

ESSAY

THE ECONOMICS OF FORM AND SUBSTANCE IN CONTRACT INTERPRETATION

*Avery Wiener Katz**

For over a century, legal commentators have debated the relative merits of formal and substantive approaches to the interpretation of contracts; in recent years, the debate has increasingly been conducted in the language of the economic approach to contract law. While this new wave of scholarship has been relatively successful in relating the traditional debates over formalism to specific transactional and institutional problems such as imperfect information, it has been less productive in terms of generating useful legal or policy recommendations. This Essay proposes a different approach: one that focuses on private rather than public legal decisionmakers as its primary audience. In general, private lawmakers are better able to make practical use of the economic analysis of contracts, in part because the detailed information that is necessary to implement such analysis intelligently is likelier to be available at the individual level. Furthermore, there are many opportunities for contracting parties to choose between relatively formal and relatively substantive interpretive regimes. What is needed is a basic taxonomy of economic considerations that can serve as an organizing framework for parties choosing between form and substance when designing contracts. The later part of the Essay sets out such a framework.

INTRODUCTION

Under the modern American law of contracts, almost all applications of legal doctrine turn on questions of interpretation; and almost all questions of interpretation implicate the tension between form and substance. In one sense, this fact is neither remarkable nor distinctive. In order for any legal directive to have an effect on human behavior, it must be applied to particular cases—an agent seeking to enforce or comply with a given regulation must determine its content and then compare it to a specific factual context.¹ Accordingly, the materials admissible at the in-

* Professor of Law, Columbia University School of Law. E-mail: avkatz@law.columbia.edu. I am grateful to Mel Eisenberg, Allan Farnsworth, Victor Goldberg, Randy Picker, Bob Scott, and workshop participants at Stanford, Boalt, Michigan, Columbia, and the University of Chicago law schools for helpful comments, and to the Dean's Summer Research Fund at Columbia Law School for financial support.

1. For instance, a contract that provides that the seller must deliver six boxes of widgets by mid-July or face liability for the buyer's lost profits requires a performing or enforcing agent to determine, inter alia, what objects count as widgets and boxes, what acts count as delivery, which dates in July count as mid-July, what flows of costs and benefits count as profits, and, as a prerequisite to all of these, whether the contract ever attained the status of a legal obligation. Each of these determinations requires the agent to gather

terpretive stage, the manner in which interpretation is carried out, and the parties' expectations regarding the interpretive process will all significantly shape the ways in which the contract provides incentives and allocates risk.

In another sense, however, interpretation looms especially large in twenty-first-century American contract law, because under doctrinal provisions and practices as they have historically developed, the prescribed interpretive process is a relatively elaborate and intensive one. The set of materials considered relevant to interpretive inquiries is broad, and reasonably thorough attention to such materials is expected from those charged with applying the law in general or the language of individual agreements. As a result, the definitive resolution of interpretive questions requires a relatively larger degree of time and effort than would be the case under a system that put stricter limits on the materials to be considered or on the resources to be devoted to their consideration. Indeed, conventional scholarly wisdom holds that contractual disputes are more difficult and expensive to resolve in the United States today than in other common law countries such as England or Canada, or than in earlier historical periods such as the early twentieth century, in part because of the greater resources demanded at the interpretive stage.²

This question—how broad and thorough should the interpretive process be?—is commonly articulated in terms of the dichotomy of form versus substance. As such, it has long been a matter of professional and academic debate, and has been widely discussed in both case law and commentary. The question also underlies longstanding controversies in public law subjects such as administrative and constitutional law, as well as in the field of theoretical jurisprudence. Viewed from a pedagogical perspective, it surely presents one of the central conceptual themes of the first-year contracts class.

More specifically, many rules³ of contract law have the effect of privileging or emphasizing certain types of potentially relevant interpretive materials, and discounting or excluding others. Such rules are often termed “formal” or “formalistic” because they confine the interpreter’s

and consider evidence, and then to engage in an act of interpretation based on that evidence.

2. See generally P.S. Atiyah & Robert S. Summers, *Form and Substance in Anglo-American Law: A Comparative Study of Legal Reasoning, Legal Theory, and Legal Institutions* (1987) (contrasting the relatively formal nature of the English legal system with the relatively substantive nature of the American); Robert S. Summers, *Instrumentalism and American Legal Theory* 136–59 (1982) (discussing instrumentalist critiques of the late nineteenth-century formalist legal method). In this Essay, in order to maintain a more fluent and conversational style, I have generally not attempted to provide exhaustive citations for propositions that would be reasonably familiar to (or at least conventionally accepted by) specialists in the field.

3. For the sake of convenience and brevity, unless the context otherwise requires, in this Essay I will use the term “rules” to denote any doctrine or principle that has recognized legal status. “Rules” used in this way thus includes principles, standards, etc.

attention to a subset of materials that may or may not give rise to the same inferences as would the universe of materials as a whole. A more “substantive” approach to contract interpretation, in contrast, would attempt to come to a more all-things-considered understanding, based on all of the materials reasonably available.

For example, the Statute of Frauds requires that certain agreements be expressed in writing before they can be enforced. The Statute is subject to many well-known exceptions, but its general effect is to confer special status on the written document as a determinant of contractual liability. The parol evidence rule, which provides that a written document that integrates the parties’ agreement may not be contradicted or varied by evidence of prior or contemporaneous oral understandings, has a similar consequence. Similarly, the law confers special significance on certain symbols or gestures such as the seal (before it was abolished in most jurisdictions), or on commercial terms of art such as “f.o.b.” (abolished in the recently revised version of Article 2 of the U.C.C.).⁴ Even the classical doctrine of consideration has been in part defended in such terms.⁵

As is well known to both students and scholars of contract law, however, for the past one hundred years or so the historical trend across the board has been to water down such formal doctrines in favor of a more all-things-considered analysis of what the parties may have meant in the individual case.⁶ The relative balance of formal and substantive approaches to interpretation varies among jurisdictions and among subfields of contract law, of course, and between statutory and common law doctrines, with Article 2 of the U.C.C. representing the epitome of contemporary antiformalism.⁷ But one sees this trend played out in all corners of the law of contracts: in the decline of the classical doctrine of consideration and the associated rise in influence of the doctrine of

4. See U.C.C. §§ 2-319 to 2-324 (repealed 2003).

5. See generally Lon L. Fuller, *Consideration and Form*, 41 *Colum. L. Rev.* 799 (1941). Also note Holmes’s famous aphorism, in *Krell v. Codman*, 28 N.E. 578, 578 (Mass. 1891), that “consideration is as much a form as a seal.”

6. See generally Eric A. Posner, *The Decline of Formality in Contract Law*, in *The Fall and Rise of Freedom of Contract* 61 (F.H. Buckley ed., 1999) [hereinafter E. Posner, *The Decline of Formality*] (surveying and evaluating the descriptive theories that have been offered to explain this development).

7. See Richard Danzig, *A Comment on the Jurisprudence of the Uniform Commercial Code*, 27 *Stan. L. Rev.* 621, 621 (1975) (relating the antiformalist jurisprudential views of that Article’s chief drafter, Karl Llewellyn, to the interpretive philosophy embodied in its text); 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, 150–51 (Jody S. Kraus & Steven D. Walt eds., 2000) [hereinafter Scott, *The Uniformity Norm*] (criticizing Article 2’s “functionalist strategy” on the grounds that it has neither encouraged the production of standardized default rules for contracting nor enhanced the predictability of contractual interpretation); see also various provisions and official comments to Article 2, including U.C.C. section 1-201(3) on the definition of “agreement.”

promissory estoppel;⁸ in the movement from traditional notions of caveat emptor and the duty to read to the modern reasonableness-based approach to adhesion contracts;⁹ in the *Second Restatement's* de-emphasis of the distinction between unilateral and bilateral contracts;¹⁰ in the development of doctrinal categories such as requirement, output and options contracts that render enforceable arrangements that half a century ago would have fallen afoul of traditional doctrines of mutuality and indefiniteness;¹¹ in the decline of the perfect tender rule in sales law and the associated expansion of the doctrine of substantial performance;¹² in the decline of the mirror image rule as a device for resolving the battle of the forms in favor of U.C.C. section 2-207's test of material difference;¹³ and in the growth of the importance of the duty of good faith.¹⁴

While legal commentators have long debated the relative merits of these doctrinal developments, in recent years the debate has intensified; in the field of contracts, this new debate has increasingly been conducted in the language of the economic analysis of law. This flourishing of scholarship has followed in the wake of a wider school of thought that some have labeled the "new formalism."¹⁵ What is new about this new formal-

8. See generally Grant Gilmore, *The Death of Contract* (1974) (discussing this development and speculating that estoppel would eventually displace classical bargain theory as the primary basis for contractual obligation).

9. See generally Todd D. Rakoff, *Contracts of Adhesion: An Essay in Reconstruction*, 96 *Harv. L. Rev.* 1174 (1983) (contending that form terms contained in contracts of adhesion should be considered presumptively unenforceable).

10. *Restatement (Second) of Contracts* §§ 32, 62 (1981).

11. U.C.C. § 2-306 (2000).

12. See, e.g., *Kreyer v. Driscoll*, 159 N.W.2d 680, 683 (Wis. 1968) (contractor who left significant amount of work unfinished was not entitled to recover on the contract on grounds of substantial performance but instead could recover on a restitution theory); *Jacob & Youngs, Inc. v. Kent*, 129 N.E. 889, 890 (N.Y. 1921) (holding that builder's failure to install the promised brand of pipe would not entitle the homeowner to withhold payment and that homeowner's measure of damages would be measured by difference in value rather than cost of replacement, and observing that the result depends on "[c]onsiderations partly of justice and partly of presumable intention").

13. Compare *Poel v. Brunswick-Balke-Collender Co.*, 110 N.E. 619, 620 (N.Y. 1915) (holding that boilerplate term in plaintiff's standard form purchase order that requested acknowledgment of the order prevented the order from being a valid acceptance of defendant's offer to sell), with *Roto-Lith, Ltd. v. F.P. Bartlett & Co.*, 297 F.2d 497, 499 (1st Cir. 1962) (early section 2-207 case interpreting section 2-207(1)'s test of material difference to find that plaintiff's response rejected defendant's original offer, and holding defendant bound by contract on plaintiff's terms), and *Dorton v. Collins & Aikman Corp.*, 453 F.2d 1161, 1165-66 (6th Cir. 1972) (interpreting section 2-207(1) more liberally so that variance between forms would count as an acceptance on offeror's terms unless the offeree expressly conditioned its acceptance on offeror's assent to the terms in offeree's form).

14. See generally Robert S. Summers, *The General Duty of Good Faith—Its Recognition and Conceptualization*, 67 *Cornell L. Rev.* 810 (1982) (surveying development of good faith doctrine in relation to its incorporation in the *Second Restatement of Contracts*).

15. See, e.g., David Charny, *The New Formalism in Contract*, 66 *U. Chi. L. Rev.* 842, 842-86 (1999); Thomas C. Grey, *The New Formalism* 1-5 (Stanford Law School, Public

ism, both in contractual scholarship and elsewhere, is that it attempts explicitly to ground formalism in functional terms; it tries to show how formal methods of interpretation help to forward practical goals such as efficiency, procedural fairness, and public accountability.

While this recent wave of economically influenced scholarship has been relatively successful in relating the traditional debates over formalism to specific transactional and institutional problems such as imperfect information, it has been less productive in terms of generating useful legal or policy recommendations. In part this is because the lessons of this scholarship are very difficult to apply at the general level at which a court or legislature must operate, and in part this is because public lawmakers have their own self-interested reasons for preferring one interpretive approach over another.¹⁶ But the fact remains that this new scholarship has not yet influenced doctrinal developments and is not likely to do so in the future.¹⁷

This Essay proposes a different approach to the problem: one that focuses on private rather than public legal decisionmakers as its primary audience. In general, private lawmakers are likely to be in a better position to make practical use of the economic analysis of contracts, in part because the detailed information that is necessary to implement such analysis intelligently is much likelier to be available at the individual level. Furthermore, there are many opportunities for contracting parties to choose between relatively formal and relatively substantive interpretive regimes. What is needed, accordingly, is a basic taxonomy of economic considerations that can serve as an organizing framework for parties choosing between form and substance when designing contracts.

The succeeding sections of the Essay develop an account of how it is possible for private contracting parties to choose between form and substance *ex ante*, catalog the main considerations relevant to that choice, and then explore how the parties can use their local knowledge regarding such considerations to improve the efficiency of their agreements. The organization of the discussion is as follows: Part I outlines and critiques the recent literature on formalism in contractual interpretation, and elaborates on my argument that legal scholarship in the area should focus on addressing private transactional lawyers, not public decisionmakers such as courts and legislators. Part II sets out my basic analyti-

Law and Legal Series Working Paper No. 4, 1999), available at <http://ssrn.com/abstract=200732> (on file with the *Columbia Law Review*).

16. See generally Alan Schwartz & Robert E. Scott, *The Political Economy of Private Legislatures*, 143 U. Pa. L. Rev. 595 (1995) [hereinafter Schwartz & Scott, *Political Economy*] (ascribing the U.C.C.'s interpretive strategy, and in particular Article 2's preference for standards over rules, to the private interests of the various groups and individuals involved in its drafting).

