Title
The Law and Economics of Costly Contracting

Permalink
https://escholarship.org/uc/item/2wh8m7bv

Authors
Schwartz, Alan
Watson, Joel

Publication Date
2001-12-01
UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

THE LAW AND ECONOMICS OF COSTLY CONTRACTING

BY

ALAN SCHWARTZ

AND

JOEL WATSON

DISCUSSION PAPER 2001-21
DECEMBER 2001
The Law and Economics of Costly Contracting

Alan Schwartz and Joel Watson*

May 2001, Revised December 2001

Abstract

In most of the contract theory literature, contracting costs are assumed either to be high enough to preclude certain forms of contracting, or low enough to permit any contract to be written. Similarly, researchers usually treat renegotiation as either costless or prohibitively costly. This paper addresses the middle ground between these extremes, in which the costs of contracting and renegotiation can take intermediate values and the contracting parties can themselves influence these costs. The context for our analysis is the canonical problem of inducing efficient relation-specific investment and efficient ex post trade. Among our principle results are: (i) The efficiency and complexity of the initial contract are decreasing in the cost to create a contract. Hence, the best mechanism design contracts can be too costly to write. (ii) When parties use the simpler contract forms, they require renegotiation to capture ex post surplus and to create efficient investment incentives. In some cases, parties want low renegotiation costs. More interesting is that, in other cases, parties have a strict preference for moderate renegotiation costs. (iii) The effect of Contract Law on contract form is significant but has been overlooked. In particular, the law’s interpretive rules raise the cost of enforcing complex contracts, and thus induce parties to use simple contracts. Worse, the law also lowers renegotiation costs, which further undermines complex contracts and is also inappropriate for some of the simpler contracts.

*Yale University and University of California, San Diego. This paper benefitted from comments received in workshops at the London School of Economics, the National Bureau of Economic Research 2001 Law and Economics Meeting, UC San Diego, the University of Florida Warrington College of Business, and the University of Southern California Law Center. We are also grateful to Ian Acres, Alvin Klevorick, Tracy Lewis, Joel Sobel, and Kathy Spier for helpful suggestions. Watson thanks the NSF for financial support under grant SES-0095207.
1. Introduction.

Contracting costs play a significant role in recent economic, finance, and law and economics analysis. Among many examples, the high costs of describing possible future states of the world in contracts, and of verifying the realized ex post state, have been cited as contributing to contract incompleteness.\footnote{Relevant work from the transaction cost and hold up literatures includes Coase (1937), Klein, Crawford, and Alchian (1978), Williamson (1979,1985), Grout (1984), Grossman and Hart (1986), Hart and Moore (1990) and Spier (1992).} Further, high contracting costs motivate the default rule project, in which a publicly supplied contract law is explained as providing private parties with terms that are not cost-justified for these parties to write themselves. The models that develop these important results have two relevant features. First, contracting and renegotiation costs are treated as exogenous parameters, commonly assumed to be either very high or very low. For example, an analyst may assume that a particular contract term is too costly to write and that renegotiation is costless, and then ask what follows. Second, the legal system affects contracting and renegotiation costs, but it is unusual for an analyst to model the effect of legally induced costs on the parties’ contracting choices.

This paper explores the middle ground between very high and very low contracting costs and it studies the connection between exogenous costs and actual legal rules. We develop a model in which the costs of writing and renegotiating a contract can take on intermediate values and are partially within the parties’ control. The model addresses the canonical problem of when parties can efficiently implement relation-specific investment and efficient ex post trade. Our analysis yields new insights on the relation between exogenous contracting costs, contractual form, and the contracting parties’ preferences over renegotiation costs. We then ask just how
costs traceable to the courts’ interpretive practices can affect contracting behavior. This paper thus joins a recent literature that formally incorporates transaction costs to explain how firms choose the type of contract to write with investors\(^2\) and how parties choose the level of contractual complexity.\(^3\)

1.1 The current legal and economic understanding regarding contracting costs.

Before summarizing our results and the outline of the paper, we give a brief overview of the current legal and economic perspectives on contracting costs. The law’s goal is to facilitate a court’s ability to ascertain and implement the parties’ intentions regarding the transaction at issue. Formalism — an almost exclusive focus on the written words, read with their dictionary meanings — now is thought to be at odds with this goal. The current legal view implies:

(a) Contextual interpretation: A court’s search for intent should reach beyond the written words to include evidence of what parties said and did during the course of their negotiations. This preference is held with sufficient strength so that party efforts in the writing to limit reference to pre-contractual evidence seldom entirely precludes the introduction of such evidence at trial.

