners to take unfavorable terms. Thus, §2-302(1) of the UCC authorizes a court to strike "any clause of the contract" if the clause is unconscionable, and it is common to hear apparently one-sided terms described as the product of "unequal bargaining power." However, when bargaining power is determined prior to contract formation, as is common in business contexts, these views are incorrect. Bargaining power instead is exercised in the division of the surplus, which is determined by the price term. Parties choose the contract terms jointly so as to maximize the surplus that the price may then divide unequally.[23]

B. Why the State Should Help Firms

We noted at the outset that there are four main objections to a single minded state pursuit of welfare maximization for commercial contract law. Part II(A) argued that the cognitive objection is weak, and Part V(A) will argue against the fairness objection. We discuss here the externality and distributional objections. The externality objection also is weak because it is descriptively true of most commercial contracts that they affect only the parties to them. A single sales contract that turns out badly is unlikely to put employees out of work or cause retailers in the firm's locality to suffer. It is a firm's *systematic* decisions that may affect third parties in material ways. For example, a firm may run a factory with disregard for the environment or the rights of

[23]This conclusion applies though parties may be uncertain about the amount of bargaining power they actually have. For example, parties can use a maximin strategy when they know the set of possible disagreement points but do not know which member of the set applies to their case. A party using this strategy evaluates uncertain gains above the status quo by taking the disagreement point that implies the smallest gain. As a consequence, some efficient deals will not be made, but those that are will be pareto efficient. See Walter Bossert and Hans Peters, *Efficient Solutions to Bargaining Problems With Uncertain Disagreement Points*, 19 Social Choice and Welfare 489 (2002). Our conclusion in the text is unaffected by this form of uncertainty because each disagreement point in the full set is exogenous. This set of models requires parties to reach agreement promptly on the basis on the possible choices available to them. When parties are optimistic about their bargaining power but can learn the truth by inference from the sequence of offers each of them makes, they will reach efficient bargains, though with delay. See Muhamet Yildiz, *Waiting to Persuade*, MIT Dept. of Econ. Working Paper #02-38 (2002). We do not claim that parties always choose efficient contract terms. The existence of asymmetric information sometimes will cause parties to make constrained efficient contracts; these contracts are not "first best" but are efficient given the information structure facing the parties. In Part V, we argue that the state seldom can improve on constrained efficient contracts because information that is unavailable to the parties is unlikely to be available to the decision maker.

its workers. Systematically inefficient or unfair behavior of this kind is subject to legal regulation grouped under the headings of environmental and employment law. Again, we take advantage of this specialization principle to assume that the transactions that a contract law regulates do not create externalities, unless there is a particular reason to believe that they do.

Distributional effects are an appropriate state concern, but there are several reasons why commercial contract rules seldom could create systematic distributional benefits for particular classes of parties. In the first place, commercial parties commonly occupy the roles both of seller and buyer (or licensor and licensee, etc.). As a consequence, a pro-seller rule would hurt firms when they buy, and a pro-buyer rule would hurt these same firms when they sell. In addition, because most commercial contract law rules are defaults, distributional benefits are hard to create even for firms that primarily buy or sell. For example, let a contract rule allocate a risk to the selling side of the market in order to create a distributional benefit for the buying side. Suppose also that contractual surplus would be maximized if buyers bore the risk at issue (because, say, they are the cheapest cost avoiders). A contract allocating the risk to the buyer then would make both the seller and buyer better off (because they would be splitting a larger surplus). Consequently, the legal rule's allocation would be unstable. Because business firms attempt to maximize contractual surplus, the default rules that constitute the bulk of commercial law thus seldom could systematically benefit either side of the market.

Moreover, it is difficult to create distributional benefits for particular firms because firms are owned by shareholders, and shareholders commonly hold diversified portfolios. A diversified shareholder often will own some firms that buy and sell, some firms that may primarily buy and others that may primarily sell. An attempt to benefit either side of the market distributionally is unlikely to create net gains for such a shareholder. Diversification also is normatively relevant. A diversified owner wants the value of his portfolio to increase, not the value of particular firms in his portfolio at the expense of other firms. Indeed, investors diversify precisely to escape firm specific risk – the risk that a particular firm in which the investor has an interest will have an unusually bad outcome. The satisfaction of this preference thus requires legal rules that maximize surplus across firms.

In sum, cogent reasons exist to justify our principal normative claim: *Contract Law should facilitate the ability of firms to maximize welfare when making commercial contracts.* The reasons set out here also imply, *for this class of contracting parties*, that it is unnecessary or futile for courts or statutory drafters to pursue distributional goals. The contract law of commercial parties is about efficiency.

III. THE ENFORCEMENT FUNCTION

A perennial question in contract law is why the state should enforce a contract against the wishes of a party to it. We exclude answers to this question that take the following form: The state should enforce a party's contractual promises the better to permit persons to enlist other persons in their projects, and thus to increase the sphere of autonomy within which persons can operate.[24] Or, the state should enforce promises to reenforce the morality of keeping them. These answers are ruled out here because the business firms that make commercial contracts are *artificial* persons whose autonomy the state need not respect on moral grounds, and whose morality is ordinarily required by positive law. The relevant question for a normative theory of commercial contract law is just when, if ever, does the goal of welfare maximization require legal enforcement of the contracts that business entities make.

A. Enforcement Often is Unnecessary

A contract has an intertemporal aspect: parties agree today to do something tomorrow.[25] State enforcement of these agreements is unnecessary when the agreements fall within the self-

---

[24]For excellent analyses of the strengths and limitations of the various autonomy-based theories of contract law, see Richard Craswell, *Contract Law, Default Rules, and the Philosophy of Promising*, 88 Mich. L. Rev. 395 (1989); Jody S. Kraus, *Philosophy of Contract Law* in The Jurisprudence and Philosophy of Law (Jules Coleman & Scott Shapiro, eds. 2001); and Jody S. Kraus, *Theories of Contact* in The Stanford Encyclopedia of Philosophy (on-line) (Liam Murphy and Joseph Raz. Eds. 2001).

[25]Agreements often are written even when the parties expect not to enforce them legally. A writing reduces disagreements over what the parties had actually agreed to do. Disagreements as to what the contract directs raise interpretation issues that are discussed in Part IV below.

enforcing range or can be enforced with reputational sanctions.[26] An agreement is said to be self-enforcing when the threat by either party no longer to deal with the other is sufficient *in and of itself* to induce performance.[27] Reputation, in turn, will induce performance when a single contract partner's boycott would not.[28] For reputation to work, however, potential contracting parties must be able conveniently to learn why the parties' deal broke down. Reputations, therefore, are difficult to establish in large economies in which particular contracting parties often are anonymous to most market participants. Rather, reputations work best in small trading communities, especially those with ethnically homogenous members, where everything that happens soon becomes common knowledge, and boycotts of bad actors are convenient to enforce.[29] Reputational sanctions also can be effective in industries that can establish trade associations; the associations become a form of collective memory regarding the contracting behavior of their members.[30] This article, however, focuses on contracts that fall outside the self-enforcing range and that cannot be enforced by reputational sanctions. We take this focus because parties often write the contracts it contemplates, and because we are interested in the role

---

[26]For good, largely informal discussions of these issues, see Benjamin Klien, *Why Hold-Ups Occur: The Self-Enforcing Range of Contractual Relationships*, 34 Econ. Inquiry 444 (1996) and Robert E. Scott, *Conflict and Cooperation in Long-Term Contracts*, 75 Cal. L. Rev. 2005, 2039-2050 (1987).

[27]Let S and B write a contract in a state that does not legally enforce contracts. B later learns that it could make $100 more by breaching the contract than by performing it. If B breaches, however, S will no longer deal with B. Let B's expected profits on these future contracts have a present value of $200. Then B will perform the contract, though it could not be sued for breach, because breach would cause it to lose $100 ($200 - $100). The contract is self-enforcing.

[28]For example, suppose that S's later refusal to deal would impose only a $10 loss on the breaching buyer, but other sellers also will refuse to deal, raising the buyer's total loss from breach in present value terms to $200 ($10 + a $190 reputational sanction). Again, B would voluntarily perform.

[29]See Janet Landa, *A Theory of the Ethnically Homogenous Middleman Group: An Institutional Alternative to Contract Law*, 10 J. Legal Stud. 349 (1981). An excellent survey of early informal enforcement mechanisms is Avner Grief, *Informal Contract Enforcement: Lessons from Medieval Trade* in 2 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND LAW 287 (Peter Newman, ed. 1998).

[30]For discussion, see Lisa Bernstein, *Private Commercial Law in the Cotton Industry: Creating Cooperation through Rules, Norms and Institutions*, 99 Mich. L. Rev. 1724 (2001); same author, *Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms*, 144 U. Pa. L. Rev. 1765 (1996).

of the law; legal rules matter when private enforcement mechanisms fail.[31]

Even so, the efficiency gains from enforcing contractual promises presuppose the existence of contracts, and commercial transactions often are conducted without them. Firms often make simultaneous exchanges of cash for goods or services rather than exchange promises for the later trade of these goods or services. In this world, property law is a sufficient encouragement to commerce because a party will only part with goods or money if the party values more highly what is offered in exchange. Thus, protecting property supports efficiency. But in contrast to simultaneous exchanges, a contract is a set of promises regarding future behavior. Such promises are costly to make and to memorialize. In order to understand the role of the state in relation to contracting behavior, it thus is necessary to explain why parties will incur these costs when contracts are legally enforceable but not otherwise.

State enforcement is helpful to contracting parties in a number of contexts but is particularly important in the two cases that Parts III(B) and C next discuss – when investment is relation specific and when a bad state realization can create serious disruption costs. We analyze these cases both because of their intrinsic economic significance and because of their relation to other aspects of contract law. For example, the need to cushion an adverse state realization and the need to avoid an adverse contract interpretation that would create similar disruption costs can cause risk neutral parties to act as if they are risk averse. The recognition of this should influence legal doctrine. The next two subparts thus treat paradigmatic cases.

---

[31]Recent theoretical and empirical work shows that when a nontrivial portion of actors in a market will act fairly toward contract partners, fairness can substitute for reputation. For example, party A will deal with party B not because party B has a reputation for fairness but because party B is likely just to be a fair person. For a review, see Ernst Fehr, Alexander Klien and Klaus M. Schmidt, *Fairness, Incentives and Contractual Incompleteness*, Working Paper #72, Institute for Empirical Research in Economics, University of Zurich (2001). In this literature, some individuals act in experiments as if they have a taste for fairness, while other individuals behave only out of self interest. Similarly, when a nontrivial portion of actors in a market will behave honestly toward contract partners, certain efficiencies can be achieved in high transaction cost environments that would otherwise be unattainable. See Yongmin Chen, *Promises, Trust and Contracts*, 16 J. Law, Econ. & Org. 209 (2000). It is an open question whether, reputational considerations aside, firms have tastes for fairness or honesty. We assume they do not for purposes of this article in order to focus on the role of the law.

B. Encouraging Relation Specific Investment

We earlier identified the canonical contracting problem as ensuring efficient ex post trade and efficient ex ante investment. We assume here, per the Coase Theorem, that parties can ensure efficient trade on their own. For example, if the parties agreed to trade ten units, but it turns out that trading twenty units would maximize joint gains, then the parties can modify the contract to provide for delivery of the larger quantity. As we will see, ensuring efficient investment is more difficult. The investments we have in mind would include the production of specialized goods, the development of human capital specific to a particular deal, or research to acquire information about future costs or prices.

We develop a simple model to explain why the state plays an essential role in encouraging investment. In the model, contracts are not legally enforceable. The sellers in our story can function in two distinct markets. A seller can produce a generic version of a particular product and sell the generic in a competitive market at a price that equals cost (including a return on the seller's investment). The seller also can produce a specialized version of the product for buyers who are willing to pay the extra cost. To be precise, a buyer's valuation for the generic product is denoted v_g and the cost of the product is simply g. Thus, the generic will sell at the price g (because price equals cost in competitive markets) and generate a contractual surplus of v_g - g, which we sometimes denote x. Let a particular buyer value the specialized version of the product at v_s and let the product cost s to produce (where s > g). Importantly, we assume that the seller's investment to make the specialized product would not be redeployable: if the seller were to spend s but the deal were to break up, the seller would lose all of s. The parties prefer a contract to produce the specialized product when that would maximize the contractual surplus. This leads to their decision rule: Produce the specialized version when $v_s - s > v_g - g$. Also, since the two parties comprise the entire relevant world, this decision rule maximizes social welfare.

To make our example more concrete, we let the buyer value the specialized product at \$80 (v_s = \$80); the product's cost is \$50 (s = \$50); buyers value the generic version at \$50 (v_g = \$50); and its cost is \$40 (g = \$40). On these values, the parties' decision rule is satisfied (and social

welfare is maximized) because the specialized product would generate a surplus of \$30 while the generic would generate a surplus of \$10.

If the parties agree to produce the specialized version, they will bargain over how to divide the expected surplus. We assume that the parties are equally patient bargainers because firms commonly can purchase capital (that is, borrow) at a similar cost. Therefore, if the parties do contract, they will split the contractual surplus equally. The parties' bargaining power, recall, also is a function of their disagreement points, and here these differ. If the parties fail to agree on a price for the specialized product, the seller would make the generic product, sell it for g and earn zero profits. The buyer, however, would earn \$10 (v_g - g) from buying on the market. Neither party would make a deal unless it was paid at least the value of its next best alternative out of the value the deal would create. Thus, any contract would have to guarantee the buyer \$10 and the seller at least zero. We denote the net surplus that is to be divided by bargaining as d_k. Then $d_k = v_s - s - (x + 0)$. This surplus is the value of the specialized product (v_s) less the anticipated cost of producing it (s) less the sum of the parties' disagreement points ($x = v_g - g$ for the buyer, and zero for the seller). On the numbers we assume here, d_k = \$20 (\$80 - \$50 - \$10). Because this surplus is split equally, the buyer receives its disagreement point x plus ½(d_k), or \$10 + ½(\$20) = \$20, and the seller receives ½(d_k) = \$10. A contract price of \$60 achieves this split.[32]

The price at which the parties will ultimately contract, however, would not be \$60 because the buyer's incentive to cooperate vanishes after the seller invests s in the subject matter of the deal. To see why, assume that the contract was made as described. The buyer would then have an incentive to demand renegotiation of the price after the seller's investment. At that point, the investment cost s would have been sunk and so would be ignored in the new bargain: the only issue for the parties would be whether to trade the specialized product at some price or not to trade. The renegotiation bargaining surplus, d_r, thus would be v_s - x: the value v_s would be

[32]The buyer earns its valuation less the price, which must equal its payoff from bargaining. Denoting the contract price as p_k, it must be that $v_s - p_k = x + ½(d_k)$. Thus, $p_k = v_s - (x + ½(d_k))$ = \$80 - \$20 = \$60. The \$60 price therefore gives the buyer a gain of \$80 - \$60 = \$20, and it gives the seller a gain of \$60 - \$50 = \$10. Notice also that bargaining power is unequal in this example: the buyer receives two thirds of the contractual surplus because the buyer has a more favorable disagreement point (\$10) than the seller has (zero).

produced if the parties traded the specialized product; the buyer would earn x if the parties did not trade; and the seller would earn zero. On our illustrative values, the parties thus would split equally d_r = \$80 - \$10 = \$70. This split would be achieved by a renegotiation price, p_r, of \$35.[33] This price would divide the contractual surplus of \$30 so that the buyer would gain \$45 (the value of the specialized product less the new price, or v_s - p_r = \$80 - \$35 = \$45); and the seller would lose \$15 (the new price less the production cost, or p_r - s = \$35 - \$50 = -\$15). The seller would accept the renegotiation price, however, because it would earn zero were there no deal; and then its loss would be the full investment cost of \$50.

The lesson that this example teaches is not that parties' ultimate transaction prices would differ from their initial contract prices; the point, rather, is that when contracts are unenforceable a sophisticated seller would reject the contract to produce the specialized product. This seller would anticipate losing \$15 under the contract as renegotiated instead of earning \$10 under the contract it initially signed. Therefore, the seller would produce the generic product. As said above, this result is inefficient: the generic generates a social surplus of \$10 while the specialized product would have generated a surplus of \$30.

The parties' would cooperate to produce the specialized product, however, if the buyer's promise to pay the contract price p_k were legally enforceable. Under the UCC (and the common law of contract) the seller could treat the buyer's demand to renegotiate rather than perform as an anticipatory breach.[34] The seller then would be entitled to recover the price if the contracted for goods could not be resold at a reasonable price.[35] Since the seller could not resell the specialized product for a positive price (its investment, recall, is assumed not to be redeployable), it thus could recover p_k = \$60 from the buyer. Knowing this, the parties would write the contract to produce the specialized product and trade it for \$60. The seller would anticipate being compensated for its investment, and the buyer would prefer to have the specialized product and

---

[33]Recalling note 32, $p_r = v_s - (x + ½(d_r))$ = \$80 - (\$10 + \$35) = \$35.

[34]See UCC §§2-610(a); 2-703(e)..

[35]See UCC §2-709.

realize a $20 surplus rather than have the generic product and realize a $10 surplus.

This analysis supports two conclusions. First, contract remedies are thought to protect injured promisees – the seller here – by awarding the expectation interest. This view is true but shallow. If contracts were not enforceable, sophisticated commercial parties seldom would put themselves in positions where they needed the law's aid. They would instead act as would the seller here, who would produce the generic product and sell it on the market rather than subject itself to exploitation. Enforcement actually empowers promisors, by enabling them to make credible promises to perform or to pay. The buyer in our example, when a contract is contemplated, thus *wants* the power to make a legally enforceable -- that is, a credible -- promise to pay the seller the $60 contract price. Enforcement, in sum, permits parties to make believable promises to each other when reputational or self-enforcement sanctions will not avail.