17. Cf. Eric Posner, *Economic Analysis of Contract Law After Three Decades: Success or Failure?*, 112 Yale L.J. 829, 830 (2003) (arguing generally that the "economic approach does not explain the current system of contract law, nor does it provide a solid basis for criticizing and reforming contract law").

cal and normative framework; it defines more precisely what I mean by form and substance, and discusses and critiques a theoretical argument that has been influential in the traditional literature on contract interpretation: namely, the argument that formalist approaches to interpretation are not coherent because all interpretation presumes some common basis of contextual knowledge between speaker and audience, and thus requires attention to the relevant context. Part III then discusses a variety of familiar transactional problems such as costly information, risk allocation, rent-seeking, agency costs, and the protection of relational investments, and explains how these problems relate to the form/substance distinction. Part IV offers some general principles regarding how these problems might be addressed in contractual planning.

I. REFRAMING THE PROBLEM OF FORM VERSUS SUBSTANCE

The historical trend away from formal and toward substantive application of contract law has been alternately celebrated and criticized. Its defenders, such as Corbin and Traynor, have emphasized the mismatch between traditional formal categories and the complexity of commercial reality, and have argued that a more substantive approach is required to do justice to actual bargains and to protect commercial expectations.¹⁸ Its critics, such as Williston and Hand, have countered that contracting parties can adapt quite well to formal categories so long as the application of such categories remains clear and stable, and that substantive approaches, especially when applied by nonspecialist judges operating at a distance from the commercial setting and susceptible to influence by a host of popular and ideological considerations, tend to undermine the certainty of exchange and to defeat the parties' intentions.¹⁹

18. See, e.g., Arthur Corbin, *Corbin on Contracts* § 572B (2d ed. 1971 & Supp. 2003) ("No contract should ever be interpreted and enforced with a meaning that neither party gave it."); *Pac. Gas & Elec. Co. v. G.W. Thomas Drayage & Rigging Co.*, 442 P.2d 641, 644-45 (Cal. 1968) (Traynor, C.J.) (holding that implementation of contracting parties' intention requires "at least preliminary consideration of all credible evidence," and observing that refusal to consider such evidence "would either deny the relevance of the intention of the parties or presuppose a degree of verbal precision and stability our language has not attained").

19. See, e.g., *James Baird Co. v. Gimbel Bros., Inc.*, 64 F.2d 344 (2d. Cir. 1933) (denying promissory estoppel claim in construction bid setting, on grounds that "[t]he contractors had a ready escape from their difficulty by insisting upon a contract before they used the figures, and in commercial transactions it does not in the end promote justice to seek strained interpretations in aid of those who do not protect themselves"); 2 Samuel Williston & Richard A. Lord, *A Treatise on the Law of Contracts* § 6:58 (4th ed. 1991):

It is theoretically possible that a contract may be formed which is in accordance with the intention of neither party. If a written contract is entered into, the meaning and effect of the contract depends on the interpretation given the written language by the court. Traditionally, courts will give the language its natural and appropriate meaning, and if the words are unambiguous, will

A. *The Economic Commentators' Views on Form/Substance Questions*

The arguments of these two camps have framed both professional and academic discussion of contract law for over a century. Until recently, however, contracts scholars influenced by the economic approach to law have had relatively little to add to the form/substance debate. Instead, they have focused their attention on direct incentives for primary behavior—such as performance, breach, and reliance investment—and on doctrines and devices governing the allocation of risk, and have generally scanted interpretative problems.

The first attempts to bring economic analysis to bear on the question of contract interpretation came in the area of default rules—that is, rules that govern what courts should do when a contract is incomplete, silent, or ambiguous with regard to a particular term of the exchange.²⁰ The default-rule literature has flourished in recent years, to the point that the concept has become a standard of contracts scholarship and teaching.²¹ The creation of a default rule, however, still leaves parties and their agents with the problems of determining when it comes into play, how to tell whether the obligations prescribed by the rule have been satisfied, what the parties must do to overcome the presumption that the rule ap-

generally not admit evidence of what the parties may have thought the meaning to be.

20. For instance, if a contract for the sale of goods makes no mention of warranties, should the court interpret the contract as containing implied warranties of merchantability and fitness, or as providing for caveat emptor? Similarly, if a sales contract makes no mention of price, should a court fill the gap with a reasonable price measured at the time of delivery (the rule under U.C.C. section 2-305), a reasonable price measured as of the time of the making of the contract (the rule under Article 55 of the U.N. Convention on Contracts for the International Sale of Goods (CISG), reprinted in 19 I.L.M. 671, 684 (1980)), or decline to enforce the contract entirely?

21. This literature has become too vast to survey in a single footnote, but leading articles include Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 *Yale L.J.* 87 (1989); Ian Ayres & Robert Gertner, *Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules*, 101 *Yale L.J.* 729 (1992); David Charny, *Hypothetical Bargains: The Normative Structure of Contract Interpretation*, 89 *Mich. L. Rev.* 1815 (1991); Richard Craswell, *Contract Law, Default Rules, and the Philosophy of Promising*, 88 *Mich. L. Rev.* 489 (1989); Charles J. Goetz & Robert E. Scott, *The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms*, 73 *Cal. L. Rev.* 261 (1985); Robert E. Scott, *A Relational Theory of Default Rules for Commercial Contracts*, 19 *J. Legal Stud.* 597 (1990); Symposium, *Default Rules and Contractual Consent*, 3 *S. Cal. Interdisc. L.J.* 1 (1993). For a survey of this literature by one of its leading contributors, see Ian Ayres, *Default Rules for Incomplete Contracts*, in 1 *The New Palgrave Dictionary of Economics and the Law* (Peter Newman ed., 1998) [hereinafter *New Palgrave Dictionary*]. Contributors to this literature have argued that default rules should be designed to minimize the direct costs of writing contracts by choosing terms that most parties would want, to encourage the private development of contractual terms of art, to discourage opportunism and rent-seeking in drafting and performance, to encourage relatively informed parties to disclose their private information up front, and to minimize the costs of ex-post Coasian bargaining.

plies, and how to interpret their efforts when they try.²² All these determinations require interpretations; such interpretation could in principle be either formal or substantive. But on this last question, the default rule literature has had little to say.²³

In the last several years, however, a number of economically-influenced scholars have turned their attention to the issue of form versus substance, and in translating some of the traditional arguments over this issue into economic language, have helped to clarify some of the traditional commentators' concerns.²⁴ In the field of contracts in particular, this has resulted in somewhat of a renaissance of formalist arguments—or what one commentator has called “anti-antiformalism.”²⁵

The most prominent of the new wave of contractual formalists is perhaps Lisa Bernstein, who, in a series of articles detailing the practices of contracting parties in a variety of specialized markets (including the diamond, grain, and cotton trades), has argued that buyers and sellers who deal regularly in a given market prefer to have their disputes governed by the private rules and procedures supplied by their individual trade organizations. Her explanation for this preference is that those rules and procedures are more formalistic, and thus provide more certainty and protection at lower cost, than those that would be applied by generalist courts applying the U.C.C.²⁶ But other stalwarts of the economic approach to contract law, including Robert Scott and Alan Schwartz, have also weighed in on the side of interpretative formalism. Scott, in particu-

22. To illustrate, the implied warranty of merchantability provides a default rule regarding product quality in cases where the seller is a merchant; absent contrary agreement, the goods are supposed to be of a quality that would pass without objection in the trade, be fit for the ordinary purposes for which such goods are used, and the like. Under U.C.C. section 2-314, however, the definition of merchantability turns on the nature of the goods the parties understand themselves to be exchanging, and under U.C.C. section 2-316, the merchantability warranty can be disclaimed by a conspicuous writing mentioning the word “merchantability,” by the buyer’s inspection of the goods, by course of dealing, course of performance, or usage of trade, or by an expression such as “as is” that “in common understanding calls the buyer’s attention to the exclusion of warranties and makes plain that there is no implied warranty.”

23. See Eric A. Posner, *The Parol Evidence Rule, the Plain Meaning Rule, and the Principles of Contractual Interpretation*, 146 U. Pa. L. Rev. 533, 564 (1998) [hereinafter E. Posner, *The Parol Evidence Rule*] (critiquing the default-rule literature on this basis).

24. Louis Kaplow, *A Model of the Optimal Complexity of Legal Rules*, 11 J.L. Econ. & Org. 150 *passim* (1995) [hereinafter Kaplow, *Optimal Complexity*]; Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 Duke L.J. 557 (1992) [hereinafter Kaplow, *Rules*].

25. Charny, *supra* note 15, at 842–46.

26. See generally Lisa Bernstein, *Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions*, 99 Mich. L. Rev. 1724 (2001) [hereinafter Bernstein, *Cotton*] (explaining operation of trade rules and tribunals in detail); Lisa Bernstein, *The Questionable Empirical Basis of Article 2’s Incorporation Strategy: A Preliminary Study*, 66 U. Chi. L. Rev. 710 (1999) (suggesting that Article 2’s attempt to incorporate trade practice into its provisions is inconsistent with preferences of commercial professionals as evidenced by industry dispute resolution practices).

lar, has argued for some years that when government lawmakers attempt to develop complex substantive regulations or default rules, or when they look deeply into context when engaging in interpretive inquiries, they discourage private actors from developing their own arrangements for dealing with the underlying transactional problem. Because state lawmakers can only operate at a general level, while private solutions to transactional problems are likely to be better tailored to the needs of individual contracting parties, Scott concludes that clear and simple interpretive rules are best, even if on their face they appear to direct less-than-efficient outcomes.²⁷ More recently, both Scott and Schwartz, drawing on work in the economic theory of incomplete contracts, have argued that many common contractual devices are designed as responses to the fact that generalist courts cannot effectively (that is, at reasonable cost and with reasonable accuracy) determine the facts necessary to enforce the parties' substantive bargain as they ideally would wish it to be enforced in a world of full and free information. For courts to ignore these limitations and to try to enforce contracts as if they operated in a full-information world, they argue, disservices the parties' bargain and reduces the expected value of their exchange.²⁸

The arguments of these new economic formalists have not gone unchallenged. With regard to lawmaking in general, Louis Kaplow has shown, using a formal decision-theoretic model, that the optimal choice between rules and standards, and the optimal level of complexity of legal rules, depends upon empirical considerations such as the relative cost of ex ante and ex post decisionmaking, the costs of information acquisition, and the probability that a dispute will arise.²⁹ While limits on judicial competence do provide a reason to follow simple rules, in general one cannot conclude that rules dominate standards or that simplicity dominates complexity for all or even most purposes. In the field of contracts in particular, Eric Posner has defended a more balanced view of formal and substantive approaches to interpretation, suggesting that under some circumstances—especially those in which the contracting parties are boundedly rational, endowed with asymmetric information, or following a suboptimal convention—courts can improve social welfare by pursuing

27. See Goetz & Scott, *supra* note 21, at 261–64; Scott, *A Relational Theory of Default Rules for Commercial Contracts*, *supra* note 21, at 614.

28. See Alan Schwartz, *Incomplete Contracts*, in 2 *New Palgrave Dictionary*, *supra* note 21, at 277, 280; Alan Schwartz, *Relational Contracts in the Courts: An Analysis of Incomplete Agreements and Judicial Strategies*, 21 *J. Legal Stud.* 271, 316–18 (1992); Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 *Yale L.J.* 541 (2003) [hereinafter Schwartz & Scott, *Contract Theory*]; Robert E. Scott, *The Case for Formalism in Relational Contract*, 94 *Nw. U. L. Rev.* 847, 864–65 (2000); Robert E. Scott, *A Theory of Self-Enforcing Indefinite Agreements*, 103 *Colum. L. Rev.* 1641 (2003).

29. See Kaplow, *Optimal Complexity*, *supra* note 24, at 150.

a liberal interpretive approach.³⁰ Posner has also pointed out that even were we to make the extreme assumption that courts were completely unable to determine the contracting parties' intentions or the underlying facts of a contractual dispute, it would still not necessarily follow that courts should take a passive or literal approach to interpretation. Under some circumstances, he suggests, even incompetent courts could promote cooperation and deter opportunism by providing a means whereby an aggrieved party could credibly invoke a mutual penalty. Increasing the complexity of legal proceedings, on this view, serves to increase the size of this threat, even if it does little to improve the accuracy of the result in those cases that actually go to trial.³¹

This new economic analysis of formalism has been relatively successful in clarifying the traditional debates over formalism, and in relating them to specific transactional and institutional problems such as imperfect information, risk allocation, rent-seeking, and bounded rationality. Where the recent commentary has fallen short, however, is along the dimension of advancing toward practical responses to the form/substance dilemma.

This is so for two reasons. First, because the difference between the formalist and antiformalist positions is a matter of degree rather than kind, resolving their arguments comes down in practice to line-drawing. Even ardent neoformalists like Bernstein or Scott would agree that courts should depart from formalist methodology in certain circumstances—for instance, when there has been a credible allegation of fraud or error in transcription. (Just as courts following the relatively formalist *First Restatement* version of the parol evidence rule made exception for cases of fraud and mistake.³²) Conversely, even advocates of a more liberal interpretive approach would acknowledge that their position demands that courts or other law-applying actors possess at least minimal interpretive competence. But the proper compromise between form and substance, if it is to be based on utilitarian considerations, depends on an empirical judgment, made over the universe of potential cases, of how the relevant informational and transactional factors balance out. The very limitations of rationality and information that lead neoformalists to conclude that courts should not engage in substantive interpretation and that legislatures should not enact vague standards that require a substantive application also prevent us from drawing the proper limits between formal and substantive approaches with any confidence. To put it conversely, if state

30. See Karen Eggleston, Eric A. Posner & Richard Zeckhauser, *The Design and Interpretation of Contracts: Why Complexity Matters*, 95 *Nw. U. L. Rev.* 91, 92 (2000); E. Posner, *The Parol Evidence Rule*, *supra* note 23, at 534.

31. Eric A. Posner, *A Theory of Contract Law Under Conditions of Radical Judicial Error*, 94 *Nw. U. L. Rev.* 749, 767–69 (2000) [hereinafter E. Posner, *Radical Judicial Error*].

