

CONTRACT DESIGN AND THE STRUCTURE OF CONTRACTUAL INTENT

Jody S. Kraus & Robert E. Scott***

INTRODUCTION

Honoring the contractual intent of the parties is the central objective of contract law. The search for intention is a key doctrinal element in determining whether the parties have made a binding agreement, what meaning attaches to the terms of that agreement, what default terms are implied in that agreement, and whether the obligations of the parties are impliedly conditioned on unstated assumptions. In addition, the obligation to honor the intention of the parties is the key linkage between efficiency and autonomy, the two major theoretical approaches to contract. Yet, despite its theoretical and doctrinal centrality, little scholarly attention has been given to the structure of contractual intention. To be sure, both scholars and courts now widely accept that intention is to be assessed prospectively and objectively: a party is taken to mean what her contract partner could plausibly believe she meant when the parties contracted.¹ Thus, the conventional inquiry under modern contract law begins by asking how a reasonable person would have understood her partner's intention at the time of contracting. However, in order to answer this question of ex ante intent, courts must be able to understand the problem of contract design that parties face at the time they form their agreements.

By focusing on contract design we can highlight the key structural components of intentionality. The first step in achieving an agreement is for the parties mutually to agree upon a contractual objective, which we will call the purpose or "ends" of their collaboration. But reaching agreement on a shared intention to realize a particular contractual objective is only the first step to designing an enforceable contract. Thereafter, the parties must choose the specific

* Robert E. Scott Distinguished Professor of Law and Professor of Philosophy, University of Virginia

** Alfred McCormack Professor of Law and Director, Center for Contract and Economic Organization, Columbia University

¹For discussion see Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 Yale L. J. 541 (2003).

terms that will serve as their *contractual means* for achieving the shared end they hope to realize. In order to bind themselves to a particular set of contractual obligations as their chosen means, parties must rely on contract doctrines to control and predict the interpretation and enforcement of their agreements. Thus, contract doctrines are essential tools for contract design. They enable parties to cope with the barriers that prevent them from writing “completely contingent contracts”—contracts that specify unambiguously each parties’ performance obligations in all possible future states of the world.

Parties face two obstacles to writing complete contracts: front end transaction costs and back end enforcement costs. Collectively, these comprise the costs of contracting. The first axiom of a theory of contract design is that contracting parties are motivated to select those contracting mechanisms that will best achieve their objectives at the least cost. Parties economize on contracting costs in two distinct ways: by choosing between formal and informal means of enforcement,² and by shifting costs between the front end and the back end of the contracting process when drafting formally enforceable agreements.³ Consider first the tradeoff between formal and informal enforcement. In experimental settings, formal enforcement has been shown to undermine informal norms. This research has demonstrated that, in many instances, informal mechanisms operate as complements for each other but as substitutes for formal enforcement.⁴ Moreover, the informal mechanisms that would be displaced by a system

²The literature has long recognized two complementary perspectives on enforcement. On the one hand, private lawsuits brought by parties dissatisfied with a contract or its performance encourage compliance with those bargains that satisfy the prerequisites for legal enforcement. These lawsuits have salience because their outcomes produce meaningful consequences: The loser faces a judgment for damages that fairly automatically leads to a payment of money, and perhaps other sanctions that also have real purchase. On the other hand, parties face an array of informal sanctions as well if they fail to honor their commitments. In many cases, especially the open-ended arrangements called relational contracts, informal sanctions may do much of the enforcement work. See ROBERT E.; SCOTT & PAUL B. STEPHAN, *THE LIMITS OF LEVIATHAN: CONTRACT THEORY AND THE ENFORCEMENT OF INTERNATIONAL LAW* (2006); Robert E. Scott, *A Theory of Self-Enforcing Indefinite Agreements*, 103 COLUM. L. REV. 1641 (2003); Robert E. Scott, *Conflict and Cooperation in Long-Term Contracts*, 75 CALIF. L. REV. 2000 (1987).

³For discussion see Robert E. Scott & George G. Triantis, *Anticipating Litigation in Contract Design*, 115 Yale L. J. 814, 822-51 (2006).

⁴ Scott, *Self-Enforcing Agreements*, supra note – at --; Robert E. Scott & Paul B. Stephan, *Self-Enforcing International Agreements and The Limits of Coercion*, 2004 WISC. L. REV. 551.

of formal enforcement are in some cases less costly and more effective than the formal alternative. In particular, informal mechanisms for motivating performance of contractual obligations are likely to be optimal whenever key information is *observable* but not easily *verifiable*: The parties themselves can at reasonable cost obtain the information, but the costs of proving to a disinterested third party that the information is correct exceeds the gains from formal enforcement.⁵ In some circumstances, however, formal contract enforcement may be optimal even though it undermines informal enforcement. The complexity of particular transactions may make it difficult for either the parties or casual observers confidently to determine whether or not one party has departed from the agreed course of conduct, but the information necessary to make that decision maybe provable to a court at reasonable cost. In these circumstances, the threat of sanctions by a court might both deter opportunistic behavior and clarify the parties' actions and responsibilities under a legally enforceable contract.

When parties determine to formalize some or all of their contractual relationship, they are motivated to write legally enforceable contracts that further optimize contracting costs. They can do this by shifting costs between the front and back end of the contracting process. Parties do this by choosing between vague and precise terms. When the parties agree to vague terms, such as the obligation to use their best efforts or to behave in a commercially reasonable manner, the subsequent adjudication of contractual disputes concerning best efforts or commercially reasonable behavior will require a court to give concrete and specific meaning to those vague phrases as part of the process of interpretation. Thus, by using vague terms, parties delegate the specification of concrete performance requirements to a court at the time of adjudication (i.e., at the back end of the contracting process). This strategy imposes on the parties the expected costs

⁵There is no sharp dichotomy between verifiable and nonverifiable conditions. As we discuss in detail in Part IIB, the question is whether the benefits of using informal norms to enforce a difficult-to-prove condition (such as the level of effort needed to comply with a contractual commitment) are greater than the alternative of verifying compliance with a less accurate but more easily established proxy for the condition in question. The important point, however, is that informal enforcement mechanisms can take into account conditions that are hard to verify even where formal mechanisms cannot. For example, parties to an agreement often can observe whether one has exercised "best efforts" to perform its obligation, but it will be quite costly to marshal the evidence necessary to demonstrate this fact to a disinterested third party. Where this is true, a move towards formal enforcement can deprive cooperating parties of mechanisms that can promote better compliance at a lower cost.

of litigation over the precise form to be given to these vague terms. However, it saves the parties the costs of specifying precise performance obligations at the time of formation. Moreover, because the court will have the benefit of information that was not available to the parties at the time of formation, it potentially allows the parties to benefit from more efficient performance standards than they could have specified at the time of formation.

Alternatively, when the parties agree to precise terms, such as the obligation to pay for a fixed quantity of goods at a price determined by a specified index, the parties reduce or obviate the need for a court to engage in the interpretive process. Instead, by including precise terms, the parties withdraw authority from courts to determine their particular performance obligations and instead direct them to enforce the obligations that the parties have explicitly specified in advance. This strategy requires the parties to specify performance obligations that rely on mere estimates of the likelihood of various future events rather than the actual knowledge of those events that would be available to a court at a later date. Thus, when parties choose between precise and vague terms, they are, in effect, selecting the means by which they intend to achieve their contractual objective. The choice between precise terms and vague terms reduces to the choice of *who* specifies the particular performance obligations of the parties and *when* those obligations will be specified: the parties at the time of contracting or the court at the time of adjudication. The parties choose between these front- and back-end specification mechanisms by comparing the informational advantage they have at the time of contracting against the hindsight advantage of the court in later litigation.

The particular mix of specific and vague terms in any contract thus constitutes the *contractual means* the parties have chosen to trade off investments in transaction costs at the front-end of the contracting process against enforcement costs at the back-end so as to best achieve their particular *contractual ends* at least cost. This trade off between front-end transaction costs and back-end enforcement costs can also be understood in terms of the analogous problem of the choice between rules and standards in legal regulation. By formulating their agreement in precise terms, the parties reduce their legal obligations to bright line *rules*. By creating a rule that specifies the content of their obligations ex ante, the parties

constrain judicial discretion ex post and thereby reduce their back end verification costs. Conversely, by framing their agreement in vague terms, the parties reduce their legal obligations to broad *standards*. Standards avoid the costs of designing precise terms but they also delegate more discretion to courts ex post and thereby increase their back end enforcement costs. This framework might seem to favor the choice of rules over standards on the grounds that the parties are better informed than are courts about their contractual purposes and have better incentives to pursue them efficiently. Nevertheless, there are good reasons for the parties under certain conditions to let courts determine the content of vague terms ex post. Where there is substantial uncertainty about the future states of the world, parties may save the costs of specifying terms governing low probability events and exploit the superior information available to courts at the time of adjudication. The parties' objective in designing their formal contract, therefore, is to select a combination of rules and standards that optimizes their total costs of contracting.⁶

Unfortunately, courts frequently fail to appreciate the key distinction between the contractual ends of the parties and the particular means they have chosen to achieve those ends. By confusing ends with means, courts often end up frustrating the parties' goal of designing efficient contracts. Errors sometimes occur when courts misunderstand the choice between formal legal enforcement and informal or self-enforcement. For example, the common law indefiniteness doctrine holds that a contract is void for indefiniteness if the parties fail to agree on material terms of the deal. The justification for this rule is that the failure to specify particular terms is evidence of a lack of contractual intent; that is, courts assume that when an agreement is indefinite, the parties have not yet chosen to pursue their contractual objectives through the means of a legally binding contract. Yet courts sometimes ignore the indefiniteness doctrine when faced with evidence that the parties intended to create a binding obligation. But the fact that the parties intend to make a binding commitment does not justify the inference that they wish to have their commitment legally enforced. Rather, the question is what contractual

⁶ Note that the parties goal is not simply to minimize contracting costs. Parties will incur additional contracting costs so long as those costs result in improvements in the parties incentives to maximize the expected surplus from the contract. Thus, the parties objective is to "maximize the incentive bang for the contracting cost buck." See Scott & Triantis, *supra* note — at —.

means were selected to achieve those ends: did the parties deliberately chose to avoid legal enforcement and to rely instead on informal enforcement? Even more frequently, courts are guilty of an analogous error – they ignore the chosen contractual means in formally enforceable contracts. Such a case occurs when courts disregard a precise contract term in a fully integrated written contract because enforcement of the precise term will frustrate the parties’ shared contractual ends. However well meaning this ex post effort to honor the parties’ shared contractual objectives, it undermines the parties’ ex ante capacity to choose the particular contractual means of achieving those objectives. In particular, this undermines the parties’ ability to use precise terms (or rules) as a means for reducing their expected back-end enforcement costs. A rule-based regime can succeed only if the rule administrators apply the rule as written.

In this Article, we argue that contractual intent should be identified not with the ends the parties intended to achieve by entering into their contract, but with the contractual means they intended to use to achieve those ends. At the time the parties form their contract, they commit to a particular contractual strategy for achieving their ends. They also realize that there is a risk that their chosen means will fail to achieve their objectives. When that risk materializes, courts should honor the contracting mechanism the parties themselves have selected, even if that means the ultimate objectives of the parties will not be realized. By interpreting contractual doctrines so as to vindicate the parties’ contractual ends, courts unwittingly undermine the ability of parties to design their own contracts, and thereby undermine both party autonomy and contract efficiency.

The Article proceeds as follows. Parts I and II each present both a theoretical overview of a particular contract design problem and an in-depth case analysis to illustrate how courts err by identifying contractual intent with contractual ends. Part I examines the strategies that inform the choice between formal and informal means of enforcement. Contract doctrines that govern the interpretation of agreements enable parties to partition their agreement into legally enforceable, or formal, components, and legally unenforceable, or informal, components. Courts can unwittingly undermine the efficacy of the parties’ chosen mix of formal and informal

commitments when they admit contextual evidence of understandings that the parties intended to exclude from legal enforcement. Part II examines the strategies of parties designing formally enforceable contracts, in particular the factors informing the choice between precise terms (rules) and vague terms (standards) as means of effecting contractual ends. Here, a misplaced focus on contractual ends can lead courts to ignore careful efforts to use particular means designed to limit judicial discretion. Part III turns to the law of conditions to identify a deep rift within the common law between doctrines requiring strict enforcement of express terms and doctrines permitting or requiring loose enforcement or abrogation of express terms when strict enforcement appears to frustrate the parties' contractual ends. Again using close case analysis, we argue that the maxim that the common law abhors a forfeiture acts as a subversive force undermining otherwise clear doctrines that provide parties with the necessary tools for designing their agreements to optimize their costs of contracting. Part IV concludes.

I. INFORMAL AND FORMAL MEANS OF EFFECTING CONTRACTUAL ENDS

A. Contractual Interpretation and Intent

1. The Function of Doctrines of Interpretation.

The classic doctrines of contractual interpretation, such as the parol evidence and plain meaning rules, are widely understood to provide parties with the means to express their contractual intent. The sole purpose of both doctrines is to give parties some control over the process courts will use to interpret their intent. The parol evidence rule enables parties to control the admissibility of certain kinds of evidence in any future adjudication of disputes over their agreement. There are a number of reasons why parties may wish to exclude evidence of certain understandings reached during the contracting process: Tentative agreements may be abandoned, agents may misrepresent their principal's commitments, or, importantly for our purposes, the parties may deliberately choose to exclude a portion of their agreement from legal enforcement. For example, by choosing to fully integrate their agreement and thereby preclude the introduction of oral and written statements made in the course of negotiations, the parties forfeit the ability to prove understandings they deliberately declined to include in the integrated writing. In return, they gain the ability to rely on self-enforcing mechanisms, such as reputation, the

prospect of future dealings and the propensity to reciprocate. Consistent with the widespread view that the decision to integrate an agreement is entirely a matter of party discretion, courts thus have no presumption for or against a finding that a written agreement is integrated. Instead, they have devised various neutral tests expressly for the purpose of determining whether or not parties intended to integrate their agreement into a final, legally enforceable writing.

Similarly, the best understanding of the plain meaning rule treats it as a device for preserving a reservoir of terms with clear meanings that cannot be contradicted in adjudication by contextual evidence supporting a different meaning. The best rationale for the plain meaning rule is that it makes available a public fund of terms with judicially “proofed” or “protected” meanings on which contractual parties can rely to effectively communicate their commitments to each other and to courts.⁷ The parol evidence and plain meaning rules therefore can be viewed as tools for parties to use, respectively, to restrict and provide the evidence courts will use to interpret that portion of their agreement that they intend to make legally enforceable.

This straightforward account of the interpretation doctrines as tools for parties to use to express their contractual intent is relatively uncontroversial. No one would be surprised to learn that the central interpretation doctrines of contract law can be understood as devices for effectuating contractual intent. But courts often perceive a conflict between the rules governing contractual interpretation and the principle of honoring the expressed intention of the parties. Courts therefore behave as if they often face a dilemma: either maintain fidelity to the language and purpose of the various interpretive doctrines even though this leads to an outcome that conflicts with the parties’ intent, or maintain fidelity to the parties’ intent even though this requires an application of the doctrine that conflicts with its language and purpose.

This dilemma is false. It conflates contractual intent with the parties’ intended contractual objectives. For purposes of interpretation doctrine, contractual intent should be

⁷For discussion, see Charles J. Goetz & Robert E. Scott, *The Limits of Expanded Choice: An Analysis of the Interactions between Express and Implied Contract Terms*, 73 Calif. L. Rev. 261 (1985), and Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 Yale L. J. 541 (2003).

identified not with the ends the parties intended to achieve by entering into their contract, but with the contractual means they intended to use to achieve those ends. Parties rely on courts to maintain fidelity to the doctrines without regard to their effect on the parties' intended contractual objectives in any particular contract. Otherwise, courts undermine the ability of parties to choose and rely on particular means for achieving those objectives. We can best illustrate the particular problems that emerge in the interaction between interpretation and intention--and in particular on the choice between formal and informal means of enforcement--by focusing on a paradigmatic case.

2. Hunt Foods v. Doliner: A Case Study in Interpretation.

Consider the case of *Hunt Foods & Industries v. Doliner*.⁸ Hunt Foods was interested in acquiring the assets of Eastern Can Company. To that end, it entered into extensive negotiations with George Doliner and his family, who collectively owned about 75% of the company's stock. The parties reached agreement on the price to be paid for the stock, but before they could reach agreement on remaining issues, including the form of the acquisition, Doliner asked for a several week recess in negotiations. Fearing that Doliner would use the break to shop the offer to other bidders, Hunt Foods requested an option to buy the Doliner stock. The parties executed a written agreement for the purchase of an option that made no mention of any conditions on the exercise of the option.