(b) Relaxed requirements of specificity: If a court finds that parties intended to contract but their writing does not settle relevant details, the court fills in the blanks with default legal terms, customary terms in the parties’ industry (if any exist) or “reasonable” clauses. The Uniform Commercial Code (the “UCC”) authorizes a court to fill such gaps as the lack of a price, a specified time for delivery, or a specified product quality.

(c) The relevance of past and current practice to interpretation: A court should consider actions under prior contracts between the parties or actions after the current contract was made when deciding what obligations the current

---

\(^2\)Hart (2000); Tirole (2000). Perhaps the earliest paper in this line is Townsend (1979), who showed that when it is excessively costly to verify a firm’s profits, the firm will reject equity contracts, which condition on profits, in favor of debt contracts, which do not.

\(^3\)Dye (1985); Anderlini and Felli (1994, 1999); Battigalli and Maggi (2001).
contract imposes. For example, a buyer’s practice of accepting nonconforming deliveries under prior contracts may persuade a court to restrict the buyer’s ability to reject under the current contract.

(d) A preference for modifications: The parties’ latest expression of intent is preferred to earlier expressions because courts should implement what parties want, not what they once wanted, and also because later intentions are likely to be better informed than earlier ones. This view sustains the rule that a term in the initial agreement prohibiting renegotiation is unenforceable.⁴

These four implications can affect contracting costs, but it is not the custom for courts to take these costs into account when creating and applying the rules. Rather, the implications best follow from an autonomy view of contract. On this view, contract law rules should require, or aid, a broad judicial search for parties’ actual intent; it is a party’s consent to be bound that legitimates the exercise of state coercion requiring the party to perform or pay damages.

The economic view regarding contracting costs follows from a commitment to efficiency. In the economic view, the costs of writing the initial contract ideally should be zero. When it is costless to contract, and also (relatively) costless to verify relevant actions and later states of the world, parties can write a complete state-contingent contract, prescribing the optimal action for each of them to take in every possible future state. When it is costless to contract, but costly to verify future actions and states, parties can write contracts that induce them to send simultaneous and independent “messages” to a third party decisionmaker.⁵ The messages’ content is a function of information that is unverifiable, yet observed by the contracting parties ex post. If the


⁵A message can be “Seller will deliver twelve units because her production costs are low” or “Buyer will pay $5 per unit because he faces high demand”, or the like. Contracts that induce such messages are generally denoted “mechanisms”, but we use a more detailed taxonomy of contract forms below.
A contractual ban on renegotiation is convenient to enforce when the trading opportunity expires before the court intervenes. In this circumstance, the court’s only role is to order the monetary transfers that the contract requires. An enforceable no-renegotiation clause then would authorize the court to reinstate the monetary transfers that the original contract required rather than enforce the transfers that the renegotiated contract directs. One of the parties commonly would do better under the original contract. In consequence, that party’s renegotiation promise would not be credible. In the standard mechanism design context, in which the court is asked to intervene before parties trade, a contractual renegotiation ban would permit a party later to ask the court to reinstate the transfers that the parties would have made had they sent the messages the original contract required. Again, at least one party would have an incentive, after trade, to petition for the original transfers.

An economic approach to contract choice would diverge from the legal approach in two important ways. First, the legal view ignores the effect of the courts’ interpretive practices on contracting costs. An economic approach should take these costs into account. Second, the current economic approach implies that when contracting is cheap, renegotiation should be costly. This is because low initial contracting costs permit parties to write sophisticated state-contingent or message-based contracts that yield ex post efficiency in every state of the world. Hence a party would want to renegotiate such a contract only “out of equilibrium”, to exploit a contract partner who has made a sunk cost investment. Very high renegotiation costs preclude this behavior. Thus, in complete contrast to the legal view, the economic approach implies that courts should enforce contractual bans on renegotiation.6

---

6A contractual ban on renegotiation is convenient to enforce when the trading opportunity expires before the court intervenes. In this circumstance, the court’s only role is to order the monetary transfers that the contract requires. An enforceable no-renegotiation clause then would authorize the court to reinstate the monetary transfers that the original contract required rather than enforce the transfers that the renegotiated contract directs. One of the parties commonly would do better under the original contract. In consequence, that party’s renegotiation promise would not be credible. In the standard mechanism design context, in which the court is asked to intervene before parties trade, a contractual renegotiation ban would permit a party later to ask the court to reinstate the transfers that the parties would have made had they sent the messages the original contract required. Again, at least one party would have an incentive, after trade, to petition for the original transfers.
1.2 A summary of our theoretical results.\(^7\)