Second, and relatedly, our example helps to explain the very small amount of foreign direct investment that private parties have made in the former Soviet states and in many third world countries. Much of this investment would have been relation specific (building a factory far from the home country, developing a mine or an oil field). Potential investors would not deal unless the host country or local firm could make credible promises to adhere to the terms originally agreed upon rather than renegotiate those terms after investments had been made. The lack of enforcement rules and honest courts in many of these countries, however, prevents the local parties from making promises that are more believable than was the buyer's promise in the example above. In response, foreign parties do not invest. The ability of a firm to make a credible promise to perform that lawyers in developed countries take for granted is a regrettably rare power in many parts of the world.[36]

---

[36]A vivid example of the second best strategies used when attempting contract enforcement in weak states is in Timothy Frye and Ekaterina Zhuravskaya, *Rackets, Regulation, and the Rule of Law*, 16 J. Law, Econ. & Org., 478 (2000):

> "Private protection rackets primarily provide two services. First, they provide basic protection from other rackets and from criminals. Second, and to a lesser extent, they help enforce agreements. These results suggest that private protection serves first as a substitute for the notoriously ineffective Russian police forces. To a lesser extent, it also serves as a substitute for the notoriously ineffective Russian courts."

C. Contracting to Avoid Disruption: The Case of Volatile Markets

Parties that function in "thick" markets[37] have a choice of making a contract for future delivery or making a spot purchase – that is, a simultaneous exchange of cash for goods. A fixed price contract for future delivery is understood to allocate the risk of price declines to the buyer and of price increases to the seller. What is less well understood is why parties write these contracts when contracting is costly and the spot option is available. We set out another simple model that states this question formally and we then attempt to give a plausible answer.

In the model, the parties can make an enforceable contract at time t^0 for delivery of goods at t^1, or the buyer can wait until t^1 to make a spot purchase. The parties believe, at t^0, that the t^1 market price could take one of three values: (i) p_k; (ii) $p_k + z$; or (iii) $p_k - z$. Each of these outcomes is thought to be equally likely; hence, viewed from t^0, the expected t^1 price is p_k.[38] This also will be the t^0 price. To see why, realize that if the t^0 price exceeded p_k, sellers at t^0 would enter the market to sell contracts for future delivery at the high t^0 price. The resultant increase in supply would cause the t^0 price to fall. On the other hand, if the t^0 price were below p_k, buyers would enter to buy contracts for future delivery at the low t^0 price. The resultant increase in demand would cause the t^0 price to rise. Hence, the unique equilibrium t^0 price must be p_k. As a consequence, if the parties did write a contract at t^0 for t^1 delivery, the contract price would be p_k. To complete the model, we assume that the seller's cost to produce or buy the goods is c # p_k and the buyer's valuation for the goods is $v = p_k + z$. On these assumptions, trading the goods would be efficient because the buyer's valuation exceeds the seller's cost. If the parties do contract, rather than have the buyer wait until t^1 to buy, the cost of writing an agreement is $w > 0$. Since, price equals cost in a competitive market, the buyer will have to pay w.

Id. at 491-92.

[37]A thick market exists when there are many sellers and buyers trading a roughly homogenous product.

[38]The expected t^1 price is $E(p(t^1)) = ⅓(p_k) + ⅓(p_k + z) + ⅓(p_k - z) = p_k$.

Actual numbers might make this story a little more concrete. We let p_k = \$100; z = \$20 (so v = p_k + z = \$120); and w = \$2. The question is whether the buyer will make the contract at t^0 or make a spot purchase at t^1.

In this model, the buyer would wait. To see why, realize that if the buyer did contract, and contracts were enforceable, the buyer would pay the price p_k and realize its net valuation z with certainty, either because the seller performed or through a damage recovery. The buyer's gain from contracting thus would be its net valuation less contracting costs, or z - w. On our assumed numbers, this gain would be \$18 (\$20 - \$2).

The buyer's *expected gain* as of t^0 if the buyer did not contract, however, would be the larger sum z, or \$20. To see why, realize that without a contract the buyer would have to purchase at the t^1 market price, whatever that price turned out to be. Since there are three equally likely future prices, the buyer's expected return, measured as of t^0, from waiting until t^1 to buy is

$$E(G_{nk}) = \tfrac{1}{3}[p_k + z - (p_k + z)] + \tfrac{1}{3}[p_k + z - p_k] + \tfrac{1}{3}[p_k + z - (p_k - z)] = z$$

The buyer's spot purchase would be either made at the high t^1 market price (p_k + z), the average t^1 market price (p_k), or the low t^1 market price (p_k - z). Given a buyer valuation of p_k + z, the buyer's expected gain from not contracting, and thus saving the contracting cost w, is just the buyer's net valuation z, or \$20.[39]

A risk neutral buyer would not pay a premium to ensure a certain gain; rather, the party would choose a higher expected but risky gain. The "premium" here would be the contracting cost (w = \$2); paying it would guarantee the buyer the sure \$18 gain, which it prefers less than the expected gain of \$20. A risk neutral seller too would reject a contract for the same reason. Spot purchases also are very common. And if parties would not write contracts for future delivery in thick markets were these contracts legally enforceable, there is no need to make these contracts legally enforceable. Thus, the question why risk neutral firms sometimes write contracts for future delivery is not trivial.

---

[39]Using our illustrative numbers, $E(G_k)$ = ⅓(\$120 - \$120) + ⅓(\$120 - \$100) + ⅓(\$120 - \$80) = \$20.

We argue that parties write these contracts (when they are enforceable) to reduce the risk of potentially disastrous outcomes. To see the argument, realize that if the buyer in the example did not contract, then one third of the time it would have to make a spot purchase at the high t^1 price of $p_k + z$, or \$120. Since the buyer's gross valuation is \$120, one third of the time a purchase of the goods would contribute nothing toward the buyer's fixed obligations. A buyer who failed to pay rent or interest, however, would incur serious disruption costs. Denoting disruption costs as f, and realizing that these would be incurred one third of the time, the comparison in the model is misleading: The buyer's actual decision rule would be to contract when the net gain from contracting would exceed the net gain from not contracting, or when $z - w > z - f/3$, or when $w < f/3$. In this example, if the inability to contribute to fixed costs would impose only a \$30 disruption cost on the buyer, it would write a contract for future delivery (\$2 < \$30/3).

This analysis predicts that a buyer who contemplates making many transactions, none of which will be large in relation to the buyer's need for cash, will act as would a risk neutral person, purchasing goods on the spot market. When a bad realization on a single contract could seriously endanger the buyer's business, however, the buyer will act as if it were risk averse, purchasing the assurance of performance at the cost of writing a contract for future delivery. To be sure, some such buyers could guard against disruption by maintaining sufficient cash reserves, but this strategy is difficult for all businesses to follow. Also, if the buyer has good business prospects, the opportunity cost of hoarding cash likely would exceed contracting costs, especially for the simple fixed price contracts we consider. The result of this analysis is that the state reduces social costs by giving parties that function in volatile markets the opportunity to make enforceable contracts for future delivery.

In sum, when exchange is intertemporal rather than simultaneous, efficiency is enhanced when parties can make enforceable contracts in two principal situations: When at least one of the parties is required under the contract to make an investment that is more profitable in the relationship than elsewhere; and when market prices are volatile and an adverse market

movement can have spillover effects.[40] In the first case, in the absence of legal enforcement the non-investing party has an incentive to renegotiate the contract price downward rather than to perform under the original contract. In the second case, the party whom a market movement disadvantages may suffer disruption costs that would much exceed its expectation interest (as that is conventionally measured). Contracts do sometimes fall within the self-enforcing range and reputation can make promises to perform credible in some sub-economies. However, nonlegal incentives can be ineffective in larger markets and in countries where social trust is low. Then, without legal enforcement the ability of parties to maximize their own and their society's welfare is severely limited.

D. Enforcement and Duress

Enforcement entails more than simply ordering a recalcitrant party to perform. A supplementary set of "enforcement rules" apparently also is needed. As an illustration, suppose that a seller sold the goods to a third party instead of delivering them to the contract buyer. A specific performance order would thus be futile. Should the buyer be able to sue the seller for damages or to impose a constructive trust on the proceeds of the sale? If the buyer makes a substitute purchase, is the seller's obligation discharged? Let the buyer instead take delivery of the goods but claim that they do not conform to the quality the seller promised to deliver. Must the buyer still pay the price and sue for damages or can the buyer cancel the sale? Finally, suppose that, after the contract is made but before delivery, a Federal agency passes a regulation that prohibits production of the product the buyer purchased the goods to make. The buyer no longer needs the goods. Must the buyer still pay the price? If not, does the seller have any other remedy?

---

[40]Recall our assumption that parties to business contracts are risk neutral. A third motive to contract is to transfer risk from more to less risk averse parties. The legal enforcement of these contracts sometimes is necessary because the transferee of risk has an incentive to breach when large risks materialize. Risk shifting contracts are not considered here in part because one of the parties to them commonly is an insurer, and insurance contracts are the subject of a distinct and heavily regulated legal field. Moreover, although many contracts have an insurance component (i.e., commodities contracts, currency hedging, etc.), these contracts tend not to give rise to litigation.

These questions illustrate the complexity of the concept "enforcing a contract". It is tempting to suggest, therefore, that a supplementary set of publicly supplied "enforcement rules" also is needed. This suggestion would be premature, however, because parties can answer these questions in their own contracts. For example, parties can write a "force majeure" clause specifying the events that would excuse the seller's obligation to deliver or the buyer's obligation to pay; these events could include the possible passage of an unfavorable administrative regulation. That the contract laws of advanced nations commonly contain sets of enforcement rules thus requires explanation, a task that we address in Part V of this article.

The duress doctrine, however, is an enforcement rule that parties cannot create on their own. The law of duress applies in two contexts. Ex ante duress occurs when a party is wrongfully coerced to make a contract.[41] Ex post duress occurs when a party is wrongfully coerced to modify an existing contract.[42] Contract law applies the same legal standard in both cases: a contract or a modification is unenforceable if a party's consent thereto was obtained by an improper threat that left the party no reasonable alternative but to submit.[43] Our focus here is on ex post duress, and we suggest that courts should ask a different question in this case than is appropriate for ex ante duress cases. In an ex post duress case, the contact was fairly obtained and the parties could have provided for the situation that later arose had they thought about the issue. The court thus should ask whether parties with sufficient foresight would have wanted the later modification agreement to be enforceable.

---

[41]An ex ante duress argument succeeds if the party proves that he would not have made the challenged contract absent the improperly coercive behavior of the other party. The key is the law's focus on behavior rather than circumstances. Thus, it is not duress on the part of an employer when a poor person accepted an offer to work at a low wage, nor is it duress on the part of a seller to charge a high price for gas when it is the only seller for many miles. The employer did not create the employee's low wealth, nor did the seller lure the buyer to the solitary spot. Contract law thus requires a distress claim to rest on the behavior of the promisee, not on the preexisting circumstances of the promisor. Chouinard v. Chouinard, 568 F.2d 430 (5th Cir. 1978). This is because absent coercion by the promisee, the promisor does better by contracting than by not contracting. See Robert E. Scott & William J. Stuntz, *Plea Bargaining as Contract*, 101 Yale L. J. 1909, 1919 (1992).

[42]See, e.g, Austin Instrument, Inc. v. Loral Corp., 29 NY.2d 124, 272 N.E.2d 533 (1971).

[43]Restatement (Second) of Contracts §175, Comments (a) and (b).

Two examples show how this test is to be applied. For the first, assume that the parties initially agreed to trade 30 units at a price of $10 each. Demand for the product turns out to be higher than the parties initially believed: it becomes efficient to trade 50 units. The seller offers to transfer an additional 20 units at a unit price, for all 50, of $12. The buyer's profit would be higher under the contract as modified, and it agrees to the new terms. The buyer cannot later claim that it was coerced to accept a price increase because, ex ante, this buyer would have wanted the court to enforce a modification that would leave it better off than performance under the original contract would have done.

For our second example, return to the relation specific investment model set out above. There, we claimed that permitting the seller to sue for the price would deter the buyer' threat to renegotiate after the seller had invested. This claim is too strong because sellers in some cases could not make a credible threat to sue. The seller's threat would be credible only if it had, or had convenient access to, the capital needed to sustain it until the buyer's performance could be replaced. The buyer then would know that it would have to pay voluntarily or involuntarily. In contrast, a seller in a short term bind (perhaps it had purchased materials on credit and is facing demands for payment) may be better off accepting a low renegotiation price than bringing a law suit. Recognizing this, the buyer may demand renegotiation even though the contract is enforceable. The renegotiated contract, however, would be a product of ex post duress. When the initial contract was made, both parties would want a court not to enforce a purely redistributional modification; that is, a modification that would create no new wealth but rather would only redistribute the contractual surplus differently than did the original contract. Parties dislike purely redistributional modifications for two reasons. As Part 3(B) showed, such modifications destroy the parties' incentive scheme for producing efficient specialized products. Further, the resources involved in negotiating the modification or guarding against it constitute a dead weight loss that reduces the parties' gain from the contract.[44]

---

[44]Ex post duress cases are largely consistent with the test we propose, though the courts formally apply the Restatement standards. See, e.g., Austin v. Loral, supra note 42, and Wolf v. Marlton Corp., 57 N.J. Super. 278, 154 A.2d 625 (1959).

The duress doctrine thus permits the seller to perform under the renegotiated contract but later to reinstate the price term in the original contract. This is because modifications made under duress are not enforceable. As a consequence, the seller in our example could pay the $35 renegotiation price, deliver the specialized goods, and later recover the difference between the $60 contract price and the lower renegotiation price. Sellers who have easier access to cash in the long run than in the short run, or who could sell their legal claim, will use the duress doctrine to recover their original expectation. Buyers, therefore, will know that promises by sellers to pay \renegotiation prices themselves are not credible. The law will permit a seller to renege on such a promise and sue for full payment. The unreliability of renegotiation promises coerced by duress reduces the incentive to extract them (that is, to behave as did the buyer in our example). The ex post duress doctrine thus is an important aspect of a publicly supplied enforcement function.[45]

In Part III, we argued that firms need state enforcement in order to permit them to make credible commitments when their promises are not self-enforcing. A court cannot enforce a contract, however, without first determining what the contract says. Thus, the parties' preference for state enforcement entails a further party preference over the set of interpretive theories that courts could use to interpret their agreements. We next attempt to identify the interpretive theory in this set that typical firms likely prefer.

[45]As may be obvious, the doctrine is symmetrical: it also applies to protect buyers who have made relation specific investments from overreaching sellers. The doctrine would be unnecessary if capital markets were perfect, or damage suits were perfectly compensatory. Then parties could finance all good law suits or recover all losses. Capital markets, however, are imperfect and damage awards are not always fully compensatory.

IV. THE INTERPRETATION FUNCTION

A. The Relevant Interpretive Question[46]

There is a consensus among courts and commentators that the appropriate goal of contract interpretation is to have the enforcing court find the "correct answer". The "correct answer" is the solution to a contracting problem that the parties intended to enact. Intention, however, is determined objectively and prospectively: a party is taken to mean what its contract partner could plausibly believe it meant when the parties contracted.

There are two justifications for the goal of finding the correct answer. The first follows from an autonomy-based view of contract law. This justification holds that the exercise of state coercion against a person must be justified. A sufficient justification is that the court is making the person do what he had agreed to do.[47] Hence, the court must ascertain just what the person had agreed to do. The second justification is consistent with an efficiency-based view of contract law. On this view, parties contract to maximize the surplus that their deal can create. This goal is unattainable if courts fail to enforce the parties' solution but rather impose some other solution. Thus, the court must ascertain the solution that the parties actually adopted. There is a dispute in the literature whether the rules that courts use when attempting to find the correct answer are mandatory in the formal sense, but there is a consensus that the rules are difficult for parties to

---

[46]A theory of interpretation has two aspects: a set of rules for determining the syntactic content of a party's utterances; and a set of rules for determining the legal significance that should attach to the syntactic content. We focus on the former set of rules here because we have already developed a normative theory: Courts should enforce business contracts as the parties to them would want the contracts enforced. Our view regarding the separability of the rules determining legal significance from the rules governing interpretation is not free from difficulty, however. Courts doing interpretation may be influenced by their view of what a good substantive outcome is when the parties' meaning is not apparent. See Richard Craswell, *Offer, Acceptance and Efficient Reliance*, 48 Stan. L. Rev. 481, 551-553 (1996). We ignore this complication on the assumption that if the syntactic content rules improve, the courts' need to give legal answers to factual intent questions will lessen. A full theory of contract interpretation would explore more seriously the actual and ideal relation between a court's substantive preferences and its choice of an interpretive style.

[47]Most autonomy-based theories are premised either on a notion of "consent" or the exercise of will, such as the making of a promise. See references cited in note 24, supra.

escape.[48]

In our view, the current consensus asks the wrong question. A commitment to party sovereignty regarding the contract's substantive terms implies the further commitment to party sovereignty regarding the interpretive style an adjudicator should use to find the substantive terms. Party preferences regarding judicial interpretive styles can differ. Therefore, interpretive styles should be defaults. The relevant question, then, is what should be the majoritarian default. Put another way, the issue is not what interpretive style is best calculated to yield the correct answer; rather, the issue is what interpretive style would typical parties want courts to use when attempting to find the correct answer. We will argue here that the majoritarian default is Willistonian: typical firms prefer courts to make interpretations on a narrow evidentiary base whose most significant component is the written contract.[49] This proposed default would reverse the UCC's interpretive rules to the extent that they are defaults; the proposal is more radical still if the Code's rules are mandatory.

B. Two Interpretive Issues: Problems of Meaning and of Language

We begin by clarifying two interpretive issues that commonly are commingled: What does the language of the contract mean? And in what language was the contract written? To see why these issues are distinct, suppose that there are two sets of linguistic communities. The first set, called M, consists of a single linguistic community. This M community is composed of judges,

---

[48]A recent commentator referred to the Code's interpretation rules as "quasi-mandatory", the idea being that the rules are and should be very difficult for parties to avoid. See David V. Snyder, *Language and Formalities in Commercial Contracts: A Defense of Custom and Conduct*, 54 S.M.U. L. Rev. 617 (2001). Because contracting has positive costs, a quasi-mandatory rule will be mandatory in practice for many parties. Other commentators believe that the interpretation rules are mandatory. See, e.g., Omri Ben-Shahar, *The Tentative Case Against Flexibility in Commercial Law*, 66 U. Chi. L. Rev. 781 (1999).