When negotiations resumed, the parties could not reach agreement on remaining issues and, following this impasse, Hunt Foods exercised the option according to its written terms. Doliner refused to tender the stock and Hunt Foods moved for summary judgment on its suit for specific performance. Doliner opposed the motion on the grounds that there was an oral understanding between the parties that the option, though unconditional on its face, was only to be used in the event that Doliner solicited an outside offer for the stock. Hunt Foods contended that there was no such condition and that, in any event, evidence of such an oral understanding

⁸ A.D. 41, 270 N.Y.S. 2d 937, *aff'd*, 272 N.Y.S. 2d 686 (1966).

was barred by the parol evidence rule. Because the underlying transaction was a purported contract for the sale of securities, Article 8 of the UCC applied and, by extension, the Code's parol evidence rule as embodied in UCC § 2-202.

a. The court's analysis. The court denied Hunt Foods' motion for summary judgment on the grounds that the alleged oral understanding was not barred by the Code's parol evidence rule. Section 2-202 provides that terms in a writing intended by both parties as final with respect to the agreed upon terms "may not be *contradicted* by evidence of any prior agreement or of a contemporaneous oral agreement, but may be explained or supplemented... (b) by evidence of *consistent* additional terms..."⁹ Such additional terms are admissible unless the court finds, in addition, that the writing is fully integrated (perhaps by a merger clause) and thus is not only final as to the agreed terms but exclusive as well.¹⁰

The court's analysis focused on the question of whether the proffered evidence was of a "consistent additional term."¹¹ The proffered condition was "clearly [an]'additional'" term because it was not set out in the writing. The question was whether the condition "contradicted" a term in the written option and whether the condition was "consistent" with the writing. Answering this question requires a court to interpret of the meaning of the statutory language. Is a term "consistent" as that word is used in §2-202(b) as long as it does not "contradict" a term in the writing? Or, is the requirement that the term be "consistent" an *independent and additional* requirement for admissibility? In other words, are the terms "may not contradict" and "consistent" synonymous or do they impose two independent tests for admissibility? The latter

⁹UCC § 2-202 (2003)[emphasis added].

¹⁰Id.

¹¹The question of whether the proffered evidence is consistent with the writing does not arise until a court has first determined that the agreement is partially integrated. The court in *Hunt Foods* never expressly found the agreement to be partially integrated. Instead, the court claimed that, under UCC § 2-202, com. 3, the parties would not have certainly included the condition in their writing, and therefore implicitly concludes the writing is not fully integrated with respect to the condition. Ideally, the *Hunt Foods* court would first have found expressly that the parties intended the written option term as their final expression of that term before turning to the question of whether the condition contradicts that term. If the court did not believe the agreement was partially integrated, then there would be no need to exclude the evidence of the condition even if it did contradict the written option term.

interpretation – that an admissible additional term must both not contradict another term in the writing *and* be consistent with the terms in the writing-- is suggested by the use of different words (with presumably different meanings) in §2-202. This interpretation would suggest that a term is “inconsistent” where it is not in “reasonable harmony” with a term of the writing even though it did not directly contradict any of the written terms.¹²

The court chose to interpret “may not contradict” and “consistent” as imposing only a single requirement for admitting the proffered additional term into evidence. To justify this conclusion, the *Hunt Foods* court introduced a novel interpretive principle: “To be inconsistent the term must contradict or negate a term of the writing [A] term or condition which has a lesser effect is provable.”¹³ In other words, even if the appropriate interpretation of the written option term is that it was unconditional, evidence that the parties agreed the option could be exercised only on condition that Doliner shopped Hunt Foods offer would not contradict the written option term because it merely “lessened” the effect of the option. Presumably, the *Hunt Foods* court would only bar evidence of a term that literally negated the written option (i.e., evidence that the parties agreed that Hunt Foods would have no option to purchase the Doliner stock).

The court never addressed explicitly the question of whether the option was an exclusive statement of the terms of the agreement and thus whether evidence of even consistent additional terms would be barred. But at least implicitly the court concluded that the writing was only a partial but not a complete integration. The court cited comment 3 to §2-202 which sets out a test for complete integration: “if the additional terms are such that, if agreed upon, they would certainly have been included in the document...then evidence of their making must be kept from the trier of fact.”¹⁴ According to the court, this standard requires only that it is not impossible,

¹²This “two step” interpretation of the Code’s test for admissibility under §2-202 was adopted by the court in *Snyder v. Herbert Greenbaum & Assoc.*, 38 Md. App. 144, 380 A.2d 618 (1977).

¹³270 N.Y.S. 2d at —.

¹⁴UCC §2-202, Official Comment 3.

even though implausible, that the parties agreed to the term but did not include it in their writing. Thus, the court concluded that the parties would not “certainly” have included the condition in the writing had they agreed to it. And, therefore, evidence of the additional term was admissible.

b. Contractual intent trumps the rules of interpretation. The interpretive gloss that *Hunt Foods* adds to the Code’s parol evidence rule threatens to undermine the purposes of the doctrine. If the parties intended the written option to be the final expression of the option term, and if it is clear that the proper interpretation of that option is that it is unconditional, then allowing evidence that the option was conditional undermines the point of the parties’ agreement that the written option term is a final expression. Similarly, the court’s interpretation of the Code’s “certain inclusion” test in Comment 3 eviscerates the notion that the terms are both final and exclusive, at least in the absence of a carefully drafted merger clause. In order to exclude evidence of an additional term on grounds that the agreement was fully integrated with respect to that term, the *Hunt Foods* court required a finding that it would have been *impossible*, not merely *implausible*, for the parties to have intended that term to be part of their agreement but failed to put it in their writing. Such a loose interpretive standard precludes a court from protecting the agreement from ex post attacks based on alleged additional terms that significantly change the allocation of contractual benefits and burdens as reflected in the final and exclusive written terms of the agreement.

Why does the *Hunt Foods* court embrace such unappealing analyses of the consistency test and certain inclusion tests? The best inference is that the court believed that the parties’ contractual objectives were inconsistent with the claim that Doliner agreed to grant Hunt Foods an unconditional option. In short, a straightforward application of parol evidence doctrine would frustrate the parties’ intent. After all, Doliner wished to adjourn negotiations without consummating the agreement. Granting an unconditional option would allow Hunt Foods unilaterally to decide to consummate the agreement. The puzzle the court does not address, however, is why Doliner would have agreed to sign an unconditional option that failed to state that it was conditional on Doliner shopping Hunt Foods’ bid. According to the record in the case, Doliner asked his attorney why the option language failed to state the condition. His

attorney responded that Hunt Foods “insisted” that the written option not state any conditions, but that the attorney had “obtained an understanding” that the option was to be used only if Doliner solicited an outside offer. Under these circumstances, why would Doliner comply with Hunt Food’s demand?

The puzzle in *Hunt Foods* can best be solved by focusing attention on the means chosen by the parties to achieve their contractual ends. This, in turn, requires an appreciation of the tools of contract design, particularly the incentives parties have to choose between formal and informal means of enforcing their respective commitments.

B. Formal and Informal Mechanisms in Contract Design.

How do parties –such as Hunt Foods and George Doliner – design commercial contracts so as to best achieve their joint objectives? Surprisingly little theoretical scholarship has focused on systematically answering this question. One reason this is so is that contract design is not merely a matter of calculating the costs and benefits of negotiating particular terms in a contract. In addition to the “front end” transactions costs of negotiating a contract, parties must also attend to the “back end” problems they may subsequently encounter in verifying and enforcing their respective obligations in court. A focus on the costs of verifying contractual obligations motivates parties to consider their choices between formal and informal means of enforcing their commitments.

1. Mechanisms for self-enforcement. In many instances, an agreement between two commercial parties will be *self-enforcing* in the sense that the desire to protect a good reputation or future business dealings will motivate the promisor to perform even in the absence of legal coercion compelling performance.¹⁵ However, reputation and future business prospects, though important ingredients to successful contracting, have their limits. Consider reputational incentives first. Social esteem and a reputation for keeping one’s word are powerful motivations whenever other potential trading partners are able conveniently to learn why the parties’ deal

¹⁵Scott 1987, 2003; Scott & Stephan 2006.

broke down. Reputations work well, therefore, in small trading communities, especially those with ethnically homogenous members or other cooperation-inducing structures, where everything that happens soon becomes common knowledge and boycotts of bad actors are readily enforced.¹⁶ Reputational sanctions also can be effective in industries that establish trade associations, because the associations can create a collective memory of the contracting behavior of their members.¹⁷ Reputations are difficult to establish, however, in large, heterogeneous economies in which particular contracting parties often are anonymous.

Even in the absence of the conditions for establishing reputations, agreements will be self-enforcing to the extent that the parties anticipate the prospect of future dealings. Specifically, when parties contemplate making a series of contracts, neither party will breach an early contract if the gains from that breach are lower than the expected profits from future contracts that a breach would eliminate. But this incentive has natural limits. Unfortunately, ongoing relationships inevitably come to an end. When parties come to realize that the relationship is soon to terminate (say, when the promisor contemplates retirement or otherwise withdrawing from the trading community), the threat that the other party will no longer deal with the promisor is insufficient in and of itself to induce performance. Thus, all repeated interactions are subject to a familiar end-game problem. Indeed, in the limiting case, the anticipation of the last transaction may cause the entire cooperative pattern to unravel as each party anticipates that the next interaction will be the last interaction and that the promisor might defect.¹⁸

The limitations of these traditional means of self-enforcement do not, by themselves, justify formal legal enforcement of promises. There are strong reasons to believe that powerful norms of reciprocity enhance and extend the reach of reputation and repeat dealings as means of

¹⁶Landa 1981, Grief 1998

¹⁷Bernstein 2001, Bernstein 1996

¹⁸Axelrod 1984, Scott 1987

self-enforcement.¹⁹ Experimental evidence shows that a preference for reciprocity – the willingness to reward cooperation and to punish selfishness-- can motivate cooperation even in arms-length interactions between complete strangers.²⁰ In sum, contracting parties frequently can (and do) turn to informal means of enforcement based on trust, or the desire to maintain a good reputation in the relevant trade or community, or on the prospect of profitable future dealings. In many contexts, reputation, repeat dealings and norms of reciprocity provide the best available means of regulating the contractual relationship so as to reduce the risk of exploitation of the parties' vulnerabilities. If parties are able to rely on these informal methods of enforcing promissory commitments, they can approximate in practice the theoretical goal of writing complete contracts that are sufficiently credible to motivate efficient investments *ex ante* and sufficiently flexible to ensure efficient trade *ex post*.

2. Tradeoffs between Formal and Informal Enforcement.

What role does (and should) formal legal enforcement play in a world where informal enforcement is pervasive and robust? Recall that the experimental evidence suggests that informal enforcement, when it is effective, is both *cheaper* and *better* than formal enforcement. Informal enforcement is *cheaper* because a party only needs to expend costs to observe the other's behavior, while formal enforcement requires the parties to expend *additional* resources (attorneys fees, court costs, etc.,) in verifying that behavior to a court. Second, less obvious perhaps but even more significant is the fact that informal enforcement is also *better*. It permits parties to make credible promises regarding observable but non-verifiable measures of performance, thus achieving contractual objectives that may not be possible with formal

¹⁹Rabin 1991, Levine 1998.

²⁰There are three key findings of a substantial body of experimental evidence on people's propensity to reciprocate. First, many people behave in a reciprocal manner that deviates from purely self-interested behavior. Reciprocity means that individuals respond cooperatively to generous acts, and, conversely, punish non-cooperative behavior. Second, individuals will repay generosity and punish selfishness in interactions with complete strangers and even if doing so is costly for them and yields neither present nor future material rewards. And third, the observed preference for reciprocity is heterogeneous. Some people exhibit reciprocal behavior and others are selfish. Taking all the experiments together – gathered from diverse countries and cultures – the fraction of reciprocally fair subjects ranges from 40 to 60% as does the fraction of subjects who are selfish. Fehr & Schmidt 1999, Fehr, Gächter & Kirchsteiger 1997.

enforcement.²¹

Consider, for example, a buyer who wishes to acquire custom goods of extraordinary quality from a willing seller. The problem confronting the parties is their inability to write a formal contract describing precisely the quality of the contract goods (and the associated price) that would be jointly optimal. This is because the attributes of high quality are not verifiable. If instead the parties write a legally enforceable contract for merchantable goods, the contract will fail to maximize joint welfare. But imagine that the buyer proposes a legally unenforceable bonus agreement as a supplement to a formal contract. Here the buyer promises to pay an unspecified bonus if the seller delivers the highest quality goods satisfactory to the buyer. In other words, the buyer offers to share a portion of the greater contractual surplus with the seller in return for the enhanced effort necessary to produce the specialized high quality goods that maximize the buyer's value. Yet under the common law indefiniteness doctrine, the buyer's promise to give a bonus if satisfied with the additional effort of the seller is not legally enforceable.²²

A preference for reciprocity that causes a party to reward a generous action and retaliate

²¹Scott, *Self-Enforcing Agreements*, supra note 4 at—.

²²*See, e.g.,* Varney v. Ditmars, 217 N.Y. 223, 111 N.E. 822 (1916); Joseph Martin, Jr., Delicatessen, Inc. v. Schumacher, 52 N.Y. 2d 105, 417 N.E. 2d 541 (1981); RESTATEMENT OF CONTRACTS § 32 (1932); RESTATEMENT (SECOND) OF CONTRACTS § 33 (1981); WILLISTON ON CONTRACTS, VOL. I, §§ 37 *et seq.* The indefiniteness doctrine is, however, in tension with the specialized doctrines that apply to the interpretation of conditions. If the condition of buyer's satisfaction is ambiguous between objectively reasonable and merely honest but subjective satisfaction, courts will interpret the condition to require objectively reasonable satisfaction if they believe the objective reasonableness of the buyer's satisfaction is practicable to determine. *See* R2d §228. Promises conditioned on objectively reasonable satisfaction would not be indefinite and so would be enforceable. But even if the agreement explicitly delegates unfettered discretion to the buyer to determine its satisfaction, or the good or service purchased has inherently subjective value, instead of or in addition to its objective value, the doctrine would still regard the condition as definite and enforceable:

Under any interpretation, the exercise of judgment must be in accordance with the duty of good faith and fair dealing (§ 205), and for this reason, the agreement is not illusory (§ 77). If the agreement leaves no doubt that it is only honest satisfaction that is meant and no more, it will be so interpreted, and the condition does not occur if the obligor is honestly, even though unreasonably, dissatisfied. Even so, the dissatisfaction must be with the circumstance and not with the bargain and the mere statement of the obligor that he is not satisfied is not conclusive on the question of his honest satisfaction.” R2d § 228, cmt. a.

against unfair behavior would enhance the performance of the bonus agreement. A fair buyer in this situation will respond to a high effort level from the seller by paying a generous bonus. Moreover, assuming that the fraction of fair types in the general population is consistent with the experimental evidence, the probability of a fair bonus being paid is sufficiently great to motivate the seller (regardless of his type) to expend the extra effort. Thus, if a substantial fraction of the population responds to opportunities to reciprocate, we would predict that the legally unenforceable bonus agreement would actually produce a better result for both parties than would a legally enforceable contract that contained no bonus.

But this hypothetical also shows that informal enforcement has some significant limitations. After all, the reciprocal outcome is not guaranteed – the experiments show only that informal enforcement between strangers works better *on average*. Moreover, common observation tells us that on-going relationships that rely on informal enforcement can break down, and when they do the parties often resort to costly litigation. The reason is that informal enforcement requires *transparency*: Each party must be able to observe and properly characterize the other’s behavior. In complex interactions, a failure to cooperate may not be observable immediately, or a cooperative response may be mistakenly interpreted as a defection from the cooperative norm. Lacking clarity, either party may mischaracterize the other’s actions. Under these circumstances, without the necessary linkage between action and response, reciprocity will be an ineffective mechanism for enforcement. When reciprocity breaks down in complex transactions due to low transparency, courts can serve a valuable function by making factual determinations that unravel complex behaviors. A disinterested adjudicator may be in a better position to sort out complex behavior and can both detect a breach and, by imposing a pre-determined sanction, forestall attempts by the aggrieved party to respond disproportionately.²³

²³ Scott & Stephan 2006. Why must courts have the power to impose sanctions? Arguably, the provision of key information to the parties is all that is necessary to maintain a robust regime of informal enforcement. One answer is that courts depend on the parties themselves to provide the information needed to adjudicate a dispute. Once the parties offer evidence, the fact finder can then *verify* information that each may lack individually. But without a sanction for non-production, no disputant in a conflict situation would have an incentive to provide truthful information to the arbiter that might harm his position. Absent a sanction, a contracting party would be motivated to conceal evidence of any defection that was known only to it.. Id at –.

3. *The “Crowding Out” Phenomenon.*

A central question remains: How do the various means of enforcement interact with each other? Here the available evidence suggests that formal enforcement is often imperialistic: an effort to superimpose formal enforcement on a regime of self-enforcement is likely to displace the informal mechanisms. For example, the experimental evidence shows that, absent a legally enforceable obligation, reciprocity – operating alone – generates high levels of cooperative behavior. But once the entire relationship, including its informal aspects, is subject to formal enforcement, voluntary reciprocity declines along with the overall level of cooperation. These experimental results suggest that formal legal sanctions and informal sanctions based on reciprocity may well conflict with each another. In other words, formal enforcement may “crowd out” behavior based on reciprocity.²⁴

A careful examination of the experimental evidence shows that this crowding out phenomenon is complex. A number of studies have confirmed the crowding-out hypothesis in single-shot interactions between strangers. In single iteration experiments, where the parties must choose either informal or formal enforcement, the choice of formal enforcement uniformly suppresses the evidence of reciprocity that is found in the alternative scenario of no formal enforcement.²⁵ But recent experiments also show that, where there is some probability that the same buyers and sellers will continue transacting in the next period, formal enforcement that is limited *only to the verifiable dimensions of the agreement* actually enhances cooperation in those dimensions of the agreement that are non-verifiable.²⁶

The experimental evidence of crowding out undermines the argument that courts should

²⁴Scott, *Self-Enforcing Agreements*, supra note 4 at —.

²⁵See, e.g., Bruno Frey & Reto Jegan, *Motivation Crowding Theory*, 15 J. ECON. SURVEYS 589 (2001); Bruno Frey & Matthias, *Motivation Transfer Effect*, University of Zurich, Institute for Empirical Research in Economics, (mimeo 2001); Iris Bohnet, Bruno Frey, & Steffen Huck, 2004; Ernst Fehr & Simon Gächter, *supra*.