Our model has two important elements. First, it incorporates contracting costs that depend on the parties’ choice of contract form. These costs are increasing in a contract’s complexity. We rank contract complexity, and hence contracting cost, as follows:

*Simple noncontingent contracts*, that specify a single price and a trading decision, are the least complex (and so the least costly to create).

*Option contracts* — that specify sets of prices and trading decisions between which one of the contracting parties can choose — are moderately complex.

*Coordinated message contracts* — that condition on messages that parties simultaneously send — are the most complex contract form and thus the most costly to write.

*Verified contingency contracts* — that directly condition prices and trading decisions on the ex post state (and thus require verification of the state) — are complex and costly, especially to the extent that verification of the state is technologically difficult.

Second, our model lets the parties’ initial contract partly control the parties’ ability to renegotiate.

In particular, the initial contract can affect the portion of the surplus that parties can realize from renegotiation.

Our major results are:

(a) Parties trade off the cost of creating complex contracts against the gain that these contracts create by inducing efficient investment incentives. When the costs of writing the initial contract are low, parties create complex (coordinated message or verified contingency) contracts that induce efficient investment and efficient trade. The higher that initial contracting

\(^7\)We inquire only into the efficiency of various contract forms because the problems we analyze face firms. These are artificial legal entities. Hence, a normative theory whose goal is to protect the autonomy of actual persons appears out of place. Part 5.2 develops further the theme that the regulation of contracts between firms should differ from the regulation of contracts between individual persons.
costs get, the more likely are parties to shift to the simpler contract forms.

(b) Parties have preferences over renegotiation costs, and will contract to affect these costs to the extent that technology and the law permit. When parties create complex contracts, they prefer very high renegotiation costs, because renegotiation interferes with efficient incentives. When parties adopt simple contract forms, they prefer low renegotiation costs because the simple contracts often are suboptimal ex post (when renegotiation allows recovery of surplus). Perhaps most interesting, when parties use the moderately complex option contract, they often strictly prefer intermediate renegotiation costs: high to retain the parties’ investment incentives, but not too high because these contracts are renegotiated with positive probability.

(c) Contract and Commercial Law affect initial contracting and renegotiation costs and so have an important, but overlooked, affect on the parties’ choice of contract form. As an illustration, parties recognize that the costs of writing a particular contract include the expected costs of enforcing that contract. Enforcement costs are increasing in contract complexity and also are influenced by the courts’ interpretive style (the more evidence courts permit a party to introduce in support of its preferred interpretation, the more costly a law suit will be).

(d) The model’s results imply that parties have preferences over what may be called “the rules of the game” (implications (a), (c) and (d) summarized in Part 1.1 above), as well over the substantive terms such as prices and quantities. The rules of the game currently are mandatory. Hence, a major normative implication of our analysis is that there are more mandatory rules in Contract and Commercial Law than there should be.8

Part 2 below begins with an example that illustrates many of our conclusions. Part 3 then sets out the model and Part 4 derives results. Part 5 discusses positive and normative implications of the analysis in more detail. Part 6 concludes.

8This conclusion is consistent with an implication of the mechanism design literature, that the court should enforce whatever the contract dictates, as a function of the messages the parties send. See Schwartz (1998). Eggleston, Posner and Zeckhauser (2001) also suggest, consistently with our analysis, that courts should obey interpretative instructions that parties give them.
2. An example.

A seller and buyer contract to trade one unit of an intermediate good. The contract specifies: (i) a court-enforced mechanism that they must play later in their relationship, and (ii) a renegotiation parameter, \( s \), which gives the share of the renegotiation surplus that the parties can capture if they renegotiate the outcome of the mechanism \( (0 \leq s \leq 1) \). The outcome of the mechanism, before renegotiation, is a determination of whether the good is traded and the price (a monetary transfer from buyer to seller). The contract is costly to write, as detailed below.