[49]See, e.g., Samuel Williston, Contracts §631 (3d ed. 1961) ("The parol evidence rule requires, in the absence of fraud, duress, mutual mistake or something of the kind, the exclusion of extrinsic evidence, oral or written, where the parties have reduced their agreement to an integrated writing."). Willistonian formalism rests on two basic claims: (i) Contract terms can be interpreted according to their plain meanings; and (ii) Written terms have priority over unwritten expressions of agreement. See Dennis M. Patterson, *Good Faith, Lender Liability, and Discretionary Acceleration: of Llewellyn, Wittgenstein and the Uniform Commercial Code*, 68 Tex. L. Rev. 169, 187-88 (1989).

lawyers, business persons and potential jurors. The members of M read and write what we call "majority talk" because it is the language that people typically use when attempting to communicate with each other. The second set is called P, and it has many linguistic communities. Each community in the set P is composed of the parties to the contract at bar and perhaps the trade in which the parties function. The set P has many member communities because there are many sets of parties and trades. The members of each set in P may write contracts in their own "party talk" or write them in majority talk.

The existence of plural linguistic communities raises the two interpretive issues just noted. Regarding the first, imagine that parties are engaged in a dispute regarding the meaning of their written agreement. Words can be vague or ambiguous.[50] If the parties agree on the language in which their contract was written, the court's interpretive task is limited to finding what the parties intended that language to say. If the parties divide on the question what language they used, the court's interpretive task expands: the court now must ask initially whether the parties wrote their contract in majority talk or in a particular private language. This question raises a separate interpretive issue because a contract's language could plainly mean m_M in community M but also plainly mean m_P in a particular community P, where m_M and m_P differ. For example, the word "wife" in a sentence in John's will reciting "I leave my money to my wife" would mean in the M community that John left his money to the woman to whom he was legally married when he died. In a well known case, however, John and another party wrote a contract in order to dispose of part

[50]Courts seldom distinguish between "vague" and "ambiguous" terms. A typical judicial definition of ambiguity, for example, includes any term or word that is "capable of more than one sensible and reasonable interpretation or has no definite significance." Ross Bros. Constr. Co. v. State ex rel Transp. Comm'n, Highway Div., 59 Ore. App. 374, 650 P.2d 1080 (1982). More narrowly, however, a word is vague to the extent that it can apply to a wide spectrum of referents, or to referents that cluster around a modal "best instance", or to somewhat different referents in different people. For classic examples, see Frigaliment Importing Co. v. B.N.S. Int'l Sales Corp, 190 F. Supp. 116 (S.D.N.Y. 1960) (does "chicken" include all types of chicken or only a subset?); Highly v. Phillips, 176 Md. 463, 5 A.2d 824 (1939) (does a requirement to remove "all dirt" from a tract refer also to subsurface sand?). In contrast, ambiguity requires at least two distinct, usually inconsistent meanings. See, e.g., Petroleum Fin. Corp. v. Cockburn, 241 F.2d 312 (5th Cir. 1957) (missing punctuation in telegram supported two different readings); Raffles v. Wickelhaus, 159 Eng. Rep. 375 (1864) (seller required to use the ship called "Peerless" sailing from Bombay but there were two such ships). Language commonly is vague in the sense that the set of objects to which a word applies is rarely delineated with absolute precision. True syntactic ambiguity occurs infrequently.

of John's estate upon his death.[51] This contract used the word "wife" to refer to the women with whom John was living when he died but whom he had never legally married; John had deserted his legal wife years before without divorcing her. The word "wife" in this example would be clear to members of the M community and would also be clear to the particular P community to which the contracting parties belonged. The two meanings differed, however.[52] That the same word or phrase can have different meanings in different linguistic communities thus requires courts to make a choice. Should courts permit parties to write contracts in the language of the parties' choosing? Or should courts create an incentive for parties to use majority talk by interpreting their agreements as if the agreements were written in that language?[53]

There are two traditional approaches to finding the correct answer to questions of contractual meaning and of contractual language. They differ in the scope of the evidentiary base each requires to make interpretations. To be sure, a minimum evidentiary base is required for any coherent interpretation. This minimum base is denoted here B_{min} and is composed of the parties' contract, a narrative concerning whether the parties performed the obligations that the contract appears to require, a standard English language dictionary, and the interpreter's experience and understanding of the world. A Willistonian, or "textualist", theory of interpretation assumes that contracts often have "plain meanings" that are apparent to judicial interpreters. Put formally, this view asserts that a court can find the correct answer on the evidentiary base B_{min}. Courts have

---

[51]In re Soper's Estate, 196 Minn. 60, 264 N.W. 427 (1935).

[52]For a modern example, see Columbia Nitrogen Corp. v. Royster Co, 451 F.2d 3 (1971) (contract specified "Minimum Tonnage Per Year" of 31,000 tons; buyer took under 2,000 tons; the court permitted the buyer to introduce evidence that, in the parties' trade, stated minimum tonnages were "mere projections to be adjusted according to market forces").

[53]The distinction between party and majority talk can blur at the edges but there are core cases. For example, a contract written in majority talk may require the seller to deliver "red sweaters". Evidence regarding usage of trade may be helpful in delimiting the permissible shades of red. On the other hand, a contract likely is written in party talk if one party gave to the other a written unconditional option to buy all of the stock in the grantor's company at a stated price by a stated date, but the parties meant "unconditional" to mean conditional. In the actual case, the option grantor thus claimed that the parties intended to give the option holder the power to exercise the seemingly unconditional option only on the condition that the grantor sought other buyers for the company. See Hunt Foods & Industries v. Doliner, 26 A.D.2d 41, 270 N.Y.S.2d 937, *affirmed* 272 N.Y.S.2d 686 (1966) (permitting the grantor's testimony). We return to the two language distinction in a moment and also in Part IVD below.

added five evidentiary categories to this minimum base: (1) the parties' practice under prior agreements; (2) the parties' practice under the current agreement; (3) testimony as to what was said during the negotiations; (4) written pre-contractual documents (memoranda, prior drafts, letters); (5) industry custom relevant to determining what the agreement's words meant to the contracting parties. When an adjudicator admits evidence in all five additional evidentiary categories, we denote the evidentiary base as B_{max}. A "contextualist" theory of interpretation holds that a court is more likely to find the correct answer if the evidentiary base expands toward B_{max}.[54]

A Willistonian theory of interpretation has the obvious effect of truncating the evidentiary base that a court uses to make interpretations, but it also has the effect of creating an incentive for parties to write contracts in majority talk. Because a word can have plain but different meanings in the M and P communities, a party seldom could establish that its contract was written in party talk unless it could introduce extrinsic evidence.[55] The Willistonian theory bars this evidence.[56] In addition, on this theory the court primarily sees the contract; hence, the parties likely can reduce interpretive errors by writing the contract in the court's language. Thus, the Willistonian theory actually resolves two distinct issues: what language the parties should speak and what evidence is admissible to show what the parties meant in the permitted language. Contextualists resolve the same issues, though differently. Their theory has courts considering all material evidence to resolve interpretive issues; and the practical effect of so widening the evidentiary base has been to permit parties to write in whatever language they choose.[57]

---

[54]Contextualist scholars do not necessarily insist on B_{max}. It is recognized, for example, that evidence in category (3) is less reliable than evidence in category (4).

[55]Extrinsic evidence refers here to evidence in the five additional evidential categories listed in the text.

[56]A court applying the Willistonian theory will admit extrinsic evidence only when the contract's language is vague or ambiguous on its face. See Pysell v. Keck, 263 Va. 457, 559 S.E. 2d 677 (2002).

[57]In a well known case, Justice Traynor thus stated: "The fact that the terms of an instrument appear clear to a judge does not preclude the possibility that the parties chose the language of the instrument to express different terms. That possibility exists ... whenever the parties' understanding of the words used may have differed from the judge's understanding. Accordingly, rational interpretation requires at least a preliminary consideration of all credible evidence offered to prove the intention of the parties." Pacific Gas & Electric Co. v. G.W. Thomas Drayage & Rigging Co., 69 Cal 2d 33, __, 69 Cal. Rptr, 561, __, 442 P.2d 641, __ (1968).

Any analysis of contractual interpretation thus should answer two questions. The first is the question of meaning: Should courts use broad or narrow evidentiary bases in determining the meaning of the contract's language? The second is the question of language: Should the "linguistic default" assume that parties write in majority talk unless their contract recites otherwise? or should courts always admit the possibility that the parties at bar wrote in a private language? The practical effect of admitting this possibility is to permit a party to introduce evidence in all of the evidential categories to show what language the parties actually used. Part C next argues that when the issue of the contract's language is settled but the meaning of that language is arguably unclear, the majoritarian default would require courts to make interpretations using the evidentiary base B_{min}. Part D then argues that the linguistic default should suppose parties to have used majority talk.

C. The Parties' Preferences Regarding Interpretive Styles

An interpretive style can be assessed on two dimensions: (1) the likelihood that the style will generate the correct answer (as defined above); and (2) the costs that the style imposes on courts and parties. We model the performance of judicial interpretive styles on these dimensions in two ways that, in combination, capture most of the relevant cases. Both models suppose that the contract is complete in the sense that the writing expresses the parties' solution to the contracting problem at issue. The parties are aware, however, that their meaning may not always be transparent to a later interpreter. They thus knowingly face the possibility that a court's interpretation may deviate from the correct answer. Our initial model considers the set of cases in which the parties' payoffs under their contract are monotonic and continuous in the space of possible judicial interpretations. A party's payoff, that is, increases as interpretations of the contract become more favorable to it and decreases as interpretations become less favorable. Firms that this set of cases describes prefer courts to make interpretations on the minimum evidentiary base B_{min} except in unusual circumstances. We then consider a set of cases in which payoffs are invariant to the extent of judicial error. Here, an unfavorable interpretation that is close to the correct answer will have negative consequences for a firm that are as severe as interpretations that are far from the correct answer. Firms that this set of cases describes have

reasons to prefer a broader evidentiary base than B_{min} but, we argue, litigation cost considerations will still incline a majority of these firms to want courts to use a narrow evidentiary base. Taken together, the two models suggest that when the question of contractual language has been settled but the issue regarding what that language means remains, the default interpretive style for courts should be textualist.

In both models, parties negotiate a contract whose object is to maximize the surplus the deal could create, and they divide that surplus through the price term. The parties want a court, should a dispute arise, to give the correct answer to an interpretive question. Anticipating that the answer will be correct permits parties to take surplus maximizing actions. Parties sign the contract when it is optimally clear, in a sense to be made more precise below.

1. The continuous payoff case.

In this case, we focus for convenience on a buyer. We denote the surplus under a deal as s*; the buyer's share is the difference between its valuation and the price: $s_b^* = v^* - p$. The judicial interpretation that gives the correct answer is denoted i*. Thus, if the court does make the interpretation i*, the contractual surplus will be s* and the buyer will realize s_b^*. The illustrative obligation we consider is the seller's duty to prepare machines it will later deliver so as to minimize the buyer's costs of adjusting the machines for their intended use. Assume that the parties later have a dispute regarding whether the seller has fully complied with its duty to prepare the machines adequately. If the court imposes a lesser preparation obligation on the seller than the obligation the contract, correctly interpreted, would require, the deal will be less profitable to the buyer and its gain will fall to $s_b^- < s_b^*$. If the court imposes a greater obligation on the seller than i*, the buyer's gain will rise to $s_b^+ > s_b^*$. Each interpretation in the space of possible judicial interpretations thus generates a particular payoff for the buyer. In this product preparation example, the buyer's possible payoffs are monotonic and continuous in the space of possible interpretations; that is, as interpretations become more favorable to the buyer, its payoffs increase, and conversely. The seller's possible payoffs also are monotonic and continuous, but these decline as the buyer's increase.

We begin at the litigation stage when analyzing the parties' preferences regarding interpretive styles, and make three assumptions: (a) A court can find the correct answer i* with positive probability on the minimum evidentiary base B_{min};[58] (b) The likelihood that a court will find i* sometimes can be increased if the court considers evidence in categories additional to $B_{min;}$ (c) Courts are unbiased. We initially support assumption (a) by observing that the written contract trumps evidence in the other evidential categories.[59] In addition, because disputes are expensive, parties benefit from specifying obligations in such fashion that subsequently it will be clear what was promised. These considerations suggest that parties have an incentive to write contracts that permit courts to discern their actual intentions. For (b), assume that the parties specified the seller's preparation obligations in considerable detail in a pre-agreement memorandum that both parties signed. Admitting the memorandum into evidence will increase the likelihood that a court will give the correct answer to interpretation questions that involve product preparation. Regarding (c), courts are assumed to be unbiased when making interpretations because there is no reason to believe that courts will systematically deviate from the correct answer i* in ways that are more or less favorable to a party or class of parties.

These three assumptions permit us to state formally a party's expectation at contracting time regarding the possible payoffs that a judicial interpretation could induce. Continuing with our illustration, if courts are unbiased, can find i* on the base B_{min}, but can err, then the buyer's expected payoff given what a court later will do can be expressed as

$$E(s_b(i)^*B_{min}) = s_b(i^*) + ,$$

The left hand side of this expression is the buyer's expected payoff given a judicial interpretation i made on the evidentiary base B_{min}. The first term on the right hand side is the buyer's payoff given the correct interpretation i*; the second term is an error term with mean zero and positive variance. That , has mean zero means that the court is unbiased: in our example, the court is as likely to make an interpretation that is more favorable to the buyer (less favorable to the seller) than the

[58]Recall that B_{min} is composed of the written contract, a performance narrative, a dictionary and the interpreter's experience of the world.

[59]UCC §1-303(e)(1) (2002) (formerly §1-205(4)) provides that when there is conflict, "express terms prevail over course of performance, course of dealing and usage of trade."

correct answer as the court is likely to make a less favorable interpretation. Judicial errors therefore cancel, in expectation. The error term has positive variance, however, because in an actual case a court's interpretation can deviate from the correct answer. The expression thus says, in English, that the buyer's expected payoff, given the evidentiary base B_{min}, equals the correct answer -- i* -- but with variance. In other words, the buyer's payoff under the contract equals in expectation the payoff that the parties intended when they wrote the contract.

When the variance in the error term is large, parties will think that the answer a court will later give can deviate widely from the correct answer. Our example of the pre-agreement memorandum regarding the seller's performance obligation thus is offered to show that adding evidence to B_{min} can shrink the variance. In ordinary language, if the parties here know that the evidentiary base in a later law suit will include pre-agreement memoranda, they will expect later interpretations to be close to the correct answer.

This specification of the relation between judicial error and party expectations implies that firms commonly prefer courts to be restricted to the minimum evidentiary base B_{min} when payoffs are continuous in the space of possible interpretations. To see why, recall our assumption (b): as the evidentiary base approaches B_{max}, the variance in the error term approaches zero. A risk neutral party cares about the mean of the interpretation distribution but not the variance, however. This is because the variance term measures risk while risk neutral parties are indifferent to risk.[60] Therefore, it is enough for a risk neutral firm that the expected interpretation (i) equals the correct interpretation i*. Put another way, a firm's preference at contract time is to have courts make interpretations on the minimum evidentiary base unless it would be costless to widen the base. But it is not costless. The wider is the permissible evidentiary base, the more evidence a party will have an incentive to introduce and the more evidence it will have to contest. Since trials are expensive,

[60]For readers unversed in statistics, variance is a measure of how far an outcome can deviate from the mean of a distribution. Thus, if a court is unbiased but the variance is large, the interpretation that the court makes in an actual case can be very far from the correct answer; conversely, when the variance is small, the court is likely to be close to the correct answer. If the contextualist claim is correct that broader evidentiary bases generate more accurate interpretations (i.e., the variance shrinks more), it follows that litigations where the court sees only B_{min} are more risky for parties than litigations where the court sees additional evidence; for when the court sees only B_{min}, its interpretation can deviate more widely from the correct answer.

risk neutral firms are Willistonians.[61]

This view should not be overstated because some firms will have a different preference. Recall the volatile markets example in Part III(B) above. There, we argued that firms functioning in volatile markets would make spot purchases unless a particular transaction would be important in relation to the size of the firm. In that event, a firm often would prefer to incur the costs of making a contract in order to avoid the costs of disruption that a bad market realization could cause. A similar preference could obtain here. Thus, when performance of a particular contract would be important to the survival of the firm – a contract with a major supplier, say –, or when the contract is new and is expected to be widely used, the firm may be unwilling to risk a seriously adverse interpretation. If contextualists are correct that larger evidentiary bases do shrink variance, then when variance is important parties likely will prefer courts to use a contextualist adjudicatory style. In the example above, if the machines were crucial to a start-up venture and the buyer had little cash, the buyer ex ante would want the evidentiary base in a later suit to be sufficiently broad so that the memorandum would be admitted into evidence. On the other hand, only unusual contracts have this "bet the ranch" quality. In the typical case, it is good enough that courts get things right on average.

We now turn to the contract creation stage to return to assumption (a). It is optimal for risk neutral firms to invest resources in drafting until the writing is sufficiently clear, in an objective sense, so that the mean of the distribution of possible judicial interpretations is the correct interpretation i*. More sketchy contracts than this would generate interpretation distributions whose mean could be anywhere. As a consequence, parties could not expect courts to protect their expectation interest in case of a dispute. As this would be inefficient, firms will attempt to write contracts with sufficient clarity to permit courts to find correct answers, though with error.

---

[61]There is considerable evidence that firms prefer a formalist adjudicatory style. See Bernstein, supra note 30.

The debate between textualists and contextualists is irrelevant to issues of interpretation for cases that resemble the performance preparation illustration. In this illustration, there was consensus regarding the language in which the contract was written but that language was arguably unclear. The contextualist in effect asserts that a larger evidentiary base shrinks variance. Indeed, if the base is large enough, the variance approaches zero, so parties will believe when they contract that a court will find the correct answer with certainty. As applied to our illustration, admitting the memorandum, perhaps together with an industry custom as to a seller's preparation obligations, would leave little doubt regarding what the seller was supposed to do. The textualist, in contrast, claims that variance does not shrink materially with a broader evidentiary base because contracts often have plain meanings. Hence, permitting parties to introduce additional evidence as to intent would generate costs in excess of gains; for the evidence seldom would shrink the variance in the error term by very much. The proponents in this debate thus disagree on the relation between the width of the evidentiary base and the accuracy of a court's interpretation. It is unnecessary for courts to decide who is right. Greater accuracy is lower variance and business parties commonly are indifferent to variance. Thus, courts that interpret contracts as typical parties would like them to would be indifferent to variance as well; these courts would make interpretations on the evidentiary base B_{min} unless parties instruct them to the contrary.