32. See cases surveyed in 2 E. Alan Farnsworth, *Farnsworth on Contracts* § 7.5 (2d ed. 1998).

actors know enough to set the appropriate boundaries on formalism, they are already significantly along the way to being able to do away with formalism entirely. Absent such knowledge, the setting of boundaries—like the application of substantive interpretation in any given case—is a matter of guesswork.

Second, as Schwartz and Scott have shown in their influential work on the political economy of the American Law Institute (and scholars in the field of positive political theory have shown generally), legislative bodies that are charged with the promulgation of generally applicable regulations and that are also institutionally responsible to a diverse set of interest groups will tend to favor standards over rules and vagueness over simplicity.³³ The positive imperatives of lawmaking thus lead naturally to interpretive conventions that disfavor formalist decisionmaking. This phenomenon may be somewhat less pronounced for common law courts, perhaps due to the influence of interjurisdictional competition and litigant initiative, but even so the process of common law development, with its continual generation of exceptions and counterprinciples, can erode the clarity and simplicity of legal doctrine. In a federalist legal system in which the choice between formal and substantive approaches can be made at a local level, furthermore, different jurisdictions may adopt different interpretive stances for reasons of their own.

B. *An Alternate Perspective: Private Ordering over Form and Substance*

Given that new economic analysis of formalism does not offer clear policy prescriptions for governmental reformers, or an operational program for implementing such prescriptions, what is the next move? In this Essay, I propose a different approach: one that focuses on private legal decisionmakers as the primary audience, rather than public ones. Note in this regard that virtually all of the above-mentioned commentators direct the bulk of their advice to governmental or quasi-governmental officials, even—indeed, especially—neoformalists like Schwartz and Scott. The advice may be that state actors should keep their hands off private contractual arrangements and restrict themselves to the relatively mechanical task of applying formal rules, but it is advice to state actors nonetheless. There are some exceptions to this blanket statement: Bernstein, for instance, in her articles on private commercial law regimes, focuses in her explicit discussions on a largely positive analysis, and is content to leave her normative critique of the U.C.C.'s interpretive approach as implicit. Posner also spends a significant amount of time in his articles discussing the likely responses of private actors to the various interpretive policies that courts and legislatures might adopt.³⁴ But this discussion—which he calls a positive analysis rather than a normative one—operates

33. Schwartz & Scott, *Political Economy*, supra note 16, at 597.

34. See generally Eggleston et al., supra note 30 (outlining the ways in which various types of parties might respond to various interpretive regimes).

in the overall context of his analysis merely as an instrument that he uses to develop guiding principles for government lawmakers. He does not consider, except incidentally, the possibility that his analysis could be useful to private actors.³⁵

As I have argued elsewhere, the almost exclusive focus in the mainstream law-and-economics literature on a hypothetical audience of public lawmakers constitutes a severe misallocation of intellectual resources.³⁶ Even if we thought the relevant officials were inclined to take our advice, and even if we thought they had sufficient ability and incentive to apply that advice fruitfully to actual policy and legal questions, we would still be ignoring the entire population of potential private lawmakers and neglecting the possibility that their efforts could also contribute to an increase in social welfare. Unless one thinks that private incentives for lawmaking are necessarily at odds with the public interest, or that private lawmakers' theoretical and practical knowledge already provides them with a fully adequate basis for enlightened lawmaking—or, more threateningly to our scholarly self-esteem, that private lawmakers would be even less inclined to pay attention to our writings than public ones—this failure to address their perspective does not make economic sense. It seems unlikely that an additional contracts article addressed to courts or legislators, on top of all such articles that have been written and published over the past decades, would have higher value added than an article or two focused on basic principles of transactional efficiency, and directed toward a hypothetical audience of private contract lawyers.³⁷ Moreover, the detailed information that is necessary to implement these principles intelligently is much likelier to be available at the individual level than at the level of the system as a whole. Private lawmakers may thus be in a significantly better position to make practical use of the economic analysis of contracts than public ones.

The value of refocusing attention on planning problems faced by private lawmakers, I think, is even greater with regard to basic problems of interpretation of the sort studied in the first-year contracts class. Most of the scholarly literature in the area is court- and case-centered, and thus tends to emphasize the aspects of interpretation that are central to the subjective experience of courts when deciding disputes. But as we know, many more contracts are written than are litigated, and the majority of our students that practice contract or commercial law will practice on the transactional side. Few of them will be judges or legislators, most of them

35. *Id.* at 131–32 (observing, as an aside, that heterogeneity among contracting parties implies that they should be permitted the freedom to choose *ex ante* between formal and substantive interpretive regimes of law).

36. Avery Katz, *Taking Private Ordering Seriously*, 144 *U. Pa. L. Rev.* 1745, 1748 (1996).

37. Cf. Ronald J. Gilson, *Value Creation by Business Lawyers: Legal Skills and Asset Pricing*, 94 *Yale L.J.* 239, 249–313 (1984) (analyzing and defending the transactional value of business lawyers and their activities).

will never be in a position to persuade a judge or legislature to change the law, but all of them might benefit from a clearer understanding of the practical consequences of formal and substantive interpretive strategies.

Taking a transactional approach to the problem of form and substance also helps emphasize the fact that, for all the discretion that courts and other arbiters may have with regard to interpretive questions, there are things that contracting parties can do *ex ante* to increase the chances that interpreting actors will follow the contractors' wishes *ex post*. For instance, if the parties want to limit courts' investigations into the history of their negotiations, they can and often do put a merger clause into their written contract, stating that the writing expresses their entire agreement and that all prior understandings or agreements have been merged into it. Such a clause will not serve as an absolute guarantee that the agreement will be enforced as written, but it does make a difference. While courts still retain the power to ignore merger clauses if they conclude that circumstances warrant, the use of the clause still tends as a practical matter to discourage courts from engaging in more free-form styles of interpretation, which is why contracting parties continue to use them even in jurisdictions that take a liberal approach to the admission of parol evidence.

There are, in fact, many opportunities for contracting parties to choose between relatively formal and relatively substantive interpretive regimes, and to have their choices matter; the merger clause is just one prominent example. No-oral-modification clauses provide another: Parties to sales transactions can provide, pursuant to U.C.C. section 2-209, that any modifications to their contract must be in writing. While such clauses do not prevent courts from using the equitable doctrines of waiver or estoppel to find that the contract has been varied, they still reduce the likelihood of such a finding, and it is possible to add additional clauses to the original writing that discourage the assertion of such claims. (Similarly, while the common law of contracts does not recognize no-oral-modification clauses as an official formal device, the presence of such a clause certainly raises the bar of persuasion for anyone who subsequently tries to claim that a contract has been so modified.)

Another common way for the parties to choose among interpretive regimes is by choosing which jurisdiction's laws govern the contract, since jurisdictions can vary considerably in their level of formalism. Virginia and Texas, for instance, continue to follow a traditionally strict version of the parol evidence rule, while California and New Jersey are widely known to take more liberal approaches.³⁸ New York and Connecticut continue to take different positions with regard to the formal effectiveness of an accord and satisfaction executed by check, even though the

38. For a survey of case law comparing common-law jurisdictions taking liberal and strict approaches to the parol evidence rule, see Scott, *The Uniformity Norm*, *supra* note 7, at 167-69.

question is ostensibly governed by a uniform statute.³⁹ In the area of international sales, the U.C.C., however antiformalist it may seem when compared to the traditional common law, is in many ways more formal than the alternative regime provided by the U.N. Convention on Contracts for the International Sale of Goods (CISG), which rejects both the Statute of Frauds and the parol evidence rule.⁴⁰ Choice-of-law clauses are common in commercial contracts, and while policy considerations such as consumer protection place some limits on their enforcement, at least in the commercial setting they are usually implemented as written.⁴¹ While there are many reasons for the parties to choose to be governed by a given legal regime, procedural simplicity and the ease of resolving disputes is a common motivation, and in that regard the level of formalism is an important factor.

Similarly, parties can often specify the forum in which contractual disputes will be heard, either by specifying a particular location where any litigation must be brought, or, as is increasingly common, by providing *ex ante* for private arbitration. Even if the substantive law to be applied to the contract is ostensibly the same, tribunals in different locations may be more or less inclined to delve into contextual matters, due to differences in jurisprudential approach, local legal culture, procedural and evidentiary rules, case loads, or other resource constraints. Private arbitrators are subject to similar variations, and face further incentives to formalize their interpretive practices in order to lower the cost of their proceedings, guard against suspicions of partiality, limit their exposure to judicial supervision, and attract future business.⁴²

39. Compare *Horn Waterproofing Corp. v. Bushwick Iron & Steel Co.*, 488 N.E.2d 56, 57–58 (N.Y. 1985) (holding that the common-law doctrine of accord and satisfaction has been superceded by provisions of U.C.C. section 1-207), with *County Fire Door Corp. v. C.F. Wooding Co.*, 520 A.2d 1028, 1031 (Conn. 1987) (holding that section 1-207 does not apply to accord and satisfaction). The subsequent revision of section 1-207 adopted the Connecticut approach, but New York, in contrast to most other states, has not adopted this revision.

40. On the other hand, in the area of offer and acceptance, the CISG's rules are more formalistic than the U.C.C.'s. Compare U.C.C. § 2-204, which indicates that a contract for sale of goods can be made in "any manner sufficient to show agreement," and § 2-206, where "an offer to make a contract shall be construed as inviting acceptance in any manner and by any medium reasonable in the circumstances," with Articles 14 through 24 of the CISG (creating elaborate framework of rules relating to offers, acceptances, counteroffers, rejections, and retractions).

41. See Erin Ann O'Hara, *Opting Out of Regulation: A Public Choice Analysis of Contractual Choice of Law*, 53 *Vand. L. Rev.* 1551, 1555–63 (2000) (discussing the use of choice-of-law clauses and observing that courts "routinely enforce them").

42. Compare Bernstein, *supra* note 26, at 1725 (discussing how low-cost private legal system is broadly regarded within the cotton industry as creating value), with Andrew T. Guzman, *Arbitrator Liability: Reconciling Arbitration and Mandatory Rules*, 49 *Duke L.J.* 1279, 1307 (2000) (emphasizing difficulty that courts face in ensuring that arbitrators apply substantive legal rules in the same way that courts would), and Stephen J. Ware, *Default Rules from Mandatory Rules: Privatizing Law Through Arbitration*, 83 *Minn. L. Rev.* 703, 719–23 (1999) (same).

Like choice-of-law clauses, forum-selection and arbitration clauses are not always enforced strictly according to their terms, and may be disregarded by courts willing to look beyond the face of the clause for interpretive evidence, or to override the clause in favor of some countervailing policy or principle.⁴³ But such clauses do receive some weight in practice; many courts enforce them presumptively, and there are self-interested reasons for even antiformalist courts to defer to them. Parties who favor a more formal interpretive approach, accordingly, have significant leeway to choose to have their disputes heard by tribunals who share their philosophy (as Bernstein's discussions of private trade tribunals suggest).

Contracting parties may also opt into a formalist interpretive regime by using a stereotypical legal device such as a negotiable instrument or letter of credit. Such commercial specialties are governed by distinct bodies of law, descending in part from the law merchant, that reflect a more formalist jurisprudential philosophy than does the common law of contracts generally. A holder in due course of a negotiable instrument, for instance, is entitled to enforce the instrument against its maker or endorser even if the defendant would have a good defense to liability on the underlying contractual obligation.⁴⁴ Similarly, the liability of the issuer of a letter of credit depends solely on whether the beneficiary presents documents that comply facially with the payment conditions provided in the letter. The issuer need not (and indeed, is not authorized to) inquire into the truth of any representations contained in the presenting documents and is entitled to demand strict compliance with all payment conditions—in marked contrast to the more liberal rule of substantial compliance that would be imposed under the ordinary law of contracts.⁴⁵ While commercial parties may have a variety of reasons for choosing to use one of these specialized devices, one important reason is the desire to contract into a more formalistic interpretive regime. A prominent illustration is provided by the rise in popularity of the standby letter of credit, which in economic terms is a close substitute for the common law suretyship or guaranty, but which in legal terms is governed by a

43. See, e.g., *Circuit City Stores, Inc. v. Adams*, 279 F.3d 889, 893 (9th Cir. 2002) (holding employee arbitration clause unconscionable); William J. Woodward, Jr., *Contractual Choice of Law: Legislative Choice in an Era of Party Autonomy*, 54 *SMU L. Rev.* 697, 780–83 (2001) (describing so-called “bomb shelter” legislation adopted by states seeking to invalidate choice-of-law clauses that purport to apply the controversial Uniform Computer Information and Transactions Act to transactions involving their residents).

44. U.C.C. § 3-305(b). The holder in due course's rights are subject to the so-called “real” defenses, which include infancy, duress, illegality, lack of capacity, discharge in bankruptcy, or essential fraud, but not subject to ordinary defenses such as mistake, misrepresentation, or failure of condition. *Id.* § 3-305 cmt. 2.