After the contract is made, the seller makes a private, unobservable investment that affects the buyer’s valuation of the good, \( v \). The seller either invests “high”, at a cost of 20, or “low” at a cost normalized to zero. If the seller invests high, then \( v = 80 \) with probability \( \frac{1}{2} \), \( v = 20 \) with probability \( \frac{1}{4} \), and \( v = -20 \) with probability \( \frac{1}{4} \). If the seller invests low, then \( v = 20 \) with probability \( \frac{3}{4} \) and \( v = -20 \) with probability \( \frac{1}{4} \). Thus, high investment shifts probability from the value of 20 to the value of 80. The trading decision is costless to the seller, given investment.

After the seller invests, the parties observe the realization of \( v \), which is unverifiable. They then decide whether to trade. Trade is ex post efficient in this example if \( v = 80 \) or 20, but is inefficient if \( v = -20 \). On the example’s parameters, the parties prefer high investment because it and the efficient ex post trade decision yield a joint payoff of

\[
\frac{1}{2}(80) + \frac{1}{4}(20) + \frac{1}{4}(0) - 20 = 25.
\]

while low investment, again with the optimal trade decision, yields a joint payoff of

\[
\frac{3}{4}(20) + \frac{1}{4}(0) = 15.
\]

The first of these expressions includes the investment cost of 20 and both expressions assume that the parties do not trade when the buyer’s valuation turns out to be \(-20\). 
We consider three contract forms in this example: (i) A simple, noncontingent contract, which specifies trade or no trade at a fixed price, and whose creation cost is normalized to zero; (ii) an option contract, under which the trade decision and transaction price depend on a message from one of the parties, and which costs $\alpha$ to write; and (iii) a coordinated message contract, under which the trade and pricing decisions depend on messages from both parties, and which costs $2\alpha$ to write. It is unnecessary to consider verified contingency contracts (which are modeled in Part 3) for the points this example makes. In order best to illustrate the parties’ preferences over renegotiation costs, we let the parties costlessly specify the value that the renegotiation parameter, $s$, will take. Finally, we assume that the parties equally split whatever surplus the contract or renegotiation permit; that is, they have equal bargaining power during renegotiation, as well as during initial contracting.

The simple, noncontingent contract: This contract cannot induce the seller to choose the high investment level (though high investment maximizes the parties’ joint payoff). To see why, let the contract provide that there is no trade ex post, but the buyer nevertheless must pay $p$. This contract would not be renegotiated when $v = -20$, but it would be renegotiated when $v$ takes on either of the higher values. The parties would prefer to set $s = 1$ (renegotiation is costless) in order to give the seller fully one half the renegotiation surplus; this expected return maximizes the seller’s incentive to invest efficiently.

---

9 A verified contingency contract would require parties to expend the resources necessary for the court to verify the “state”, which we could call H when $v = 80$, L when $v = 20$, and N when $v = -20$. The contract would then specify prices and trade contingent on the state.

10 The assumption that parties can costlessly control $s$ is relaxed below. Note that when $s = 1$, renegotiation is costless, so the parties can capture the full renegotiation surplus; when $s = 0$, renegotiation would entirely exhaust this surplus.
A simple noncontingent contract that does set $s = 1$ yields to the seller that invests high the expected payoff of

$$\frac{1}{2}(p + 80/2) + \frac{1}{4}(p + 20/2) + \frac{1}{4}(p) - 20 = p + 2.5.$$  

This seller receives the price $p$ plus half the renegotiation surplus when $v = 80$, which occurs with probability $\frac{1}{2}$, and half the renegotiation surplus when $v = 20$, which occurs with probability $\frac{1}{4}$; the parties do not renegotiate, and so the surplus is zero, when $v = -20$. Subtracting the seller’s investment cost of 20 yields the seller’s expected payoff. A seller who instead chooses the low investment level under this simple contract realizes an expected payoff of

$$\frac{3}{4}(p + 20/2) + \frac{1}{4}(p) = p + 7.5.$$  

With probability $\frac{3}{4}$, the seller now receives $p$ and splits the 20 surplus, and with probability $\frac{1}{4}$ the seller receives only the price $p$.