2. The invariant payoff case.

We illustrate this case by recalling the specialized product example in Part IIIA. Denote the quality that the contract required the seller to produce as q^k. The seller produces q^k but the buyer, possibly sincerely, claims that the contract required a higher quality level. The product's actual quality is not in dispute; the issue is what the contract said. We denote an interpretation that the contract required a lower quality level than that actually delivered as the answer $q^- < q^k$; an interpretation that the contract required a higher quality level than the level the seller delivered is the answer $q^+ > q^k$. Just as before, we assume that the court is unbiased. Therefore, the mean of this three point distribution is q^k: the seller's expectation regarding a later judicial interpretation of the contract can be written as $E(i) = q^k$.

The seller's payoffs here, however, are not continuous in the space of the court's possible interpretations because of the perfect tender rule.[62] If a court finds that the contract required a quality level that was less than or equal to the quality the seller delivered, the seller can recover the price; if the court finds that the contract required a higher quality level, the buyer can reject and credibly threaten a suit for damages. The seller then will realize a renegotiation payoff that likely is less than the price because the damage threat reduces its bargaining power. That judicial interpretations are unbiased does not matter for this seller because the seller does not gain from judicial errors on the low side (it cannot recover an amount greater than the price if it delivered a quality level that is higher than the level the court requires), while the seller is harmed by judicial errors on the high side, however small (any quality shortfall from an incorrectly interpreted contract would permit the buyer to reject). The seller thus wants the court find the correct answer with certainty, even though the seller is risk neutral.

We formalize this preference in the following way. Let 0 # $ # 1 be the probability that the court finds the correct answer with certainty; with probability (1 - $), the court makes an unbiased interpretation of the contract's required quality level. The contract price is p; the seller's renegotiation payoff is denoted r # p. On this formulation, when the parties' contract the seller's expected payoff is

$$E(B_s) = \$(p) + (1 - \$)\,[\tfrac{2}{3}(p) + \tfrac{1}{3}(r)].$$

With probability $, the seller receives the payoff that the contract's terms direct, which is p, because the court will interpret the contract correctly. With probability (1 - $), the court will make an unbiased interpretation of the quality level the contract required. Each of the three possible interpretations the court could make is assumed to be equally likely; hence, when the seller is found to have complied or over complied, which in total occur two thirds of the time, the seller receives the contractual payoff of p, and when the seller is found to have under complied, which occurs one third of the time, the seller receives the renegotiation payoff of r. Note that if the parties expect $ to equal zero, they will expect a court to make only an unbiased interpretation, with a positive probability of error; if the parties expect $ to equal 1, they will expect a court to give the correct

[62]UCC §2-601 provides that "if the goods or the tender of delivery fail in any respect to conform to the contract, the buyer may reject the whole ...."

answer with certainty.

The seller's expected payoff under the contract – $E(B_s)$ -- ordinarily will be *lower* than the seller's expectation interest, which is the contract price of p. This is because $ commonly is less than one -- the court is expected not to find the correct answer with certainty -- and r commonly is less than p -- the seller expects to be disadvantaged in a renegotiation. Since protecting the expectation interest generally is efficient, the parties can increase surplus by raising the seller's expected payoff. Expanding the evidentiary base may achieve this goal.[63] Our assumptions regarding the relation between the width of the evidentiary base and the accuracy of a court's interpretation imply that the probability of a court finding the correct answer -- $ -- rises as the evidentiary base the court sees expands beyond $B_{min.}$ Because the seller's expected payoff under the contract increases as $ increases -- as the court becomes more accurate -- the parties, though risk neutral, thus have reason to prefer a broader evidentiary base than B_{min}. Expanding the evidentiary base is not costless, however. The parties, therefore, actually face a *tradeoff* between increased accuracy and increased contract enforcement cost.[64]

Two factors influence how parties will make this tradeoff. The first factor is the bargaining power the seller will have in a renegotiation: the greater is the seller's bargaining power, the more closely the renegotiation payoff r will approach the contractual payoff p. Then, the closer the seller's expected return will be to the contract price. Hence, when the seller has bargaining power, the parties will be less inclined to spend resources in a later litigation in order to increase $ – the likelihood that the court will make a correct interpretation. Recall that bargaining power is a function of the parties' relative discount rates, which commonly are the same, and the parties'

---

[63]In addition to expanding the evidentiary base, the parties could also increase surplus by changing the contract's payoff structure. See TAN 64-65, infra. Since the buyer would benefit, through the price term, from actions that increase surplus, the buyer would agree to either method when the method would be cost justified.

[64]A reader may note that the first model assumes that error always is possible while the second model assumes that courts sometimes will find the correct answer with certainty. The distinction between representations is adopted for convenience and does not affect our conclusions. To see why, we can rewrite the initial model so that courts will sometimes find the correct answer with certainty. The buyer's expected payoff then can be expressed as $E(s_b{*}B_{min}) = \$(s_b(i^*) + (1 - \$)[s_b(i^*) + \ ,] = s_b(i^*)$ because , has mean zero. That is, the buyer's expected payoff equals in expectation the correct payoff. The models thus differ in their payoff structures, not the particular representations of judicial error that are chosen.

disagreement points. The seller's disagreement point is largely a function of how redeployable is its investment. Its bargaining power is minimized when, as in the illustration in Part IIIA, its investment is not redeployable at all. The seller's bargaining power thus increases as the subject of sale becomes less specialized or experimental.

The second factor influencing the parties' tradeoff between accuracy and cost is how productive a piece or category of evidence would be in increasing the likelihood that a court will make a correct interpretation. In the example above, the introduction of a single piece of evidence – the pre-agreement memorandum – was assumed to increase materially the probability that the court would find the correct answer. The memorandum thus was very productive, generating a large increase in accuracy at a low cost. When evidence is expected to be less productive, parties will be less inclined to have courts make interpretations on a broad evidentiary base.

This analysis of the relevant tradeoff suggests that parties whom the payoff invariant case describes would more commonly prefer a narrow evidentiary base for interpretations. Initially, many more deals are for relatively standard goods that a party's specialized investment will improve than they are for goods that would be worthless in uses other than those the parties contemplate. Thus, parties such as the seller here ordinarily could redeploy a significant fraction of their investment, and so would have a nontrivial amount of bargaining power in a renegotiation. Further, the choice that parties commonly face in contract actions is not between admitting a particular piece of evidence or not admitting it; the choice is between a summary judgment motion on the contract – on B_{min} – or a trial, often before a jury.[65] Trials can be very costly. Thus, parties apparently would prefer broad evidentiary bases either when their performance would be highly specialized or when an evidential category in addition to B_{min} would be very productive -- a clear

[65]Courts frequently consider extrinsic evidence when interpreting the written contract. A commentator recently explained: "Extrinsic evidence includes both evidence about trade customs and evidence about interchanges between the parties – concerning the course of performance of the current contract, the course of dealing in prior transactions, or the bargaining history of the current contract. Very often this extrinsic evidence will not be solely documentary and will require evaluation of oral testimony about conversations between the parties." See William C. Whitford, *The Role of the Jury (and the Fact/Law Distinction) in the Interpretation of Written Contracts*, 2001 Wisc. L. Rev. 931, 937. Over 50% of tried contract cases in Federal courts are to juries, and between 25 and 30% are tried to juries in state courts. See Marc Galanter, *Contract in Court: Or Almost Everything You May or May Not Want to Know about Contract Litigation*, 2001 Wis. L. Rev. 577.

custom, for example.[66] Since these illustrations apparently capture only a minority of cases, we suggest that the majoritarian preference in the payoff invariant case also is for courts to use narrow evidentiary bases when making interpretations. Turning again to the contract creation stage, given this preference typical parties would invest resources in drafting just to create the evidentiary base -- B_{min} -- that would permit a court to make an unbiased interpretation.

Where payoffs are discontinuous in the interpretation space, however, parties also have the ability to protect the seller's expectation without expanding the evidentiary base. Their method is to change the contract's payoff structure. The structure of the payoff invariant case creates the possibility that a party will incur a large loss because of a small judicial error. The seller in the example here thus could suffer substantially if the court held that it to had to produce only a slightly higher level of quality than the correctly interpreted contract required. It is this "knife-edge" property of the perfect tender rule that reduces the seller's expected payoff below the price and so creates an incentive for parties to contract out of the rule. The common way to do this is to use the customary warranty term, which contains a repair and replacement clause that eliminates the buyer's right to reject but requires the seller to repair or replace defective tenders or parts of tenders. When the buyer cannot reject, the seller is entitled to the price, with damages deducted from it.[67] The customary warranty thus eliminates the knife-edge feature of the perfect tender rule, thereby ensuring that both parties' expectation will be protected if they believe that their payoffs will be determined by the correct interpretation of the repair and replacement clause.

This condition will be satisfied because the repair and replacement clause is analytically similar to the product preparation term considered in Part IV(B)(1). Under that term, a party's expected payoff, $E(s_b(i))$, equaled the correct interpretation, i*. To see why the repair and replacement clause has the same property, recall that UCC §2-719 permits a seller to limit the buyer's remedies to repair and replacement, while §2-608(1) permits the buyer to "revoke his

---

[66]Richard Craswell shows, however, that courts evaluate the probativeness of customs by evaluating the purposes that the customs are meant to serve. See Richard Craswell, *Do Trade Customs Exist?* in THE JURISPRUDENCE OF CORPORATE AND COMMERCIAL LAW 118 (2000). This suggests that custom evidence is costly to admit and to contest.

[67]See UCC §2-717.

acceptance" only if a "non-conformity substantially impairs its [the good's] value to him". Section 2-719(2) authorizes a court to refuse enforcement to a repair and replacement clause if "circumstances cause ... [the] limited remedy to fail of its essential purpose...." Comment 1 to §2-719 explains that this authorization exists because "there must be at least a fair quantum of remedy for breach of the obligations or duties outlined in the contract." In light of these sections, a court's interpretive task under a repair and replacement clause is not to see if the goods fail to conform to the contract in any respect. Instead, the initial question for a court is whether the degree of success the seller achieved in the repair or replacement task was such as to give the buyer "a fair quantum of remedy." If not, the clause becomes inoperative and the question shifts to whether the goods are "substantially impaired" in value. The buyer's possible payoffs under a repair and replacement clause thus increase as the court requires the seller to exceed the correct interpretation and decrease as the seller is permitted to fall short, just as the payoffs in the product preparation case. And since the seller's payoffs here are a function of the buyer's payoffs, the seller's possible payoffs also are continuous in the same way. Both parties thus will expect the payoff under a repair and replacement clause to equal in expectation the payoff the parties intended. This suggests that parties choosing a repair and replacement warranty will want courts to interpret the warranty using the minimum interpretive base B_{min}. This is because, as we have argued above, it is ordinarily enough for business parties that courts are correct in expectation, not that they are always correct.

3. Summary

The case in which the parties' payoffs are continuous in the space of a court's possible interpretations covers a lot of the ground. This is because the case often arises "naturally", as in the product preparation example, and can arise "artificially", as when parties contract to create continuous payoffs in order to increase the efficiency of their payoff structure. Firms that the case describes ordinarily prefer courts to follow a textualist interpretive style. The case in which the parties' payoffs are invariant to the degree of judicial error probably is less common. Even in this case, however, litigation cost considerations will cause a majority of parties to prefer courts to use a narrow evidentiary base. Therefore, the best interpretive default for firms is textualist when the issue is what their contract language meant.

We conclude this section with a comment concerning judicial practice. Courts making common law adjudications commonly take a Willistonian approach[68], while the UCC strongly urges a contextualist interpretive style.[69] Courts in general, however, treat interpretation rules as mandatory. Judges are reluctant to invoke the coercive machinery of the state to require a party to perform a contract (or to pay damages) unless the judge is satisfied that the contract actually directed what the party failed to do. It seemingly follows that courts, not parties, should choose the rules that determine how contracts are read. This view is understandable but misguided. The law in general permits persons and firms to make litigation relevant choices even though a consequence of those choices is that courts may act on less than full information. For example, persons and firms may waive the right to counsel, agree to stipulated findings of fact, and use summary arbitration procedures whose results courts are required to enforce. In sum, the law generally sacrifices accuracy in adjudication to persons' self interested choices. Similarly, parties should be permitted to realize the cost savings from contract interpretations on minimal evidentiary bases even if, in any given case, the odds of an accurate interpretation would be higher with a broader base.[70]

D. Private Languages, Linguistic Defaults and the Parol Evidence Rule

1. The preferred linguistic default.

We now consider the case in which the parties have written a complete contract in some language. The issue is whether, if the contract is silent on the matter, a court should take the parties to have written in majority talk. The alternative judicial assumption would hold that the parties

---

[68]See Peter Linzer, *The Comfort of Certainty: Plain Meaning and the Parol Evidence Rule*, 71 Fordham L. Rev. 799 (2002); Robert E. Scott, *The Uniformity Norm in Commercial Law: A Comparative Analysis of Common Law and Code Methodologies* in THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW 149 (J. S. Kraus & S. D. Walt, eds., 2000).

[69]See Comment 1 to UCC §1-205: "This Act rejects both the "lay-dictionary" and the "conveyancer's" reading of a commercial agreement. Instead, the meaning of the agreement of the parties is to be determined by the language used by them and by their actions, read and interpreted in light of commercial practices and other surrounding circumstances."

[70]If parties preferred courts to use evidentiary bases so broad as to foreclose judicial time for deciding other categories of case, then courts should override the parties' preference. This danger would not occur when parties in general want adjudications to be briefer than they now are, and there seem no other possible negative third party effects from judicial deference to the parties regarding interpretive styles.

preferred, should a dispute arise, to have the opportunity to introduce extrinsic evidence that relevant parts of the contract were written in the parties' private language. We next set out four reasons why the "linguistic default" should hold that the contract was written in majority talk. The practical implication of this proposal is that, when a contract does not speak to the issue, the court should not go beyond the evidentiary base B_{min} when attempting to identify the language of the contract.[71]

Many parties would prefer the linguistic default we propose because the default would (a) reduce contracting costs; (b) minimize the opportunities for strategic behavior; (c) reduce the risk of judicial error; and (d) expand the set of efficient contracts parties could write. Beginning with the issue of contracting costs, the comments to UCC §2-202 adopt a strong contextualist linguistic default. Thus, Comment 2 recites: "writings are to be read on the assumption that the course of prior dealings ... and the usages of trade were taken for granted when the document was phrased [that is, were meant to be aids in interpretation] *unless carefully negated*...."[72] But if the default were reversed, parties could contract out of a plain meaning linguistic default at the same or a lower cost. For an example, consider: "This agreement is to be read in light of the customs of the widget trade". Parties seldom would describe the actual customs in the agreement because contracting costs are incurred today with certainty while dispute resolution costs are incurred tomorrow and probabilistically. Thus, it ordinarily will be cheaper just to tell courts to consider custom should an irreconcilable difference later arise.[73] In addition, a minority of contracts are written largely in

---

[71]The position that a party always should be able to show that its contract was written in party talk sometimes is justified on autonomy grounds. Thus, Justice Traynor explained in Pacific Gas and Electric Co., supra note 53,: "In this state, ... the intention of the parties as expressed in the contract is the source of contractual rights and duties. A court must ascertain and give effect to this intention by determining what the parties meant by the words they used." at __. This argument is a non sequitur as stated. No one would claim that the Statute of Frauds interferes with party autonomy in a normatively serious way, though the Statute sometimes requires parties to cast their agreements in written form. It apparently would not be a serious interference with autonomy for courts also to assume that parties cast their agreements in majority talk if good reasons exist to create the incentives that this assumption implies. Thus, it takes more argument than now exists to show that an autonomy based view of interpretation would justify the rule Justice Traynor stated, even if such a view were appropriately applied to firms.

[72]UCC §2-202, Comment 2 (emphasis added).

[73]Some scholars argue that if courts are Willistonian, and thus implicitly adopt majority talk as the linguistic default, parties will incur additional drafting costs translating their private language into the majority language in order to make their intentions clear to judicial interpreters. See Jody S. Kraus and Steven D. Walt, *In Defense of the Incorporation Strategy* in THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL

private languages, and the parties that write them frequently enforce them with expert arbitrators, not lay judges. Thus, fewer parties would have to contract out of a default that supposed them to be writing in majority talk than would have to contract out of the contrary default.[74]

A plain meaning linguistic default -- that is, a default which restricts the court to the interpretive base B_{min} -- also would reduce strategic behavior. To see why, focus first on the ex post case, and consider a contract between party A and party B whose relevant provisions were written in party talk. This contract, suppose, turns out badly for party B. It would like to raise an interpretation issue strategically, claiming that the contract was written in majority talk, in order to improve its bargaining position. B *could* raise such an issue only if it could plausibly show that words in the parties' private language had a clear but different meaning in the majority linguistic community. B *would* raise the interpretive issue only if the different plain meaning would also relieve it from performance. These two conditions for avoiding the effect of a private language are difficult to satisfy because while there are many private languages there is only one majority language. Hence, it would only be coincidence for words in a particular private language also to have a clear but different meaning in the majority language that favored a litigant such as B. Now assume that the contract was written in majority talk. The multiplicity of possible private languages would permit party B more easily to assert a helpful private meaning. A common move is to claim that stated prices or quantities are only "estimates" or "projections" in the private language the

---

LAW 193 (2000). This argument assumes that parties can only opt out of the plain meaning linguistic default with costly translations. But as the discussion in the text has shown, parties can cheaply opt into a private language by agreeing in their contract to have evidence admitted regarding that language should a dispute arise. We can expand this argument with the observation that even when contracts contain technical party talk, most of their words will be written in the majority language. For example, parties may attribute a private meaning to the phrase "two by four" (wood supports so described in construction contracts are meant to state dimensions of 17/8" x 33/4"), but such parties seldom would use a private language to describe the delivery date, the place of delivery, the price and so forth. These parties would want words with trade language meanings to be read with trade understanding, but would not want words written in the majority language to be read as if they were special. If the linguistic default is whatever language parties choose, then parties who wish to exclude party talk interpretations of majority talk must identify all of the majority terms and explicitly negate the use of extrinsic evidence for interpreting them. On the other hand, by opting into only the technical party talk they wish to incorporate, parties can more readily unbundle the two types of language. Hence, if courts are Willistonian just when parties want them to be, parties will not incur unnecessary writing costs.