45. Compare U.C.C. § 5-108 (detailing terms of strict compliance to issuer's conditions), with Restatement (Second) of Contracts, *supra* note 10, § 241 cmt. a (explaining that the question of whether a party's failure to perform is material “turns on a standard of materiality that is necessarily imprecise and flexible”).

substantially more formalistic body of legal doctrine.⁴⁶ Experts in the field commonly assert that it is this very formalism that makes the standby credit an attractive commercial alternative.⁴⁷

Finally, contracting parties can often implement a more formal interpretive regime with regard to particular aspects of their agreement through the use of specific stipulations. The most familiar case of such a stipulation is the standard liquidated damage clause. By liquidating damages in their initial agreement, parties reduce the likelihood that a court will engage in a substantive inquiry into the actual state of damages *ex post*. A reduced likelihood is not a guarantee, of course, since courts will still supervise a liquidated damages clause to ensure that it does not work a penalty, and some courts remain resistant to the use of liquidated damages in cases where damages are amenable to *ex post* calculation. But by adopting such a clause, the parties do buy themselves somewhat greater formality, and in practice perhaps a presumption of enforceability.⁴⁸ Similar stipulations regarding other terms of the agreement have an analogous effect.⁴⁹

Because there are so many ways for contracting parties to influence the interpretive regime under which their agreements will be enforced, the existing literature's emphasis on advising public lawmakers whether to restrict or liberalize their interpretive approach is to a significant extent beside the point. The interpretive regime should be understood as a sort of default rule, which parties can opt out of with careful planning. Different parties, depending on their circumstances, will prefer different

46. See Peter A. Alces, *An Essay on Independence, Interdependence, and the Suretyship Principle*, 1993 U. Ill. L. Rev. 447, 465–76 (discussing and comparing defenses under Article 5 and the common law). For specific examples, consider U.C.C. § 5-102(a)(7) (defining good faith as mere “honesty in fact,” in contrast to the more equitable standard of “reasonable commercial standards of fair dealing” found in § 2-103(1)(b)), § 5-108(a) (substituting a strict compliance principle for doctrines of waiver and estoppel that would otherwise apply under § 1-103), § 5-111(a) (eliminating consequential damages and, more strikingly, the duty to mitigate damages), and § 5-111(e) (providing that courts must award attorneys’ fees to prevailing party in any dispute that arises over issuer’s duty to pay).

47. See Alces, *supra* note 46, at 450–52; Henry Harfield, *Guaranties, Standby Letters of Credit, and Ugly Ducklings*, 26 UCC L.J. 195, 196 n.2 (1994); Rufus James Trimble, *The Law Merchant and the Letter of Credit*, 61 Harv. L. Rev. 981, 1006 (1948).

48. For a list of suggestions to drafters of liquidated damage clauses seeking to maximize the chances of enforcement, see 3 Farnsworth, *supra* note 32, § 12.18a (recommending, *inter alia*, avoiding the word “penalty,” allowing damages to vary with the amount of delay or quantity of defective goods, reciting the types of loss that are intended to be compensated, and arranging for deposits or bonuses as an alternative enforcement device).

49. See, for example, U.C.C. § 1-102(3), which provides that although duties of good faith, reasonableness, diligence, and care cannot be disclaimed entirely, the parties can by stipulation determine the standards by which those duties are to be measured, so long as “such standards are not manifestly unreasonable.” The upshot is that courts retain the power to supervise the parties’ stipulations, but parties can, through careful planning, make it less likely that their choices will be second-guessed in practice.

tradeoffs between form and substance, and helping the parties to choose the correct balance in this regard is one of the main tasks the transactional lawyer faces. Judicial formalism may wax and wane, but this planning problem will remain important for lawyers and for their clients.

What is needed, accordingly, and what the economic analysis of contracts can provide, is a basic taxonomy of substantive considerations that can serve as an organizing framework for parties choosing between form and substance when designing contracts. A good commercial lawyer needs to understand the functional underpinnings of the transaction in order to help plan it—and in commercial settings, these underpinnings are economic. I am not claiming that actual transactional attorneys do not already take such considerations into account—of course they do; a working familiarity with such factors is one of their main stocks in trade. But organizing such insights into a more systematic conceptual framework helps us to integrate and synthesize disparate bodies of practical knowledge relating to various commercial and legal fields: negotiable instruments, letters of credit, choice of law, sales, and so on. Such a synthesis enables insights from one field to be translated and analogized for the purposes of critiquing and improving transactional planning in others. Additionally, it serves an important pedagogical function in the training of law students, because young lawyers beginning legal practice will be able to assimilate conventional wisdom more quickly and effectively if they are first equipped with its implicit theoretical underpinnings.

In the succeeding sections of this Essay, therefore, I turn to the question of when and why contracting parties should choose formal methods of interpretation over substantive ones, or vice versa. My analysis thus will implicate questions such as whether the parties should write a merger clause into their agreement or whether the parties should opt into Virginia law, rather than questions such as whether a court should admit a given item of parol evidence, or change its doctrines so as more closely to resemble Virginia law. Of course, if we are able to develop a framework for answering the former set of questions, that will likely help courts to answer the latter set of questions as well, and possibly to ask those questions differently. Instead of asking what substantive terms the parties intended to have in their agreement, courts might begin to ask what interpretative method the parties wished to have.

II. A MODEL OF THE CHOICE BETWEEN FORM AND SUBSTANCE

A. *Normative Considerations*

In this section, I set out my operational definitions of form and substance in the interpretive context, as well as the normative goals that my framework is designed to pursue. The latter question is more quickly addressed. This Essay is intended to follow in the tradition of functional-

ist accounts of formalism such as the one put forward by Lon Fuller.⁵⁰ But because I am focusing on those functional considerations that are most relevant to decisionmaking by contracting parties at the planning stage, my analysis is limited almost exclusively to issues of economic efficiency. Specifically, I concentrate on the question of what administrative arrangements will maximize the total expected value of the underlying exchange, with adjustments for risk aversion but not for distribution or procedural fairness (except insofar as the parties are willing to sacrifice exchange value in order to promote such other goals).

I defend this focus on the pragmatic grounds of brevity and specialization of scholarly effort, which are standard in the economic literature on contracts. The normative appeal of the efficiency criterion has been thoroughly discussed by other scholars (indeed, there has been a recent resurgence of debate over the criterion) and I have nothing to add to this discussion at present. Justifying the efficiency criterion as a matter of fundamental principle is beyond the scope of this Essay, and the usual admonitions will apply.⁵¹ Second, this Essay attempts in particular to develop general principles that can be used to further the aims of private contracting parties. Such parties, especially those operating in the commercial context, generally engage in exchange for instrumental purposes that typically include the goal of material profit. Any analysis that did not give a central place to maximizing contractual value would not address these needs. Third, as long as the transaction in question is an arms-length one, the parties have the option not to enter into it, and they are informed of the relevant business risks and legal consequences, there are no clear distributional consequences flowing from any change in legal rules.⁵² As a general matter, the surplus from exchange tends to be divided among contracting parties in proportion to their relative eagerness to enter into the bargain. Any efficiency gains or losses resulting from a change in regime, accordingly, will tend to be shared.⁵³

50. See Fuller, *supra* note 5.

51. For general discussion of the efficiency criterion, see Jules L. Coleman, *Markets, Morals and the Law* 95–112 (1988); Richard A. Posner, *The Economics of Justice* 48–115 (1981); see also Symposium on Efficiency as a Legal Concern, 8 *Hofstra L. Rev.* 485–770 (1980). More specifically, Louis Kaplow and Steven Shavell argue that even if distributional equity is an important social objective, it is more effectively promoted by using direct public instruments such as tax and transfer payments, rather than by use of the rules of private law. Louis Kaplow & Steven Shavell, *Why the Legal System Is Less Efficient than the Income Tax in Redistributing Income*, 23 *J. Legal Stud.* 667, 669 (1994).

52. One exception to this general statement is the doctrine of silence as acceptance. See Avery Katz, *Transaction Costs and the Legal Mechanics of Contract Formation: When Should Silence in the Face of an Offer be Construed as Acceptance?*, 9 *J.L. Econ. & Org.* 77 (1993) (construing silence as acceptance raises the cost of declining an exchange, shifting value from the offeree to the offeror).

53. See Richard Craswell, *Passing on the Costs of Legal Rules: Efficiency and Distribution in Buyer-Seller Relationships*, 43 *Stan. L. Rev.* 361, 365–66 (1991) (explaining why, in general, any legal regime change will result in sharing of benefits between consumers and producers).

In the bulk of this Essay, furthermore, I also treat the interests of the contracting parties as paramount. This approach is equivalent to assuming that there are no important third-party effects attaching to the principal parties' decisions. If there are such third-party effects, then we can stipulate that from an efficiency standpoint courts and other public officials ought to watch out for them, and ought to refuse to give effect to any contractual provisions—including those that deal with the form/substance issue—that impose negative externalities. Similarly, a complete analysis of any policy issue in contract or commercial law would obviously have to include consideration of potential market and contractual failures (such as monopoly power, bounded rationality, and imperfect information) that would justify overriding the parties' contractual freedom, but in this Essay I will put aside such problems, for reasons generally analogous to those I have given above for ignoring issues of distribution.⁵⁴

It is worth making one last prefatory remark regarding the relevance to my analysis of the liberal norm of personal autonomy. Some contracts scholars, including Randy Barnett, have argued in favor of formalist modes of interpretation on grounds of autonomy, reasoning that clear and predictable rules help to facilitate the free exercise of individual will and operate as a safeguard against state agents illegitimately infringing on individual choice.⁵⁵ Other autonomy theorists, such as Charles Fried, have instead claimed that deference to parties' freely exercised choices may sometimes require courts to pay closer substantive attention to what choices the parties actually intended to exercise.⁵⁶ I take no position on this controversy, and indeed have little to say about autonomy. It does seem to me, however, that a principled liberal should be in favor of allowing people entering into contracts to choose between formal and substantive modes of contractual interpretation, based on what seem to them to be good and sufficient reasons, unless there is some reason such as force or fraud that justifies overriding the parties' will. In this regard, my goals here are consistent with those of a liberal or libertarian, in that I focus on clarifying the considerations that would be relevant to such a choice in the individual case. Parties are more likely to have their contractual decisions respected when they have engaged in well-considered

54. For an excellent (and perhaps definitive) general discussion of such problems, see generally Michael J. Trebilcock, *The Limits of Freedom of Contract* (1993).

55. Randy E. Barnett, *A Consent Theory of Contract*, 86 *Colum. L. Rev.* 269, 291–319 (1986); Randy E. Barnett, *The Sound of Silence: Default Rules and Contractual Consent*, 78 *Va. L. Rev.* 821, 857–60 (1992); cf. Geoffrey Brennan & James M. Buchanan, *The Reason of Rules* 19–23 (1985) (grounding contractarian philosophical perspective in premise that individuals are “ultimate sources of value and of valuation”).

56. See Charles Fried, *Contract as Promise* 57–73, 85–91 (1981) (pointing to interpretative moments where courts must presume intent of parties and perform a gap-filling role).

deliberations, and so to the extent that my analysis fosters better private decisionmaking, it also forwards liberal values.⁵⁷

B. *A Descriptive Model*

With these normative preliminaries out of the way, we are now in a position to turn to the main analysis. There are numerous accounts of the distinction between form and substance in the scholarly literature.⁵⁸ One sees the dichotomy expressed in terms of rules versus standards, rules versus discretion, textual versus contextual modes of interpretation, static versus dynamic interpretation, simplicity versus complexity, determinacy versus flexibility, objective versus subjective standards, and so on. Each of these opposed pairs highlights different functional aspects of the formalism problem, but what they have in common is that the first member of each opposed pair connotes an interpretive approach that focuses on a more limited set of authoritative or evidentiary materials, and the second member connotes an approach that embraces or allows for the consideration of a more expansive set of materials. A rule-based theory of interpretation, for instance, directs the interpreter to limit his or her attention to the specific considerations set out by the lawmaker at the time that the rule was promulgated, while a standard-based theory also allows the interpreter to consider factors that may not become apparent until the moment that law is applied to facts.⁵⁹ Similarly, an objective standard of interpretation directs the interpreter to limit attention to factors that would be accessible to all individuals who can be categorized as being in the relevant agent's position (with the category being defined widely or narrowly depending on the prescriptions of the standard), while a subjective interpretive standard directs the interpreter additionally to consider factors that might be accessible only to the individual parties to the contract.⁶⁰

57. Compare in this regard the normative tradition of corrective justice, which in some of its versions might be thought to suggest that moral or legal wrongs need to be substantively righted regardless of the wishes and intentions of the victim, and that this obligation is inalienable. While my efficiency-oriented approach is compatible and even consonant with the autonomy norm, it conflicts with the corrective justice approach to this extent. See generally Louis Kaplow & Steven Shavell, *Fairness Versus Welfare* (2002) (contrasting their welfare-based approach to normative theories that evaluate actions or policies as right or wrong without regard to the consequentialist effect on individuals).

58. See, e.g., William N. Eskridge, Jr., *Dynamic Statutory Interpretation* 109, 174–204 (1994); Frederick Schauer, *Playing by the Rules* 207–32 (1991); Kaplow, *Rules*, *supra* note 24, at 618–20; Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 *Harv. L. Rev.* 1685, 1687–1701 (1976); Carol M. Rose, *Crystals and Mud in Property Law*, 40 *Stan. L. Rev.* 577 (1988); Antonin Scalia, *The Rule of Law as a Law of Rules*, 56 *U. Chi. L. Rev.* 1175, 1177 (1989); Grey, *The New Formalism*, *supra* note 15, at 19–21.

59. See, e.g., Kaplow, *Rules*, *supra* note 24.

60. See, e.g., *Restatement (Second) of Contracts*, *supra* note 10, §§ 20, 201 (contrasting objective and subjective interpretive standards, as well as standards emphasizing the alternative interpretive positions of speaker and listener).