²⁶Sergio Lazzarini, Gary J. Miller & Todd R. Zenger, *Order with Some Law: Complementarity versus Substitution of Formal and Informal Arrangements*, 20 J. L. ECON. & ORG. 261 (2004).

imply reciprocal “relational” duties in formal contracts.²⁷ For example, assume that changed circumstances materialize after the parties have made the contract and threaten to impose severe losses on one party unless the court grants relief from specific contractual obligations. Scholars have argued that the severity of the consequences of strictly enforcing such a contract according to its precise terms implies that the disadvantaged party did not fully consent to the risk of the losses that enforcement would cause. In this situation, they maintain, courts should create broad contractual obligations – such as the obligation to adjust the contract terms so as to treat one’s contracting partner “fairly.” But this understandable instinct to relieve the burden on a disadvantaged party may well prove counter-productive in the long run. The evidence suggests that an attempt to extend formal enforcement to non-verifiable contract terms – such as imposing an obligation to agree to an equitable adjustment of precise terms – is likely to impair the efficacy of those informal means of enforcement that rely on reciprocity norms.

Significantly, common law contract doctrine has resisted this invitation to imply broad standards of fairness or equitable adjustment. Common law doctrine makes contractual liability hard to assume and hard to escape once assumed. If a promise falls within the scope of legal enforcement, contract law fills only a few gaps and uses simple, verifiable default rules when it does.²⁸ The evidence that there are informal means of enforcing commitments that courts cannot readily verify supports this approach. The more general lesson for courts, therefore, is that an effort to judicialize preferences for fairness and reciprocity may well destroy the very informality that makes these mechanisms so effective in the first instance.²⁹

²⁷See e.g., Robert Hillman, *Court Adjustment of Long-Term Contracts: An Analysis Under Modern Contract Law*, 1987 DUKE L. J. 1.

²⁸The common law’s approach stands in sharp contrast to the much more expansive approach of contemporary commercial statutes such as Article 2 (Sales) of the Uniform Commercial Code and the Convention on the International Sales of Goods (CISG). These statutes invite courts to modify express agreements in light of postcontractual information subject to soft, generally unverifiable, standards. For discussion, see Robert E. Scott, *The Death of Contract Law*, 54 U. TORONTO L. J. 369 (2004); Robert E. Scott, *The Case for Formalism in Relational Contracts*, 94 NW. U. L. REV. 847 (2000); Robert E. Scott, *A Relational Theory of Default Rules for Commercial Contracts*, 19 J. LEGAL STUD. 597, 615 (1990).

²⁹It is important to keep in mind that legal disputes only arise when the informal modes of enforcement have broken down. Litigated cases, therefore, give us no clue of the power of reciprocity, reputation and other informal mechanisms in enforcing the many agreements between parties to relational contracts that never reach litigation.

C. A Contract Design Explanation of the Hunt Foods/Doliner Contract

1. Reconstructing the Parties' Motivations.

Understanding the relationship between formal and informal enforcement provides the key to resolving the puzzle posed above concerning the Hunt Foods contract: Why did Doliner agree, on the advice of counsel, to sign an unconditional option? One can imagine two scenarios in which Doliner valued a break in the negotiations. In the good faith scenario, he was not yet satisfied with all the elements of the agreement, anticipated he would be able to work them out in negotiations, but had a pressing demand that required him to adjourn negotiations for several weeks. In the bad faith scenario, Doliner wanted the adjournment because he planned to shop Hunt Foods' bid. Either way, Hunt Foods would be understandably worried that Doliner might shop their bid, so it insisted on an option to protect it in the event Doliner did so. But Hunt Foods understood that an option conditional on proving that Doliner shopped their bid would be of little value in court because it would have great difficulty carrying the burden of proof. Doliner could simply reconvene negotiations, terminate negotiations on ostensibly independent grounds, and then consummate the deal with another bidder to whom he had secretly shopped Hunt Foods' bid during the adjournment. Even if Hunt Foods officers believed they could *observe* Doliner's bad faith (e.g., they could find out through sources if he shopped the bid), they knew they could not readily verify his bid-shopping to a court.

In short, the costs of verification precluded the choice of including an explicit bid shopping condition in the written option held by Hunt Foods. But the parties had an alternative: Hunt Foods' commitment not to exercise the option unless Doliner solicited another bid was credible because it was subject to powerful informal sanctions. Not only was the reputation of Hunt Foods jeopardized if it behaved opportunistically in acquisition negotiations, but more important, counsel for both sides could bond the understanding with their own reputations. Since attorneys in merger and acquisition deals are drawn from homogeneous communities, their respective commitments are credible and they can "lend" their reputations to enhance the credibility of their principal's commitments.

2. Any Condition Was Self-Enforcing and Not Legally Enforceable.

The preceding reconstruction of motivations is consistent with the facts as drawn from the record of the case. One of two things occurred. Either Hunt Foods never agreed to a condition and Doliner's testimony was untruthful, or, as he testified, the parties agreed *informally* not to exercise their option unless Doliner shopped their bid. But they insisted *formally* on a legally unrestricted option. Doliner would agree to these terms if either (1) he had no intention of shopping their bid (i.e., he was acting in good faith) and trusted Hunt Foods to abide by their informal agreement (because reputational incentives and natural inclination would lead them to do so), or (2) he planned to shop the bid and then render the option useless by using precisely the gambit in evidence in this litigation: claiming that the option was conditional and requiring Hunt Foods to prove the condition was met (knowing they would be unable to bear that burden of proof). Hunt Foods then exercised the option either because Doliner did, in fact, shop the bid (they observed that fact even if they could not verify it to a court) or because its officers acted in bad faith by taking advantage of their legally unrestricted option and breaching their informal agreement.

Regardless of which scenario better conforms to the underlying facts, the court would have been wise to find the written option fully integrated and therefore to rule inadmissible the evidence of an oral condition on the option. Given that Doliner knowingly signed the written option without requiring the condition to be stated in the writing, the strongest inference is that the contractual means chosen by the parties was to regulate their relationship with a dual regime of formal and informal norms. The difficulty of proving bid-shopping forced Doliner to grant an unrestricted option to Hunt Foods (as its demand made clear) or lose the bid entirely (or forgo adjournment and continue negotiations). Hunt Foods made clear that the price for adjournment was an unrestricted, written option (why else would they refuse to put the restriction in writing and why else would Doliner agree?).

Thus, the court could have reconciled enforcement of the unrestricted written option with the belief that the parties contractual objective (their intended end) was to grant only a conditional option. Given the limits of proof and the availability of informal enforcement,

Doliner would have quite reasonably chosen to effect those ends by granting Hunt Foods a legally unconditional option as the price for adjourning negotiations without losing the offer. On this account, Doliner chose to rely on informal means to enforce the agreement that the option was to be exercised only if he shopped the offer. Nonetheless, self-enforcing mechanisms occasionally break down. Despite this fact, the evidence is overwhelming that contracting parties prefer to retain the option to enforce their commitments through two sets of rules: an explicit (and rigid) set of rules for those parts of their relationship that require legal enforcement and an implicit (and flexible) set of rules for those aspects that respond best to self-enforcement. By correctly identifying the parties' objectives, but failing to understand why those objectives could be achieved only through a particular means, the court in *Hunt Foods* undermined the ability of parties in the future to regulate their relationship with a combination of formal and informal norms.

II. RULES AND STANDARDS AS MEANS TO EFFECT CONTRACTUAL ENDS

A. *Contractual Intent and the Law of Mistake and Excuse.*

In the preceding Part, we examined the tradeoffs between formal and informal enforcement that parties confront when designing their contracts. In this Part, we assume that parties have determined to formalize their commitments in a legally enforceable agreement. We begin our analysis of the trade offs in the design of formal contracts by focusing on the contract dispute in a particular case—*Aluminum Co. of America v. Essex Group, Inc.*³⁰ The *ALCOA* case is a paradigmatic illustration of how a focused search to recover the intention of the contracting parties can go awry if courts are not attentive to the nature of contract design and the structure of contractual intent.

1. *The ALCOA/Essex Contract.*

ALCOA and Essex entered into a long term tolling contract whereby ALCOA undertook

³⁰499 F. Supp. 53 (1980).

to convert alumina supplied by Essex into aluminum.³¹ The agreement was to last for sixteen years with Essex having an option to renew for five years, and it specified a fixed quantity of 75 million pounds per year. Along with its obligation to take 75 million pounds per year, Essex had the concomitant obligation to deliver sufficient alumina for ALCOA to convert into aluminum, which Essex could then dispose of as it pleased.

The contract contained a detailed price indexing provision. The initial contract price was 15¢/pound to be adjusted according to a complex formula: Five cents of the charge were designated as a “demand charge.” This reflected ALCOA’s capital costs for the smelting capacity at its Warrick, Indiana plant. The demand charge was then multiplied by a demand charge index consisting of the arithmetic average for the previous six months of the Engineering News Record Construction Cost-20 Cities Average Index, as published in the Engineering News Record. In addition, Essex agreed to pay a ten cent production charge of which four cents were fixed. Of the remaining variable costs, three cents corresponded to the labor cost and three cents reflected the non-labor production cost. The former was indexed by a labor index, the average hourly labor cost at ALCOA’s Warrick Works as determined by ALCOA. The latter was indexed by the Wholesale Price Index, Industrial Commodities (WPI-IC). At Essex’s insistence, the parties agreed that the contract price would be capped at 65% of the price of a standard grade of aluminum ingot as published in American Metal Market, a trade publication. Although ALCOA did not choose to similarly protect itself against the possibility that the WPI-IC would fail to track costs accurately, the contract did reflect the risk that the indexes would no longer be available.³² In short, at Essex’s insistence, the index related to ALCOA’s costs included a circuit

³¹The contract between ALCOA and Essex is analyzed in detail in VICTOR GOLDBERG, FRAMING CONTRACT LAW; AN ECONOMIC PERSPECTIVE 348 (2007).

³²Clause 28 of the ALCOA/ Essex contract provides:

In the event that any index referred to in paragraph 8 and 9 hereof is discontinued, no longer published in the sources indicated or become unavailable, the parties shall agree on a comparable substitute index and if they are unable to agree the selection of a substitute index, the selection of a comparable substitute index shall be submitted to arbitration in accordance with the rules of the American Arbitration Association. In the event that the price of unalloyed primary aluminum ingot referred to in paragraph 9 hereof is no longer published in the American Metal Market, Alcoa’s published price for such product shall be used for purposes of subparagraph (c)(5) of paragraph 9 hereof.

breaker if the index rose too fast relative to the underlying market price, but ALCOA did not require a corresponding “booster” if the index moved too slowly. Thus, the index had a ceiling but not a floor.

The contractual purposes for the index were clear from ALCOA’s perspective: The five cent demand charge represented four cents for financial costs (depreciation, interest, etc.) and one cent “guaranteed profit.” The fixed portion of the production charge included three cents profit and one cent to cover general administrative and selling expenses. Of the variable costs, three cents represented labor costs and three cents reflected material costs, about 75% of which was for power. As Victor Goldberg suggests, the contract was thus designed “to replicate a situation in which Essex invested in a smelter with a capital cost of forty cents per pound with a ten percent (four cent) return. The demand charge would be paid by Essex regardless of whether it took any aluminum, so in that sense it was taking the risk of ownership.”³³

2. The Problem.

The contract performed satisfactorily until 1973. Unfortunately for ALCOA, the index moved too slowly relative to the actual market owing, in part, to the under-representation of energy costs in the basket of inputs that comprise the WPI-IC relative to the costs of converting alumina into aluminum. More specifically, three things happened which caused the contract price to be out of line with both ALCOA’s costs and the price of aluminum. The OPEC oil embargo caused coal prices to rise far more rapidly than the WPI-IC, so the non-labor variable production cost component failed to track the actual costs. In particular, ALCOA’s non-labor

³³GOLDBERG, FRAMING CONTRACT LAW, supra at —. The contract also included a “most favored nation” provision in clause 29:

If, at any time during the term of this Agreement, Alcoa enters into an agreement with any other party, who competes with Essex in the United States in the sale of aluminum electrical conductor, other than any subsidiary or affiliate of Alcoa, whereby Alcoa agrees to toll convert into Aluminum smelting grade alumina of substantially the same specifications as Alumina to be toll converted hereunder on substantially the same terms and conditions as provided in this Agreement, but for toll conversion charges and/or payment terms more favorable than those afforded Essex in this Agreement, such more favorable charges and/or payment terms shall become a part of this Agreement effective as of the next succeeding calendar month after such agreement was entered into by Alcoa with such other party but shall be a part of this Agreement only so long as such more favorable charges and/or payment terms remain in effect in such other agreement. [cl 29]

production costs rose from 5.8 cents in 1973 to 22.7 cents in 1978, while the WPI-IC less than doubled. In addition, the 1970's experienced extraordinary rates of inflation, so that the nominal price for aluminum rose significantly. Finally, there was an increased demand for aluminum and thus the underlying or “real” price of aluminum rose as well.³⁴

By 1979, the per pound market price of aluminum ingot was around 73 cents while ALCOA's costs were around 35 cents and the indexed contract price was at 25 cents. Since Essex's costs for the alumina and for transportation to the Warrick smelter were about eleven cents, it could capture a profit of 37 cents per pound if it chose to resell the ingot on the market. And, in fact, ALCOA claimed that Essex had exploited this favorable structure and had resold 23% of the aluminum on the open market.

3. The Litigation in ALCOA v. Essex.

ALCOA sought to renegotiate the contract, but Essex refused. ALCOA then brought suit asking for equitable reformation of the contract. The crucial issue for the court was the rapid increase in the cost of power at the Warrick smelter and the failure of the index to capture that increase. The decision hinged on one factual finding: the failure of the price index to track costs. The court granted ALCOA relief on the grounds of mutual mistake, commercial impracticability and frustration of purpose.³⁵ Rather than excuse ALCOA, however, the court chose to reform the contract by rewriting the price term. The price, recalculated quarterly, would be the lesser of the 65% cap and a second price: either the price as defined in the contract or ALCOA's costs plus one cent whichever was greater.³⁶

³⁴The price term also failed to index properly the fixed demand charge representing Alcoa's capital investment in the smelter. See GOLDBERG, FRAMING CONTRACT LAW, *supra* at —.

³⁵499 F. Supp. at —. The court based its analysis of the three relevant doctrines on the relevant provisions in the Restatement (Second) of Contracts and the Uniform Commercial Code. See, e.g., Restatement § 151-154 (mistake); § 271 (discharge by supervening impracticability), and §265 (discharge by supervening frustration); and UCC § 2-615 (excuse by failure of presupposed conditions).

³⁶Ultimately, that would have meant that in 1979 Alcoa would have received about 11 cents per pound more than it would have absent the modification. See GOLDBERG, FRAMING CONTRACT LAW, *supra* at —.

In its mutual mistake analysis, the court concluded that there was a mistake as to a fact—the belief that the index would work as the parties expected. It also found that this mistake was mutual because mutuality is a question of understanding, not motivation (it thus rejected Essex’s argument that the mistake was not mutual since Essex was not concerned about keeping ALCOA’s cost in check). The court emphasized the strict test required to establish excuse on the basis of mutual mistake: the mistake doctrine applies to a mistaken belief about a particular fact only if, as Corbin argued, a court decides “the parties made a definite assumption that it existed and made their agreement in the belief that *there was no risk with respect to it.*”³⁷ According to the court, the alleged mutual mistake in *ALCOA* was the parties’ belief in “the suitability of the WPI-IC as an index to accomplish the purposes of the parties.” The court found that “each [party] assumed the Index was adequate to fulfill its purpose. This mistaken assumption was essentially a present actuarial error.”³⁸

Because the court found the doctrine of mutual mistake applicable to the case, it would seem to follow that it believed the parties “made their agreement in the belief that there was no risk” that the price index would fail to reflect ALCOA’s actual costs. This is precisely the claim

³⁷[cite Corbin]

³⁸499 F. Supp. at —. The requirement that the mistaken belief concern a fact that exists at the time of the agreement serves to distinguish the law of mistake from the law of excuse. While mistake doctrine voids agreements based on mistaken assumptions about matters of fact that exist at the time of formation-- for example, the fertility of a cow or the authenticity of a painting-- it is designed not to apply to cases in which parties make a mistaken predictions about the future, such as an erroneous assumptions about future market conditions or their future financial situations. “A mistake of both parties does not make the contract voidable unless it is one as to a basic assumption on which both parties made the contract. . . . Market conditions and the financial situation of the parties are ordinarily not such assumptions, and, generally, just as shifts in market conditions or financial ability do not effect discharge under the rules governing impracticability, mistakes as to market conditions or financial ability do not justify avoidance under the rules governing mistake.” R2d § 152, cmt. b. The *ALCOA* court therefore could use the mistake doctrine as a ground for reformation only by characterizing the parties’ mistake about the future functioning of the price index as a mistake concerning a fact in existence at the time of formation. Thus, the court characterized the parties’ mistake as a “present actuarial error” to shoehorn a mistake about the future into a present mistake to which the mistake doctrine would apply. Of course, if all mistaken beliefs about future events could be transformed into mistakes about present facts simply by re-describing them as “present actuarial errors,” there would no longer be a difference between mistake and excuse doctrine.