The seller thus optimally chooses the low investment level. In sum, the best that parties can do under the simple contract is to specify trade (or, equivalently, no trade) and set $s = 1$. This permits them to avoid the trade outcome when $v = -20$ (or realize the entire 20 renegotiation surplus when $v = 20$, if their contract specifies no trade). The parties’ joint expected payoff is

$$Simple\ contract:\ \frac{3}{4}(20) = 15.$$  

The option contract: A contract that lets the trade outcome depend on a message from the buyer sometimes will induce efficient investment.\textsuperscript{11} Suppose the contract permits the buyer to trade at price $p$ or not trade but pay price $p'$, with these prices set so that the buyer will trade only when $v = 80$. Otherwise, the buyer sends the “no trade” message. This message leads to

\textsuperscript{11}One can easily show that a seller-option contract does no better than a noncontingent contract.
renegotiation when \( v = 20 \), and parties let the no trade outcome stand when \( v = -20 \). Such prices must satisfy all of

\[
\begin{align*}
& (i) \quad 80 - p \geq -p' + 80s/2, \\
& (ii) \quad 20 - p \leq -p' + 20s/2, \text{ and} \\
& (iii) \quad -20 - p + 20s/2 \leq -p'.
\end{align*}
\]

Inequality (i) states that, when \( v = 80 \), the buyer does better trading and paying \( p \) than it would do by sending a “no trade” message, paying \( p' \), and then renegotiating to split the surplus of 80.

Inequalities (ii) and (iii) provide that, when \( v = 20 \), the buyer prefers renegotiating from the no-trade outcome rather than sending the “trade” message; and, when \( v = -20 \), the buyer prefers sending the “no trade” message and letting it stand. Rearranging these inequalities yields the following bounds on the contract prices:

\[
(*) \quad 20(1 - s/2) \leq p - p' \leq 80(1 - s/2).
\]

This option contract gives the seller an expected payoff from high investment of

\[
\frac{1}{2}(p) + \frac{1}{4}(p' + 20s/2) + \frac{1}{4}(p') - 20.
\]

With probability \( \frac{1}{2} \), \( v = 80 \) and the parties trade under the contract, the buyer paying \( p \); with probability \( \frac{1}{4} \), \( v = 20 \), the buyer pays \( p' \), and the parties split the trade value through renegotiation; and with probability \( \frac{1}{4} \), \( v = -20 \) so the parties let the contract result stand, the buyer again paying \( p' \). By a similar logic, a seller who chooses the low investment level would receive a payoff of

\[
\frac{3}{4}(p' + 20s/2) + \frac{1}{4}(p') = p' + 15s/2.
\]

The seller thus invests efficiently if and only if

\[
(**) \quad \frac{1}{2}(p - p') - 5s \geq 20.
\]
Condition (**) illustrates that the seller’s incentive to invest efficiently is increasing in the difference between the two prices, because the likelihood that the seller will capture this difference is higher when the seller chooses the high investment level. From condition (*), this price difference cannot exceed $80(1 - \frac{s}{2})$. Substituting this value into (**) yields

$$40(1 - \frac{s}{2}) - 5s \geq 20,$$

which simplifies to $s \leq \frac{4}{5}$.

Regarding the intuition, the parties face a tradeoff regarding the renegotiation parameter $s$. Since the parties renegotiate with positive probability (when $v$ turns out to be 20), they prefer zero renegotiation costs ($s = 1$) in order to capture the full surplus. But when the renegotiation surplus is reduced by positive renegotiation costs, the wedge between what the parties obtain when $v = 80$ and when $v = 20$ widens; this encourages the seller to choose the high investment level and thereby increase the probability that $v = 80$ occurs. Combining these incentives, the best option contract, on the parameters in this example, sets $s = \frac{4}{5}$. Recalling that an option contract is assumed to cost $\alpha$ to write, the parties realize an expected joint gain of

$$\text{Option Contract}: \quad \frac{1}{2}(80) + \frac{1}{4}(20)(\frac{4}{5}) + \frac{1}{4}(0) - 20 - \alpha = 24 - \alpha.$$

The coordinated message contract: As is well known, any contingent split of the investment surplus can be achieved under this type of sophisticated contract, so long as renegotiation is or can be made to be infinitely costly ($s = 0$). These contracts rely on messages that the parties send after uncertainty has dissipated, and punish parties jointly if their messages regarding the ex post state differ. Since the contracts achieve efficiency in every state of the world, renegotiation would only disrupt the mechanism. If $s = 0$ and the other parameters of the example are retained, and recalling that a coordinated message contract costs $2\alpha$ to write, we thus
have that parties to this contract can realize a joint expected gain of

\[ Coordinated\ Message\ Contract: \ \frac{1}{2}(80) + \frac{1}{4}(20) + \frac{1}{4}(0) - 20 - 2\alpha = 25 - 2\alpha. \]