Following this general distinction, accordingly, in this Essay I will model the concept of formality as a function of the set of materials that an interpreter considers in arriving at an interpretation. Formalism entails restriction to a smaller set of decisional materials (for example, the presence or absence of a wax seal, as it relates to the enforceability of a written promise); while substantive interpretation permits and sometimes directs attention to a larger set of decisional materials (for example, the underlying facts of a business relationship, as they relate to the presence or absence of contractual consideration). I am hoping that this account of formalism will seem both intuitively appealing and familiar; it resembles and draws on, for instance, the concept of exclusionary reasons put forward by jurisprudential writers such as Raz and Schauer.⁶¹ In order to highlight the connection of my approach with economic analysis in general and decision theory in particular, however, I denote the set of permissible materials associated with a given interpretive regime as the regime's *information set*, and model the parties' choice between formal and substantive regimes as an *ex ante* choice of the information set *I*.

Note that this definition of formalism can itself fairly be called formalistic, since it suppresses other factors that some people might consider relevant to an account of the distinction between form and substance. The definition can be interpreted quite generally, however, to include most of the issues discussed under the rubric of formalism by traditional scholars of jurisprudence. For instance, according to this definition, both traditional legal positivism, which distinguishes between moral and legal considerations and claims that only the latter provide an appropriate basis for legal decisionmaking, and Ronald Dworkin's contemporary theory of judging, which distinguishes between considerations of policy and of principle and claims that only the latter provide an appropriate basis for judicial decisionmaking, are formalist theories.⁶²

On such a definition, it is not strictly possible to rank all interpretive regimes in order of their formality, since the information sets associated with two regimes may overlap. For instance, U.C.C. section 2-202 directs courts interpreting an integrated contract to consider trade usage, course of dealing, and course of performance in interpreting the meaning of the contract, but not to consider evidence of prior agreements or contemporaneous oral agreements to the extent that they are inconsistent with the written contract terms.⁶³ The standard for determining inconsistency, however, is not prescribed, and many courts have applied it liberally.⁶⁴

61. See Joseph Raz, *The Authority of Law: Essays on Law and Morality* 19–33 (1979); Schauer, *supra* note 58, at 88–93.

62. On positivism and its relation to Dworkin's theory of judging, see generally Jules Coleman, *Negative and Positive Positivism*, 11 *J. Legal Stud.* 139 (1982).

63. U.C.C. § 2-202 (2000).

64. This liberal interpretation reflects the spirit of official comment 3 of section 2-202, which appears to suggest a presumption of consistency: "If the additional terms are such that, if agreed upon, they would *certainly* have been included in the document in the view

This approach is in contrast with the traditional common law, which took a stricter stand on the admission of parol evidence and did not explicitly confer official status on course of performance. In this regard, the U.C.C. is less formal. But section 2-209(2) also provides that if the parties to a signed contract adopt a no-oral-modification clause, attempted oral modifications will be ineffective.⁶⁵ This device was not recognized at common law, and in this regard the U.C.C. is more formal. Thus, if we are being absolutely precise, only if the information set associated with a given regime is entirely contained within the information set of another regime (i.e., the first information set is a proper subset of the second) can we say that the first regime is strictly more formal than the second. But speaking more casually, it will be useful to call a regime relatively formal if its information set is relatively more restricted than another's, or if its information set contains relatively little that is not contained within the other information set, and omits a significant amount of material that is contained within the other information set.

While the set-theoretic definition I have given may appear to suggest a bright-line distinction between formal and substantive modes of interpretation, it can also be understood in probabilistic terms. Some regimes—indeed, probably most—may admit certain types of material into their permissible information sets, but only some of the time, or only for limited purposes, or with less weight, or only if the material is weighty enough to overcome a presumption against admissibility. Accordingly, a regime that allows the consideration of more interpretive material more of the time or with greater probability is more formalistic, other things being equal, than a regime that uses such material less of the time or with lower probability. Similarly, a regime that establishes a hierarchy of influence and that treats certain types of material as more weighty or more privileged than others is more formalistic than one that accords all types of material equal consideration. To illustrate, under the *Second Restatement's* approach to the parol evidence rule, the court may consider parol evidence for the purpose of deciding whether the written contract is an integrated one or not. If the court decides on the basis of the evidence that the writing is an integration, then the parol evidence is not supposed to be used to interpret the writing further, and must be withheld from the trier of fact.⁶⁶ Similarly, under U.C.C. sections 1-205 and 2-208, factfinders are directed to interpret trade usage, course of dealing, course of performance, and express contractual terms as consistent with one another if they can reasonably bear such a reading. If they cannot, express terms are to take precedence over the other categories of material,

of the court, then evidence of their alleged making must be kept from the trier of fact.” U.C.C. § 2-202 cmt. 3 (emphasis added).

65. Under section 2-209(4), however, attempts at modification may nevertheless operate as a waiver under certain circumstances.

66. Restatement (Second) of Contracts, supra note 10, § 209.

course of performance is to take precedence over trade usage and course of dealing, and course of dealing is to take precedence over trade usage.

It is important to note that the information set associated with a given interpretive regime is not the same thing as the information set that is actually used by any particular interpretive agent within that regime when making an interpretive determination—and similarly, that this latter information set I_a (call it the *agent-specific information set* for individual agent a) may vary among agents within the same regime. In this case, the effective formality of a regime will depend on the distribution of information sets across all agents within it, and as such must be understood in statistical terms.

For example, a given regime might permit individual judges to make use of their experience in previous cases when making an interpretive decision, or might allow courts to take judicial notice of particular facts. Under that regime, judges with different backgrounds or levels of experience would have different agent-specific information sets available when making their decisions. A regime that prohibited judges from considering this sort of background information would make their individual agent-specific information sets more similar. If we compare the information available to judges to the information available to contracting parties, however, a ban on judicial notice could make the *expected information set*, as averaged over the set of all judges, either more or less similar to the expected information set, as averaged over the set of all contracting parties. The direction of the outcome would then depend on how much the experience of judges overlapped with the experience of contracting parties.

In sum, the effective degree of formalism achieved by the regime should be understood as a function of its agent-specific information sets, each of which are themselves functions of the regime's general information set, as chosen by the parties or by default rules of contract interpretation.

Note that the expected outcome depends not only on the distribution of agent types, but also on the specific way in which individual agents combine their own agent-specific set I_a with the information specified in the contractual set I . For instance, we could imagine that some agents might simply combine all of their agent-specific information with the contractual information set I , using both types of information with equal weight; alternatively, some agents might base their interpretations only on information that appears both within their personal information set as well as in the contractually-specified set. Similarly, one could imagine that agents also vary in the ultimate decision rule they apply after engaging in an interpretation (so that, for instance, some agents could pursue what they saw as the parties' intentions, others could pursue what they saw as economic efficiency, and still others what they saw as distributional justice). The above specification, accordingly, is general enough to encompass various types of interpretive approaches and agency problems.

This discussion implies, therefore, that the parties' optimal choice of the contractual information set I will thus depend on how agents make use of that information, and thus on particular features of the legal regime and on the identities of the interpretive agents.⁶⁷

The foregoing remarks also imply that within this theoretical framework, contracting parties can opt into a more formal interpretive regime in two ways: first, by placing limits on the overall information set permitted by the regime (for example, by excluding parol evidence or evidence of oral modifications), and second, by limiting the set of eligible interpretive agents (for example, with a choice-of-forum or arbitration clause). They could also combine these methods, by specifying different information sets for different types of interpreters—for example, one interpretive regime could be used when the contract is interpreted by the parties themselves, another regime when interpreted by fellow traders, yet another when interpreted by arbitrators, another when interpreted by courts. Thus it would be possible, if the law allowed and the parties so wished, to use a relatively standard-based approach when applying nonlegal sanctions among the parties and their trading community, and a relatively formalistic approach when litigating in the public courts.⁶⁸

C. *The Contextualist Argument*

This way of framing the problem helps to clarify and rebut a common argument against using formal methods of interpretation, to the effect that all interpretation depends upon a common basis of contextual knowledge between speaker and audience and that formalism mistakenly supposes that this is not the case. The argument typically goes as follows: As a matter of social practice, words have no fixed or plain meaning, and communications are not self-executing. A tribunal faced with a communicative text of potentially legal significance must always make a contextual interpretation, based on its experience, on its stereotypes about parties such as these and their likely purposes, and on the linguistic conventions it regularly participates in and in which it thinks the parties

67. For mathematically-inclined readers, it is useful to represent the foregoing conceptual model in formal notation as follows: First, each interpretive agent combines her agent-specific information set I_a with the contractual information set I , using a particular aggregation function $G(I, I_a)$. Second, the agent applies some decision rule $D(\bullet)$ to the resulting aggregated information. Third, this same process is carried out by all types of agents, so that the ex ante expected outcome from the viewpoint of the contracting parties can be written as:

$$\int D(G(I, I_a)) f(a) da$$

where $f(a)$ represents the probability distribution of agents of type a . As indicated in the text, both the aggregation function $G(\bullet)$ and the decision rule $D(\bullet)$ may be agent-specific. For example, some agents may combine their personal information with the contractually specified information using the union function $\cup(I, I_a)$, while other more skeptical agents may use the more restrictive intersection function $\cap(I, I_a)$.

68. Indeed, this is the very arrangement that Bernstein claims contracting commercial parties will typically prefer. Bernstein, Cotton, *supra* note 26, at 1739–45.

participate. As the legal and literary critic Stanley Fish has put it, "A sentence is never not in a context. We are never not in a situation. . . . A set of interpretive assumptions is always in force. A sentence that seems to need no interpretation is already the product of one."⁶⁹ Accordingly, if the tribunal interprets a contract formally—that is, without fully inquiring into the actual context out of which it arose—there is no guarantee that it will apply the contract as the parties subjectively intended. The parties may have meant "chicken" to mean "broiler chicken," they may have meant "minimum quantity" to mean "at buyer's option," they may even have intended "buy" to mean "sell." One cannot know for sure without inquiring; if the court does not inquire, it is interpreting by its own lights, not the parties'. The choice for the court, therefore, is not whether to rely on context and substance, but which context and substance to rely on: the parties' or its own.

This argument—call it the *contextualist argument*—has been very influential in the contracts literature in the last fifty years; its advocates have included such luminaries as Wigmore,⁷⁰ Corbin,⁷¹ and Justice Traynor.⁷² In its claim that all interpretation requires some context, it seems plainly

69. Stanley Fish, *Is There a Text in This Class? The Authority of Interpretive Communities* 284 (1980).

70. "The fallacy [of plain meaning] consists in assuming that there is or ever can be *some one real* or absolute meaning." 9 John Henry Wigmore, *Wigmore on Evidence* § 2462(1) (Chadborn rev. 1981).

71. "[S]ome of the surrounding circumstances always must be known before the meaning of the words can be plain and clear; and proof of the circumstances may make a meaning plain and clear when in the absence of such proof some other meaning may also have seemed plain and clear." 3 Arthur Linton Corbin, *Corbin on Contracts* § 542 (1960) (footnote omitted). Also note the Second Restatement's position on contextual meaning:

It is sometimes said that extrinsic evidence cannot change the plain meaning of a writing, but meaning can almost never be plain except in a context. Accordingly, the rule stated in Subsection (1) is not limited to cases where it is determined that the language used is ambiguous. Any determination of meaning or ambiguity should only be made in the light of the relevant evidence of the situation and relations of the parties, the subject matter of the transaction, preliminary negotiations and statements made therein, usages of trade, and the course of dealing between the parties.

Restatement (Second) of Contracts, *supra* note 10, § 212, cmt. b.

72. A rule that would limit the determination of the meaning of a written instrument to its four-corners merely because it seems to the court to be clear and unambiguous, would either deny the relevance of the intention of the parties or presuppose a degree of verbal precision and stability our language has not attained.

Pac. Gas & Elec. Co. v. G.W. Thomas Drayage & Rigging Co., 442 P.2d 641, 644 (Cal. 1968) (Traynor, C.J.).

Words are used in an endless variety of contexts. Their meaning is not subsequently attached to them by the reader but is formulated by the writer and can only be found by interpretation in the light of all the circumstances that reveal the sense in which the writer used the words. The exclusion of parol evidence regarding such circumstances merely because the words do not appear ambiguous to the reader can easily lead to the attribution to a written instrument of a meaning that was never intended.

right. Where the argument goes wrong, however, is in concluding that this claim, together with the goal of carrying out the parties' intentions, commits one to a substantive approach to interpretation; such a conclusion does not follow. As Eric Posner has observed, the argument neglects the possibility that the parties can have intentions regarding how their intentions are to be interpreted.⁷³

Translated into the framework of our model, the contextualist argument simply states that interpretation is always carried out with reference to a particular information set. A tribunal's information set is made up of various elements, including the judges' experience and training, the text of the contractual agreement, as well as any additional material presented by the parties in litigation. Given its information set, the court can carry out its interpretation with the goal of forwarding the intentions of the parties, or it can pursue some other goal, such as forwarding its own view of the best social policy. Whatever goal the court pursues, however, it must make its best guess based on the information available to it (in the way that courts knowingly make decisions based on limited information when deciding a motion for summary judgment or judgment on the pleadings). While the quality of the guess, accordingly, depends on the available information set, which information set to use and which goal to pursue are independent questions. A broad information set can be used to pursue goals other than the fulfillment of the parties' intentions, and a narrow information set can be used to pursue the parties' intentions, however roughly.

This translation into information-theoretic language shows that not only does the contextualist argument not prove that plain meaning is incoherent, it actually provides us with an operational definition of plain meaning—and an economic one at that. Namely, for a given audience or interpreter, plain meaning corresponds to the interpretation associated with the interpreter's ordinary or zero-cost context—that is, the context that the interpreter can apply with minimal work. Under more substantive interpretative doctrines, the tribunal deliberately seeks out an augmented context. Under more formalist interpretive doctrines, the tribunal deliberately restricts its context. Note that on this definition, what meaning is plain will be agent-specific and context-specific. If I make a pun or employ irony, for instance, my plain meaning will be one thing to an audience that catches the irony and another to one that does not.