Essex denied.³⁹ The court’s response was to demonstrate that the parties “plainly sought to limit the risks of their undertaking.” Chief among the pieces of evidence cited to demonstrate the parties’ efforts to limit risk under their agreement was the fact that ALCOA hired Dr. Alan Greenspan to advise it on the drafting of the objective price index.⁴⁰ As to why ALCOA did not include a price floor, just as Essex included a price ceiling, the court responded: “the absence of an express floor limitation can only be understood to imply that the parties deemed the risk too remote and their meaning too clear to trifle with additional negotiation and drafting.”⁴¹ The court argued that “the proper question is not simply whether the parties to a contract were conscious of uncertainty with respect to a vital fact, but whether they believed that uncertainty was effectively limited within a designated range so that they would deem outcomes beyond that range to be highly unlikely. . . . Both consciously undertook a closely calculated risk rather than a limitless one.”⁴² Thus, the court moved from Corbin’s requirement that the parties believe there is *no risk* that they are wrong about their shared factual belief, to a more liberal requirement that the parties believe it is *highly unlikely* that their shared factual belief is wrong.

But when the parties are conscious of a particular risk, and yet form a belief that the risk has been reduced or eliminated, it is no longer plausible to argue that it did not occur to them that their belief might be wrong. In such a case, it is far more difficult to conclude that the parties did not intend to allocate this risk. Nevertheless, the court argued that the parties’ demonstrable efforts to reduce the risk showed that they intended not to allocate the reduced risk. In its view, the parties considered that risk so remote that it was not worth allocating. This explains and justifies ALCOA’s failure to bother with drafting a floor to protect it against the remote risk that

³⁹Id. At —. As the court noted, “Essex first asserts that ALCOA expressly or implicitly assumed the risk that the WPI-IC would not track ALCOA’s non-labor production costs. Essex asserts that ALCOA drafted the index provision . . . and that ALCOA’s officials knew of the inherent risk that the index would not reflect cost changes.” Id.

⁴⁰Alan’ Greenspan’s actual role in developing the price term is a matter of some dispute. In its brief to the 3d Circuit, Essex claimed that “Contrary to the trial court’s finding, George [Alcoa’s key negotiator] testified that Alan Greenspan . . . was not consulted by Alcoa in connection with the Contract.” Goldberg at —.

⁴¹499 F. Supp. at —.

⁴²Id. At ----.

their price index would not function as it anticipated. Thus, the court concluded, ALCOA did not assume the risk through any of the four ways one can do so.⁴³

The court's conclusion that the risk was not allocated to ALCOA is certainly possible, but it is highly improbable. Even if both parties – perhaps because of Greenspan's role in drafting the index – believed the probability that the price index would malfunction was extremely low, it does not follow that ALCOA, let alone Essex, understood the contract to be subject to the condition that the index function as the parties anticipated it would. Indeed, if both parties believed the contract was, in effect, conditional on the price index functioning to limit the range of price variation to no more than three cents per pound, it is difficult to explain why Essex insisted on a price ceiling (or circuit breaker) on the index. And given that Essex did include a price ceiling, it is difficult to explain how ALCOA could reasonably conclude that Essex believed the contract was conditional on the price index working as predicted.⁴⁴

⁴³Restatement (Second) § 154 provides that:

A party bears the risk of a mistake when

(a) the risk is allocated to him by agreement of the parties, or

(b) he is aware, at the time the contract is made, that he has only limited knowledge with respect to the facts to which the mistake relates but treats his limited knowledge as sufficient, or

(c) the risk is allocated to him by the court on the ground that it is reasonable in the circumstances to do so.

In applying this provision to the facts of the case, the court first rejected the argument that, by not including a floor in the price term, ALCOA expressly or implicitly agreed to bear the risk that the escalator might rise too slowly. Rather, as we note above, the court found that ALCOA considered the possibility that a floor might be necessary too remote to put into the contract. The court also rejected Essex's argument that the contract should be interpreted against its drafter (ALCOA) because, according to the court, that interpretive maxim is only appropriate when there is ambiguity or a policy concern, neither of which were present in the case. As to whether ALCOA assumed the risk by proceeding in the face of conscious ignorance of the risk, the court said that the test is not whether they were aware at some level that uncertainty existed, but rather whether they believed the uncertainty was limited (which ALCOA did). And, the court found that there was no reason for the court to allocate the risk to ALCOA, nor did customary dealing call for ALCOA to assume the risk.

⁴⁴The court also considered impracticability and frustration, which have the same basic doctrinal requirements as mistake, but focus on the hardship imposed on the plaintiff. It found for ALCOA on both of these grounds based on its \$60 million expected loss resulting from the failure of the index. The court applied its findings on the mistake issue to the impracticability issue, finding that the non-occurrence of the event (and the non-existence of the fact) that caused the impracticability was a basic assumption on which the contract was made. Also, the court found there was no assumption of risk by ALCOA. The difficulty envisioned in the impracticability doctrine, according to the court, must be extreme and unreasonable; a change in degree is not sufficient unless it is well beyond the normal range. In this case, the loss was so significant in absolute size and proportion to the value of the contract (\$60 million) that it altered the essential nature of performance (it was beyond the limit on the risk of market change that was part of the purpose of the contract).

4. The Ex Post Justification: Honoring the Parties Intent.

Critical to the court's analysis on all of the doctrinal grounds used to justify reformation was its focus on contractual intention. The price index, the court reasoned, was designed to reduce, if not eliminate, the deviation between contract price and market price. Thus, this contract was unlike a standard fixed price contract where the risk of price increases is allocated to the seller and the risk of price declines assigned to the buyer. Here the objective of the parties was to "avoid the full risk of future economic changes" by using the complex price indexing mechanism. In this way, the intent was to limit each partner's risk in a particular way: to provide Essex "an objective pricing formula" and to give ALCOA "a formula which would cover its out of pocket costs over the years and yield a return of around four cents a pound."⁴⁵ As a consequence, the failure of the price index to achieve the contractual ends sought by this contract justified invocation of each of the three doctrines used by the court to relieve ALCOA from full liability under the contract. In short, the failure of the price index to achieve the parties' objectives was: a) a mistake of both parties as to a basic assumption of the contract justifying rescission on the grounds of mutual mistake,⁴⁶ b) the occurrence of an event whose non-occurrence was a basic assumption of the contract justifying excuse on the grounds of commercial impracticability,⁴⁷ and c) the occurrence of an event whose non-occurrence was a basic assumption thus frustrating Alcoa's principal purposes under the contract.⁴⁸

Viewed ex post, the decision in *ALCOA* may seem justifiable. Indeed, Victor Goldberg who offers much criticism of the transactional lawyers who wrote the *ALCOA/Essex* contract and the litigators who argued the case, nonetheless concludes that "the judge imposed a cost-plus gloss on the contract which was not explicit in the contract but most likely comported with the

⁴⁵Id. at —.

⁴⁶Restatement (Second) of Contracts § 152.

⁴⁷Restatement § 261.

⁴⁸Restatement § 265.

parties' intentions."⁴⁹ But this understanding of the parties' intentions is flawed. While the court may well have recognized accurately the parties' objectives, it failed to appreciate the doctrinal mechanisms that serve as essential means to achieving those objectives. This point can best be appreciated by turning the focus of analysis away from the ex post litigation perspective assumed by the court to the ex ante perspective of the parties charged with designing contracts. From the vantage point of contract design, the *ALCOA* decision takes on a much different cast.

B. Rules and Standards in Contract Design.⁵⁰

Contract theorists who analyze legally enforceable contracts tend to focus their attention either on the front end transactions cost obstacles to writing efficient contracts or on the back end verification obstacles. This focus misses the fact that parties can (and do) shift costs between the front and back end by their choice in formulating either specific terms or vague terms.

1. The Trade Offs Between Front End and Back End Contracting Costs. The tradeoff between front-end transaction costs and back-end enforcement costs can best be understood in terms of the analogous problem of the choice between rules and standards in legal regulation. Thus, we can frame the choice between specific terms (*rules*) and vague terms (*standards*) as the decision to give content to legal obligations either on the front-end or back-end of the contracting process. The former are designed to define ex ante the chosen means to achieve the parties contractual ends while the latter delegate broad discretion to a court to choose the best means of achieving those objectives ex post. In making this choice, the parties trade off the direct and error costs of ex ante contracting against the direct and error costs of ex post delegation. The particular mix of specific and vague terms yields a combination of front end and back end contracting costs that determines the total costs of contracting. Since the parties' goal is to optimize contracting costs (i.e., to get the biggest "bang for the buck"), the ex post judicial

⁴⁹GOLDBERG, FRAMING CONTRACT LAW, supra at 369.

⁵⁰The following analysis of the elements of contract design draws on the discussion in Robert E. Scott & George G. Triantis, *Anticipating Litigation in Contract Design*, 115 Yale L. J. 814, 822-51 (2006).

system for enforcing contracts necessarily has significant feedback effects on the strategies parties use in writing contract terms ex ante.

To appreciate this point, we might begin by examining the determinants of front-end transaction costs and back-end enforcement costs respectively. Front-end transaction costs are relatively straightforward: The parties invest in foreseeing future contingencies, determining the obligations for each contingency, bargaining over the share of the surplus, and drafting the contract language that communicates their intent to the court. These costs are constrained by the parties shared incentives to optimize contracting costs.⁵¹ Back end enforcement costs include the direct costs of litigation and the effect of uncertainty and the risk of legal error. Here the parties' incentives are no longer aligned and other forces affect these costs. Nonetheless, back end enforcement costs are also constrained by a number of factors: Proof in the adversarial system is relative, and is based on the evidence presented by each party. Moreover, a court reaches a judgment even if the information provided by the parties is poor and litigation costs are confined by judges' incentives as well as by burdens of proof and presumptions.⁵² Thus, there is no a priori reason to believe that either ex ante transaction costs or ex post enforcement costs will preclude parties from using any particular type of contract term. Rather, their choice will be influenced by the relative costs they anticipate in effecting their contractual purpose.

2. Factors Determining the Choice Between Precise and Vague Contract Terms.

Parties shift costs between the front and back of the contracting process depending on exogenous factors such as the degree of uncertainty and the relative costs of writing the contract. As suggested above, they do this by changing the character of the terms in the contract. To see

⁵¹Charles J. Goetz & Robert E. Scott, *Principles of Relational Contracts*, 67 Va. L. Rev. 1089, 1092-95 (1981).

⁵²Among the constraints on litigation costs are that: a) judges have incentives to confine costs as do the parties; b) the fact-finding process in litigation is governed by burdens of proof and presumptions that serve to abbreviate trials; and, finally, c) truth is easier to prove than lies; in many cases, the truth is revealed simply by the decision of the person alleging the alternative story to avoid a costly litigation that she will not win. In sum, direct verification costs are likely to be significantly smaller than contract theorists appear to assume. See Scott & Triantis, *Anticipating Litigation*, supra at —.

how, recall that courts do not directly observe the effects of a contingency – such as the oil shock in *ALCOA* - on which a contract is implicitly conditioned. Rather, they rely on evidence or *proxies*. Consider a simple example of how these evidentiary proxies work. Units of evidence (say, that the oil shock caused carbon-based energy costs to increase by 40%) prove directly whether a precise contract term (or *rule*) is satisfied (e.g., “the contract price is subject to adjustment if ALCOA’s energy costs increase by more than 35%”). Now consider a vague term (or contractual *standard*) such as, “the contract price is subject to adjustment if there are extraordinary increases in production costs caused by extrinsic factors beyond ALCOA’s control.” Vague terms are one step further removed from the evidentiary units: the evidence establishes whether a proxy is satisfied (i.e. it establishes that the oil shock caused the cost of coal power to increase by 40%) but then the court must determine how much importance to give to that proxy compared to alternatives (labor disputes, management’s decision to shift from hydro to coal, etc.).⁵³

A precise rule is “noisy:” it mirrors the contractual objective less perfectly. But it also restricts the court’s discretion more severely and thus reduces ex post enforcement costs. A vague standard has less noise but since it gives courts more freedom to choose evidentiary proxies, enforcement costs will increase because the parties will be motivated to introduce competing bits of evidence as alternative proxies for the broader standard. With this framework, we can evaluate the tradeoff between contractual rules and standards in terms of two dimensions: a) the relative costs of contracting between the front and back end, and b) the informational advantage the parties may have at the time of contracting relative to the hindsight advantage of courts determining proxies in later litigation.

Parties limit the space within which a court selects proxies in litigation by investing in precise terms that define proxies at the front-end. Here, the parties are increasing their

⁵³“Warrick was the only Alcoa smelter that relied on coal for its electric power. The remainder used hydroelectric power, power provided by TVA, or natural gas. The costs of all of Alcoa’s other smelters did not rise with the price of oil. So, while Warrick was Alcoa’s low cost plant when it was built, the changing fuel prices made it far and away the highest cost plant, post-1973.” GOLDBERG at —.

investment in front-end transaction costs, but by reducing the role of the court in litigation, back-end enforcement costs will decline. At the same time, the parties are exploiting their informational advantage (they know their contractual objectives and have the right incentives to choose the best means to achieve them) but they are sacrificing the hindsight advantage that a court might have. Alternatively, the more vague the term, the more work the parties leave for the back end by delegating the proxy selection to the court. Since the parties can't foresee all contingencies, by adopting a vague contract term they can delegate to a court the task of completing the contract ex post. At the time of litigation, uncertainty has been resolved and the court sees realized facts rather than probability distributions. It follows that parties will choose a precise term when their private information is more important than the effect of contingencies on the choice of proxy. Alternatively, when the best proxies are highly contingent on the realized state of the world and less dependent on the parties private information, parties will be more inclined to use vague standards to delegate the proxy choice to a court.

3. The Function of Combinations of Specific Rules and Vague Standards.

The options available to the parties are even broader than the stark choice between a rule or a standard. With the aid of interpretation maxims, parties can design *combinations* of specific and vague terms that more precisely define the “space” within which a court has discretion in selecting proxies. If a contract uses precise terms exclusively, the maxims of interpretation caution the court against considering other proxies at the time of trial. For example, under the maxim *expressio unis est exclusio alterius*, the expression in the contract of one or more things of a class implies exclusion of all that is not expressed.⁵⁴ The inference is that the specification of particular items impliedly excludes other items relating to the same general matter and that all

⁵⁴See *Smart v. Gillette Co. Long-Term Disability Plan*, 70 F.3d 173, 179 (1st Cir. 1995) (where contract listed specific benefits, any other benefit not so listed is excluded); *Eden Music Corp. v. Times Square Music Publications Co.*, 514 N.Y.S. 2d 3 (1987); *United States v. First Nat'l Bank of Crestview*, 513 So. 2d 179 (Fla. App. 1987).

omissions should therefore be understood as exclusions.⁵⁵ Moreover, when a contract provides that a thing should be done in a certain way, it is presumed to be exclusive.⁵⁶

Alternatively, the parties can combine precise terms with vague standards so as to increase the space within which a court can select a proxy. Consider once again the precise term in our earlier example (“the contract price is subject to adjustment if Alcoa’s energy costs increase by more than 35%”). Frequently, parties add to specific terms such as this an additional standard (e.g., “...or for any other causes not within ALOCA’s control”). The parties thus use a precise term requiring adjustment of the contract price together with a vague standard that catches the residual factors that are not covered by the precise term alone. But why don’t the parties cover all excusing factors with the broad standard alone? After all, the 35% increase in energy costs likely falls within the scope of events beyond ALCOA’s control. The reason is that the parties want to contain the proxy-choosing discretion of the court. The parties combine their description of the standard with precise terms so as to define the constraints, or space, within which the court can choose proxies *ex post*. In enforcing this vague term, therefore, the court—aided by the interpretation maxim of *ejusdem generis* (the general is limited by the specific)⁵⁷—will only choose evidentiary proxies that invoke verifiable factors that are similar in kind or related to the precise terms.

The fact that parties can (and do) fine tune the design of their contract argues even more for judicial caution in any attempt to interpret contracts so as to recover the parties “intention.” The theory of contract design shows that intention is a multifaceted inquiry. The parties have an intention to pursue a particular contractual objective. But they also have an intention to select

⁵⁵*See, e.g.*, *Tate v. Ogg*, 170 Va. 95 (1938) (enumeration which included “any horse, mule, cattle, hog, sheep or goat” excluded turkeys).

⁵⁶NORMAN J. SINGER, *SUTHERLAND ON STATUTES AND STATUTORY CONSTRUCTION* §47.26 (6th ed. 2000).

⁵⁷Where the parties combine standards and precise terms that relate to the same subject matter, the *ejusdem generis* canon applies, whether the general language is preceded or followed by the enumerated precise terms. The meaning of the general language is then limited to matters similar in kind or classification to the enumerated precise terms. *See Liberty Mutual Ins. Co. v. East Cent. Okla. Elec. Coop.*, 97 F. 3d 383, 390 (10th Cir. 1996) (when interpreting a general word that follows a series of specific words, the specific words restrict the meaning of the general-encompassing only action of the same general type).

particular terms as the best means of achieving those objectives. To ignore the chosen means is to undermine the fundamental principle of contractual intent. Sometimes the chosen means will fail fully to achieve those objectives, but recall that the parties are optimizing contracting costs so as to maximize *expected* contractual surplus net of those costs. Thus, they are motivated to choose a less ambitious objective if they can achieve an even greater savings in contracting costs and, conversely, they will increase their investment in contracting costs if that expenditure achieves an even greater expected return.

This contract design framework argues for judicial restraint in filling perceived contractual gaps with broad standards in cases where the parties themselves are silent. As in *ALCOA*, a court may be tempted (with the encouragement of one of the parties) to see gaps between the precise terms and therefore to read into the contract implied standards, such as mistake, excuse or frustration. But this is often an error. Commercial parties can include standards in their contract at relatively low cost *and* they enjoy superior knowledge of the context of their contractual relationship to determine the optimal mix of precise and vague terms.⁵⁸ As a rule of thumb, therefore, courts are wise to assume, without substantial evidence to the contrary, that the absence of vague standards in commercial contracts are instructions from the parties to limit their interpretation to the precise terms of the contract.