[74]The current interpretive rules are mandatory (or "quasi mandatory"). For the purpose of our discussion here, which focuses on party preferences, it is clarifying to treat the rules as defaults.

parties used.[75] Turning to the ex ante case, on the linguistic default we favor, a party would have to propose to its contract partner that the contract be interpreted in a particular private language. The partner would agree only if there were such a language, and its meanings would be accessible to courts. Thus, requiring parties to say they are writing in a private language would largely ameliorate the concern that a party would attempt to rescue itself from a bad deal by claiming that the contract was written in a mythical private language.

The two linguistic defaults at issue now can be reconsidered in light of this analysis. If a court would permit parties always to prove that they wrote in a private language, then a disappointed party would have a strong incentive to attribute a fictional favorable private meaning to a majority talk contract. A default that supposes parties to be writing majority talk *unless* their contract recited otherwise would eliminate this move. And then if parties actually did write in a private language and contracted out of the Willistonian default, it would be difficult for one of them later to behave strategically regarding interpretation because, we have just shown, the conditions for avoiding particular private languages are so stringent. The linguistic default that supposes parties to be writing majority talk thus would reduce strategic behavior more effectively than the rules in the comments to today's Uniform Commercial Code.

A plain meaning linguistic default that assumes parties to be writing in majority talk also would reduce the risk of judicial error. This is because, as said, there is one majority linguistic community but many possible private communities. Hence, when private languages are permissible, as they often are today, a court has two interpretive tasks: to ascertain the parties' language and to ascertain what the parties said in that language. A court that picks the wrong language necessarily will pick the wrong meaning. This danger is real because a private linguistic community can be as small as the parties to the contract. Hence, a disappointed party may plausibly claim that the parties' course of dealing or their oral negotiations showed that, in the parties'

---

[75]As an example, a contract required the buyer to take "approximately 70,000 cubic yards of cement" and also recited that "no conditions which are not incorporated in this contract will be recognized." The buyer took a little over 12,000 yards in a falling market and successfully claimed that parties in the trade understood explicitly specified quantities to be estimates. See Southern Concrete Services, Inc. v. Mableton Contractors, Inc., 407 F. Supp. 581 (N.D. Ga. 1975).

language, "all" meant "some"[76]; agreeing to take a "minimum" quantity meant that the buyer could take much less[77]; an unconditional option is conditional.[78] When such a claim is false but found to be true, the court necessarily will misinterpret the contract.

To see why, recall our argument in Part IVC that firms are content to have courts be right on average, not right every time. The error here is different from the error described there. For example, let a contract use the word "red", and let a party persuade the court, wrongly, that the parties used a private language in which the word "red" meant "green". Both red and green are vague. The space of possible judicial interpretations of the illustrative contract would likely center around some modal instance of the concept "green", but the court here would not be right on average; for the parties actually intended the court to be construing the word red. When courts are mistaken regarding the contract's language, that is, their constructions must be inefficient because it is only efficient to protect a party's expectation interest under the contract it actually wrote. Hence, parties face a heightened risk of inefficient interpretations when courts always entertain claims that a contract was written in a private language. Typical parties thus would prefer courts to assume that they wrote in majority talk.

Finally, contextualist interpretation in general, and especially contextual interpretation that permits the use of private languages, can truncate the set of efficient contracts that parties can write. To understand this point, recognize that the cost of writing a contract includes the expected cost of enforcing it. We have just argued that contextualist interpretation can create moral hazard. The more complex the contract, it is plausible to believe, the easier it will be to create disputes regarding what the contract says and what language it was written in. This creates the possibility that contextualist interpretive regimes encourage parties to use simpler but possibly less efficient contracts.[79]

---

[76]Pacific Gas & Electric Co. v. G.W. Drayage & Rigging Co., supra note 53.

[77]Columbia Nitrogen Corp. v. Royster Co., 451 F.2d 3 (4th Cir. 1971).

[78]Hunt Foods & Industries v. Doliner, supra note 50.

[79]A formal treatment of the relation between contracting costs and the parties' choice of contractual form is Alan Schwartz and Joel Watson, *Economic and Legal Aspects of Costly Contracting*, Mimeo (2003).

To illustrate, recall the relation specific-investment example in Part III(A), where the seller chooses between producing a generic product at a zero gain or a specialized product at a gain of $10. We now assume that this contract costs the parties $5 to write, split evenly, while a contract for the generic product would cost $2 to write, which the buyer pays.[80] Begin by assuming that the parties do write the contract to produce the specialized product. The buyer again rejects on quality grounds. We now suppose, as in the invariant payoff case of Part IV(C)2, that if the court uses a contextualist interpretive style, it will find the correct answer (so that the seller will realize $10) with a probability of .7. Conversely, the court will make an unbiased interpretation of the contract's required quality level with a probability of .3. In this event, the seller will receive the contractual payoff of $10 two thirds of the time, and the low renegotiation payoff of -$15 one third of the time. The seller's expected return under this interpretive style thus is $9[81] minus its share of the contracting cost ($2.50) minus enforcement cost. In the textualist regime, the court is assumed to find the correct answer only .5 of the time, so the seller's expected return is $8.34 minus contracting and enforcement costs.

We use the argument that a contextualist interpretive regime creates a greater incentive for moral hazard to suppose that the seller could enforce the contract on summary judgment in a textualist interpretive regime, but would be forced to a trial in a contextual regime. Putting numbers to this assumption, we let the expected cost of a summary judgment action be $4 and the expected cost of a trial be $8. As a consequence, the seller would have a negative expected return of $.50 in a contextualist regime and a positive expected return of $1.84 in a textualist regime. Therefore, if the court's expected interpretive stance were contextual, the seller would only agree to produce the generic product. Note too that this result obtains when the seller has the least bargaining power that it could have; for we assumed above that its investment in the specialized product is not redeployable at all.

---

[80]Recall that price equals cost in competitive markets.

[81]The seller's expected return is $.7(10) + .3[\frac{2}{3}(10) + \frac{1}{3}(-15)] = 9$.

When parties compare possible interpretive regimes, they will not only consider ex post enforcement costs and gains; they will add the present value of the net to the cost of writing a contract. In the example here, doing that would cause parties to prefer the simple but less efficient contract for the generic product to the more complex but more efficient contract for the specialized product. To be sure, an example cannot show that contextualist interpretive regimes always induce parties to use less efficient contracts. The lesson rather is that an overlooked cost of these regimes is that they sometimes buy greater accuracy at a cost of less efficient contracting. This disadvantage reenforces our view that business parties generally prefer a plain meaning linguistic default, which implies the use of a Willistonian interpretive style.

The analysis here permits further us to clarify the debate between contextualists and textualists and also suggests a new understanding of the function of merger clauses. Contextualists claim that interpretations made on broad evidentiary bases are more likely to be correct than interpretations made on narrow bases. Textualists dispute this claim. The dispute between these camps, we have shown above, is irrelevant when the issue is what the contract says: business firms are content with interpretations of their language that are correct on average, not always correct, and so prefer narrow evidentiary bases to broad ones. The interpretation dispute matters a great deal, however, when the interpretive issue concerns the language in which the parties wrote. In such a case, we have argued, a broad evidentiary base affords a disappointed party the opportunity to raise the language issue strategically. Broad evidentiary bases also increase the risk of judicial error and truncate the set of efficient contracts parties can write. Our analysis thus supports the conclusion that courts should interpret business contracts on minimal evidentiary bases whether the issue is what the contract language means or what language the contract was written in, unless parties explicitly instruct the court otherwise.[82]

[82]This conclusion may be thought to raise a regress problem. If parties can give interpretive instructions, then those instructions will themselves have to be interpreted, as would instructions as to how to interpret the instructions and so forth. This problem does not appear to be serious. The default we advocate would require parties to tell courts to widen the evidentiary base. For example, the contract would recite: "Use custom". There seems little reason for parties to say "Use custom sympathetically". Courts seldom would need instructions as to how to interpret simple directives that refer the courts to evidentiary categories now in use.

Merger clauses are understood to have the function of restricting the evidentiary base available to courts when making interpretations.[83] Because courts today often search broadly for the correct answer, merger clauses are difficult for parties to enforce.[84] These clauses now can be seen to have an additional function. A merger clause, if honored, would limit the court to the evidentiary base B_{min}. When courts see only this base, parties have a strong incentive to write the contract in majority talk. Hence, when parties do adopt a merger clause, they are telling the court that they are speaking its language. The realization of this should incline a court that is not persuaded by the arguments here to permit a merger clause to bar the admission of extrinsic evidence offered to show that the contract at issue was written in party talk.[85]

---

[83]A merger clause recites that the written agreement is the parties' final expression of their intentions. A common example states:

> This contract contains the final understanding between the parties, and represents the final agreement on all terms. There are no verbal agreements or representations in connection therewith. The writing is a merger of all proposals, negotiations and representations with reference to the subject matter and provisions.

See, e.g., Luther Williams, Jr., Inc. v. Johnson, 229 A. 2d 163 (D.C. App. 1967); UAW-GM Human Resource Center v. KSL Hotel Corporation, 228 Mich. App. 486, 579 N.W. 2d 411 (1998).

[84]See Franklin v. White, 493 N..E. 2d 161, 166 (Ind. 1986) ("An integration clause is only some evidence of the parties' intentions. The trial court should consider an integration clause along with all other relevant evidence on the question of integration."); Sutton v. Stacey's Food Mart, Inc., 431 A.2d 1319, 1322 (Me. 1981) (Citing Restatement (Second) of Contracts §216, comment e, which provides that "A merger clause does not control the question whether the writing was assented to as an integrated agreement."); Restatement (Second) of Contracts §209, comment b: "Written contracts may include an explicit declaration that there are no other agreements between the parties, but such a declaration may not be conclusive."). Professor Corbin is regarded as the most persuasive advocate of this position: "It can never be determined by mere integration of the words of a writing whether it is an 'integration' of anything, or whether it is the 'final and complete expression of the agreement." 3 Arthur Corbin, Contracts §581 at 442.

[85]We argue that courts should adopt a Willistonian linguistic default because parties prefer it, but there also are "external" efficiency reasons to justify the default. First, the state subsidizes the judicial system and thus has an independent reason to reduce the likelihood of disputes. For the reasons given, our default would do better at this than current law. Second, if contracts are written in majority talk, courts can create standard vocabularies in which commercial transactions can be conducted. When a phrase has a set, easily discoverable meaning, parties who use it will know what the phrase requires of them and what courts will say the phrase requires. Courts that insulate the meaning of terms in the majority language from deviant interpretations – that interpret the same words in the same ways across cases – thus create a collective good: a set of terms with meanings that are already understood by a large majority of potential contracting parties. It follows that courts should encourage parties to use majority talk. See Robert E. Scott, *The Case for Formalism in Relational Contract*, 94 Nw. L. Rev. 847, 853-56 (2000); id., *The Uniformity Norm*, supra note __ at 157-58; Alan Schwartz, *Contract Theory and Theories of Contract Regulation*, 92 Revue D'Industrielle 101 (2000).

2. The parol evidence rule.

A typical statement of the parol evidence rule provides that when terms are "set forth in a writing intended by the parties as a final expression of their agreement", the terms "may not be contradicted by evidence of any prior agreement or of a contemporaneous oral agreement but may be explained or supplemented by course of performance, course of dealing, or usage of trade."[86] This rule comes in a "hard" and a "soft" version. Courts that adopt the hard version of the parol evidence rule decide whether a written contract is ambiguous from the document itself; that is, the court makes an interpretation on the evidentiary base B_{min}. Courts that adopt the soft version hear all extrinsic evidence before deciding whether there is an ambiguity.[87] The UCC and a number of common law courts are taken to have adopted the soft version.[88] As should be apparent, this version of the parol evidence rule is justifiable if courts should consider *all* evidence that may bear on what the parties meant, but it is not justifiable if courts should consider *only* the evidence that parties, ex ante, want courts to see.

Contrary to the conventional understanding, however, the UCC version of the parol evidence rule actually creates a strong incentive for parties to write in majority talk. Under §2-202, extrinsic evidence cannot "contradict" a writing but can "explain or supplement" the writing. Suppose that a contractual phrase has the plain meaning m_M in majority talk and the equally plain but different meaning m_P in party talk. If a party can introduce extrinsic evidence explaining that the contract actually had the meaning m_p, then extrinsic evidence never could contradict the writing. The meaning m_p would not "contradict" the meaning m_M because the writing, properly understood, never meant m_M. Put more vididly, if a party may introduce extrinsic evidence to show that the parties meant "green" when they wrote "red", extrinsic evidence could "explain" but never "contradict" their contract. A contradiction could arise only if the contract *were* written in majority talk and the extrinsic evidence was offered to show a different meaning in that language than m_M. Therefore, unless the word "contradict" is read out of the statute, section 2-202 must be taken to

---

[86]UCC §2-202(a).

[87]See Whitford, supra note 65, at 939; Eric Posner, *The Parol Evidence Rule*, *the Plain Meaning Rule, and the Principles of Contractual Interpretation*, 146 U. Penn. L. Rev. 533 (1998).

[88]See Ben Shahar, supra note 48; Snyder, supra note 48; and etc.

presuppose that contracts are written in majority talk, but to permit the introduction of extrinsic evidence to clarify ambiguities in this language. In contrast to the comments to section 2-202, the Code itself thus adopts the linguistic default for which we argue.[89]

3. Course of performance evidence.

The UCC explicitly invites courts to consider the parties' course of performance under a contract because such evidence is said to be "always relevant" to the contract's meaning.[90] If the state is to provide the interpretive theory that the parties want, however, then supplementing contracts with course of performance evidence would frequently be a mistake. The parties' course of performance under a contract differs from evidence in the other evidential categories because it can be offered not only to show what the parties originally meant, but also to prove that the parties' meaning had changed. Thus, evidence that the buyer accepted shipments at quality levels below those the contract specified may show that the parties modified the contract's quality requirement. Admitting course of performance evidence to prove a change in meaning is consistent with the traditional contract law rule that the parties' agreement may be inferred from acts or silence.[91] Nevertheless, courts should be reluctant to admit act or acquiescence evidence to show a change in the meaning of a written contract. The existence of a writing indicates that the parties once believed that the gains from writing things down exceeded the costs. The best inference to be drawn from the absence of evidence that this calculus has changed is that parties want modifications to be written as well. And if this inference is not compelling, it should become so when the contract

---

[89]This proposed interpretation of §2-202 actually was made in Southern Concrete Services v. Mableton Contractors, Inc., 407 F. Supp. 581 (N.D.Ga. 1975). The contract there required the defendant to take "approximately 70,000 cubic yards" of concrete. The defendant purchased 12,542 cubic yards. It defended the subsequent law suit with the claim that trade custom and supplementary agreements between the parties would show that "the quantity stipulated in the contract was not mandatory ... and that both quantity and price were understood to be subject to renegotiation." The court excluded the evidence because it believed that an explicit quantity requirement would be contradicted by an understanding that the requirement was not mandatory. It explained: "To admit evidence of an agreement which would contradict the express terms of the contract would clearly eviscerate the purpose of §2-202." Id. at __.

[90]UCC §§ 2-202(a); 2-208. Comment 2 to §2-208 provides that "a course of performance is always relevant to determining the meaning of the agreement."

[91]See e.g., UCC §2-208, Comment 1: "The parties themselves know best what they have meant by their words of agreement and their action under that agreement is the best indication of what that meaning was."

contains a term requiring modifications to be in writing. Courts, however, accepting the UCC's invitation, often hold that conduct can waive a "no oral modification" clause.[92]

The Code and the courts' use of course of performance evidence to establish a change in meaning reflects a misunderstanding of the parties' likely intentions. To see why, consider the quality level illustration in the preceding paragraph. Suppose that the contract price was $2000 per lot delivered and the contract contained a "no oral modification" clause in addition to the quality specification. The parties expect that deviations from the specified contract quality will sometimes occur. The expected loss that the *average* deviation (the probability of a deviation times the cost) imposes on the buyer is $100. Thus, sometimes a deviation will create a cost that approaches zero (perhaps the buyer has forgiving customers or the quality shortfall can be quickly corrected). At other times, a deviation can impose a loss whose expected value equals or exceeds $500 (say the buyer has a new, potentially large customer for whom quality is important, or the particular deviation would be slow to correct in a high demand period). The parties also know that it would not be cost justified to litigate against the average quality shortfall (the litigation cost, say, would be $150 per deviation).[93]

The contract in this example has three salient features: (i) The price does not fall with declines in the level of quality supplied; (ii) There is a written specification of the quality the seller is to deliver; (iii) There is a "no oral modification clause". These contractual features support two conclusions respecting the interpretive relevance of course of performance evidence. First, the parties do not expect to litigate the average quality deviation; rather, the expected cost of these over the life of the contract is reflected in the fixed price. Second, high cost deviations in product quality

---

[92]See UCC §2-209(4): "Although an attempt at modification or recission does not satisfy the requirements of subsection (2) [excluding modifications except by a signed writing], it can operate as a waiver." Comment 4 explains that "subsection (4) is intended , despite the prohibition of subsection (2) ..., to prevent contractual provisions excluding modification except by a signed writing from limiting in other respects the legal effect of the parties' actual conduct...."

[93]This example was stimulated by the model in Ben-Shaher, supra note 48, but his model apparently assumes that deviations impose the same loss in each period rather than the same loss on average. On his assumption, parties are indifferent to whether courts admit course of performance evidence to show a change in meaning or they do not. Ben-Shaher later relaxes the same loss assumption, however, and then concludes, as we next do, that the soft version of the parol evidence rule will disadvantage typical parties.

are prohibited. From this it follows that the seller could not plausibly infer from the buyer's acceptance of nonconforming deliveries causing average losses that the buyer also will accept nonconforming deliveries that would impose large losses. And if this inference is prohibited to the seller, it also must be prohibited to the court.