This example illustrates two of our positive themes: the parties’ preferences over contract form are partly a function of trading off the costs of contract writing against the gains of inducing more efficient investment; and the parties’ preferences over renegotiation costs are partly a function of their choice of contract form. In the example, when \( \alpha \leq 1 \), the parties write a coordinated message contract, set \( s = 0 \), and maximize their joint gain. When \( \alpha \geq 9 \), the parties write the simple noncontingent contract, set \( s = 0 \), choose the inefficient investment level, and realize the lowest possible joint gain. And when \( 1 \leq \alpha \leq 9 \), the parties write an option contract, set \( s \) at the intermediate value of \( 4/5 \), induce efficient investment, and achieve an intermediate joint gain.

The example also illustrates our normative theme that the law’s mandatory rules sometimes yield inefficiency. For example, the law attempts to reduce renegotiation cost to zero (see Implication (d) in Part 1.1). When the renegotiation parameter \( s \) does equal one, coordinated message contracts and option contracts seldom could create efficient investment incentives.

Further, this example helps to show that contract law’s interpretive rules can create inefficiency. To see how, recall that the costs of contract creation included the expected costs of enforcing the contract that is written. When a court that is called on to enforce an agreement does not restrict itself largely to the written words, but rather considers contextual evidence (see Implications (a) and (c) in Part 1.1), a party cannot easily win a contract action on summary judgment. The primary evidence in a summary judgment motion will be the written contract.
Context evidence, in contrast, is often verbal and contested. The ability of a party to introduce and contest evidence, especially testimonial evidence, is increasing in the complexity of the contract. Hence, parties to the more complex contracts may anticipate needing expensive trials to enforce their deals. To illustrate the effect of this expectation, suppose that the contract creation cost $\alpha$ in the example would be below 1 (or below 9) when courts are formalist (they largely reject context evidence). Then $\alpha$ could be caused to rise above 1 (or above 9) when courts make extensive reference to context. That is, the chance that the parties will use the efficient contract form is decreasing in the intensity of the courts’ ex post search for the true contractual interpretation.

Parties often would prefer judicial formalism even if a court’s accuracy were increasing in the size of the evidentiary base the court considers.\textsuperscript{12} The usual way of modeling judicial inaccuracy is to suppose that accurate and inaccurate courts both are unbiased, but there is more variance in the expected findings of inaccurate courts.\textsuperscript{13} Risk neutral firms would then prefer a formalist interpretive style, with its greater variance, to a contextual interpretive style that causes them to shift to the simple but less efficient contract forms.

\textsuperscript{12}This position is contested in the law and economics literature.

\textsuperscript{13}See Calfee and Craswell (1986); Che and Schwartz (1998).
3. The model.

We analyze a straightforward extension of the standard model of mechanism design with an external enforcement authority (the court), who acts to implement the parties’ contract.\textsuperscript{14} Contractual mechanisms prescribe trading outcomes as functions of information that the court can access. Contracting and renegotiation are costly, but parties can influence these costs by their choice of contractual form.\textsuperscript{15}

\subsection*{3.1 Model details.}

The relationship between the buyer and seller takes places over five time periods:

\textbf{Time 1:} The parties make a contract, denoted $f$, with two components. The externally enforced component specifies a mechanism that the parties are to play at time 4. The outcome of the mechanism is a tuple $(d, p, s)$, as explained below. The self-enforced component specifies an equilibrium of the mechanism (for each contingency) on which the parties coordinate. A contract $f$ costs $\alpha(f)$ to write.

\textbf{Time 2:} The seller makes an unverifiable and private investment decision $x$, that is chosen from a finite set $X$ at an immediate cost of $\sigma(x)$. The buyer does not observe $x$.

\textbf{Time 3:} A random event determines the state of the relationship $\theta$, which is an element of a finite set $\Theta$. The distribution of the state partly depends on the seller’s investment choice.

\textsuperscript{14}In Anderlini, Felli and Postlewaite (2001), the court maximizes expected ex ante gains from contracting. Our court plays a more passive role, as is commonly assumed.

\textsuperscript{15}Thus, in a nutshell, we add contracting and renegotiation costs to the standard “mechanism design with ex post renegotiation” model (Maskin and Moore 1999; Segal and Whinston, 2001). This is the “complete contract” approach in the sense that mechanisms are permitted, but it is the “incomplete contract” approach in the sense that contracting entails a cost. Tirole (1999) discusses the two polar approaches in the contract theory literature. An accessible review is Schmitz (2001).
The probability that state $\theta$ occurs is denoted $q(\theta, x)$. The value the buyer places on trade and the cost of trade are partly a function of the ex post realized state, which the parties observe at this time.