C. A Contract Design Explanation of the ALCOA/Essex Contract.

1. Reframing the Question of Intent.

In a case like *ALCOA*, where both parties were clearly aware of the risk, the fundamental question is whether the parties conditioned the contract on the risk not materializing, or whether they allocated the risk between them. The court argued that the parties' significant efforts to reduce the risk is evidence that they conditioned their agreement on the risk not materializing. The court relied on these efforts to demonstrate that the parties were motivated, and intended, to reduce the risk that the contract price and the market price of aluminum might deviate

⁵⁸*See, e.g.,* Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 YALE L.J. 541, 545 (2003)

significantly in the future thus imposing large losses on one party or the other. But the issue is not whether the parties intended to reduce this risk; clearly, they did. That they intended to devise an indexing methodology that reduced the risk of deviation between market price and contract price does not, however, provide any evidence of their intent to condition their agreement on the residual risk that the index would malfunction.

To see this point, it is helpful to return to the distinction we have drawn between two kinds of contractual intent: the parties' intended contractual means and their intended contractual ends. The evidence is that the parties intended to create a pricing mechanism that tracked ALOCA'S costs (and thus implicitly tracked the market price for the smelting services subject to a built in price discount). One alternative would have been to agree to a cost plus contract. Cost plus contracts have many problems, however. They reduce the seller's incentives to economize on costs, they are hard to monitor and they require revelation of confidential information. So, the parties were motivated instead to create a verifiable proxy for a cost plus contract that avoided these difficulties. The agreed upon price index was thus intended to allow ALCOA to recover its capital costs in the Warwick plant plus a return on its investment and to permit Essex to obtain a favorable price for the smelting services over the life of the contract. These were the parties' intended contractual ends. But the question before the court concerns the parties' intended contractual means: What instruments did the parties select in order to achieve their goal of having the contract price track the market price of smelting services?

2. The Choice of Contractual Means: Rules v. Standards.

As we showed in Part IIB, there are two basic approaches to designing a price term in relational contracts such as the ALCOA/Essex deal. One option is to specify the contract price in terms of a broad standard. For example, the parties in *ALCOA* could have agreed that the contract price would be the market price for the smelting services at the date of delivery less a "reasonable" discount for Essex. In the extreme, the parties could have omitted any express price term at all, thereby delegating to a court the task of imputing a reasonable price at the time and place of delivery (*see e.g.*, UCC § 2-305). By choosing a standard, the parties delegate the selection of the best proxy for the contract price to a court and thus shift contracting costs from

the front end to the back end.⁵⁹ To optimize the total costs of contracting, however, the parties must compare the back end costs (and benefits) of standards to the front end costs (and benefits) of a precise and conclusive price term. The alternative strategy requires parties to incur *ex ante* the bulk of their costs of determining the pricing mechanism for their contract.⁶⁰ Here the parties would choose a precise price term, one that achieves the parties' objectives subject to an acceptable risk of deviation from that objective. Here the parties economize on back end enforcement costs and also exploit their informational advantage in understanding the contractual objective. But unlike the standards approach, when the parties reduce the price term to a rule, they anticipate no role for a court *ex post* in selecting a proxy in the event the precise term does not function as predicted.

Thus, parties face the familiar trade-off of standards versus rules. Under either approach, the parties might fail fully to achieve their contractual objectives. Courts might misinterpret a standard and select a poor proxy, thereby undermining the parties' objectives. Alternatively, a price term framed as a rule might fail to function as predicted once the state of the world is realized. The trade-off, as noted above, is between the parties' informational advantage in knowing their objectives and the court's informational advantage in knowing what future states materialized. Contract design theory argues for the approach that maximizes the expected value of the contract.

⁵⁹As noted above, this approach has both advantages and disadvantages. The advantages are that: 1) standards are relatively inexpensive to draft *ex ante* so that the parties can economize on front end transactions costs, and 2) standards permit courts to custom-tailor contract terms to the parties' objectives in light of actual circumstances determined *ex post* rather than predictions of circumstances made *ex ante*. There are several disadvantages. First, standards subject the parties' agreement to the risk of misinterpretation by a court. A court might misinterpret the parties contractual objectives and choose the wrong proxy (e.g., whether they intended the price to track retail or wholesale prices) or the court might misapprehend the relevant market circumstances obtaining at the time adjudication (e.g., what the relevant costs and market prices are in a given industry). Second, standards increase the direct costs of adjudication.

⁶⁰The chief advantage of contractual rules are that the parties themselves know better than the courts what their objectives are and presumably can specify a proxy that best enhances contractual incentives. Moreover, by making proof of the price term trivial, a rule reduces the expected costs of adjudication. The chief disadvantage of a rule is that it might not achieve the objective the parties intended. Like all rules, a precise price term, no matter how sophisticated, is likely to be over- or under-inclusive over a large range of possible future circumstances.

3. *Determining the Chosen Contractual Means.*

Returning to *ALCOA*, we can now focus on the central question in the case: What were the parties' intended contractual means for achieving this objective? The key point is that *both approaches are consistent with the parties' intended contractual objective of having the contract price track the market price*. So the question is simply which strategy – rule-based or standard-based – did the parties, in fact, adopt? In *ALCOA*, the evidence strongly suggests that *ALCOA* and *Essex* both understood that they were using the rule-based approach: the parties intended the precise price term to be conclusive. The parties incurred substantial front end drafting costs. These costs are rational to incur only if they exploit the parties' informational advantage and/or create off-setting savings in expected enforcement costs (including the strategic consequences attending adjudication). If the parties intended an *ex post* determination of the contract price, they would have chosen to forgo the substantial investment in drafting a price term designed to replicate a cost-plus contract.⁶¹ If the goal was to exploit a court's hindsight advantage, the parties should have drafted a vague standard that clearly gave the court discretion to determine the appropriate proxy.⁶² Instead, *ALCOA*'s efforts seem designed to achieve the benefits, and incur the costs, of a rule-based approach. Absent compelling evidence to the contrary, a court should presume that *ALCOA* and *Essex* chose a contractual rule to achieve their contractual ends.

In sum, by choosing a precise price term, the parties allocated to *ALCOA* the risk that the

⁶¹As Goldberg notes, the contract does not explicitly state that it was intended to specify proxies that would mimic a cost-plus contract while avoiding familiar difficulties with cost-plus agreements. “Nonetheless, the structure of the agreement makes it quite clear that the intent was to make the contract cost-based.” Goldberg at —. To be sure, the agreed upon indexing formula in the contract had many other problems, in addition to the failure of the WPIC to track carbon-based fuel costs. Over 60% of the contract costs - namely *Alcoa*'s capital costs invested in the Warrick smelter were not indexed at all. So, the index was probably doomed to fail to achieve its objectives under any circumstances. See *Id.* at —. But the mere fact that commercial parties design an *ex ante* rule badly does not justify judicial action to rewrite a precise contract term *ex post*. Indeed, one of the major reasons for inserting caps and floors in such indices is to guard against other risks in addition to the risk of low probability- high impact states of the world. Among those risks is the risk of formulation error by the parties themselves.

⁶²For example, the parties could have specified that “the price shall be a reasonable price that guarantees *ALCOA* a profit of not less than \$.01 nor more than \$.03 per pound . . .” Or, alternatively, the contract might have provided that “the price shall be a reasonable price as determined by two factors: a) the evidence of market prices for similar tolling contracts and b) the provision of a reasonable discount from market price that guarantees a favorable price to *Essex* and a return on its investment to *ALCOA*.”

index would malfunction and increase too slowly and allocated to Essex the risk that the index would malfunction and increase too rapidly. This does not imply, however, that a claim of mutual mistake *necessarily* should be rejected whenever parties choose a contractual rule rather than a standard. For example, contrast *ALCOA* with *Eastern Air Lines v. Gulf Oil Corp.*⁶³ a case involving a malfunctioning price index in which the contract price term was both precise and low cost. The parties entered into a long term contract for the sale of jet fuel at designated locations. They wished to set a price for the jet fuel in order to allocate the risk of exogenous changes in the input price of crude oil to Eastern Airlines and the risk of fluctuations in production cost to Gulf. They selected a contract proxy that adjusted the contract price according to an easily verifiable indicator of crude oil price--West Texas Sour crude “as listed...in Platts Oilgram Service.”⁶⁴ Subsequently, as a result of governmental regulation following the oil crisis in the 1970s, this proxy failed. The court declined to choose a substitute proxy. The parties might have anticipated the failure of the indicator by stating explicitly in the contract that the price either would be adjusted to the price of crude oil (a standard) or that it would be tied to Platt’s or “any other appropriate index.”⁶⁵ As noted above, parties can (and do) adopt a blended strategy by providing for a specific proxy and then delegating to the court the choice of a replacement if the specific proxy should fail.

The question for a court in a case such as *Eastern Airlines*, however, is whether to select a substitute proxy when the parties fail to combine the precise rule with a general standard. Clearly, this was not a risk the parties considered but deemed too remote to address: Neither party considered the possibility that oil would be deregulated *and* that Platts would not report that unregulated price. This risk simply never occurred to them. Because the price index was neither complex nor expensive to create, there is less reason than in *ALCOA* to believe the parties’ intended contractual means for replicating input prices was a conclusive price term. On

⁶³415 F. Supp. 429 (1975).

⁶⁴*Id.* at —. For discussion, see ROBERT E. SCOTT & JODY S. KRAUS, *CONTRACT LAW AND THEORY* 864-69 (Rev. 3d ed. 2003).

⁶⁵Note that the parties in *ALCOA* included such a clause in their contract. See note —*infra*.

the other hand, that fact also suggests that it would have been relatively simple for the parties to have specified a substitute proxy. This argues against excusing performance based on mistake or reforming the agreement. But given that the cause of the price term's malfunction was not even remotely contemplated by the parties, and reformation of the term to conform with the parties' contractual objectives would have been straightforward, the case for mutual mistake and reformation is stronger in *Eastern Airlines* than in *ALCOA*.

III. CONDITIONS AND THE INTERPRETATION OF EXPRESS AND IMPLIED TERMS

Before a court can make any decision governing a contract dispute, it must first interpret the alleged agreement. All other contract doctrines are therefore logically dependent on a prior application of the interpretation doctrines. These doctrines not only determine the content of enforceable agreements, but necessarily precede the application of the excuse doctrines that permit a court to refuse to enforce part or all of a contract. The most fundamental distinction recognized by the interpretation doctrines is between express and implied terms. In general, the interpretation doctrines direct courts to respect the parties' express terms. Express terms can specify both primary terms governing the parties' performance obligations and secondary, or meta, terms governing the interpretation of their agreement. Express terms therefore form the foundation of establishing contractual intent: in principle, they not only allow the parties to communicate to each other and to courts the precise content of certain terms they wish to include in their agreement, but also allow them to control the extent to which courts may imply additional terms into their agreement. Under the parol evidence doctrines, express terms can constrain the court's discretion to imply terms into an agreement by, for example, directing the court not to admit prior written or oral evidence of implied terms. Express terms therefore provide the most powerful tool available to parties for reducing the risk that courts will undermine their chosen contractual means by interpreting their obligations so as to promote their contractual objectives.

However, express terms themselves require interpretation. The doctrines governing the interpretation of express terms present yet another occasion on which courts can undermine the

parties' intended contractual means by impressing a meaning on express terms that best promotes the parties' intended contractual objectives. For example, courts have long been divided on the question of whether express terms should be given a so-called "contextual" or "plain meaning" interpretation. On the contextual view, the interpretation of express terms inevitably requires judicial speculation about what meaning the parties were likely to have attached to the express terms they used. Thus, courts that identify contractual intent with the parties' contractual ends will assign express terms a meaning that, at the time of adjudication, best promotes those ends, instead of seeking to identify the meaning that reflects the contractual means the parties most likely intended at the time of formation. On the plain meaning view, however, courts ultimately base contract interpretation on their common knowledge of word meaning rather than context-based speculation about intent. Yet even under a plain meaning regimes, the interpretation of express terms can be significantly influenced by a court's erroneous effort to use the process of interpretation to vindicate contractual ends instead of means. First, the plain meaning rule does not apply if a court finds an express term to be ambiguous. Thus, a court's threshold determination of whether an express term is ambiguous can be influenced by its perception of the effect various interpretations would have on the parties' contractual objectives. Second, even if a court concludes that an express term has a plain meaning, it might narrow or expand the plain meaning it assigns to that term in order to vindicate the parties' contractual ends, at the expense of undermining their intended contractual means.

Thus, even though express terms constitute the means by which parties can control the judicial interpretation of their agreement, the judicial tendency to identify contractual intent with the parties' intended contractual ends can lead them to impress unintended meanings on the parties' express terms or imply unintended terms into the parties' agreement. Such judicial practices further undermine the reliability of the most basic doctrinal tools essential for contract design. In this Part, we explain how the law of conditions embodies a deep-seated conflict between legal doctrines dedicated to vindicating the ex ante perspective in interpretation, and equitable doctrines driven by the ex post perspective in interpretation. We argue that courts err when they accept equity's invitation to set aside clear express terms in order to vindicate

contractual ends *ex post*. We also argue that a basic understanding of contract design is essential for courts to interpret accurately the meaning of ambiguous express terms and to identify the implied terms in agreements. In particular, we draw on extended case analysis to illustrate how legal and equitable doctrines driven by the *ex post* perspective lead courts astray when deciding whether to recognize, imply, or enforce conditions in agreements.

A. The Law of Conditions

The Restatement (Second) of Contracts defines a condition as “an event, not certain to occur, which must occur, unless its non-occurrence is excused, before performance under a contract becomes due.”⁶⁶ Since parties incur duties in contracts by making promises, a party who makes an event a condition of its promise is under a duty to perform that promise only if the event occurs.⁶⁷ A common example is a contract for the purchase of land in which the buyer’s promise to pay is made expressly conditional on approval of his mortgage application. The buyer’s duty to pay does not arise until his mortgage application is approved, and never arises if his application is denied. Another common example is an insurance contract that imposes on the insurer a duty to pay if the insured brings a claim within a specified time period after the insured suffers a covered loss. The insurer’s duty to pay arises when the insured suffers a covered loss, but that duty is discharged if the insured fails to bring the claim within the specified time period.⁶⁸ The law of conditions explicitly endorses the principle of freedom of contract by

⁶⁶R2d §224.

⁶⁷The event on which the promise is conditioned sometimes is “largely within the control of the obligor (the homeowner’s honest satisfaction with the paint job), the obligee (the insured’s furnishing proof of loss), or a third person (the bank’s approval of the mortgage application), or is largely beyond the control of anyone (damage as a result of fire). Farnsworth, Vol. II, §8.2, p. 417.

⁶⁸The First Restatement of Contracts describes the land buyer’s duty as subject to a “condition precedent” because the duty never arises if the event never occurs, and describes the insurance company’s duty as subject to a “condition subsequent” because the duty arises when the buyer incurs a qualifying loss and the condition serves only to discharge the duty if the buyer fails to bring his claim within the specified time period. The Second Restatement of Contracts rejects this terminology and instead defines “condition” as what the First Restatement called a condition precedent. The Second Restatement characterizes what the First Restatements refers to as conditions subsequent as events that “terminate an obligor’s duty of immediate performance or one to pay damages for breach.” Restatement Second §230. The distinction between conditions precedent and subsequent served to allocate the burdens of pleading and proof by requiring the promisee to prove the occurrence of a condition precedent and the promisor to prove the non-occurrence of a condition subsequent. The Second Restatement seeks to decouple the allocation of

committing to the strict enforcement of all express conditions.⁶⁹ Yet it is also home to the hoary equitable maxim that the law abhors a forfeiture.⁷⁰ The Restatement (Second) defines a “forfeiture” as the denial of compensation for losses the promisee incurs when the non-occurrence of a condition of the promisor’s duty causes the promisee to lose his right to the agreed exchange after he has relied substantially on the expectation of that exchange by preparing to perform or by performing.⁷¹ The anti-forfeiture norm suffuses the law of conditions, which therefore reads like a schizophrenic text, in one sentence insisting on the sanctity of strict construction and enforcement of conditions in spite of forfeiture,⁷² while in the next sentence admonishing courts, whenever possible, to avoid the conclusion that the promisor’s obligation is subject to an enforceable condition if enforcement of the condition would raise the specter of

burdens of pleading and proof from the sometimes difficult question of whether a term makes the occurrence of an event a condition or termination of a duty. Restatement Second §224, Reporter’s Note. *See also* Farnsworth, Vol. II, §8.3, p. 421.

⁶⁹ *See* Farnsworth, Vol. II, §8.3, pp 422-23. “If the occurrence of a condition is required by the agreement of the parties, rather than as a matter of law, a rule of strict compliance traditionally applies. As the Supreme Court of New Hampshire explained, “when the parties expressly condition their performance upon the occurrence or non-occurrence of an event, rather than simply including the event as one of the general terms of the contract, the parties’ bargained-for expectation of strict compliance should be given effect.” (citing *Renovest Co. v. Hodges Dev. Corp.*, 600 A.2d 448, 452-53 (N.H. 1991).”