Note also that according to this definition, plain meaning is not the most formalistic interpretive mode possible. The tribunal can ignore or throw away information that is part of its ordinary context, as when a court applying constitutional rules of criminal procedure deliberately excludes evidence that is the fruit of an illegal search or coerced confession.

Universal Sales Corp. v. Cal. Press Mfg. Co., 128 P.2d 665, 679 (Cal. 1942) (Traynor, J., concurring).

73. E. Posner, *The Parol Evidence Rule*, *supra* note 23, at 570–71.

Because some effort is involved in the exclusion, however, it only makes sense to do this if there is some cost associated with using the excluded information.

The same analysis holds, by the way, for all types of interpretation. If I plan to attend a Shakespearean play, for instance, I could read the play in advance so as not to miss intricacies of language that would not otherwise be familiar to me, or I could just go unprepared and enjoy the play as best as I can. If I do read the play in advance, I could buy the pocket-book version, which is cheaper and easier to carry around on the subway, or I could at somewhat greater cost buy and read the annotated edition. I could read the introductory essays in that edition, or not; I could go to the library and read secondary literature or do historical research, or not; I could even go to graduate school and get a Ph.D. in English. Assuming I know of the existence of the annotated edition, the secondary literature, and the available graduate programs, however, my choice is a deliberate and informed one, influenced by the relative costs and benefits of the alternatives.

Conversely, creators of communicative texts also make choices about how much context to provide, and this choice is also influenced by the costs and benefits. An author could spell out additional meaning in a fuller and longer text. This will increase printing and shipping costs as well as the time required for reading; it will also tend to reduce spontaneity and creative experience for the reader. Thus in literary (and especially poetic) communications this is not usually done, but it can be (consider T.S. Eliot's *The Waste Land*, with its extensive annotations). The same is true in music, painting, arts, letters, and law. The decision whether to provide more context, however, depends on purposes of the interpretation—or, stated in economic terms, the marginal costs and benefits of context.

The trick is to identify the relevant costs and benefits and how to trade them off against each other. In the example of the Shakespearean play, for instance, the tradeoff is relatively straightforward: more time and effort versus a deeper enjoyment of the play. Furthermore, since both the costs and benefits accrue to me personally, there is little reason not to let me decide how I wish (putting aside paternalistic situations such as high school English class). In cases where contracts and other texts of legal significance are being interpreted, however, the problem is more complicated, and the costs and benefits are more varied. For example, errors in determining whether or not one party owed another a legal duty or whether such a duty was breached can undercut incentives to comply with such duties.⁷⁴ Errors in determining the standard of care implied by a legal duty, the amount of care that the parties actually took

74. See Louis Kaplow & Steven Shavell, Accuracy in the Determination of Liability, 37 J.L. & Econ. 1, 10–14 (1994) [hereinafter Kaplow & Shavell, Accuracy] (examining the social value, as measured by the principles of standard welfare economics, of reducing errors of liability).

in their particular case, or the damages resulting from a breach of duty can encourage either inadequate or excessive caretaking.⁷⁵ In contractual settings, the parties are also interested in incentives for information exchange and for investment in the relationship. To the extent that the tradeoffs among these various considerations differ among contractual and commercial settings, accordingly, different contracting parties might have different preferences about how their agreements should be interpreted.

A committed contextualist might offer an objection to the foregoing approach, one that we can call the “bootstrapping” argument because it suggests that any private interpretative direction by the parties depends for its enforcement on a public and mandatory interpretative regime. Specifically, the contextualist would argue, while such tradeoffs might well exist, it is not possible to determine what interpretative regime the parties intended to choose without investigating the context in which they contracted. Suppose for instance that a court is confronted with a standard form contract that contains, as one of its printed terms, a standard merger or arbitration clause. The mere fact that the clause appears as part of the printed text does not by itself imply that the parties agreed to it; it could be a boilerplate or fine-print term to which no one attended, the result of fraud or force, and so on. In order to apply the clause properly, the court must first decide whether it is part of the contract. In short, the court must make its own prior interpretation, and this interpretation, in principle, cannot be the subject of the parties’ choice.

The proper answer to the bootstrapping argument is that it is overstated. It is true that in order for the parties to choose a contractual information set, they must find some way of communicating to their interpretative agent that they have done so. It is also true that an interpretative agent faced with an apparent contract must find some way of determining whether the item in question actually is a contract, and what information set she is supposed to use to interpret it. But the interpretation that is required to make this threshold determination is rather more limited and bounded than the interpretation that would be necessary to establish all the parties’ substantive rights and duties. It is possible to separate out the specific question of interpretative approach and to determine that question in an initial and narrower inquiry; all formal regimes of interpretation have managed to do that. The relatively formal version of the parol evidence rule contained in the *First Restatement of Contracts*, for example, admitted various doctrinal exceptions based on fraud, condition precedent, and the like.⁷⁶ These exceptions made the doctrine more complicated, but they did not change the fact that the *First Restate-*

75. See Richard Craswell & John E. Calfee, *Deterrence and Uncertain Legal Standards*, 2 J.L. Econ. & Org. 279, 298–99 (1986); Louis Kaplow & Steven Shavell, *Accuracy*, supra note 74, at 192–93.

76. *Restatement of Contracts* § 238(b) (stating proof of fraud can be made out using parol evidence); id. § 241 (stating oral condition precedent to effectiveness of contract is

ment regime was relatively more formal than the *Second Restatement* regime that succeeded it. Similarly, the law of negotiable instruments and of letters of credit admits various real defenses—duress, forgery, fraud in the inducement—that can prevent a holder in due course from recovering from the drawer, and to this extent, courts ruling on claims must make contextual inquiries.⁷⁷ But these inquiries are rather more limited than those that would be allowed in a full-blown suit on the underlying contract between drawer and drawee, and they do not undercut the relative formality of these specialized legal regimes. Thus, it is possible and coherent for courts to follow relatively formal interpretative methods, and for contracting parties to direct them to do so, even if there is some slippage in the process.

In summary, the problem of form versus substance in contract interpretation can be assimilated to the problem of optimal information acquisition. From an economic viewpoint, a fuller or broader context can be purchased—but only at a cost of time and trouble, and of exacerbating certain incentive problems—so it pays to stop at some optimal point. The next section of this Essay surveys the main types of considerations that determine the costs and benefits of formalism, and thus the optimal stopping point.

III. CHOOSING BETWEEN FORM AND SUBSTANCE

The standards for measuring contractual liability and damages for breach influence contracting parties' behavior in many respects: with regard to decisions to breach, to take advance precautions, to mitigate damages, to gather and communicate information, to allocate risk, to make reliance investments, to behave opportunistically, and to spend resources in litigation, among others. The regime of contract interpretation, because it determines how liability and damages will be assessed *ex post*, has similarly widespread incentive and efficiency effects. Accordingly, given the purposes for which I am writing, it does not make sense to try to develop a unitary theory for choosing between form and substance, since the answer in any particular case will turn on a comparison of various types of transaction costs. Instead, I will list and discuss the main categories of these transaction costs, with the hope that a systematic consideration of these issues will help individual parties address the formality problem in specific contexts.

binding if there is nothing in writing which is inconsistent therewith); *id.* § 240 (1932) (stating oral promise enforceable if contained in collateral agreement).

77. See U.C.C. §§ 3-305(a)(1), 3-305(b) & cmt. 1 (2000) (noting, *inter alia*, that the court must account for all relevant factors, “including the intelligence, education, business experience, and ability to read or understand English of the signer” to determine whether the signer satisfied the “excusable ignorance” test of the fraud defense).

A. *A Survey of Economic Criteria Relevant to the Choice*

1. *Direct Transaction Costs.* — The most obvious consideration relevant to choosing an interpretive regime is the direct costs of writing contracts and litigating contractual disputes, and these costs can be affected in various ways by formality. Other things being equal, an agent that bases its interpretive decisions on a smaller set of materials should require, whether it is performing or enforcing a contract, less time and effort to carry out its task. Thus, formal modes of interpretation will be appropriate whenever the ex post costs of time and effort are especially large—for example, when time is of the essence or when the opportunity costs of the enforcing or performing agent’s time is high. Thus, in letter of credit transactions, where the viability of the letter of credit as a payment device depends on the speed and administrative efficiency with which payment can be processed, formal methods of interpretation are favored.⁷⁸ Issuing banks are not supposed to look past the face of submitted documents when determining whether documentary conditions have been satisfied, and the rules for determining compliance with such conditions are strict. Similarly, in markets where a high level of trust among the participants is necessary to support cooperation with regard to the performance of nonverifiable aspects of the contract, and where extended disputes can undermine such trust, disputes can be kept short and relatively painless through the application of more formal decision procedures.⁷⁹

It should be recognized, however, that if the contracting parties anticipate that formal decision procedures will be applied at the performance or enforcement stage, they may be induced to put greater effort into specifying additional considerations or supplying additional interpretive materials at the contract-writing stage, in order to address some of the issues that are discussed below, such as risk or performance incentives. For example, the anticipation that issuing banks will not look beneath the surface of any supporting documents when processing a letter of credit may induce the issuer to provide a more elaborate set of documentary conditions up front. (Conversely, the prospect that any such conditions will be enforced strictly may induce parties to provide a less elaborate list of requirements.)

78. See U.C.C. § 5-103 cmt. 1 (noting important difference between formal letters of credit and “secondary,” “accessory,” or “suretyship” guarantees); John F. Dolan, *The Law of Letters of Credit: Commercial and Standby Credits* § 3.07 (Warren, Gorham & Lamont 2d ed. 1991) (detailing advantages of modern formal methods of interpretation with regard to commercial credit); Boris Kozolchik, *The Financial Standby: A Summary Description of Practice and Related Legal Problems*, 28 UCC L.J. 327, 334–35, 358–59 (1996) (“The need for a speedy and abstract payment . . . requires that the issuer . . . rely exclusively on documentary terms and conditions.”).

79. See, e.g., Lisa Bernstein, *Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry*, 21 J. Legal Stud. 115, 119–30 (1992) (describing trust relations and dispute resolution procedures in the uncut diamonds market).

If this effect is a significant one, then the cost of considering additional interpretive materials *ex post* needs to be weighed against the costs of increased contract-writing costs *ex ante*.

2. *Risk*. — Variation in interpretive outcomes introduces risk into the contractual relationship. Since contracting parties usually dislike risk and are willing to expend resources to avoid it, they may choose between form and substance as a risk management device. The choice will be different, however, for parties with different attitudes toward risk or different abilities to spread or diversify it. In the interpretive setting, outcome risk derives from variations in the distribution of agent-specific information sets. A widely dispersed distribution of information sets means that factfinders or performers will interpret the same materials differently. To the extent that including additional interpretive material reduces this variation, it will reduce the resultant interpretation risk.⁸⁰

For instance, suppose that judges vary in their background experience with regard to commercial matters. They will accordingly differ in their reading of particular documents, or of legal standards such as reasonableness or good faith. If the judges are directed to inquire more deeply into the commercial context before deciding on their interpretation, this inquiry will (at a cost) reduce variance by making the less experienced judges' information sets more closely resemble the more experienced judges' information sets.⁸¹

The value of risk reduction may help to explain Lisa Bernstein's observation that industry tribunals tend to follow relatively formal regimes of interpretation, even though their cost of inquiring into substance is less than that of generalist judges.⁸² To the extent that the judges are already expert in the subject of the contract, the variance among their information sets is likely to be low. Thus, the marginal value of risk reduction that is purchased by the consideration of additional information is likely to be smaller, and less likely to justify incurring the additional costs. The fact that contracting parties in such settings prefer relatively formal rules when litigating in front of expert tribunals, accordingly, does

80. The conclusion would depend on the precise functional form of the information aggregation function G . For example, if G were the union function $\cup (I, I_a)$, increasing the size of the contractual information set would reduce the variance of G . If G were the intersection function $\cap (I, I_a)$, conversely, increasing the size of the contractual information set would have no effect on the variance of G .

81. One might speculate that differences in perspective will lead the experienced judges to evaluate the new material differently, thus increasing the interpretive variance rather than decreasing it, but such an outcome is unlikely so long as there are diminishing returns to expertise, or so long as the variations in judges' evaluation of individual items of interpretive material are less than fully correlated, so that expanding the basis of decision will reduce total variance through diversification and the law of large numbers. Again, stated in terms of formal mathematical notation, this would depend on the precise functional form of the information aggregation function G .

82. Bernstein, Cotton, *supra* note 26, at 1735.

not imply that they would have a similar preference when litigating in front of generalist judges and juries.

Additionally, to the extent that substantive interpretation reduces the variance of interpretive outcomes, it is more valuable to relatively risk-averse parties, other things being equal. Conversely, contracting parties who are less risk-averse or who have other methods of risk reduction available to them should be less willing to incur the costs of substantive interpretation. This latter category includes larger or more diversified businesses and other contractual repeat players, who can diversify interpretation risk over a greater number of transactions, as well as agents such as middlemen who are likely to be on both the buying and selling side of transactions with equal frequency. The risk factor thus provides an additional explanation for why such actors tend to use standard forms, in addition to the more obvious reason of economies of scale.

3. *Performance Incentives.* — Variation in interpretive outcome is not just a matter of risk, of course, because the parties' anticipation of what enforcers will do can affect their incentive to perform their contractual duties. For instance, legal error in assessing contractual damages following breach may induce either inefficient performance (if the tribunal tends to overestimate damages) or inefficient breach (if the tribunal tends to underestimate damages). Variations in the assessment of substantive duties may have similar effects. For instance, if the tribunal tends to overestimate (underestimate) the promised level of product quality by reading an express or implied warranty more broadly (narrowly) than the parties intended, this may induce the seller to provide too much (little) quality from an efficiency point of view. The distributional consequences of these sorts of errors can be priced out on average, but the efficiency consequences may remain. Lower variance in interpretive outcome, accordingly, can provide the parties with more precise performance incentives.