This example captures an important feature of contracting behavior. Business parties incur costs to cast obligations in written form partly in order to permit a party to stand on its rights when standing on the contract matters. Course of performance evidence therefore commonly will be irrelevant to show what the contract originally meant or what it currently means. This is because the parties' amicable behavior after the contract likely evidences only their view regarding how the average case should be treated. Courts, however, see the unusual case that the contract was, in considerable part, written to govern.[94] This example thus illustrates an ambiguity in the principle that what matters is to ascertain the parties' intentions. Parties may have multiple intentions. Their actions under the contract often evidence their intentions for typical cases, but seldom evidence their intentions for the atypical case. Thus, a court is likely to make a category mistake when it infers the parties' preferences regarding how a litigated case should be treated from an evidentiary base an important part of which is the parties' behavior in non-litigated cases.[95]

V. THE LEGAL DEFAULT PROJECT

In Part IV, we argued that sophisticated parties commonly prefer a default theory of interpretation that instructs courts to use narrow evidentiary bases and to presume that business

---

[94]Our argument here is similar to Llewellyn's view of custom. He believed that customs govern ordinary cases but seldom are relevant to the unusual cases that cause parties to litigate. See Alan Schwartz, *Karl Llewellyn and the Origins of Contract Theory* in THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW __ (J. S. Kraus and S. D. Walt, eds., 2000).

[95]Arbitrators obey the parties' interpretive instructions but courts do not. Our analysis thus identifies a reason for parties to use arbitration, but it cannot support the inference that parties use courts because they prefer the interpretive styles in current use. There are other reasons for using courts, such as the ability to get discovery, to have appeals, to have the substantive law applied by experts in it, to create effective precedents and the like. See Kenneth S. Abraham and J.W. Montgomery, *The Lawlessness of Arbitration*, Mimeo 2002; Charles L. Knapp, *Taking Contracts Private; The Quiet Revolution of Contract Law*, 71 Fordham L. Rev. 761 (2002). The widespread use of merger and no oral modifications terms in contracts intended for courts also suggests, in line with our views, that parties who prefer judicial enforcement also prefer enforcement under a different interpretive style.

contracts are written in the majority language. Courts can only interpret what is said, so our analysis assumed that the parties' writing was complete for the subjects at issue. Contracts, however, are often incomplete in relevant respects. The question we take up here, therefore, is whether firms would prefer the state to complete these contracts with default legal terms. The somewhat surprising answer we derive from contract theory is that most state-created defaults will be useless or inefficient. Firms would prefer the state not to create inefficient defaults because firms will contract out of them; thus, the only effect these defaults will have is to increase transaction costs.[96]

Commercial parties commonly will want to condition their obligations on the nature of their contract partners or on the state of the world that will materialize after the contract is written. For example, a seller would like to condition its warranty obligation on the buyer's intensity of use: the seller would charge more or warrant less for more intense users because these users are more likely to make claims. A buyer also would like to condition price or quantity on the nature of demand ex post. If demand turns out to be high, the buyer would be willing to pay more or to take more product; but the buyer would want a lower price or less product if demand turned out to be low. These considerations imply that contracts will inevitably be incomplete. This is because there are an infinite number of possible future states and commonly a very large set of possible partner types. When the sum of possible states and partner types is infinite and contracting is costly, contracts must contain gaps. Parties cannot write contracts about everything.

Incomplete contracts sometimes produce law suits because parties will not always agree ex post regarding the treatment of omitted contingencies.[97] Courts in such cases no longer can do interpretation because, by definition, the contracts they see contain gaps. The courts' task thus shifts to the development of rules to resolve gap cases. Hence, some default rules are judicially

---

[96]See note 128, infra.

[97]Parties tend to omit low probability states. For example, if very low demand is thought quite unlikely, parties will not incur the contracting cost to specify their obligations in the very low demand state. A law suit could arise if this state materializes and the parties cannot agree on a new price.

created.[98] Restatement and statutory drafters also create defaults when, in their view, certain gaps are likely to recur. Though our analysis has relevance for courts, we focus here principally on the Restatement and the UCC, to ask just when the state can create good defaults for business parties.[99] In particular, we derive criteria for efficient defaults, and then argue that these criteria are difficult for drafters to satisfy.

A. The Case for Defaults

The commonly held view is that but for the cost parties would write complete contracts. The state may increase efficiency in cases when contracting costs prevent parties from solving contracting problems. To see how, consider the problem of developing a damages rule -- a contract term specifying the transfer, or the method of computing the transfer, that a party in breach must pay to its contract partner. Any particular set of parties would bear the full cost of solving this problem but likely could not capture the full gain because it can be difficult to prevent later parties from copying successful solutions that have precipitated into contract terms.[100] In this circumstance, the total social gain from having a rule -- a solution to a contracting problem-- may exceed the social cost but parties themselves will not create the rule. There is a case, therefore, for a publicly supplied contract law that contains efficient solutions to common contracting

---

[98]Since these rules are facilitative, courts should permit future parties to vary them. Judicial creation of such gap-filling defaults is not inevitable. A court instead can refuse enforcement when gaps cause a contract to be obligationally incomplete: that is, when the terms the contract does contain provide an insufficient basis on which to ground a remedy. The common law rule is that contracts leaving material terms incomplete or indefinite are not legally binding. It may be paradox that the indefiniteness rule is itself a default. Parties can opt out of it by more completely specifying their obligations in the contract.

[99]Restatement defaults differ from statutory defaults because a restatement is not self-executing. A restatement provision cannot become a legal default until it is both promulgated by the ALI and then adopted by a court when resolving a dispute. Thus, our analysis here applies both to the private legislatures that promulgate defaults and to the courts that adopt them.

[100]For analyses of how legal practice innovations spread, see Gerald F. Davis and Henrich R. Greve, *Corporate Elite Networks and Governance Changes in the 1980s*, 103 Am. J. Sociology 1 (1997); Gerald F. Davis, *Agents Without Principles? The Spread of the Poison Pill through the Intercorporate Network*, 36 Admin. Sci. Q. 583 (1991).

problems.[101]

This conclusion implies that contract law rules should be defaults. This is because parties are heterogenous in large economies so that not everyone will want the same thing. Any particular default rule is itself justified as doing for parties would they would have done for themselves had their contracting costs been lower. Hence, parties who will do for themselves -- create their own solutions -- should be free to do so. While this conclusion may seem obvious, courts sometimes create barriers to contracting out of Restatement or UCC rules, on the ground that these rules either reflect better or fairer solutions than those that parties could develop.[102] To the contrary, we will next argue that commercial parties themselves are the best judges of what constitutes a good contracting solution, and that default rules should not be chosen on fairness grounds.

A perhaps more precise way to state the ground on which drafters should choose defaults is that a good default minimizes contracting costs. Parties, if left free, will supplant or modify state supplied terms that they dislike. In a world of free contracting, unpopular defaults thus will raise contracting costs but not otherwise affect parties' behavior. As a consequence, the state's power is limited to reducing contracting costs, which it best does by enacting popular defaults. This leaves open the question what makes a default popular. This is a difficult question to answer in general but, as we have seen, firms prefer contract terms that maximize joint surplus. Good defaults for firms thus will be popular if they maximize joint surplus and unpopular otherwise.

It is this result that implies the error of choosing defaults just because they are fair. Parties have an incentive, and sometimes the ability, to contract out of even fair defaults that do not maximize surplus.[103] As a good example of this response, commentators and courts once thought

---

[101]See Charles J. Goetz and Robert E. Scott, *The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms*, 73 Cal. L. Rev. 261, 291-93 (1985).

[102] Id. at 283-86. The courts' behavior explains the intensive lobbying by firms regarding the proposed revisions to Article 2, especially the warranty terms. See Robert E. Scott, *The Rise and Fall of Article 2*, 63 La. L. Rev. 1009 (2002). Firms know that UCC defaults can be sticky in practice.

[103]If the state wants a particular rule actually to control behavior, it should make the rule mandatory. Raising costly but legally surmountable barriers to contracting out arbitrarily disadvantages firms whose gains from having their own contract are low in relation to contracting costs, yet because many parties do contract out, the

(and may still think) that it is fair for merchant sellers to make implied warranties of quality. Merchant sellers, however, routinely attempt to disclaim the implied warranty of merchantability in UCC §2-314 and their merchant buyers routinely consent.[104] The effect of §2-314 in business contexts thus often is limited to the increasing of transaction costs. In short, drafters and courts should ask what parties would like, not what parties should want.[105] We argue in the next section that this is a difficult question to answer.

B. The Cost Concern

1. Default Rules.

A good default rule will condition on few possible states of the world, be relatively simple in form, and be efficient for a wide variety of contract parties. The first of these criteria is necessary because there can be an infinite number of possible future states of the world, but the state's drafting costs also are finite. Thus, it likely would not be cost justified to create a publically supplied rule if the rule had to contain directions for parties in many possible future states. The criterion of simplicity is a function of institutional competence. While legal rules can be complex, contract law rules are created by courts and drafters. Courts cannot conduct investigations into the efficiency properties of possible rules and rule combinations. Drafters also have limited resources.

---

policy behind the rule is not fully realized.

[104]Part VC below will argue that the UCC's implied warranty term is inefficient because it provides a seller with too little guidance regarding the performance obligation and also creates moral hazard.

[105]Scholars have identified a set of default rules that are termed "information forcing" because the rules create an incentive for a party to disclose relevant information to its contract partner. See Ian Ayres and Robert E. Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 Yale L. J. 87 (1989). The rules providing that damages cannot be recovered unless they are foreseeable and reasonably certain are information forcing; they create an incentive for a promisee that fears breach to disclose in advance to the promisor the loss the promisee would suffer. The promisor then will be able to price more accurately and to take optimal precautions against breach. Information forcing rules fall within the analysis here because parties can create them on their own. For example, sellers routinely propose contracts that disclaim liability for consequential damages. This creates an incentive for buyers who may suffer these damages to disclose them to the seller and then to purchase insurance – i.e., the deletion of the disclaimer. An information forcing default thus will be popular if parties would agree to disclose in the circumstances to which the rule applies. It has been shown that the foreseeability default may not always be popular in this sense. See Jason S. Johnston, *Strategic Bargaining and the Economic Theory of Default Rules*, 100 Yale L. J. 615 (1990) (Parties facing a monopolist may not want to disclose the profits they would earn from successful transactions because the monopolist then may price discriminate against them.)

The American Law Institute, which creates restatements, and the National Conference of Commissioners on Uniform State Laws, which together with the ALI created the UCC, does not hold legislative style hearings on proposed rules, cannot hire neutral economic or industry experts to help in rule creation, and generally rely on the part time labor of law professors and private attorneys.[106] The drafters thus cannot, and do not, write rules for business contexts that match the tax code's complexity; rather, contract law necessarily is restricted to the consideration of problems whose solutions can be embodied in simple rules.

The efficiency criterion is perhaps the most difficult to satisfy because parties in large economies are heterogeneous. Default rules would be too expensive to create if efficient solutions were party specific. Then there would need to be as many legal rules as there are sets of contracting parties. The task, then, is to find rules that would be efficient – surplus maximizing – in a wide variety of contexts.

There actually are very few default rules[107] in contract and commercial law. In part this is because drafters recognize, at least implicitly, that the three criteria set out above are very difficult to satisfy. We illustrate this point by considering two variants of the investment example analyzed in Part III.A above. Assume first that after the parties contracted for the seller to produce the specialized product, but before the goods were produced, the seller's factory is destroyed in a fire. The seller no longer can produce the goods, but it can pay money damages.[108] The common law default rule nevertheless excuses the seller when an exogenous event makes the contemplated, specialized performance impossible to render.[109] This excuse rule appears to satisfy the criteria for an efficient default. First, rule creation costs are low. There are only two possible future states of the world: an exogenous event either prevents performance or it does not; and there is need for a

---

[106]See Alan Schwartz and Robert E. Scott, *The Political Economy of Private Legislatures*, 143 Pa. L. Rev. 595, 607-37 (1995).

[107]Recall that a rule specifies the relevant conduct in advance – not to drive above fifty five miles per hour.

[108]Recall that the buyer's expectation interest would be $v_s - p_k = \$80 - \$60 = \$20$.

[109]See, e.g., Taylor v. Caldwell, 3 B. & S. 826, 122 Eng. Rep. 309 (King's Bench, 1863); Restatement of Contracts (Second) §263.

rule in only one of them (when the seller cannot perform). The rule also is simple in form: the seller is excused or not, depending on whether there was a fire or not. Finally, the rule likely is efficient for a wide variety of parties. This is because buyers in general are better insurers against lost valuations of specialized investments than are sellers; buyers usually are better informed than sellers about the consequences of seller breach. Excusing the seller requires the buyer either to insure on the market or to reveal its valuation to the seller.

This example is deceptively simple, however. The set of possible causes of a seller's incapacity to perform can be large, ranging from fire or flood, to a temporary or permanent supply shortage, to a government regulation, and so on. A seller's ability to anticipate and to take precautions against the causes of incapacity thus will differ across sellers and among causes. The law will excuse only when the seller's inability to perform resulted from causes that were difficult to anticipate and prevent. Thus, the excuse doctrine raises difficult causation issues. The seller also may be able to perform in part or to perform in full only for some contract buyers. There is a question how a seller should make the resultant allocation decisions. A buyer's ability itself to take precautions or to insure efficiently also may differ across buyers. Drafting *rules* to cover all of these possible causes, effects and parties would be quite costly. As a consequence, the excuse case today is regulated by a standard, not by a set of rules. Section 2-615(a) of the UCC provides that a seller's failure to perform "is not a breach if performance as agreed has been made impractical by the occurrence of a contingency the non-occurrence of which was a basic assumption on which the contract was made...."[110] Courts decide after the fact whether a performance would have been "impractical" and whether the parties had "basically assumed" that the allegedly excusing cause would not occur or would not materially affect performance if it did occur.

To appreciate the second illustration, recall that the investment example implicitly assumed that the parties' contract specified what a specialized product would do, so that if the seller delivered something else breach would be clear. A default rule also could specify the illustrative seller's quality obligation, but recall that the product was assumed to be *specialized to the buyer's use*. A legal rule that set out the quality obligation the seller was to meet thus would apply to one

[110]Section 261 of the Restatement contains the same rule in almost the same language.

case. It is inefficient to draft for one case. A particular seller's quality obligation, however, sometimes can generalize to an industry. It may be, for example, that book publishers would agree on what requirements a good printing press must meet. Even so, the costs to an Article 2 drafter of describing the quality that sellers in every industry must deliver to every industry buyer would likely exceed the social gain. And unsurprisingly, the law also regulates the seller's quality obligation with standards. Under UCC §2-314(2), goods "to be merchantable must be at least such as (a) pass without objection in the trade ... or (c) are fit for the ordinary purposes for which such goods are used...." Courts after the fact decide whether it was enough for a seller to satisfy subsection (a) or (c) ("at least such as"); whether the goods would pass without objection; what "ordinary" seller purposes are; whether the goods "are fit" for such purposes.

Contract and commercial law thus contain very few default rules because parties are heterogeneous in modern economies, good rules sometimes must be complex, and the efficient rule may have to take many possible future states of the world or party types into account. Rather, the Restatement of Contracts and Article 2 of the UCC primarily contain standards; these texts are replete with provisions requiring parties to behave "reasonably", "conscionably", "fairly", "in good faith", and the like.[111] These codifications fail to address important contracting problems. For example, parties to long term contracts face the problem of keeping each party's gain under the contract above the opportunity cost of performance in every period. This problem is best addressed with index clauses, that tie the current period price to current cost and demand conditions. Efficient indices are party or industry specific, however, and so are too costly for contract law to create.[112] Instead, the law ignores this long term contracting problem in favor of letting parties solve it for themselves. So in sum, the cost concern forces contract law commonly to regulate with standards

---

[111]See, e.g., UCC §§2-205, 2-206, 2-207, 2-208, 2-209, 2-302, 2-305, 2-306, 2-307, 2-308, 2-309, 2-311, 2-312, 2-313, 2-314, 2-315, 2-317, 2-503, 2-504, 2-508, 2-511, 2-513, 2-602, 2-603, 2-604, 2-605, 2-606, 2-607, 2-608, 2-609, 2-610, 2-612, 2-614, 2-615, 2-704, 2-705, 2-706, 2-709, 2-710, 2-712, 2-714, 2-715, 2-716, 2-718, 2-719, 2-723. The UCC also adopts standards in order to avoid resolving controversial issues. See Schwartz and Scott, supra note 106. Our analysis here is meant to show that even if the ALI and NCCUSL were bolder, the cost concern would generate many standards.

[112]An index clause links the transaction price between the parties to verifiable external indices, such as the Consumer Price Index, that correlate with the economic conditions facing the parties. The degree of correlation is a function of the type of party (some firm's costs will move more closely with the CPI than other firm's costs will, for example) so that index clauses vary with particular parties or party types.

rather than rules, when it regulates at all.

2. *Default Standards*

The project of creating publicly supplied default standards is difficult to defend. A publicly supplied contract law is justified as solving problems for parties that contracting costs prevent parties from solving on their own. It is relatively costless for parties to write standards in their contracts, however. Thus, for parties to provide that the seller should deliver a "reasonable" quality level, deliver within a "reasonable" time, but be excused if its performance has become "impractical", would exhaust few drafting resources. When a court observes the absence from a contract of a standard that would have resolved the issue at bar, then, the court should not infer that the contract has a gap that typical parties would want the court to fill. Instead, given how easy it is for parties to create standards, the best inference commonly will be that a standard is not "missing" from the contract, but rather its use has been rejected. The court therefore should not imply a standard to decide the case. And for the same reason drafters should be reluctant to enact sets of seemingly missing standards into statutes or restatements. To fill nonexistent gaps is to do for parties what parties do not want done.