⁷⁰ *See* Murray on Contracts, §102(A), p. 641 (“One of the maxims often repeated in contract law is that the law abhors forfeitures,” citing *Stevenson v. Parker*, 25 Wash. App. 639, 608 P.2d 1263, 1267-68 (1980): “This court has held the general doctrine that forfeitures are not favored in the law, and that courts should promptly seize upon any circumstance arising out of the contract or relations of the parties that would indicate an election or an agreement to waive the harsh, and at times unjust, remedy of forfeiture, a remedy which is oftentimes too freely granted by those who have taken no account of the misfortunes and disappointments which conditions, unforeseen and beyond a party’s control, have raised as a bar to performance, however honest may be his intent . . .” (quoting *Spedden v. Sykes*, 51 Wash 267, 272 P. 752, 754 (1908)).

⁷¹ *See* R2d §227, com. b. The intuition underlying the maxim that the law abhors a forfeiture is that, courts should promptly seize upon any circumstance arising out of the contract or relations of the parties that would indicate an election or an agreement to waive the harsh, and at times unjust, remedy of forfeiture, a remedy which is oftentimes too freely granted by those who have taken no account of the misfortunes and disappointments which conditions, unforeseen and beyond a party’s control, have raised as a bar to performance, however honest may be his intent . . . *Stevenson v. Parker*, 25 Wash. App. 639, 608 P.2d 1263, 1267-68 (1980)(quoting *Spedden v. Sykes*, 51 Wash. 267, 272, 98 P. 752, 754 (1908).

⁷² “The policy favoring freedom of contract requires that, within broad limits . . ., the agreement of the parties should be honored even though forfeiture results.” R2d §227, cmt. b; “[T]o the extent that the parties have, by a term of their agreement, clearly made an event a condition, they can be confident that a court will ordinarily feel constrained strictly to apply that term.” R2d §226, cmt. c; “[I]f the term that requires the occurrence of the event as a condition is expressed in unmistakable language, the possibility of forfeiture will not affect the interpretation of that language.” R2d §229, cmt. a;

forfeiture.

The anti-forfeiture norm takes both an ex ante and ex post form. In its ex ante form, the anti-forfeiture norm controls the interpretation of whether an obligation is subject to a condition. Although courts use general principles of interpretation to identify and distinguish conditions and duties,⁷³ they have developed specialized principles that resolve only interpretive questions regarding conditions.⁷⁴ The ex ante version of the anti-forfeiture norm applies when the language of an agreement fails to make clear whether it is intended to subject the promisor's duty to a condition or impose a duty on the promisee, or both. For example, an employment contract might provide that the employer shall not terminate the employee without just cause, but also provide that the employee "will, within thirty days of termination, give written notice to the employer of any claim of wrongful termination and will not take any legal action based on the claim within six months of such notice." Suppose that the employee violates this provision by either failing to provide written notice of his claim or taking legal action in less than six months after providing notice. A court could interpret the clause as a promise by the employee to provide written notice and not take legal action within the specified time periods. On this interpretation, the employer remains subject to suit for wrongful discharge but would be entitled to recover damages from the employee for breach of her promise. A court might instead interpret this language as creating a condition on the employer's promise not to terminate without just cause. On this view, the employee's failure to satisfy the condition would relieve the employer of any liability for wrongful discharge. Finally, a court might interpret the

⁷³R2d §226, cmt. a ("Whether the parties have, by their agreement, made an event a condition is determined by the process of interpretation. That process is subject to the general rules that are contained in previous topics of this Chapter. For example, as in other instances of interpretation, the purpose of the parties is given great weight . . . , and, in choosing between reasonable meanings, that meaning is generally preferred which operates against the draftsman). Farnsworth, Vol. II, §8.2, p. 416 ("A court determines whether an agreement makes an event a condition by the process of interpretation."); Farnsworth, Vol. II, §8.4, p. 430 ("The process of interpretation by which these questions are resolved is essentially the same as that by which other questions of interpretation are resolved. The same emphasis is put on purpose, maxims, prior negotiations, usage of trade, course of dealing, and course of performance." (citations omitted)).

⁷⁴R2d §226, cmt. a ("There are also some special standards of preference that are of particular applicability to conditions, and these are set out in §227.") Farnsworth, Vol. II, §8.4, p. 430 (When considering whether parties intended to create a condition, a duty, or both, "[a]lthough courts apply traditional techniques of interpretation in such cases, they have displayed two preferences in interpretation that merit special attention.).

language as creating both a promise and a condition, in which case the employee's failure to satisfy the contract terms not only relieves the employer of liability but also subjects the employee to liability for breach of her promise.

Thus, when a court decides that a term merely imposes a duty on *A*, the non-occurrence of the event entitles *B* to sue *A* for damages for breach but does not relieve *B* of her obligation to perform. However, if the event is deemed a condition of *B*'s promise, its non-occurrence not only deprives *A* of his expectancy, but also subjects him to any losses incurred in performing his duties under the agreement. Following the anti-forfeiture maxim, contract law regards *B*'s reliance as a forfeiture to be avoided,⁷⁵ whenever the contract's express language and circumstances allow, by favoring an interpretation of a term as creating a promissory duty rather than a condition. Even when express language clearly creates a condition, the doctrine directs courts to choose the least expansive interpretation of the condition if doing so will avoid a forfeiture.⁷⁶ The anti-forfeiture rationale underlying these interpretive rules is not that the law should intervene to prevent forfeitures that have materialized under an agreement, but rather that the promisee is unlikely to have agreed to a risk of forfeiture at the time of formation.⁷⁷ Thus, in

⁷⁵“The non-occurrence of a condition of an obligor's duty may cause the obligee to lose his right to the agreed exchange after he has relied substantially on the expectation of that exchange, as by preparation or performance. The word “forfeiture” is used in this Restatement to refer to the denial of compensation that results in such a case.” R2d 227, cmt. c; “As the term *forfeiture* suggests, the obligee's loss of the reliance interest is regarded as more serious than is the obligee's loss of the expectation interest.” Farnsworth, Vol. II, §8.4, p. 434.

⁷⁶See R2d §227: “(1) In resolving doubts as to whether an event is made a condition of an obligor's duty, and as to the nature of such an event, an interpretation is preferred that will reduce the obligee's risk of forfeiture, unless the event is within the obligee's control or the circumstances indicate that he has assumed the risk. (2) Unless the contract is of a type under which only one party generally undertakes duties, when it is doubtful whether (a) a duty is imposed on an obligee that an event occur, or (b) the event is made a condition of the obligor's duty, or © the event is made a condition of the obligor's duty and a duty is imposed on the obligee that the event occur, the first interpretation is preferred if the event is within the obligee's control.” See also R2d 227, cmt. c (“When . . . it is doubtful whether or not the agreement makes an event a condition of an obligor's duty, an interpretation is preferred that will reduce the risk of forfeiture. . . . In determining the nature of the event that is made a condition by the agreement, as in determining whether the agreement makes an event a condition in the first place . . . , it will not ordinarily be supposed that a party has assumed the risk of forfeiture. Where the language is doubtful, an interpretation is generally preferred that will avoid this risk.”

⁷⁷“Although [R2d §227] is consistent with a policy of avoiding forfeiture and unjust enrichment, it is not directed at the avoidance of actual forfeiture and unjust enrichment. Since the intentions of the parties must be taken as of the time the contract was made, the test is whether a particular interpretation would have avoided the risk of forfeiture viewed as of that time, not whether it will avoid actual forfeiture in the resolution of a dispute that has

principle, if a finding that a promise is subject to a condition would impose a forfeiture on the promisee but the parties would not have understood the condition to create such a risk at the time of formation, then the ex ante forfeiture doctrines allow the courts to interpret the contract to contain the condition.

In its ex post form, however, even if the parties succeed in writing an express term that unequivocally creates a condition, the anti-forfeiture norm strongly encourages courts to exercise their equitable discretion to excuse the condition whenever its enforcement would create a forfeiture and the court deems the condition not to have been a material part of the agreement at the time of formation.⁷⁸ In addition, even if a court agrees that a contract contains an unexcused, express condition, the ex post anti-forfeiture norm invokes principles of equity to permit the court to hold that the promisor implicitly waived the condition, either retrospectively or prospectively, or both. Equity favors the finding of a waiver whenever enforcement of the condition would create a forfeiture.

In sum, the law of conditions explicitly stacks the deck heavily against the finding and enforcement of conditions on the ground that the law abhors a forfeiture. In this Part, we argue that the anti-forfeiture norm is based on two false assumptions. The first is that parties are unlikely to select terms that create the risk of forfeiture. This assumption underlies the ex ante forfeiture norm, which underwrites the doctrines directing courts to avoid finding a condition absent express language that unmistakably creates it. The second assumption is that express conditions are sometimes not material at the time of formation. This assumption underlies the ex post forfeiture norm, which underwrites the doctrines directing courts to avoid enforcing even clear, express conditions. We have argued that considerations of contract design commonly favor the selection of precise terms in order to create rule-like obligations that are easy for the parties to understand and easy to enforce in court. Our claim is that conditions serve just this purpose. In addition, they afford the promisor absolute protection from the risk of certain losses

arisen later.” R2d §227, cmt. b.

⁷⁸See R2d §229.

he wishes to avoid, instead of the more costly and less reliable protection afforded by the alternative of requiring the promisor to prove his losses in a suit for damages. The price of such protections for the promisor is an increased risk of forfeiture for the promisee. But especially when parties include express conditions, there is no systematic reason to doubt that the promisee understood that risk and charged accordingly by requiring compensating contractual benefits from the promisor. Like all contract terms, conditions play a vital role in defining the means by which parties intend to achieve their contractual ends. They are therefore always material from the ex ante perspective because they allocate risks between the parties, the contract compensates each party for bearing those risks, and the parties inevitably rely on that allocation of risks. Since materiality is determined by the parties' intent at the time of formation, conditions will always be material. Thus, the doctrines inviting courts to set aside conditions ex post only if the conditions were not material ex ante should never be applied.⁷⁹

1. The Law Governing Conditions of Satisfaction

We have seen how the anti-forfeiture norm gives rise to interpretive rules favoring promises over conditions. But it also influences the interpretation of contracts that subject a promise to the express condition of the promisor's satisfaction with the promisee's performance or something else. If the express condition does not specify whether the promise is subject to the promisor's objectively reasonable satisfaction or subjective but honest satisfaction, the law directs courts to prefer the former unless it is not practicable to determine objective reasonableness of satisfaction under the circumstances.⁸⁰ The rationale of this interpretive rule is that a promisee ordinarily is unlikely to agree to subject himself to the idiosyncratic tastes of the promisor in circumstances where objectively reasonable satisfaction would be practical to

⁷⁹We argue below that the anti-forfeiture norm should play no role in the law of conditions. This means we reject all interpretive doctrines that bias courts against the implication, recognition, or enforcement of conditions. However, we set aside the doctrines directing courts to avoid the implication of conditions into agreements and to interpret ambiguous terms as promises rather than conditions. Instead, we focus our attention below on the doctrines urging courts to avoid recognizing or enforcing terms that unmistakably create express conditions.

⁸⁰R2d §228 ("When it is a condition of an obligor's duty that he be satisfied with respect to the obligee's performance or with respect to something else, and it is practicable to determine whether a reasonable person in the position of the obligor would be satisfied, an interpretation is preferred under which the condition occurs if such a reasonable person in the position of the obligor would be satisfied.")

ascertain. The rule is also supported by the rationale that promisees would be unlikely to agree to a subjective satisfaction condition because they would thereby be incurring a significant risk of forfeiture.⁸¹ By its own terms, this rule applies only in the absence of clear language creating a condition of honest, subjective satisfaction or circumstances in which objectively reasonable satisfaction would be impracticable. Yet the judicial instinct to avoid perceived forfeitures is so strong that it often overrides the doctrinal caveats directing courts to enforce express conditions even when they do create a forfeiture. Courts are especially vulnerable to this error when they fail to understand why the promisee would agree to subject himself to the risk of forfeiture. To illustrate this error, we turn to a case in which the court must decide the effect to give to a condition that appears to make the promisor's obligation illusory.

In *Corthell v. Summit Thread Company*,⁸² Robert Corthell, an employee of Summit Thread, promised to turn over patents for three of his current inventions as well as all of his future inventions for a period of five years, in return for which Summit Thread promised to increase Corthell's annual salary by \$620 for five years, make Corthell a one-time payment of \$3,500, and provide Corthell additional "reasonable recognition" for his future inventions, "the basis and amount of recognition to rest entirely with the Summit Thread Company at all times." The written agreement between the parties additionally specified that the terms of the contract were "to be interpreted in good faith on the basis of what is reasonable and intended and not technically." During the term of the contract, Corthell turned over four additional inventions to Summit Thread and requested but never received any additional payment for them as "reasonable recognition" from Summit Thread. Corthell brought suit seeking payment for the four inventions. Summit Thread argued that its promise was unenforceable because it was illusory and therefore indefinite. A promise is illusory if it is subject to a condition that is entirely within the promisor's control and the promisor incurs no detriment by deciding not to

⁸¹R2d §228, cmt. b ("When, as is often the case, the preferred interpretation will reduce the obligee's risk of forfeiture, . . . there is an additional argument in its favor.")

⁸²132 Me. 94 (1933).

satisfy that condition.⁸³ The illusion of an enforceable promise is created in the first part of Summit Thread’s promise, in which it promises to pay reasonable recognition for Corthell’s inventions. Had the promise stopped there, standard interpretive canons would have lead the court to attempt to enforce it implying a term requiring Summit Thread to pay Corthell the value of his inventions according to an objective standard. But the agreement destroys the illusion of an enforceable promise when it stipulates that the amount of recognition rests “entirely” within Summit Thread’s discretion. Because Summit Thread’s promise to pay reasonable recognition expressly grants Summit Thread unfettered discretion to determine the amount, if any, of recognition that is reasonable, it imposes an obligation on Summit Thread to pay compensation to Corthell only if Summit Thread chooses to make such a payment. It therefore qualifies as a classic illusory promise. Historically, courts refused to enforce illusory promises on the ground that they are not supported by consideration. Had the court in *Corthell* followed suit, it would have declared unenforceable Summit Thread’s promise to pay reasonable recognition. If Corthell had not yet performed, he also would have been free from any legal obligation to turn over his inventions. But because Corthell had already performed by turning over his inventions, he would have been deemed to have made a gratuitous transfer and therefore would have been denied recovery in contract.⁸⁴

The court followed instead more contemporary decisions that imply a term imposing objective duties on a promisor whose promise otherwise would be unenforceable because illusory. Although it noted that “[reasonable recognition was] coupled with the reservation that the ‘basis and amount of recognition (was) to rest entirely with’ the Company ‘at all times,’” it

⁸³“One of the commonest kind of promises too indefinite for legal enforcement is where the promisor retains an unlimited right to decide later the nature or extent of his performance. This unlimited choice in effect destroys the promise and makes it merely illusory.” §43 Williston on Contracts (1920). “Words of promise which by their terms make performance entirely optional with the ‘promisor’ do not constitute a promise. . . . Where the apparent assurance of performance is illusory, it is not consideration for a return promise.” Restatement of Contracts (Second), § 77, cmt. a.

⁸⁴Corthell could have sued in restitution for a *quantum valebant* recovery of the value of the inventions he turned over to Summit Thread. But we argue below that a court that properly ruled against recovery in contract would for the same reasons deny recovery in restitution on the ground that Corthell accepted the risk that Summit Thread would retain his inventions without paying him for them.

noted that “[n]evertheless, the contract was ‘to be interpreted in good faith on the basis of what is reasonable and intended, and not technically.’” On the basis of these provisions, the court concluded that

the parties continued to exhibit a contractual intent and a contemplation of the payment of reasonable compensation to the plaintiff for his invention. The Company was not free to do exactly as it chose. Its promise was not purely illusory. It was bound in good faith to determine and pay the plaintiff the reasonable value of what it accepted from him.

Thus, the court argued that the clause directing it to interpret the contract “in good faith on the basis of what is reasonable and intended, and not technically” justified it in over-riding the express language granting Summit Thread “entire” discretion to determine the amount it will pay Corthell for his inventions. Yet the court offers no reason to believe that the literal interpretation of the word “entirely” (as meaning entirely) is a technical, rather than simple plain meaning, analysis of the language the parties used.⁸⁵ The court’s justification for its interpretation appears to be that the clause granting Summit Thread “entire” discretion to set the amount of compensation is, on its face, inconsistent with the promise’s express language requiring Summit Thread to act in good faith. Faced with the choice between enforcing the two clauses the court believed were inconsistent, it gave priority to the good faith term and effectively set aside the clause granting Summit Thread unfettered discretion to determine the amount of reasonable recognition to which Corthell would be entitled.

Even when contracts contain no express language requiring parties to act in good faith, contract law implies such a duty into all contracts. *Corthell* then might be taken to suggest that promises can never be illusory, for the implied duty of good faith would always cure an otherwise illusory promise by imposing enforceable duties on the promisor. But this is clearly not the case. The illusory promise doctrine still operates to render promises unenforceable when their clear language unequivocally subjects the promise to a condition that is entirely within the

⁸⁵As Defendant argued in its brief, “if the basis and amount of recognition lies entirely with the defendant at all times, it lies with the defendant now, not with the plaintiff and not with any third party or parties, not even with a court of law.” (15 D’s Brief)

promisor's control and which the promisor can avoid satisfying without incurring a detriment.⁸⁶ Instead, courts use doctrines such as the implied duty of good faith to impose duties on the illusory promisor, despite express language disavowing them, only if they find the promisee reasonably believed the promisor's discretion was implicitly constrained. Here, courts see the application of the consideration doctrine as enabling the promisor, in effect, to fraudulently induce the promisee into making a gratuitous transfer, which would of course violate the implied duty to act in good faith. Thus, the implicit premise of the *Corthell* decision is that it was reasonable for Corthell to believe Summit Thread was undertaking an obligation to pay an objectively reasonable amount of compensation for his inventions, and therefore intended to be legally bound by its promise, despite its express language that the amount of compensation to be paid to Corthell rested "entirely" with Summit Thread.