The value of such increased precision, however, depends upon the parties' circumstances, including the information available to them at the time they make performance or precaution decisions and their ability to renegotiate the contract *ex post*. The fact that tribunals vary in their potential assessments of damages, for example, should not lead to inefficient breach or performance so long as the assessment is correct on average, since the contracting parties are unlikely to know the particular characteristics of their tribunal at the time they have to decide whether to perform. In making their decisions, parties will be in a position of uncertainty and will only be able to compare the costs of performing against the expected costs of paying damages for breach, averaged over the set of all potential tribunals. A reduction in the tribunal's variance, accordingly, does not purchase any efficiency gains up front, so long as the parties are not risk-averse.

On the other hand, if parties do have information about the likely direction of tribunal error at the relevant time of decisionmaking, their

incentives to perform or take precautions against breach will be inefficient. For instance, suppose that a sales contract contains a clause that requires the seller to deliver goods by June 1, but the parties as a matter of trade usage understand the delivery date to be interpreted flexibly, and in their understanding the seller has the option to deliver as late as June 15 if market conditions make it unduly expensive to meet the June 1 date. If not all courts would recognize this implicit understanding absent an inquiry into the commercial context, a seller who does not meet the June 1 date will expect to be found in breach of contract with some positive probability. Depending on the damages that might be assessed and the expected costs of any litigation, accordingly, he might be led to take inefficiently costly precautions to guard against late delivery. If the inefficiency losses are high enough, it would be worth directing courts to inquire into trade usage before finding any liability.⁸³

The extent of these inefficiency losses, however, depends on the ex post costs of disputing and renegotiation. Just because the original contract does not provide efficient incentives for performance does not mean that an inefficient outcome will result, since the parties can modify the contract after the fact to reach an efficient outcome. This renegotiation may entail an additional payment from one of the parties to another, but the expected cost of this payment can be calculated up front and included in the original contract price.⁸⁴

Parties with relatively low ex post renegotiation costs, accordingly, should tend to favor formalistic methods of interpretation, other things being equal. This category includes parties who anticipate a continuing relationship, parties who engage in many similar transactions or do business together regularly, and parties who expect to have symmetric information ex post regarding the costs and benefits of performance. Parties for whom the transaction is an unusual one, parties in one-shot contracts, and parties who expect there to be asymmetric information ex post should tend to favor more substantive methods of interpretation as a substitute for their own ability to bargain to an efficient outcome.⁸⁵ Similarly, parties who have available to them other methods of ensuring effi-

83. They could also write their trade usage explicitly into the contract, but this has transaction costs of its own and is likely to be cost-justified only for usages that are unfamiliar to a sufficient number of courts, or that govern contingencies that are especially likely to arise. If the contingency in which trade usage becomes relevant is of sufficiently low probability, or if the likelihood of a given court being unfamiliar with it ex post is low, then it will be cheaper for the parties to remain silent and to take the risk of an incorrect interpretation.

84. See Richard Craswell, *Contract Remedies, Renegotiation, and the Theory of Efficient Breach*, 61 *S. Cal. L. Rev.* 629, 633–36 (1988) (discussing how ex post negotiations tend to lead to efficient outcomes).

85. Compare Jason Scott Johnston's argument that legal regimes that use standards to determine liability do a better job at encouraging bargaining under asymmetric information than regimes that use rules. Jason Scott Johnson, *Bargaining Under Rules Versus Standards*, 11 *J.L. Econ. & Org.* 256, 257 (1995).

cient performance, such as nonlegal or reputational sanctions administered through membership in a commercial subcommunity, are less likely to want to incur the expenses of substantive interpretation, and other things being equal should prefer a formalist approach.

4. *Reliance Incentives.* — Much of the economics-of-contracts literature has emphasized the role of contractual liability in promoting investment in relationship-specific assets. Absent legal protection for such investments, rational contracting parties will underinvest in them from the efficiency viewpoint, for fear of losing some or all of their value in an ex post holdup. The standard intuition here is that because the asset is worth little outside the specific relationship, the party who invests in it becomes vulnerable to threats to terminate the relationship. Such threats provide the noninvesting party with the bargaining power to obtain a unilaterally favorable modification. But investors' ability to anticipate such opportunism reduces their incentive to make such investments in the first place. In contrast to the problem of inefficient performance and breach, ex post renegotiation cannot address this efficiency problem, since it is precisely the prospect of such renegotiation that creates the threat of holdup.⁸⁶ It can only guarantee that whatever investments are made are put to efficient use ex post.

The need to encourage specific investments will influence the form/substance decision whenever the value of the investment turns on the nature of contractual interpretation. Many investments, even if they are relationship-specific, will not depend on interpretation in this way. For example, suppose that a supplier of complex industrial machinery must invest substantial time and effort acquiring expertise about the specific production process of a particular customer. This expertise is only partially transferable to relationships with other customers and is thus relationship-specific; in order to be induced to acquire it, the supplier must be persuaded that a relationship with this particular customer is in the offing. An important way to commit to such a relationship is through a binding contractual promise. However, the value of the supplier's investment in expertise need not turn on the specific content of the contract; it may be that knowing the customer's needs reduces the cost of providing the customer with machinery of all sorts. In this case, the supplier need not worry about unexpected contractual interpretations that leave the basic contract in place (for example, requiring the delivery of a machine with this set of characteristics rather than that), since its investment in expertise is equally sunk with all interpretations. The possible variation in contract requirements does involve some risk, of course, but this can be priced out or dealt with using the other methods described above.

86. See generally Benjamin Klein et al., *Vertical Integration, Appropriable Rents, and the Competitive Contracting Process*, 21 *J.L. & Econ.* 297 (1978) (explaining that specific investments determine appropriate scope for vertical integration of the firm); Oliver E. Williamson, *Transaction-Cost Economics: The Governance of Contractual Relations*, 22 *J.L. & Econ.* 233 (1979) (discussing specific investments as a type of transaction cost).

On the other hand, there are some investments whose value turns on the specifics of the task to be performed. A supplier who contracts to supply goods within a narrow time window may need to take special precautions in storing inventory and arranging for timely shipment; it may conversely fail to invest in facilities that would provide it with greater flexibility to deliver outside the window. If the contract is subsequently interpreted to provide the customer with greater discretion in specifying the time of delivery, the supplier can become vulnerable to holdup; it may have to agree to a substantial reduction in price in order to induce the buyer to take delivery during the originally anticipated window.⁸⁷ A second example is provided by the case of a supplier who promises to supply finely milled machine parts of a particular specification. If the contract is interpreted to allow the buyer more leeway to alter the specifications (or, conversely, to insist on strict rather than substantive compliance with the specifications), then the supplier will be relatively vulnerable to holdup if its retooling costs are large, but relatively less vulnerable to holdup if its retooling costs are small.

To the extent that a substantive interpretive approach improves the quality of the enforcing tribunal's estimate of the parties' expectations (that is, to the extent that it reduces the expected difference between the interpretive outcomes *ex post* and the parties' interpretations at the time they must sink their specific investments), it will reduce the potential for such holdups. Other things being equal, parties who find it relatively important to undertake interpretation-specific investments, or whose investments are especially vulnerable to changes in contractual interpretation, will therefore be more likely to want to opt into regimes of substantive interpretation. Parties who do not need to make such investments, or whose investments are more flexible, or who have other methods at their disposal for dealing with contractual opportunism, will have less need for interpretive accuracy and should tend to prefer relatively formalist regimes.

5. *Rent Seeking*. — The discussion so far presumes that the costs of writing and litigating contracts is exogenous to the parties' behavior, but more generally this is not the case. Depending on the legal regime, the parties can do various things *ex ante* or *ex post* to turn the bargain in their favor. Under a regime of substantive interpretation, for instance, parties may be tempted to invest substantial resources in litigation in order to maximize the chance of a favorable outcome. From the perspective of the contracting parties together, such behavior is wasteful, except to the extent that it improves incentives for primary behavior. From the point of view of litigants *ex post*, however, it is individually rational even if there is no such incentive effect.

87. Conversely, if the contract is interpreted to require delivery within a window when one or both of the parties understood the window to be more flexible, the party who is caught short may have to pay a substantial ransom in exchange for being released from this unanticipated obligation.

Formality, by limiting the scope for *ex post* interpretive disputes, probably reduces the marginal productivity of litigation expenditure, and thus reduces the amount of such expenditure. To the extent that it conditions the outcome of litigation on publicly available information, and reduces the variations of litigants' expectations regarding that outcome, it probably also encourages settlement.

On the other hand, rent-seeking can take place at the contract-writing stage as well. For example, I have elsewhere argued that one cost of enforcing standard form contracts according to the plain meaning of their written provisions is that those who write such contracts will be tempted to sneak one-sided but inefficient terms into the fine print.⁸⁸ Non-drafting parties will generally not find it worthwhile to examine standard forms with the care required to unearth such self-serving terms, so they are likely instead to assume that such terms have been included and discount the price they are willing to pay accordingly. (Even if parties do examine the standard forms of their contractual partners, such an examination is costly.) Accordingly, both parties would find it useful to have a way of committing to abstain from such behavior, and an interpretive regime that de-emphasizes the text of the agreement in favor of less manipulable considerations, such as market expectations, may provide such a commitment device.

But there is, at least in theory, a similar risk of such manipulation with regard to contextual materials, such as parol evidence. Just as parties may be tempted to sneak self-serving terms into the contractual text under a formalist interpretive regime, they may be tempted to fill the negotiating history with self-serving proposals and offers under a more substantive interpretive regime, in the hopes of influencing the ultimate result.⁸⁹ I am inclined to regard this latter risk as relatively less important, since in most cases the parties will have more symmetric and effective access to their common negotiating history than they will to each others' standard forms. Blatantly self-serving attempts to manipulate parol evidence are more likely to be observed and parried during negotiations, while self-serving form terms are more likely to escape notice until the contingencies they relate to have materialized.⁹⁰ Still, there is a theoretical tradeoff, so contracting parties concerned about such rent-seeking will want to choose between relatively formal and relatively substantive interpretive regimes depending upon whether they believe rent-seeking is a more significant problem *ex ante* or *ex post*, and whether it is a more

88. Avery Katz, *Your Terms or Mine? The Duty to Read the Fine Print in Contracts*, 21 *RAND J. Econ.* 518, 533 (1990).

89. Eric Posner offers this risk as a possible reason in favor of a relatively formalistic parol evidence rule. See E. Posner, *The Parol Evidence Rule*, *supra* note 23, at 564–65. Compare the analogous arguments that advocates of strict statutory construction have made with respect to incentives to create self-serving legislative history.

90. Again, this may contrast with the case of using legislative history in statutory interpretation, since relevant actors may find it easier to introduce self-serving statements into the official record without being noticed by their political adversaries.

significant problem with respect to contractual text than with regard to context. Certain types of contextual evidence may be more or less subject to such manipulation; for instance, trade usage seems fairly immune to rent-seeking ex ante, since no individual agent is in a position to have a significant effect on it. On the other hand, to the extent that usage is diffuse and there is room to argue about its substantive content, it is relatively susceptible to rent-seeking ex post. The restrictions on the admissibility of trade usage provided by U.C.C. section 1-205(6) may thus be viewed as a way to lessen the rent-seeking costs associated with this type of evidence.

6. *Agency Problems.* — Another commonly-cited reason for privileging textual over contextual material when interpreting contracts is the need to control the behavior of imperfectly loyal agents. Here it is worth distinguishing between two kinds of agency problems: problems in controlling the behavior of enforcing agents ex post, and problems in controlling the behavior of contracting agents ex ante.

a. *Controlling Enforcing Agents Ex Post.* — While implementing the contracting parties' intentions is a major and perhaps primary consideration when courts interpret their agreements, it is far from the only factor. Courts and other tribunals may be tempted to tailor their interpretations, if only marginally, in furtherance of other goals such as distributional equity, risk sharing ex post, or corrective justice. Since none of these goals are in the ex ante interests of the contracting parties, the parties would like to arrange the interpretive process so as to minimize the influence of such considerations, to the extent that they can do so at reasonable cost.

One obvious way to do this is to choose in advance a tribunal that is expected to give greater weight to the expected value of the contract, and lesser weight to the tribunal's own countervailing values. Choice of law and, especially, arbitration clauses are straightforward ways of implementing such a choice. But it has often been suggested that restricting the scope of admissible interpretive materials has a similar constraining effect.

Whether this is the case is not clear and depends on a closer study of the particular agent in question and the professional community to whom the agent looks for validation. Certainly, the expansion of the informational universe provides additional opportunities for a court seeking to promote its own values to find justification for its actions.⁹¹ But as the Legal Realists famously argued, a court bent on ignoring the parties'

91. A possible example of this phenomenon may be found in *Nanakuli Paving & Rock Co. v. Shell Oil Co.*, 664 F.2d 772, 779-80 (9th Cir. 1981), in which the court directed the admission before a local jury of trade usage evidence purporting to show that Shell had promised Nanakuli, a small local paving company to whom Shell was selling asphalt, that it would protect Nanakuli against any price increases that came after the placing of Nanakuli's order and before delivery, notwithstanding a written contract term providing that the price would be "Shell's Posted Price at time of delivery," and notwithstanding the

intentions in favor of its own policy values also has substantial freedom to do so under a more formal interpretive regime.

b. *Controlling Contracting Agents Ex Ante.* — Many contracting parties are not individuals, but organizations that can only act through individual agents. Because the incentives of the agents are imperfectly aligned with those of the organization, agents may not behave in a value-maximizing way when entering into a contract or engaging in actions that may affect contractual terms. For example, an insurance agent may make inaccurate representations about policy coverage or about an applicant's insurability in order to earn a commission, or a manager charged with supervising a supplier's obligations may shirk by failing to object to defective performance, thus providing the supplier with a plausible claim of waiver.