**Time 4:** The parties play the mechanism their contract specifies. The outcome of the mechanism is a joint *trade decision* $d$, a price $p$, and a recontracting parameter $s$. The decision $d$ is an element of a finite set $D$, and the parties’ preferences over the trade decisions are partly a function of $\theta$. Thus, in some ex post states it may be efficient to trade in a certain way while in other states the same trading decision would be inefficient. The price $p$ is a transfer from the buyer to the seller.

**Time 5:** The parties may recontract to change the outcome of the mechanism. The disagreement point for renegotiation is the outcome that the mechanism specifies. The recontracting parameter $s$ specifies the share of the gains from recontracting that transaction costs do not exhaust. We assume $s \in [0,1]$. For example, if $s = \frac{1}{2}$, then renegotiation dissipates one half of the contractual surplus. The outcome of recontracting is a new trade decision $d'$ and a new price $p'$.

The parties’ payoffs depend on the state, the seller’s investment, the trade decision and price, and the costs of contracting and recontracting. Let $v(d, \theta)$ be the buyer’s value from trade, and let $c(d, \theta)$ be the seller’s cost of producing the traded goods. For example, if the time 4

---

16For simplicity, we focus on proportional renegotiation costs here; our results would not change substantially if we specified more a general cost (including a fixed element). See Brennan and Watson (2001) for an analysis of general renegotiation costs (without initial contracting costs).
decision, $d$, specifies “no trade,” then $v(d, \theta) = 0$ and $c(d, \theta) \leq 0$. Payoffs are linear in the price transfer. Thus, the buyer’s payoff from trade is $v(d, \theta) - p$, and the seller’s payoff is $p - c(d, \theta)$. The \textit{ex post optimal trade decision} in state $\theta$, denoted $d^*(\theta)$, maximizes the joint value of the trading decision, $v(d, \theta) - c(d, \theta)$, by the choice of $d$. We assume that $d^*(\theta)$ is unique for each state $\theta$, and make the following

\textbf{Assumption A}: For each $x$, there exist at least two states $\theta, \theta' \in \Theta$ such that $d^*(\theta) \neq d^*(\theta')$ and $q(\theta, x), q(\theta', x) > 0$.

Assumption A requires that at least two different trading decisions will be optimal with positive probability, no matter the level of investment the seller chooses. This assumption ensures a role for contractual flexibility.

At time 5, parties renegotiate to $d^*(\theta)$ if the mechanism would yield a suboptimal decision $d \neq d^*(\theta)$. The \textit{renegotiation surplus} is given by

$$r(d, \theta) = [v(d^*(\theta), \theta) - c(d^*(\theta), \theta)] - [v(d, \theta) - c(d, \theta)].$$

The first bracketed term on the right hand side is the gain from making the optimal trading decision; the second bracketed term is the \textit{lower} gain that would have been realized had the parties allowed the outcome of the mechanism to stand. There is no gain from recontracting when the mechanism specifies the efficient outcome $d^*(\theta)$; then $r(d^*(\theta), \theta) = 0$.

Uncertainty is resolved by time 4, so the renegotiated contract that replaces the original contract always takes the simple noncontingent form, specifying a price $p'$ and a trade decision $d'$. It must be that $d' = d^*(\theta)$, and parties choose $p'$ to divide the fraction $s$ of the renegotiation

\footnote{It is possible to have $c(d, \theta) < 0$ because the seller could incur a “negative cost” from selling the intermediate good to another party on the spot market (an outside option).}
The generalized Nash bargaining solution has this representation, as do other standard bargaining solutions. We normalize the cost of writing a simple noncontingent contract to zero. Therefore, renegotiation is costly only when the renegotiation friction parameter \( s \) < 1.

Renegotiation is resolved according to fixed bargaining weights \( \pi_B \) and \( \pi_S \) for the buyer and seller, respectively.\(^{18}\) Thus, in state \( \theta \), if the outcome of the parties’ initial contract is \((d, p, s)\), then from time 5 the buyer obtains

\[
z_B(d, p, s, \theta) = v(d, \theta) - p + s \pi_B r(d, \theta)
\]

and the seller obtains

\[
z_S(d, p, s, \theta) = p - c(d, \theta) + s \pi_S r(d, \theta).
\]

The parties’ total payoffs are these amounts minus the seller’s investment cost \( \sigma(x) \) and the initial contracting cost \( \alpha(f) \). How the parties split \( \alpha(f) \) does not affect the analysis.