The *Corthell* decision, therefore, provides no plausible account of why the parties expressly stipulated that the amount of recognition was to rest entirely with Summit Thread. As an interpretation of the parties' agreement, *Corthell* fails the basic requirement of providing at least a minimally adequate account of the express language in the parties' writing. The court simply treats the writing as if it did not contain the word "entirely." But why would the *Corthell* court insist on an interpretation directly contradicted by the writing's express language? The answer, we believe, is that it could not otherwise explain why Corthell would have been willing to transfer his inventions to Summit Thread. If the parties intended Summit Thread to have unfettered discretion to decide the amount, if any, of compensation to be paid to Corthell after he turns over an invention, then Corthell would be subjecting himself to the risk of giving something in return for nothing. Since the court was unable to explain why a minimally rational party would promise to turn over potentially valuable inventions without requiring that the beneficiary pay for any benefit it received, the court simply dismissed the possibility that the parties could have meant the word "entirely" to be taken seriously. The express clause directing the court to interpret the contract language in good faith and not technically provided convenient linguistic cover to justify an interpretation of the contract that eviscerated the meaning of the

⁸⁶Cites.

word “entirely.”

Like the court’s failure in *Hunt Foods*, the *Corthell* court’s inability to understand why Corthell rationally might be willing to subject himself to the risk that Summit Thread will pay him nothing in return for a valuable invention results from its failure to appreciate the role that information theory and informal norms play in designing contracts. Information theory teaches us that the more difficult information is to verify, the less likely it is that parties will subject their contractual obligations to conditions whose satisfaction can be proved only by verifying that information.⁸⁷ This basic principle suggests that Summit Thread would not be willing to condition its performance obligation on the objective value of Corthell’s inventions unless that value was likely to be verifiable. But that information is not likely to be verifiable. First, as a general matter, the economic value of intellectual property is often difficult to determine. In this case, the value of Corthell’s inventions is especially difficult to determine. Like the previous inventions he sold to Summit Thread,⁸⁸ the inventions for which Corthell sought compensation were relatively minor variations on the design of thread-handling devices, such as spools, bobbins, and shuttles, that Summit Thread and its customers used to store and sew with the thread it manufactured. Summit Thread manufactured and used spools, bobbins and shuttles, but it did not sell them.⁸⁹ Thus, the value, if any, of these inventions to Summit Thread consisted

⁸⁷Recent scholarship argues against the standard assumption in the contract theory literature that information is either verifiable or not. In fact, information falls along a multi-dimensional continuum of verifiability, and where an item of information falls on that continuum is a function of both its inherent character and the contingent rules of contract interpretation and evidence law. See Robert E. Scott & George Triantis, *Anticipating Litigation*, *supra*.

⁸⁸Corthell’s first inventions were “bobbin control adjuncts” and “guarding attachments for thread caps.” (P’s Brief, 4).

⁸⁹Corthell sought compensation for four inventions. His first invention consisted of “a series of corrugations or ridges to be made on the base of the King Spool,” a “tube with a wooden, tapered base fastened to it, the idea being that the thread should run freely from the top without catching.” (16 D’s brief). This invention was a minor variation on the King Spool on which Summit Thread spooled the thread it sold. Corthell maintained that the principal value of this invention was that it “effected a continuation of the old King Spool patent, which patent would expire . . . unless a continuation was granted,” and that “the chief purpose for which the invention was conceived, namely, to prevent any of the Defendant’s competitors from using any type of King Spool, was accomplished— so whether it was ever used or not is immaterial because no other thread concern ever used this King Spool at any time.” (P’s Brief, 20). He did not claim that the value of this invention was attributable to any increase in the spool’s practical value. Indeed, Corthell was prepared to concede that it hadn’t ever been used, let alone sold.

exclusively in the value placed on these inventions by Summit Thread's customers. That value consisted either in the increased ease of using the spools on which Summit Thread's thread was wound when delivered, or the increased effectiveness of the sewing machine parts, like bobbins and shuttles, which Summit Thread supplied free of charge to its customers. Presumably, the easier it was for Summit Thread's customers to use Summit Thread's thread, and the more valuable Summit Thread's free machine parts were to its customers, the more likely they would be to buy their thread from Summit Thread and to pay a higher price.

Clearly, Summit Thread believed that Corthell's previous inventions were sufficiently valuable to warrant paying him \$3,500 for them and negotiating a contract renewal that included a clause entitling Summit Thread to Corthell's future inventions during the contract term. Summit Thread is in the best position to assess the value of these inventions because it has the greatest access to and control over information regarding its costs and revenues. Yet Summit Thread would likely have difficulty verifying an assessment of the discrete value attributable to such inventions. Since neither Summit Thread nor Corthell expected that Summit Thread would sell Corthell's inventions (and in fact none of them were ever sold), no objective measure of their value would be available. Instead, their value-added to Summit Thread's revenue would have to be factored out of the many inputs that contribute to its sales revenues, a calculation that is inevitably subjective and difficult, if not impossible, to make, let alone prove. Moreover, the complexity and subjectivity of such an evaluation would invite litigation based on genuine differences of opinion on the actual value of Corthell's inventions, as well as strategic attempts to extort payments for inventions of little or no value to Summit Thread. Given the practical and

The second invention consisted of "an adjunct to sewing machine shuttles to be attached to bobbins for certain styles of sewing machines." (20 D's brief) However, "[Summit Thread] never sold bobbins." (D's brief, 21). The third invention consisted "of celluloid discs containing an indenture or boss, so called, to be attached to certain sewing machine shuttles." According to Summit Thread's brief, "[i]nsofar as the discs were distributed to the public it was a gratuitous distribution." (D's brief, p 30). And the fourth invention consisted of "paper or celluloid discs to be attached to all bobbins used by [Summit Thread]." (31 D's brief). Corthell maintained that the only value of this invention was its unrealized potential to allow Summit Thread to attach a disc to the bobbins it manufactured and distributed to its customers without violating a patent on a disc, which served the same function, by one of Summit Thread's competitors, the American Thread Company. (P's Brief, pp. 22-23). The invention was never patented and neither Summit Thread nor any other company manufactured or used it, apart from one made for an experimental trial which failed.

theoretical difficulties of cost-effectively determining and proving the indirect value of these inventions to Summit Thread's business, it is plausible that the expected costs of litigation would exceed the expected value of the inventions themselves. Information theory therefore provides a plausible explanation for why Summit Thread would intend a literal interpretation of the language resting discretion to determine the reasonable recognition for Corthell's inventions "entirely" with Summit Thread.

Defending the literal interpretation of the word "entirely," however, requires more than a demonstration of why Summit Thread would have been likely to have meant the word to be interpreted literally. It also requires a demonstration that a reasonable person in Corthell's position would have given the word "entirely" a literal interpretation as well. Corthell had four plausible reasons for agreeing to turn over his future inventions without requiring Summit Thread to make a legally enforceable promise to pay for them. First, Summit Thread had already demonstrated its willingness to pay him for such inventions. In the same agreement in which he promised to turn over his future inventions, Corthell received a payment of \$3,500 for inventions he had already turned over to Summit Thread. Second, the agreement between Corthell and Summit Thread constituted a repeat play game in which Summit Thread's defection from its non-legal commitment to pay Corthell the reasonable value of his inventions could be punished by Corthell's refusal to produce more inventions during the remainder of the contract period. Because the contract placed him under no obligation to create inventions, Corthell could protect himself through self-help.⁹⁰ Third, if the reasonable value of Corthell's inventions is observable to Corthell and third parties, then Summit Thread would incur a reputational harm by breaching its informal agreement to pay that value to Corthell. Fourth, recent research confirms that most people place an intrinsic value on fairness norms and are willing to incur significant costs to punish unfairness even when doing so holds out no prospect of inducing future cooperation to compensate them for those costs. In short, as a general matter, Corthell had good reason to believe that Summit Thread would honor its commitment even if not legally required to do so,

⁹⁰Moreover, if Summit Thread's promise is interpreted as an illusory and therefore unenforceable promise, any return promise made by Corthell would likewise be unenforceable.

and Summit Thread had given him no reason to think otherwise.

In sum, Summit Thread plausibly might have believed that the expected costs of litigation exceeded the expected value of Corthell's inventions, and Corthell had good reason to believe that Summit Thread would pay the reasonable value of his inventions even if it were not legally required to do so.⁹¹ While it is clear that the shared contractual objective of the parties was for Summit Thread to pay Corthell the reasonable value of the inventions he created during the contract period, it seems equally clear that the parties had good reason to choose non-legal means to achieve that objective. There is, therefore, no good reason to doubt that the parties meant precisely what they said when they expressly agreed that the determination of reasonable recognition was to rest "entirely" with Summit Thread.

2. The Law Governing Excuse of Conditions

Corthell illustrates how courts can misapply interpretive doctrines to avoid application of the illusory promise doctrine. We have argued that the court sought to avoid the illusory promise doctrine because it felt bound to interpret the parties' agreement in a way that was consistent with a plausible view of the parties' intentions at the time of formation. This reconstruction bases the *Corthell* decision on a commitment to maintain fidelity to the parties' intentions at the time of formation. The problem with the court's decision, on this view, is that it lacked both a sufficiently rich understanding of how and why parties might use a mix of formal and informal norms to achieve their contractual objectives, and a commitment to enforce the plain meaning of the express terms of an agreement even when it does not understand why parties would have agreed to it.

But the *Corthell* decision could also be explained as an application of the ex post forfeiture norm as embodied in the doctrine governing excuse of conditions. On this view, the question of whether Corthell knowingly took the risk of forfeiture when he entered into the

⁹¹Moreover, Corthell could limit the risk of Summit Thread's breach of its informal agreement by reducing his efforts to develop future inventions until Summit Thread paid him the reasonable value of inventions as he produced them *seriatim*.

agreement would be irrelevant. As we've seen, the anti-forfeiture norm not only creates a strong bias against the finding of conditions at the interpretive stage of adjudication, it also grounds an independent bias against them at the enforcement stage. Formal excuse doctrine permits a court to invalidate even express conditions "to the extent that the non-occurrence of a condition would cause disproportionate forfeiture, . . . unless its occurrence was a material part of the agreed exchange."⁹² Thus, according to this ex post version, the court simply refused to enforce a term that clearly created a condition of the promisor's obligation because the court believed that the condition was not a material part of the parties' agreement and its enforcement would both impose a substantial reliance loss on Corthell and create a windfall for Summit Thread. Courts regard such losses as forfeitures and such gains as "unjust enrichment."

The excuse account of the *Corthell* decision requires an explanation for why the court would determine that the express condition was not a material term of the parties' agreement. Although the court does not provide an explicit account of how the excuse doctrine applies to the case, the same reasons that explain the ex ante account also would explain why the court would have, or implicitly did, reach the same conclusion under the excuse doctrine. If the court fails to understand the information barriers to formal enforcement and the role that informal norms can play in enforcing non-legally binding commitments, it would understandably fail to see why, assuming that Summit Thread planned to act in good faith, it would regard as material a term granting it unfettered discretion to determine the reasonable value of Corthell's inventions. As we've explained above, the *Corthell* court appears to treat that term as anomalous, which means it certainly does not regard it as material.

The above re-characterization of *Corthell* provides one illustration of how the doctrine governing excuse of conditions is premised on reasoning that fails to use and appreciate the ex ante design perspective in contract interpretation. However, the best illustrations of this flawed reasoning are provided by the five official illustrations included in §229 of the Restatement (Second) of Contracts. Unlike *Corthell*, each of these illustrations expressly embraces the

⁹²R2d § 229.

excuse doctrine. The first illustration is based on the celebrated case of *Jacob & Youngs, Inc. v. Kent*.⁹³

A contracts to build a house for B, using pipe of Reading manufacture. In return, B agrees to pay \$75,000 in progress payments, each payment to be made “on condition that no pipe other than that of Reading manufacture has been used.” Without A’s knowledge, a subcontractor mistakenly uses pipe of Cohoes manufacture which is identical in quality and is distinguishable only by the name of the manufacturer which is stamped on it. The mistake is not discovered until the house is completed, when replacement of the pipe will require destruction of substantial parts of the house. B refuses to pay the unpaid balance of \$10,000. A court may conclude that the use of Reading rather than Cohoes pipe is so relatively unimportant to B that the forfeiture that would result from denying A the entire balance would be disproportionate, and may allow recovery by A subject to any claim for damages for A’s breach of his duty to use Reading pipe.⁹⁴

The key to the application of the excuse doctrine in this illustration is the conclusion that the use of Reading pipe is “relatively unimportant” to B. Note that under §229, the court can excuse the condition only if the parties regarded the condition as non-material at the time of formation. The mere fact that the actual failure to satisfy the condition does not materially affect the promisor’s interests ex post is irrelevant.⁹⁵ So the question presented in this stylized version of *Jacob & Youngs* is whether the court is correct in concluding that the condition was not material when the parties included it in their agreement. Implicit in the reasoning of the Restatement is the view that the condition is self-evidently non-material on the facts as stated: No reasonable person would believe that the condition of using Reading brand pipe was material to Kent because the Cohoe brand pipe that was in fact used “is identical in quality and distinguishable only by the name of the manufacturer which is stamped on it.” But then the question is why the parties chose to make Kent’s payment obligations expressly conditional on Jacob & Youngs’ installation of Reading brand pipes only, instead of simply including a promise by Jacob & Youngs to use Reading pipe, thereby subjecting it only to damages for any losses caused by its use of non-Reading pipes rather than a forfeiture of any unpaid balances promised

⁹³230 N.Y. 239 (1921).

⁹⁴R2d §229, ill. 1.

⁹⁵R2d §229, cmt. c. The question is not whether “the actual non-occurrence [of the condition that Reading pipe must be used] happened to involve a departure that was not a material part of the agreed exchange[, but rather] if the occurrence of the condition was a material part of that exchange.”

by Kent.

It is, of course, possible that the choice to draft the requirement to use Reading pipe was careless or made in ignorance of the legal implications of making it a condition rather than a promise. But such a conclusion would require objective evidence to override the strong presumption that a commercially sophisticated merchant such as Jacob & Youngs exercises reasonable care in executing its agreements and knows or should know the legal implications of the express contractual language to which it agrees. Instead, the Restatement illustration invites the court to excuse the condition if it believes that the promisor did not regard the condition as relatively important at the time of agreement, even though the court concedes that the parties deliberately chose to create a condition instead of a promise. When parties use express language to communicate their obligations to each other and a court, however, they should not have to rely on courts to intuit their motivations in order to expect their terms to be enforced as written. In this case, it is easy to imagine reasons why Kent might have placed a high value on the use of Reading brand pipe, even though he knew other brands of equivalent quality were available. We have argued that parties faced with the challenge of specifying performance standards must choose between precise, rule-like terms and vague, standard-like terms. In doing so, they trade off front-end specification costs against back-end verification or enforcement costs. Even if Kent's sole objective were to insure the installation of Reading-quality pipes in his house, he might have intentionally conditioned his payment obligations on the installation of Reading *brand* pipe, rather than Reading *quality* pipe, in order to lower the expected costs of enforcing that requirement. A term requiring Reading quality pipe sets out a standard which allows the builder to use any pipe he pleases and places on Kent the costly burden of proving that the pipe installed by the builder does not conform with the Reading quality standard. A term requiring Reading brand pipe sets out a precise rule which allows Kent to verify performance or non-performance at relatively low cost. It is also possible that Kent had other, atypical reasons for preferring the brand. Perhaps he was skeptical of claims that other brands were "just as good." Or he had an emotional, reputational, or business reason for preferring Reading brand

pipe over pipe of equivalent quality.⁹⁶

Illustration two of R2d §229, based on *Delaware Steel Co. V. Clamar Steamship Corp.*,⁹⁷ reads as follows:

A, an ocean carrier, carries B's goods under a contract providing that it is a condition of A's liability for damage to cargo that "written notice of claim for loss or damage must be given within 10 days after removal of goods." B's cargo is damaged during carriage and A knows of this. On removal of the goods, B notes in writing on the delivery record that the cargo is damaged, and five days later informs A over the telephone of a claim for that damage and invites A to participate in an inspection within the ten day period. A inspects the goods within the period, but B does not give written notice of its claim until 25 days after removal of the goods. Since the purpose of requiring the condition of written notice is to alert the carrier and enable it to make a prompt investigation, and since this purpose had been served by the written notice of damage and the oral notice of claim, the court may excuse the non-occurrence of the condition to the extent required to allow recovery by B.

This illustration implies the general principle that express conditions should be subject to a "no prejudice" standard: they should not be enforced if their purpose has been served by other means. We do not believe this principle is defensible even if current doctrine permitted courts to set conditions aside based on a purely ex post analysis of whether the promisor was harmed by their non-occurrence. Such a doctrine would allow the court to create an exception to material performance obligations to which the parties expressly agreed at the time of formation. Those obligations reflect the parties' design strategy. Changing those obligations ex post therefore

⁹⁶ In the actual case, Kent's payment obligations are conditioned not just on the use of Reading brand pipe, but on Jacob & Youngs' exact conformity with *all* the specifications in the building plans. That fact makes both of the above accounts questionable. Although use of a precise, rule-like term might well make sense as a means of specifying certain dimension of a builder's performance obligation in a construction contract which the owner regards as relatively important, the use of such a term for specifying all of the builder's obligations, down to the last nut and bolt, would predictably expose the builder to strategic fly-specking by the owner. Indeed, because the condition applied to all specifications rather than just a few, *Jacob & Youngs* might usefully be read actually to have enforced the condition as written but to have interpreted the express condition to require the installation of Reading quality, rather than Reading brand, pipe. So interpreted, Jacob and Youngs actually satisfied the condition. The facts in the illustration, however, state only that Kent's payment obligation is conditioned on the use of Reading brand pipe. On the facts so described, there is no justification either for interpreting the condition to require installation of Reading quality pipe or for setting aside the condition.