The parties reject standards for two reasons. First, firms often need specific guidance regarding the performance obligation. For example, a seller commonly needs to know what quality level to produce. Thus, telling the seller that its product must "at least" satisfy a buyer's "ordinary purposes" is generally unhelpful. In practice, sellers of complex products thus disclaim the implied warranty of merchantability in favor of making an express warranty that describes important aspects of the seller's performance obligation. As another illustration, when an exogenous event induces the seller to consider breaching, both parties need to know whether breach would be legally permissible. An erroneous decision could expose one or the other party to a damage judgment. Telling the seller at this point, with reference to §2-615, that breach is permissible if performance is "impractical" also is unhelpful. And telling the buyer that its return performance (i.e., a payment of an installment of the price) can be suspended under §2-609 if suspension would be "commercially reasonable" is similarly useless. In response to this difficulty, parties commonly ignore the

standards in §2-615 and §2-609 in favor of writing force majeure and insecurity clauses that specify precisely the exogenous causes that will excuse the seller's performance obligation, and specify the permissible responses that the buyer can take when insecure about the seller's performance. Standards thus are useful only when parties can predict accurately the behaviors that courts will find sufficient to satisfy a vague descriptor. This criterion is difficult to meet even for "ancient standards", such as those governing excuse, that have generated extensive, and largely inconclusive, litigation. Firms thus are unlikely to accept new default legal initiatives that would require them to behave "reasonably" or "fairly".[113]

Standards also increase the likelihood of moral hazard. When it is unclear what a party must do, contracting parties have an incentive to interpret ambiguous circumstances in their favor. Thus, a buyer has an incentive to claim that a product is unsuitable for one of its "ordinary" purposes when the market price drops.[114] And a seller has an incentive to claim that its performance has become "impractical" when a rise in input cost makes the contract's pricing term outmoded; the seller will argue that neither party "assumed" that a rise of the actual magnitude could occur.[115] A standard is efficient, therefore, only when the party on whom it confers discretion has the incentive to maximize joint returns in the course of maximizing its private gain. We think that few UCC standards could pass this test.[116] We know that few UCC standards have been shown to pass this

---

[113]An interesting summary of the effect in England of adopting an excuse standard is given by an English lawyer (Andrew Rogers, QC) in his introduction to a book on the excuse doctrine: "It is only in relatively recent times that the English courts relented in their demand that the strict words of the contract be adhered to, notwithstanding changes in circumstances. The resulting doctrine of frustration of purpose has not worked satisfactorily. The courts found it difficult to determine ... the limits for its application." "... frustration under the general law brings in its train automatic discharge. To avoid such results, the parties to the contract are required to draft their own particular code. That in turn means a substantial increase in transaction costs. Notwithstanding the costs involved, in an effort to meet the difficulty many contractors have undertaken fairly detailed contractual risk allocations." Force Majeure And Frustration of Contract v-vi (2d Ed. 1995, Ewan McKendrick, Editor).

[114]See, e.g., T.W. Oil, Inc. v. Consolidated Edison Co. of New York, 457 N.Y.S. 2d 458, 443 N.E. 2d 932 (1982).

[115]See, e.g., Aluminum Co. of America v. Essex Group, Inc., 499 F. Supp. 53 (W. D. Pa. 1980).

[116]See, e.g., Victor P. Goldberg, *Discretion in Long-Term Open Quantity Contracts: Reining in Good Faith*, 35 U.C. Davis L. Rev. 319 (2002) for a detailed analysis of the ways in which courts have used the good faith standard that regulates discretion in output and requirements contracts under §2-306(1) to undermine the controls on discretion that the parties had set for themselves.

test.[117] As evidence for this view, the customary product warranty for machines contracts out of the quality standard (§2-314); the cure standard (§2-508); and the revocation of acceptance standard (§2-608)[118]. The common force majeure clause contracts out of the excuse standard (§2-615); and many contracts contract out of the consequential damage standard (§2-715), the warranty damage standard (§2-714), and the standards governing insecurity and anticipatory repudiation (§§2-609; 2-610)[119]. In the absence of more rigorous evidence, it seems safe to say that contract law's standards seldom affect actual contracting behavior.[120]

In sum, the project of creating default rules founders on the costs of rule creation for heterogenous parties that function in complex commercial environments. These costs essentially preclude the creation of all but a few default rules.[121] The project of creating default standards for parties founders on the need of parties for specific guidance as to what they are supposed to do, and

---

[117]The test is not impossible to pass. For example, the standard merger agreement contains a term permitting the buyer to exit without penalty if a "material adverse change" occurs in the interim between signing the contract and concluding the merger. This term does not specify the events that could constitute a material adverse change because there can be many such events whose affects will vary with the ex post state of the world. A standard nevertheless is efficient here because the buyer will exit only if the seller's value has fallen materially. The threat of exit creates an incentive for the seller to take synergy increasing actions that lower the probability that its value to the buyer will fall materially. See Ronald Gilson and Alan Schwartz, *Understanding MACs*", Mimeo (2003). This example is offered to suggest that some statutory standards may be efficient, but the example also may support the inference that parties will write a standard in their contracts when a standard would be efficient.

[118]ALAN SCHWARTZ AND ROBERT E. SCOTT, COMMERCIAL LAW: PRINCIPLES AND POLICIES 204-224 (1991).

[119]Before 1978, parties routinely contracted out of §2-609 with ipso facto clauses, that stated precisely just when a party could cancel the agreement in consequence of its partner's financial difficulty. The current Bankruptcy Code outlaws these clauses on the (erroneous) ground that they permit solvent parties to deplete the bankrupt estate. For an analysis, see Yeon-Koo Che and Alan Schwartz, *Section 365, Mandatory Bankruptcy Rules and Inefficient Continuance*, 18 J. Law, Econ. & Org. 441 (1999). A concrete example of contracting out is in Northwest Lumber Sales v. Continental Forest Products, Inc., 261 Ore. 480, 492, 495 P.2d 744, 749 (1972).

[120]Ian Ayres has suggested that standards may be good defaults because precedent can crystalize around them, thereby providing parties with guidance. See Ian Ayres, *Making a Difference: The Contractual Contributions of Easterbrook and Fischel*, 59 U. Chi. L. Rev. 1391 (1992). This argument assumes that legal standards will become operative implied terms in parties' contracts, thereby giving courts the opportunity to apply them. The premise commonly fails, however, because parties contract out of the standards.

[121]A rare example of a successful default rule is the requirement that a breacher pay the other party the difference between the contract and market prices. The rule applies in only one state of the world, when there is breach; it is simple to apply because the court only compares the contract and market prices; and it is efficient for many parties because the rule protects the expectation interest. Other good default rules are hard to find.

on the property of standards to create moral hazard. These gloomy conclusions hold even when the information that a rule or standard would require is assumed to be known to the parties and accessible to the courts. For example, the parties and a court will know whether a fire destroyed the seller's factory in whole or in part. Information often is asymmetric, however. Parties may not know relevant things about each other and courts may not know relevant things about parties. The existence of asymmetric information exacerbates the moral hazard concern and thus makes the creation of efficient defaults even more difficult.

C. The Asymmetric Information Concern

The existence of asymmetric information truncates the set of contracts that parties can write. As a consequence, it also truncates the set of defaults that drafters can write. To see why, it is helpful to begin with an information taxonomy from the contract theory literature. A datum of information is "unobservable" if a party cannot observe it. Buyers ordinarily cannot observe a seller's production cost. A datum of information is "observable but not verifiable" if a party can observe it, but cannot verify the information's existence at an acceptable cost to a third party such as a court. For example, an employer usually can know which employees sometimes shirk, but it would be expensive relative to the gains to prove to a court that a particular employee shirked 20% of the time. A datum of information is "verifiable" if a party both can observe it and establish its existence to a third party. Information is asymmetric when it is either unobservable or unverifiable.

To understand the constraints on contracting behavior caused by asymmetric information we change the relation specific investment example of Part III(B) in two ways. First, we let the seller's costs be stochastic. In some states of the world, the seller's cost to produce the specialized product will be lower than the buyer's valuation, as in the original example; in other states of the world, the seller's costs will exceed the buyer's valuation, so that trade would be inefficient. Second, we let the buyer invest to increase the value of the seller's performance to it. These changes generate the

well known overinvestment problem.[122] The buyer's damages if the seller breaches will be the difference between its realized valuation and the price. The buyer thus will invest until the marginal cost of further investment will equal the marginal increase in value. This is too much because in those states of the world in which it is inefficient for the parties to trade, the buyer's investment has no social value; it serves only to increase the damages the seller would pay.

This problem would vanish if the seller could observe the buyer's production function -- the functional relation between the possible investment levels the buyer could choose and the values associated with those levels. The parties then could write a liquidated damage clause that would award the buyer the difference between the value the goods would have had were the buyer to invest optimally and the price. This value would be lower than the unconstrained value that would be produced if the buyer never took into account that in some states of the world its investment would generate no value at all (because the parties will not trade then). Such a contact seldom can be written, however, because sellers seldom could observe buyer production functions. As a consequence, the buyer would have an incentive to specify as liquidated damages the unconstrained value of the seller's performance to it, and then invest to realize that value. In addition, the court can observe only what the parties can observe. Thus, were there no liquidated damage clause, the court could not award the buyer its valuation had it invested optimally, in lieu of awarding the value the buyer actually lost, because the court will not know the value that optimal investment would have generated. The over-investment problem therefore exists because information is asymmetric.

To be sure, the modified investment example does not show that parties never could induce efficient investment when information is asymmetric. The contract theory literature has developed a number of contracts that, in theory anyway, sometimes would induce the parties to invest when they should.[123] These contracts, however, are "parameter specific"; that is, the prices the contracts

---

[122]See Steven Shavell, *Damage Measures for Breach of Contract*, 11 Bell J. Econ. 466 (1980); William P. Rogerson, *Efficient Reliance and Damage Measures for Breach of Contract*, 15 Rand J. Econ. 39 (1984).

[123]Concise but moderately technical discussions of these contracts are Patrick W. Schmitz, *The Hold-Up Problem and Incomplete Contracts: A Survey of Recent Topics in Contract Theory*, 53 Bulletin of Econ. Research 1 (2001) and Alan Schwartz, *Incomplete Contracts* in 2 The New Palgrave Dictionary of Economics and the Law 277-83 (1998). A nice though also technical example is Benjamin Hermalin and Michael Katz, *Judicial Modification of Contracts Between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach*, 9 J. L.,

set depend on the values, costs and probability distributions that particular parties face.

The lesson is that the state cannot help when asymmetric information prevents parties from writing the efficient contract. Parties thus would reject default standards that permitted the seller to obtain "a commercially reasonable price in proportion to its cost of production", or that permitted the buyer to recover "damages in proportion to commercially reasonable reliance investments made in good faith." Under the first proposed standard, the seller always would claim to have high costs; while under the latter proposed standard, the buyer always would claim that all its investments were commercially reasonable and were made in good faith. As another example, the UCC provides in §2-715 that a seller must either perform or pay the buyer consequential damages measured by the difference between the contract price and the value of the goods to the buyer.[124] This requirement is superficially efficient because it induces the seller to perform when performance would increase value and to breach otherwise. Valuations, however, often are difficult to verify. As a consequence, buyers have an incentive to overstate their valuations, thereby inducing sellers to perform even when the option of breaching and paying true damages would be less costly.[125] Commercial parties respond to this problem by routinely contracting out of §2-715. In place of the law, parties create complex repair and replacement provisions that strive for efficiency in other ways.[126] An appreciation of the information problems parties confront compels this conclusion: A good default does for parties what parties would have done for themselves had their contracting costs been lower. When asymmetric information prevents parties from writing certain types of contracts (even when their contracting costs are zero), a state-supplied default serves no purpose.

Econ. & Org. 98 (1993).

[124]Under §2-7152)(a), the buyer can recover (in addition to direct damages under §§2-712, 2-713 or 2-714), "any loss resulting from general or particular requirements and needs of which the seller at the time of contracting had reason to know...."

[125]The prospect of inefficient performances also would induce excessive precautions against breach.

[126]These clauses disclaim the seller's liability for consequential damages and restrict the buyer's right to reject, but obligate the seller to repair or replace defective tenders. See SCHWARTZ & SCOTT, supra note 118, at 204-09.

D. Summary

Much of what commonly is called contract law are the sets of default rules and, much more frequently, default standards contained in the Restatement and UCC Article 2. We argue here that the efforts of restatement writers and UCC drafters in creating this law have been largely wasted.[127] Default rules have proved too expensive to write for large heterogeneous economies and default standards founder over the parties need for ex ante guidance and the property of standards to create moral hazard. The moral hazard difficulty is exacerbated by the presence of asymmetric information. Defaults that condition on behavior that parties or courts cannot observe would be exploited for private ends. Thus, parties contract away from them. As a consequence, inefficient defaults only raise transaction costs unnecessarily.[128]

What then is the proper role of courts in resolving disputes over incomplete contracts? If a contract contains a gap, a court sometimes must decide whether to create a standard or to declare the contract incomplete and thus void for indefiniteness. It would be appropriate to create a standard if the standard could satisfy the criteria set out above. If such a standard does not exist but the contract has been partly performed, so that declaring it void would create hardship, courts

---

[127]Some Restatement sections embody the (few) simple, binary default rules that have emerged through the common law process. Parties have been relatively more receptive to these sections than to the numerous sections that reflect the efforts of Restatement drafters, following in the UCC tradition, to create additional standards to fill other perceived contracting gaps.

[128]This claim may appear vulnerable to the objection that inefficient defaults are "harmless bromides" because parties are already writing the contract; their marginal cost of avoiding a bad legal rule thus is slight. This objection is mistaken on two levels. First, the total social cost of avoiding an inefficient legal rule is not slight. The small cost of changing the rule for a particular contract must be summed over all of the transactions that the rule affects in the decades that the rule will exist. Second, it is actually costly to change a bad rule when writing a particular contract because courts tend to regard state-created defaults as presumptively fair or efficient. This institutional bias raises the cost of contracting out. See, e.g., Hayward v. Postma, 312 Mich. App. 720, 724, 188 N.W. 2d 31,33 (1971) (parties must use clear and unequivocal language to shift liability for risk of loss from seller to buyer); Caudle v. Sherrard Moters Co., 525 S.W. 2d 238, 240 (Tex. Civ. App. 1975) (same); Davis v. Small Business Inv. Co., 535 S.W. 2d 740, 744 (Tex. Civ. App. 1976) (contractual provision purporting to allocate to debtor the burden of "all" expenses incurred in preserving collateral not an "agreement otherwise" sufficient to opt out of UCC §9-207(2)(a)). Moreover, judicial interpreters may be reluctant to give the express language of the contract a meaning that conflicts with the relevant default. See Nanakuli Paving & Rock Co. v. Shell Oil Co., 664 F.2d 772 (9th Cir. 1981) (merger clause that excludes evidence of prior dealings does not bar evidence of usage of trade to alter the price term in the contract); Legnos v. United States, 535 F.2d 857, 858 (5th Cir. 1976) (despite express term "F.O.B. vessel" in contract, international context requires that "the intention of the contracting nations, rather than definitional niceties, must be given controlling weight).

understandably attempt to read the contract to reach a reasonable result. It does not follow from these understandable efforts that drafters should infer default standards from the courts' holdings. Rather, our argument holds that UCC or Restatement drafters commonly should do nothing, and that courts should be hospitable to attempts by later parties to alter or avoid earlier rules of the case.

We have argued in Part V that firms would reject the "default project" if they could because the project generates many inefficient terms that contracting parties must incur costs to avoid. There is, however, a welfare maximizing role for the state in creating certain default structures. While parties often will incur the costs of creating contracts, they almost never will find it cost justified to create a structure, such as a close corporation law or a bankruptcy act. Structures that create default procedures for parties to follow when conducting (or terminating) different forms of enterprise, and that specify the legal consequence of following these procedures, have created very large welfare gains. Substantive contract law, therefore, may simply function on the wrong level of generality: business parties commonly can create their own contracts, but the state best creates the broad structures within which to fit the parties' contracts. These structures have been largely unexamined from the perspective of contract theory.

VI. MANDATORY RULES

Mandatory contact law rules ban terms that parties choose; hence, these rules are inconsistent with the commitment to party sovereignty that we have defended. The rules nevertheless are justifiable on two grounds. The first is to prevent externalities, the classic example of which is price fixing.[129] The second ground is to ameliorate a market failure that disclosure cannot cure. As an illustration, consumers are thought to be poorly informed about the odds of product defects, but it is very difficult to communicate probability information in the format of a product label. The law's response is to ban disclaimers under the doctrine of strict liability in tort.

[129]The rules prohibiting fraud and duress also function to prevent a party from externalizing costs. Fraud and duress do not create social wealth but rather redistribute it (unfairly) between the parties. Were these practices permitted, the costs of enforcing the redistributions would be externalized to society. Thus, not only are fraud and duress grounds for declining to enforce a contract (because the resulting deal does not maximize joint welfare), but neither can two parties mutually agree to coerce or defraud each other.

Neither of these traditional grounds can justify the mandatory interpretation rules that we discussed in Part IV, however. Terms such as merger clauses and no oral modification provisions affect only the parties to the contract and there is no market failure to which the rules barring their enforcement respond.

This suggests that there is a third, largely unremarked, ground for mandatory contract law rules such as the rules that govern contract interpretation. To illustrate this claim, recall that today's interpretation rules are justifiable if the law's goal is to discern the parties' intentions regarding the contract's terms, but the rules are not justifiable if the law's goal is to follow the parties' preferences regarding judicial interpretive styles. The third ground for mandatory rules, then, is a form of paternalism whose logic, in the interpretation context, runs like this:

(A) There are good reasons for courts to enforce the parties' intentions.
(B) Parties sometimes use contract terms, such as merger clauses or no oral modification clauses, that restrict the ability of courts to find the parties' intentions.
(C) There is *no good reason* for the existence of such "intention blocking" terms.
(D) The terms thus are inconsistent with the parties' true or deep intention, which is to have the court enforce their actual deal; hence, courts should enforce the actual deal.