In such contexts, formalism can be used by one group of organizational actors to disable other actors from binding the organization on terms that might be in the latter actors' private interests, but not the interests of the organization.⁹² Merger clauses, for example, can be used to take contracting power away from the sales and purchasing agents who orally represent the organization in its dealings with outsiders, and to consolidate that power in the managers and legal professionals who control the official texts of company documents. Similarly, anti-waiver clauses can be used to protect an organization against shirking by its enforcement agents.

It is important to recognize, however, that the individuals who control the formal text of an organization's contractual agreement are no less agents than those who control the less formal context. Most commercial form contracts are drafted by lawyers, either in-house or not, whose compensation structure provides them with incentives that are not identical or even proportional to the benefits and costs to the firm that employs them. For example, a company lawyer charged with drafting terms in a standard form is probably more likely to be punished for omitting a term that turns out to lead in some remote contingency to a loss for the firm, than he or she is to be rewarded for the time saved by the omission in the far more probable event that the term is unnecessary. The asymmetric nature of the payoff will lead the lawyer to overdraft the contract. Similarly, the lawyer is unlikely to be rewarded for creating terms that increase the value of the contract to customers or, if he or she works in isolation

text of U.C.C. section 1-205(4), which provides that when express terms and trade usage cannot be read together consistently, express terms control contractual meaning.

92. See, e.g., Kevin Davis, *Licensing Lies: Merger Clauses, the Parol Evidence Rule and Pre-contractual Misrepresentations*, 33 *Val. U. L. Rev.* 485, 534 (1999) (arguing that in the interests of minimizing agency costs, sophisticated commercial parties should be allowed to disclaim liability for their agents' extra-contractual representations); Avery Wiener Katz, *On the Use of Practitioner Surveys in Commercial Law Research: Comments on Daniel Keating's 'Exploring the Battle of the Forms in Action'*, 98 *Mich. L. Rev.* 2760, 2769–70 (2000) (offering agency-cost interpretation of the battle of the forms); E. Posner, *The Parol Evidence Rule*, *supra* note 23, at 563–64 (suggesting that rational firms consider the costs of shirking by agents when adopting merger clauses).

from the sales department, even to know what such terms might be. For that reason, he or she will likely draft terms that are inefficiently favorable to the company, in that they shift risks and duties to customers that from a business viewpoint are more cheaply borne by the company. Adjusting the contract so that it better fits the needs of the customer and makes him or her willing to buy, accordingly, is a task usually left to sales agents, not lawyers.

Thus, an organization should favor formal over substantive methods of interpretation if it has established relatively efficient incentive structures for controlling the behavior of its legal department, and relatively inefficient incentive structures for controlling the behavior of its sales and purchasing departments. If the agency problems are greater with respect to the organization's lawyers, conversely, it should prefer a less formal interpretive regime. The optimal choice may depend on the administrative tools available and on other agency problems that the organization faces. For example, due to difficulties in monitoring the effort level of sales agents, it may be desirable to give them high-powered incentives by providing the bulk of their compensation in the form of commissions. Given the incentives set up by the commission system, it may then make sense to limit the sales staff's ability to vary contract terms by use of standard forms including a merger clause, especially if the lawyer's compensation is reasonably well tied to the overall profits of the firm. On the other hand, if it is feasible to establish a chargeback system whereby sales agents' earnings are reduced in an amount proportional to the number of disputes arising out of their sales, or by the extra costs necessary to service the special terms promised to their customers, and if the lawyer drafting the contract is an independent contractor rather than an ongoing member of the organization, then the sales agents may be in a better position to balance the costs and benefits to the organization when making informal promises and representations.

7. *Liquidity and the Cost of Other Complementary Services.* — Finally, the optimal choice between form and substance may depend on the importance of third-party contributions to the value of the contract. For example, a buyer of commercial machinery may need to borrow funds in order to pay for its purchase, and will usually be able to borrow at a lower rate if it pledges the machinery as collateral. Similarly, a seller of consumer goods will be able to make credit sales at a lower price if it can sell its customer accounts to a commercial factor, who specializes in buying such accounts and can service them and bear default risk at a lower cost. Another example would be a buyer who purchases two specialized pieces of equipment from two different suppliers when the two items are intended to be used together; in this case the terms of the buyer's arrangement with one supplier (e.g., terms granting the supplier the discretion to alter the specifications) will affect the terms it can get from the other supplier.

In such cases, the third party's ability to provide such complementary services at low cost will depend in part on the cost it faces when determin-

ing the terms of the supported contract. If a factor has to worry about an account debtor asserting defenses when it comes time to collect on the account, the amount it is willing to lend against the account will be reduced accordingly. If the factor can effectively assess the risk of such defenses from an examination of the underlying contract's text, however, it can price the risk out, hold back an amount in reserve that corresponds to the expected value of uncollectible accounts, and lend the balance. If the existence of such defenses depends upon more contextual factors, such as oral communications between the account debtor and the seller's sales agents, the factor will need to be more conservative and will not be able to lend at the same rate or in the same amount. Casting the account in the form of a negotiable instrument, however, or including a holder-in-due-course clause in the original sales agreement, lowers the factor's costs by reducing the risk of nonpayment resulting from a cause that the factor could not assess *ex ante*. Other formal devices, including merger clauses, antiwaiver clauses, and the like, similarly lower the costs to third parties of providing supporting services, and thus enable those parties to provide the services at lower cost.

In general, contracts whose value depends significantly on the participation of or the purchase of complementary services from third parties will have higher value when interpreted under a relatively formalistic regime, other things being equal. But this conclusion need not always hold. In some cases, the relevant third party may face higher costs when assessing contractual text than when assessing contextual factors such as trade usage. If the third party is not a legal specialist, for example, as in the case of a third-party guaranty supplied by a friend or family member of the primary obligor, or in the case of a specific investment undertaken by a trade creditor or senior employee who finds it cheaper to observe the parties' ordinary business actions and informal commercial reputations than to examine the details of their written contract, then a more substantive approach to interpretation will protect the third party at lower cost, and thus will make him or her willing to provide the complementary services on more favorable terms.

B. *Choosing Between Form and Substance: A Survey of Economic Criteria*

Because the list of economic and commercial considerations relevant to the choice between formal and substantive interpretation is long, and because the various considerations may well cut in opposite directions in individual cases, drawing specific conclusions regarding how to apply the above framework must be tentative at best. Indeed, it is for these very reasons that I argue that public lawmakers are not in a particularly good position to issue strong prescriptions regarding the proper balance between form and substance, and that private parties should be allowed the leeway to choose their favored interpretative regime. Nonetheless, the foregoing discussion does suggest some general rules of thumb in this regard. The summary table lists a number of such heuristic principles,

Transactional issue	Formal interpretation favored when:	Example	Substantive interpretation favored when:	Example
Direct transaction costs	Ex ante negotiation costs are relatively low	Repeat players, high-value transactions, contracts negotiated by junior agents	Ex ante negotiation costs are relatively high	One-shot players, low-value transactions, contracts negotiated by senior principals
	Renegotiation costs are relatively low	Repeat players, socially-connected parties	Renegotiation costs are relatively high	One-shot players, isolated parties
	Chances of dispute are relatively high	High variation in ex post value of exchange	Chances of dispute are relatively low	Low variation in ex post value of exchange
Risk	Tribunals are consistent in their interpretation of form ex post	Expert tribunals, parties trading within a single jurisdiction	Tribunals vary in their interpretation of form ex post	Generalist tribunals, parties trading across jurisdictions
	Parties are consistent in their interpretation of form ex ante	Parties who trade using a common form	Parties vary in their interpretation of form ex ante	Parties who trade with different forms, parties unfamiliar with forms
Performance incentives	Parties have access to effective non-legal sanctions	Socially-connected parties	Parties are relying on legal sanctions to motivate performance	Isolated parties
	Parties expect courts to be biased in interpreting contract	Jurisdictions where tribunals have distinct policy preferences	Parties expect courts to interpret contract accurately on average	Jurisdictions where tribunals do not have distinct policy preferences
Reliance incentives	Specific investments do not depend on context	Thick markets, traders of relatively fungible commodities	Specific investments depend on context	Thin markets, custom-made goods or services
Rent-seeking	Legal outcomes are relatively sensitive to litigation expenditure	Law is unsettled	Legal outcomes are relatively insensitive to litigation expenditure	Law is settled
	Ex post stakes are high	High variation in value of exchange	Ex post stakes are low	Low variation in value of exchange
Agency problems	Principals have relatively weaker control over negotiating agents	Large firms with extended contracting networks, in-house legal services	Principals have relatively weaker control over drafting agents	Small firms with limited contracting networks, for-hire legal services
	One or both parties subject to bias in litigation	Deep-pocket parties, unpopular parties, out-of-jurisdiction parties	Little risk of bias in litigation	Jurisdiction with strong reputation for equal treatment
Third-party investment	Exchange requires complementary input from distant strangers	Negotiable instruments, letters of credit	Exchange requires complementary input from mercantile insiders	Guaranties by affiliated parties, specific investments by senior employees or trade creditors

together with an illustrative example for each principle that shows which kinds of contractors might be in a position to make use of it.

As we can see from the table, some aspects of the conventional wisdom regarding form and substance are borne out by our framework. The framework suggests, for instance, that small and infrequent traders will tend to benefit from a more substantive interpretative regime for a variety of reasons: they are relatively less well-placed to undertake the fixed cost of detailed *ex ante* negotiation; they have relatively poor access to reputational networks *ex post*; they are likely to do their own contract negotiating but to contract out when acquiring legal services; they are less likely to be able to recover specific investments in other exchanges; and they are possibly less likely to face bias in *ex post* judicial tribunals. Conversely, large and experienced mercantile traders should prefer their contracts to be governed by relatively formalistic rules of interpretation—and this prediction is consonant with the observation that, in general, it is such traders whom we observe contracting into relatively formal enforcement regimes through devices such as arbitration, choice of law, and forum selection clauses. Such preference could stem from such traders' ability to amortize the fixed cost of detailed *ex ante* negotiation over a series of transactions, from their relatively good access to nonlegal enforcement via reputational networks, from the fact that their large size weakens their control over their sales and purchasing agents, but strengthens their control over their lawyers, from their greater ability to recover specific investments in substitute exchanges, and from their greater wariness of biased tribunals and juries. Thus our analysis lends some support to the common claim that contracts between merchants and commercial specialists should be interpreted more formally and with less judicial supervision than contracts between less experienced parties or between merchants on the one hand and consumers or employees on the other.⁹³

CONCLUSION

Interpretation is an essential aspect of all fields of law—statutory, common law, and constitutional—but it looms especially large in the area of contracts. This is so for two interrelated reasons. First, from an *ex post* perspective, judicial officials called on to enforce an asserted private agreement as law face special difficulties in determining whether the agreement was actually established, what obligations it provides, and what to do if the agreement's terms appear incomplete, ambiguous, or contradictory. Contractual lawmaking is typically decentralized, acts of legal significance commonly take place in private, the participants are often legal amateurs, and their purposes and methods of communication are highly

93. See, e.g., Schwartz & Scott, *Contract Theory*, *supra* note 28, at 544–45 (conceding that their strong normative recommendations in favor of formal interpretation should be limited to contracts among experienced commercial parties).

varied. In contrast to a professional legislature with its public records and voting procedures, or a court with its official rulings and published opinions, individual contracting parties can regard themselves as having created legal obligations over a period of time without being able to identify the precise moment at which such obligations came into force. It should be no surprise that disagreements over interpretation are a primary cause of litigated contract disputes.

Second, from an *ex ante* perspective, contracting parties have substantial leeway to influence subsequent interpretation by the manner in which they conclude their agreement. They can take more or less care to identify their underlying assumptions and to communicate their intentions to each other; they can anticipate possible interpretative disputes and settle them in advance; and they can create and preserve evidence of their understandings through the use of writings and other permanent documents, independent witnesses, and terms of art. Helping the parties to translate their underlying bargain into something that can actually be applied to guide (if not to bind) their subsequent behavior is the main professional task a transactional lawyer faces.

Most scholarly discussion of interpretive problems, however, especially that dealing with the tension between form and substance, has been addressed to courts and other public lawmakers, and not to private contracting parties. The participants in this discussion have argued for formal and for substantive approaches to interpretation, and have based their recommendations on grounds of efficiency, fairness, and party autonomy. But even those who have disagreed on policy recommendations or normative commitments have found common ground in their choice of audience.

This Essay has argued that the traditional scholarly approach to form and substance founders on a lack of information about the likely consequences of formal and substantive modes of interpretation. From an efficiency viewpoint, the information available at the general level at which courts and legislatures must operate is inadequate to determine the relative magnitude of the relevant transaction costs. From an autonomy viewpoint, the traditional stance of the court system neglects the possibility that different parties in different contexts might prefer—or ought to be delegated the power to choose—one interpretive approach over other. One does see distinctions drawn in the case law and in the commentary between different sorts of contracts; it is generally acknowledged that formalism is relatively more important to experienced commercial actors, and substantive interpretation better suited to transactions involving consumers and other amateurs. But as far as I know there has been no systematic attempt to determine, using the standard tools and methods of the economics of contracts, in which contexts and for which parties formalism is most useful and in which contexts and for which parties a substantive approach is most useful. This Essay has aimed to lay out a basic framework within which such a systematic analysis could take place.