⁹⁷378 F.2d 386 (1967).

would subject parties to risks and losses they were not paid to bear and relieve parties of risks and losses they were paid to bear.

However, R2d §229 does not, in fact, permit courts set conditions aside based on a purely ex post “no prejudice” standard. As we have seen, it requires that the condition not have been regarded as material by the parties at the time of formation. Even if the court is correct that the non-occurrence of the condition was, in fact, not prejudicial ex post to the interests of the promisor, that fact alone is irrelevant if the condition was material ex ante. In this case, the parties had ample reason to regard the condition as a material component of their agreement. Assuming the court’s claim is correct that the sole “purpose of requiring the condition of written notice is to alert the carrier and enable it to make a prompt investigation,” from the carrier’s ex ante perspective, a term creating a rule enforceable by forfeiture may well constitute the most cost-effective means of insuring either that it receives the notice it needs to avoid certain losses or that it is held harmless whenever it does not receive such notice. Although both parties might anticipate that notice could be provided in many ways, they deliberately chose expressly to condition the carrier’s liability on written notice within a specified time period. Thus, the parties chose a clear, rule-like term to address the carrier’s need for notice rather than the standard-like term requiring actual notice by any means within a reasonable time. Compliance with the precise term is easy to prove and the promisee’s failure to satisfy it relieves the carrier of any liability, thus obviating the need to prove any losses caused by the promisee’s failure to give it notice.

Like all rules, such a term is almost certain to be under- and over-inclusive. In this case, it may well be that the non-occurrence of the written notice did not cause the carrier to suffer any loss because the actual notice it received sufficed to accomplish the same purpose as the written notice was designed to serve. In that case, it is possible that the carrier is simply using the failure of the condition to enjoy a windfall at the expense of the promisee. Although such behavior offends the equitable principles of forfeiture and unjust enrichment, that offense is premised on a purely ex post evaluation of the carrier’s behavior. The offense cannot be sustained if both parties knowingly subjected themselves to a rule-like regime for notice, and the

promisee has already been paid to bear the risk of forfeiture. Likewise, the promisor has paid for the right to invoke the rule even when it can not prove any losses because has suffered none. Similarly, it is possible that the carrier might suffer harm from lack of adequate notice even when the shipper provides it with the required written notice within the time required by the contract. But the under- or over-inclusiveness of rules can provide no ground for refusing to apply the rule. Otherwise, all rules would immediately devolve into standards. Any time a rule was applied to the detriment of one party, it would always have the option to argue that the rule should not apply because its underlying purpose was served anyway. If rules always could be thus avoided, parties could no longer use precise terms to shift contracting costs to the front end of the transaction. In short, the “no prejudice” standard has no place in a regime that recognizes the value of choosing between rules and standards in contract design: conditions serve as the principal means by which parties can choose to regulate their contractual obligations according to a rule-based regime.

Finally, it is important to bear in mind that the court’s surmise of the purpose served by the written notice requirement in this case is not only crucial but fallible. It may seem obvious, but appearances can be deceiving. It is always dangerous for a court to speculate about the particular purpose served by a particular contract term. For example, it is possible that the carrier put in place internal procedures designed to insure that it took prompt and adequate account of all complaints seeking compensation for goods damaged in transit. Those procedures would have been designed on the assumption that the carrier will receive a particular form of written notice within a particular time period for any loss for which it might be potentially held liable. By making the required written notice a condition of its obligation, the carrier sought to obviate the need to prove that a failure to satisfy the condition caused it to suffer a loss. The “no prejudice” standard shifts that burden back to the carrier, even though by including the express condition, the carrier likely paid for that right in the form of a contractual concession to the shipper. Courts cannot avoid speculating on the parties’ purposes when interpreting ambiguous language or implying terms into agreement. But there is no need for such speculation when interpreting express terms that unambiguously create conditions. In almost every case, the intended purpose of such a term is simply to create an enforceable condition and thereby avoid

the need to prove a party's ex ante purpose or ex post loss to a court.

The third, fourth and fifth illustrations of Restatement (Second) §229 each constitute variations on the theme of excusing the timing of the occurrence of a condition. The Restatement states that a court may excuse the non-occurrence of a condition “during the period of time in which it would otherwise have to occur, if it concludes that the time of its occurrence is not a material part of the agreed exchange. This conclusion is sometimes summed up by the phrase that “time is not of the essence.”⁹⁸ Here are the illustrations:

Illustration 3: A contracts to make repairs on B's house, in return for which B agrees to pay \$10,000 “on condition that the repairs are completed by October 1.” The repairs are not completed until October 2. A court may decide that there are two cumulative conditions, repair of the house and completion of the repairs by October 1, and that the non-occurrence of the second condition is excused to the extent of one day.⁹⁹

Illustration 4: On July 1, A makes an option contract with B, under which B has the right to buy land for \$200,000, on condition that he exercise it no later than June 30 five years later. B makes an initial payment of \$10,000 and agrees to make additional \$10,000 payments on or before June 30 of each of the four succeeding years, unless he has already exercised the option, his right being “conditional on his paying the \$10,000 on or before the prescribed date.” These payments are not to be applied to the purchase price. After paying for two years and building on adjacent land, substantially increasing the value of the land subject to the option, B mails a \$10,000 check for the third year on June 30. A receives it on July 1 and returns it to B, stating the option contract is cancelled. A court may decide that there are two cumulative conditions, payment of \$10,000 and payment on or before June 30, and that the non-occurrence of the second condition is excused to the extent of one day.

Illustration 5: The facts being otherwise as in Illustration 4 [above], B makes the payments on June 30 of each of the four succeeding years, but does not exercise the option by tendering the \$200,000 until July 1, following the June 30 expiration date. Even if a court decides that there are two cumulative conditions, payment of \$200,000 and payment on or before June 30, it may not decide that the non-occurrence of the second condition is excused to the extent of one day because that would give B a more expensive option than that on which the parties agreed.

⁹⁸R2d §229, cmt. c.

⁹⁹R2d §229, ill. 3.

Let's begin with Illustration 3, which provides that a court "may" bifurcate the condition into a repair requirement and a timing of repair requirement, and excuse the latter by extending it one day. As always, the example is designed to illustrate how courts can and should refuse to enforce conditions when their enforcement would create a forfeiture. But recall that the question under the doctrine is not whether enforcement would create a forfeiture without imposing significant losses on the promisor at the time of adjudication. Instead, the court must ask whether the promisor would have regarded as significant the prospect of loss that might be created by performing absent the occurrence of the condition. In doctrinal terms, the condition cannot be excused unless it was material to the parties at the time of formation. So whether a court should excuse the timing condition depends on how it determines whether the condition was material at the time of formation. The illustration seems to imply that courts should use their common sense understanding of particular types of contractual settings to assess the likely importance of the condition to the promisor at the time of formation and compare that to the likely forfeiture to the promisee at the time of adjudication. Here, courts properly might conclude that one day usually does not make an appreciable difference to the value a home owner receives from having its home repaired. Presumably, however, the illustration would encourage courts to allow parties to introduce evidence to override their common sense presumption of non-materiality. For example, if the homeowner could establish that both parties knew that one day's tardiness in the repairs would subject the owner to highly speculative, and thus unrecoverable, reputational or other losses, then the court could find that time was of the essence in the contract and that the timing condition was therefore enforceable.

But once again, it is important to note that the parties ordinarily should be presumed deliberately and knowingly to have chosen to frame the timing requirement as a condition rather than a promise. If the homeowner either anticipated minimal damages from delays or that any significant damages would be easy to prove in court, it presumably would not have insisted on subjecting its payment obligation to the timing condition and instead would have been content to put the timing requirement in the form of a promise from the contractor. By including it as a condition rather than a promise, the home owner increased the risk to the contractor and thus raised the price that the contractor would charge him for the repairs. Thus, it is not clear that in

any particular class of contracts, let alone across all contract types, a presumption, or default rule, that time is not of the essence is justified. If anything, courts might well be justified in adopting the default rule that the time is of the essence whenever parties create express conditions that by their plain meaning require events to occur by a date certain. Absent evidence to the contrary, the presumption should be that a date setting a deadline is enforceable as written. A contrary presumption simply increases the costs the parties must incur to communicate their intended contractual means to a court. In this case, the parties could avoid the implied default rule by including an express “time is of the essence” clause in their agreement.

Illustration 4, like Illustration 3, presumes that a court can competently and reliably determine that a one-day late payment to renew an option to purchase real estate is not a material term of the parties’ agreement. The previous discussion of Illustration 3 explains why this presumption is ill-considered. But the contract in the case on which Illustration 4 is based, *Holiday Inns of America v. D.Manley Knight*,¹⁰⁰ contained a clause stating that “it is mutually understood that failure to make payment on or before the prescribed date will automatically cancel this option without further notice,”¹⁰¹ which the court held constituted an express “time is of the essence” clause.¹⁰² Thus, *Holiday Inns of America* cannot be rationalized on the assumption that the court was relying on an implied default rule that deems time not to be of the essence in agreements that fail expressly to state otherwise. Instead, the court excused the condition making time of the essence.¹⁰³

The implication of *Holiday Inns of America* is therefore far more profound than the

¹⁰⁰70 Cal. 2d 327 (1969).

¹⁰¹*Id.* at 328.

¹⁰²*Id.* at 331.

¹⁰³ The court relied on §3275 of the California Civil Code, a rough analogue to Restatement (Second) §229. Section 3275 provides that “Whenever, by the terms of an obligation a party thereto incurs a forfeiture, or a loss in the nature of a forfeiture, by reason of his failure to comply with its provisions, he may be relieved therefrom, upon making full compensation to the other party, except in the case of a grossly negligent, willful, or fraudulent breach of duty.”

Restatement's Illustration 3. Even if parties expressly condition an obligation on a timing requirement and expressly state that the timing requirement is of the essence, *Holiday Inns of America* holds that a court is still permitted to set aside the condition to prevent a forfeiture. Rather than merely second-guessing the parties' contractual design, the court literally overrides the parties' clearly expressed intention to subject the promisor's obligation to a strictly enforced condition of a deadline and substitutes its own judgment that the parties should not be allowed to create such a condition. The regulation of contract terms through interpretation that deliberately abrogates clear terms to which parties have agreed would appear to be a species of the substantive unconscionability doctrine. The comment to Restatement (Second) §229 confirms this view:

Although both this Section and [Restatement (Second)] §208, on unconscionable contract or term, limit freedom of contract, they are designed to reach different types of situations. While §208 speaks of unconscionability "at the time the contract is made," this Section is concerned with forfeiture that would actually result if the condition were not excused. It is intended to deal with a term that does not appear to be unconscionable at the time the contract is made but that would, because of ensuing events, cause forfeiture.¹⁰⁴

But if the term is not unconscionable at the time of contract, why would the forfeiture it creates ex post provide grounds for refusing to enforce it? By abrogating freedom of contract, the rule vitiates party autonomy and undermines the parties' ability to design efficient contracts.

Interestingly, Illustration 5 claims that a one-day late payment to exercise precisely the same option should not be excused "because that would give B a more extensive option than that on which the parties agreed." Illustration 5 follows dicta in *Holiday Inns of America*, which claims that courts may excuse a condition when "the right to exercise the option in the future was forfeited by a failure to pay the consideration for that right precisely on time," but insists that "the time within which an option must be exercised . . . cannot be extended beyond that provided in the contract. To hold otherwise would give the optionee, not the option he bargained

¹⁰⁴R.2d §229, cmt. a.

for, but a longer and therefore more extensive option.”¹⁰⁵ There is, however, no relevant distinction between a deadline condition on a promise to renew an option and a deadline condition on a promise to sell pursuant to the exercise of the option. *Holiday Inns of America* argued that having paid \$10,000 per year for two years, the option holder had “paid a substantial portion of the \$30,000 for the right to exercise the option during the last two years. Thus, they have not received what they bargained for and they have lost more than the benefit of their bargain. In short, they will suffer a forfeiture of that part of the \$30,000 attributable to the right to exercise the option during the last two years.”¹⁰⁶

But if the promisee’s previous payments to renew the option establish a forfeiture that grounds excuse of condition when payment is one day late, then why wouldn’t the full payment of \$50,000 over five years establish a forfeiture that grounds excuse of the condition when the option is exercised one day after the five year deadline? The court argues that the latter would extend the option beyond the term that was bargained for. But by the same token, extending the deadline for renewing the option by one day extends that option—the option to renew the option—by one day as well. Such reasoning can only be explained only by a deep-seated per se policy against the enforcement of conditions that create a forfeiture ex post. This policy serves only to undermine both party autonomy and the efficiency of contract design. Parties have sound reason for creating express conditions and courts have good reason to enforce them.

3. The Law Governing Waiver of Conditions

The final doctrine we consider allows courts to avoid enforcement of express conditions that courts have acknowledged are undeniably contained in the parties’ agreement and are not subject to excuse. The law governing waiver of conditions, nonetheless, allows a court to set aside conditions by holding that the promisor has waived the condition. A promisor can waive a condition by explicitly promising or representing that he will not to enforce it, even if he receives no consideration in return. Or a promisor can waive a condition implicitly by failing to

¹⁰⁵*Holiday Inns of American, supra n. ?* at 330.

¹⁰⁶*Id.* at 331-32.

object if the condition is supposed to be satisfied by the promisee but the promisee fails to satisfy it, or by the promisor performing despite the non-occurrence of the condition. Waiver by election operates to bar the promisor from defending its own failure to perform on the ground that the promisee had previously failed to satisfy a condition of the promisor's prior performance. Waiver by estoppel operates to bar the promisor from refusing to perform because of the non-occurrence of an express condition that the promisor had, implicitly or explicitly, waived not only in the past but in the future as well.

The doctrine of waiver subjects parties to the otherwise unnecessary risk of inadvertent forfeiture of contractual rights and thereby undermines the reliability of the parties' agreed contractual design. We illustrate the problems with waiver law by considering a case in which the parties' sought to eliminate the risk of such inadvertent forfeiture by including an express anti-waiver clause in their contract. In *Fritts v. Cloud Oak Flooring*,¹⁰⁷ a landlord and tenant entered into an agreement that granted the landlord the right to terminate the lease upon the expiration of 10 days following the tenant's default and the landlord's written notice to the tenant of its default. The lease also expressly provided an anti-waiver clause and a time-is-of-the-essence clause.¹⁰⁸ The tenant paid the landlord on time with a check that bounced because funds from a check the tenant had previously deposited in his checking account had not yet cleared. By the time the tenant submitted payment, the deadline for the rental payment had passed. The landlord gave notice of termination and the tenant sought relief. The court found that the landlord had accepted and cashed "ten consecutive tardy monthly rental payments without ever warning the tenant it would seek a forfeiture of the leasehold estate for any future tardiness in payment."¹⁰⁹ The court then held that, notwithstanding the time-is-of-the-essence

¹⁰⁷478 S.W.2d 8 (1972).

¹⁰⁸The contract expressly stated that "no failure of the landlord to exercise any power given him under the lease, or to insist upon strict compliance by the tenant with its obligations thereunder, and no custom or practice of the parties at variance with the terms of the lease should constitute a waiver of the landlord's right to demand exact compliance with the terms thereof. And *paragraph 26* declared that 'time is of the essence of this agreement.'" *Id.* at 10.

¹⁰⁹*Id.* at 13.

and anti-waiver clauses, this conduct constituted a waiver of the landlord's right to terminate upon late payment without first providing advance notice of his intention to strictly enforce the payment deadline. The court held that "[b]y the mere act of including an essence provision . . . the landlord did not immunize or insulate himself from the legal effect and consequence of that course of conduct."¹¹⁰ It then cited Corbin for the proposition that "a provision that an express condition of a promise or promises in the contract cannot be eliminated by waiver, or by conduct constituting an estoppel, is wholly ineffective. The promisor still has the power to waive the condition, or by his conduct to estop himself from insisting upon it, to the same extent that he would have had this power if there had been no such provision."¹¹¹

Fritts provides a dramatic illustration of the role that the law governing conditions, and in particular the law governing waiver of conditions, plays in frustrating the parties' ability to design optimal contracts. Even when parties include not only express terms governing their primary obligation, but also draft meta-terms providing clear instructions to courts about how they intend their agreement to be interpreted, the law of waiver permits courts to ignore the parties expressed intentions. When parties create express conditions, and then include an express clause unmistakably prohibiting waiver of those conditions by subsequent conduct short of modification, there is no reason for courts to refuse to strictly enforce those conditions. As we have explained, conditions serve as critical terms in defining the parties' obligations to cohere with an efficient design strategy. The law of waiver provides an extreme example of how equitable principles can erode, and ultimately eviscerate, the usefulness of the most basic and important means parties have at their disposal to structure their agreements to respond effectively to the complex ex ante and ex post problems of contractual design. Between sophisticated commercial parties, equitable exceptions to the strict enforcement of conditions is counter-productive and unjustified.

..... under the §3275 of the California Civil Code, a rough analogue to Restatement (Second) §229.

¹¹⁰*Id.*

¹¹¹*Id.* at 14.

IV. CONCLUSION

[TO BE SUPPLIED]