The interpretation rules are paternalistic because courts override the parties' expressed preferences out of a concern with the parties' welfare. Courts, that is, do not conceive themselves as imposing an agreement that parties would reject had they considered the matter under ideal deliberative conditions. Rather, courts believe themselves to be supplying the agreement that those conditions would have produced.[130]

The difficulty lies with the premise specified in step (C), that parties have no good reason to write contracts with the suspect terms. Our discussion in Part IV shows that rational, well informed

[130]Traditional accounts of paternalism require the paternalist – here a judge – to aim at the agent's welfare. For discussion, see Seana V. Shiffrin, *Paternalism, Unconscionability Doctrine and Accomodation*, 29 Philosophy & Public Affairs 205 (2000); Richard Arneson, *Mill v. Paternalism*, 90 Ethics 470 (1981).

and uncoerced parties have good reasons to use the intention blocking contract terms that the Code disfavors and that many courts do not enforce. If the premise in (C) is rejected, however, then the conclusion in (D), that courts are implementing the deep intentions of the parties, must fall as well. Paternalism with respect to the interpretation rules thus is misplaced because it frustrates rather than advances the parties' welfare: parties choosing under ideal conditions would want what the state prohibits. In this Part, we discuss three other contexts in which the rules are mandatory. In each case, the rules must be justified on the premise that parties have no good reason to use the terms that the legal rules prohibit, but in each case that premise fails. These rules too thus reflect a misplaced paternalism.

A. Parties Cannot Ban Modifications

Parties are formally free to require modifications to be written but they are not even formally free to prevent themselves from modifying their contract.[131] In the courts' view, the best inference from the existence of a modification is that the parties' original intention to prevent modifications has become outmoded. Expressed more fully, parties would not change their minds regarding the deal's substance without good reasons, so parties can have had no good reason to prevent themselves from changing their minds regarding the deal's substance. Contracting parties, however, do have good reasons to freeze the original deal. We next give an example to show that modification bans can enable parties to induce efficient relation- specific investment.[132] Since a major reason for enforcing contracts is to encourage relation-specific investment,[133] not to enforce a contractual modification ban is inconsistent with a basic justification for the existence of a contract

---

[131] Restatement (Second) of Contracts §311 comment. A (1979); Beatty v. Guggenheim Exploration Co., 122 N.E. 2d 378, 387 (N.Y. 1919) ("Those who make a contract can unmake it. The clause that forbids a change can be changed like any other ... Whenever two men contract, no limitation self-imposed can destroy their power to contract again." Cardozo, J); See also, Zumwinkel v. Legget, 345 S.W. 2d 89 (Mo. 1961); Christine Jolls, supra note 1. Civil law codes, on the other hand, discourage renegotiation; these codes thus are more consistent with the argument in text than the common law. See Eric Brousseau, *Did the Common Law Bias the Economics of Contract ... And May it Change?* In 6 LAW AND ECONOMICS IN CIVIL LAW COUNTRIES __ (B. Deffains & T. Kirat, eds. 2001).

[132] A more extensive formal discussion of the parties' preferences regarding modifications is in Schwartz & Watson, supra note 79.

[133] See Part III(A), supra.

law itself.

The example involves the production and sale of a product. The finished product can have two values for the buyer – high or low – depending on the ex post state of the world and how the product is made and delivered. The seller can make a relation specific-investment that would increase the likelihood that the product will have a high value. We assume that the parties can observe the value the product turns out to have, but that values are unverifiable. Also, the buyer cannot observe whether the seller made the investment or not. The parties could, however, write a contract that would induce the seller to invest efficiently even in this impoverished information environment. The contract has the seller make an up-front payment to the buyer that divides the expected surplus from the deal; the seller, however, later makes a take it or leave it offer to the buyer that requires the buyer to pay a price equal to the realized value of the product ex post.

To see how this contract would work, we assume that if the seller invests, at a cost of $c > 0$, the product will have a high value for the buyer with probability .8. If the seller does not invest, its cost is zero but the product will have a low value with certainty. After the seller invests (or not), it produces the product. We normalize the production cost to zero, so that it always is efficient for the parties to trade. In the example, a high value is $40 and a low value is $20. It would be efficient for the seller to invest if the expected gain less the cost would exceed the low value: i.e., if .8($40) + .2($20) - c $ $20, or if c is $16 or less. We let c = $12 so that investment would be efficient. Recalling that the buyer must pay $40 if the value is $40 and $20 if the value is $20, the seller expects to earn .8($40) + .2($20) - $12 = $24. Suppose for convenience that the parties share the surplus the contract creates. On these assumptions, the buyer will accept the contract if the seller makes an up-front payment of half the expected surplus, or $12, and the seller will accept the contract, and invest, because its expected gain also is $12.

The difficulty is that the buyer will reject the take it or leave it offer, and this will defeat the contractual solution. Just as in the example in Part III.(B), the seller's investment cost is sunk. Thus, after the take it or leave it offer is rejected, the parties will agree to renegotiate the price. If the realized value is $40, the buyer thus will offer $20, which the seller will take; and if the realized

value is $20, the product will change hands for a price of $10.[134] To see how renegotiation defeats the contract, realize that if the seller does not invest, the product will be worth $20 with certainty and the seller's payoff would be $10. Given renegotiation, the seller's expected gain if it invests is .8($20) + .2($10) - $12 = $6. Hence, the seller will not invest though investment would have been efficient. Perhaps a more vivid way to put this result is that when renegotiation is a possible outcome, our illustrative parties would do as well without a contract as with one. If there were no contract, the seller would decline to invest; instead, it would offer the finished low value product to the buyer for $10 in a spot transaction. But this is exactly the payoff a non-investing seller would expect to receive under the contract, given renegotiation, so it would be pointless for the parties to contract.[135]

The lesson is that parties sometimes cannot write contracts that will induce efficient relation-specific investment because renegotiation – a contract modification – cannot be banned. An enforceable ban on modifications would permit the seller to reinstate the contract's original price term. As applied to this case, if the buyer paid $20 a court could later order the buyer to pay another $20 because it can observe both the written contract and any monetary transfers the parties make. And if the buyer refused to pay anything, the seller would deliver nothing and sue to recover the $12 it paid up front. The buyer's payoff thus is positive if it complies with the contract and zero if it does not, so it can be expected to comply and the seller will invest. In short, parties can have good reasons for banning modifications.[136]

---

[134]Recall that the seller cannot sue for the price because values are unverifiable.

[135]The investment in this example is termed "cooperative" because the seller invests to increase the buyer's value. When renegotiation is permitted, cooperative investment can be impossible to motivate by contract. See Yeon-Koo Che and Donald Hausch, *Cooperative Investment and the Value of Contracting: Coase v. Williamson*, 89 Amer. Econ. Rev. 125 (1999). Irit Segal, *Complexity and Renegotiation: A Foundation for Incomplete Contracts*, 66 Rev. Econ. Stud. 57 (1999). Our earlier examples involved self investment (the seller invests to lower her costs; the buyer invests to increase his value). Efficient self investment is easier to motivate with appropriate contracts. Both kinds of investment are common.

[136]Bans on modification also can be efficient when one of the parties is risk averse. To see why, consider a possible contract between a risk neutral principal and a risk averse agent who is supposed to do a task. The principal, it is commonly assumed, cannot verify the agent's behavior to a court so the contract must motivate the agent. This will require the agent to bear risk; that is, her pay must be contingent on the outcome so she will try to produce a good result. The agent bears risk because the outcome is a function both of her efforts, which increase the likelihood of a good result, and chance. After the agent acts, but before the result is known, a pareto improving deal between the parties becomes possible. Since the agent has acted, she no longer needs to be motivated but she still bears risk

B. Parties Must Accept Substantial Performance

Courts will generally require parties to accept substantial rather than full performance unless, in the court's view, the deviation is material.[137] Parties sometimes try to opt out of this substantial performance default by making full performance an express condition of the promisee's duty to pay. Courts, however, frequently refuse to require exact compliance with the express condition on the ground that "the law abhors a forfeiture."[138] The basic premise is that the performing party would not agree to a contract that would penalize it severely for minor deviations. Since the paying party would know this, neither party would think that the contract required the equivalent of a forfeiture for a slight nonconformity, in spite of what the written words appear to say. The parties, courts believe, have no good reason to require forfeitures; and thus courts strictly construe express conditions that require full performance.

We offer an example, drawn from the famous case of Jacob & Youngs v. Kent[139], to show that parties have good reasons to construct deals that make forfeitures possible. Consider a contract that requires an owner to make progress payments to a builder as construction progresses. The last payment, which is sizable, is due when construction is completed. We assume that values are unverifiable. Thus, a court cannot observe the value of a building completed in accordance with the contract to the owner. On this assumption, there can be moral hazard on both sides. If the builder

---

because nature has yet to act. The deal will transfer risk from her to the risk neutral principal by paying the agent a fixed fee that will lie somewhere between the contract's good and bad state payoffs to her. If the parties anticipate this renegotiation, however, they will know that the agent's payoff will not be a function of the outcome; rather, it will be the fixed fee the agent expects to get in the renegotiation. Thus, the agent cannot be motivated to try hard because her payoff actually is noncontingent. When trying hard is efficient, permitting renegotiation thus is inefficient. The parties, ex ante, again have good reasons to prevent themselves from changing their minds.

[137]If the performance is substantial, the default rule is that the promisor must transfer the price less a sum that would compensate it for the deviation from full performance. Restatement (Second) of Contracts §§237 and 241. As we will see below, when the promisor's valuation is not verifiable the right to deduct damages is hollow.

[138]See Restatement (Second) of Contracts) §227, comment b: "The policy of the law requires that, within broad limits ..., the agreement of the parties should be honored even though forfeiture results. When, however, it is doubtful whether or not the agreement makes an event a condition of the obligor's duty, an interpretation is preferred that will reduce the risk of forfeiture." See also, SCOTT & KRAUS, CONTRACT LAW AND THEORY,supra note __, at 718-21.

[139]230 N.Y. 239, 129 N.E. 889 (1921).

has the burden of proof, the owner may claim that defects in the final version reduce her value substantially though they do not. The unverifiability of valuations would make it difficult for the builder to disprove this claim. Conversely, if the owner has the burden of proof, the builder may deliberately render a defective performance but claim that, in fact, it had substantially complied.

To understand the potential effect of such moral hazard, assume that the builder has the burden of proof and the owner will cheat by withholding the entire final payment if the building is less than perfect. Since perfection is difficult to achieve, the builder will expect not to receive the final payment. Therefore, it will not render the final performance. But then the owner will know that the contract's penultimate performance will be the final one, and it will cheat then. The builder, anticipating this, will not render the penultimate performance. In equilibrium, therefore, the builder will render no performance at all. The parties' contracting problem thus is to induce the builder to perform when the unverifiability of values makes strategic behavior likely.

The parties' solution follows from the contextual nature of verifiability. A datum of information may not be verifiable to a court because explaining matters to a generalist judge or a lay jury can be costly in relation to the gains. The same datum of information may be verifiable to an arbitrator, however. The arbitrator's expertise makes her cheaper to inform; she acts in an informal setting; and she has a reputational stake in not appearing to be biased in favor of builders or owners. To ensure that parties cooperate with the arbitrator, the parties will make her decision conclusive in the absence of fraud, bias or mistake.

In the actual case, the parties had adopted this common solution to the moral hazard problem, making an architect the arbitrator. The architect refused to certify that the builder had fully complied, though the defect appeared trivial. The seeming disjunct between the size of the withheld final payment and the nature of the noncompliance suggested possible fraud or mistake by the architect. The builder, however, did not attempt to impeach the architect's decision. Rather, the builder asked a court to hold that perfect compliance was not a condition to receiving the entire last payment, and the court agreed. It believed both that forfeiture of the entire last payment would have been unfair, and that the parties could not have intended this result. The issue the case posed,

however, was not whether the parties had or lacked good reasons to permit forfeitures. Instead, the issue was whether the parties had good reasons to make the architect's findings conclusive. Requiring a promisee to accept substantial performance of a construction contract when the parties had adopted a sophisticated governance scheme on the ground that the parties lacked good reasons to require a perfect tender thus is mistaken paternalism.

C. Parties Cannot Agree to Penalties

Contracting parties are permitted to specify the damages the breaching promisor must pay, provided that the specified damages represent a reasonable estimate of the promisee's lost expectation.[140] Parties, however, are not permitted to specify damages that exceed a reasonable estimate of the promisee's expectation. The logic here is the same as in the above examples. The parties made the contract in order to give the promisee a particular performance for a price. They know that breaches sometimes occur, so the parties also have good reasons to ensure that the promisee will receive a monetary substitute for performance in the event of breach. But contracting parties do not have good reasons to award the promisee much more than its lost expectation when the promisor fails to perform. The parties' deep contractual goal is advanced, therefore, by a mandatory rule declaring penalty clauses unenforceable.

This premise fails for two reasons. First, courts will sometimes implement the penalty rule inaccurately. Courts regulate liquidated damage clauses by comparing the difference between anticipated damages under the expectation damages default rule to the stipulated damages in the contract. Expectation damages, in turn, are based on the verifiable losses that the promisee anticipates from breach. Any liquidated damage clause that incorporates observable but nonverifiable values thus will be vulnerable to a penalty claim even when the clause accurately measures the promisee's lost expectation.[141] Moreover, courts sometimes find compensatory

---

[140]See UCC §2-718(1) ("Damages for breach by either party may be liquidated in the agreement but only at an amount which is reasonable in the light of anticipated or actual harm caused by the breach .... A term fixing unreasonably large liquidated damages is void as a penalty.")

[141]See Charles J. Goetz & Robert E. Scott, *Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on a Theory of Efficient Breach*, 77 Colum. L. Rev. 554 (1977).

liquidated damage clauses to be penalties in complex cases because the courts have failed to understand just how the clause protected the promisee's expectation.[142] These difficulties with current law permit promisors to invoke the penalty doctrine strategically. As a consequence, sophisticated parties are discouraged from using liquidated damage clauses when these clauses would otherwise be optimal.

The second reason why current law should change is directly relevant to the argument advanced above. The premise that parties have no good reason to contract for penalties is itself mistaken. Rather, penalties can permit parties to induce efficient relation-specific investments in certain asymmetric information environments. Commercial parties thus have good reasons to agree to penalties in these circumstances. Banning a liquidated damage clause because, and only because, it requires the breaching promisor to make a transfer that exceeds the promisee's expectation thus wrongly interferes with the parties' sovereignty.[143]

To summarize, contract law contains a number of mandatory rules that apply in the absence of an externality or a market failure. These rules override contractual terms that appear to be inconsistent with the intentions that rational, informed and uncoerced parties would have in the circumstances. It appears to courts and the drafters of the UCC that actual parties have no good reasons for choosing these terms. A defensible paternalism thus would not enforce them. In the cases we have analyzed, however, commercial parties turn out to have good reasons for the things they do. Put another way, the contract terms that courts and the UCC refuse to enforce actually advance the parties' welfare. A paternalist justification for contract law's mandatory rules therefore fails, a result that should not be surprising. In Part II, we showed that commercial parties pursue a

---

[142]See Alan Schwartz, *The Myth that Promisee's Prefer SupraCompensatory Remedies: An Analysis of Contracting for Damage Measures*, 100 Yale L. J. 376 (1990).

[143]See Aaron Edlin and Alan Schwartz, *Optimal Penalties in Contracts*, 78 Chicago-Kent L. Rev. 101 (2002). When the seller has market power, parties may use penalties to deter the entry of competitors into the seller's market. See id. Under current law, however, courts strike what they perceive to be penalty terms whether those terms were used to increase investment or to impede entry. We argue here that it is only a misplaced paternalism that inclines courts not to restrict themselves to the banning only of inefficient penalties. Another way to put this claim is that a party should always be free to argue that any term would create a negative externality or perpetuate a market failure. But it is a mistake to treat as a sufficient proxy for these inefficiencies a liquidated damage clause that would overcompensate the promisee in expectation.

goal -- joint welfare maximization – that the state supports, and generally can choose the means that best implement this goal. The rules regulating contracts between business firms thus should be mandatory only when the parties' contract creates an externality or is the product of market failure.

VII. CONCLUSION

This article does not ask the conventional normative question: What contract law should the state provide. Rather, it asks what contract law do business firms want the state to provide? A contract law for firms, we answer, would be smaller and more deferential to contracting parties than the contract law we now have. Of first order importance, firms want the state to enforce the contracts that they write, not the contracts that a decision maker with a concern for fairness would prefer them to have written. Enforcement of written agreements presupposes a theory of interpretation. The commitment to party sovereignty that we defend in this Article requires courts to delegate to parties both the choice of a contract's substantive terms *and* the choice of the interpretive theory that will be used to enforce those terms. Commercial parties, we show, commonly prefer adjudicators to be accurate on average in ascertaining the meaning of their agreements rather than accurate in every instance; therefore, these parties want courts to make interpretations on the smallest evidentiary bases that will support on average accurate interpretations. Courts that defer to party preferences regarding interpretation thus will use a textualist interpretive style, one that restricts the evidentiary base to not much more than the written agreement. Moreover, most firms prefer courts to interpret their contracts with the presumption that the contract is written in what we call majority talk, the language that firms and courts usually speak. For this reason, too, a textualist interpretive theory is the best default.

Much of today's contract law is in the form of default rules and standards. These defaults cause more harm than good. An efficient default rule – one firms will accept – is simple in form, conditions on few states of the world and maximizes joint gains in a wide variety of contexts. A default standard is efficient only when parties can live with vague definitions of their contracting obligations. Because standards confer considerable discretion on parties, a standard will be unsatisfactory if, as a result of that discretion, parties are likely to behave strategically under it. As

we show, parties are heterogeneous, drafting costs are finite even for public decisionmakers, rules must sometimes be complex, parties commonly will exploit standards to redistribute rather than to maximize joint surplus, and information often is asymmetric. All of these factors create obstacles to the creation of efficient default rules and standards that the state seldom has been able to overcome. Unsurprisingly, business parties often contract out of the current failed attempts. The effective domain of state supplied contract law thus is necessarily smaller than is widely believed.

The welfare maximization goal that we advance justifies courts in refusing enforcement to unconscionable contracts, contracts affected by fraud or duress and contracts that create externalities. This goal, however, cannot support many of the mandatory rules that today govern much contracting behavior between firms. Those rules bar enforcement to contract terms that efficiently cope with problems of hidden information and hidden action.

A normative theory of contract law that takes party sovereignty seriously shows that much of the expansion of contract law over the last fifty years has been ill advised. Contract law today is composed of a few default rules, many default standards and a number of mandatory rules. The mandatory rules should at most be defaults and most of the defaults should vanish from the law. Advocating freedom of contract for firms is uncontroversial. Taking freedom of contract seriously, however, would radically truncate current contract law. A law merchant appropriate to our time would be a merchants' law; and for merchants the less publically supplied law the better.

April 1, 2003