The mechanism played at time 4 is static: Each party sends a message to the court, which then prescribes the outcome \((d, p, s)\) that the contract dictates given these messages. Let \( M_B \) denote the buyer’s message space and let \( M_S \) denote the seller’s message space. In addition to sending unrestricted messages, the parties also can directly verify none, some, or all aspects of the ex post state to the court. \( M_D \) denotes the set of variables that the parties can verify.

The model collapses verification costs into initial contracting costs for convenience. We suppose that \( M_D = \Theta \), so that the court can process information that directly reveals the ex post state. Courts, however, only know what parties are able to prove. This institutional fact implies that when parties cannot verify the state to the court, any contract \( f \) that conditions directly on \( \theta \)

\(^{18}\)The generalized Nash bargaining solution has this representation, as do other standard bargaining solutions.
would have a cost $\alpha(f) = \infty$ to create; that is, $f$ cannot be written. This modeling strategy also permits analysis of cases when parties make the state verifiable by installing a monitoring technology. In such cases, a contract $f$ that conditions directly on $\Theta$ would cost $\alpha(f)$ to write, where $\alpha(f)$ includes the cost of the technology.\(^{19}\)

A message profile is denoted $m = (m_B, m_S, m_D)$, where $m_B$ is the buyer’s message, $m_S$ is the seller’s message, and $m_D = \emptyset$ is what the court can directly verify. For any message profile $m$, the parties’ initial contract prescribes the outcome $(d'(m), p'(m), s'(m))$. Thus, from time 5 in state $\emptyset$, the parties receive the payoffs given by

$$z_B(d'(m), p'(m), s'(m), \emptyset) \text{ and } z_S(d'(m), p'(m), s'(m), \emptyset).$$

These payoffs, along with the messages spaces, define a game the parties play at time 4. We assume that a Nash equilibrium is played in each state and that, if there is more than one Nash equilibrium in any given state, the parties’ initial contract specifies the Nash equilibrium on which they coordinate.\(^{20}\)

By the revelation principle, we can restrict attention to direct revelation mechanisms and

\(^{19}\)We do not explicitly separate ex post and ex ante costs or address strategic aspects of evidence disclosure. For research on these, see Bull and Watson (2001) and Bull (2001).

\(^{20}\)In technical terms, this is “weak implementation.” Existence of equilibrium is assured because $\Theta$ is finite. However, it is generally not the case that, for a given state, any two equilibria of the message game are equivalent (yield the same payoffs). Equivalence holds in models with free renegotiation (see Segal and Whinston (2001), for example), but may not hold here because renegotiation is costly. We do not allow the contract to specify arbitrary randomization over the outcomes (other than by using the state) for three reasons. First, randomization schemes can be costly to set up; implicitly, we are assuming that the set up costs are prohibitively large. Second, with positive contracting costs, detailed randomization schemes may be of little use. Third, the law also imposes constraints. For example, the rule in UCC §2-716 that conditions a court’s ability to award specific performance on the occurrence of “proper circumstances” may prevent parties from conditioning outcomes on random events that a court would consider irrelevant to the contractual relationship.
equilibria with truthful reporting. Thus, we assume that $M_B = M_S = \Theta$ and look for equilibria in which, in state $\theta$, the parties actually report that $\theta$ is indeed the state. Letting $m_B(\theta)$ and $m_S(\theta)$ denote the messages sent by the parties in state $\theta$, truthful reporting means $m_B(\theta) = m_S(\theta) = \theta$ for each state. Thus, in state $\theta$, the equilibrium message profile is $m(\theta) = (\theta, \theta, \theta)$. To establish an equilibrium with truthful reporting, we must analyze what would happen if players unilaterally deviate, leading to such message profiles as $(\theta', \theta, \theta)$ or $(\theta, \theta', \theta)$.

Let $u_B(x \mid f)$ and $u_S(x \mid f)$ denote the parties’ expected payoffs from time 3, under contract $f$ and investment level $x$.

$$u_i(x \mid f) = \sum q(\theta, x) z_i(d'(m(\theta)), p'(m(\theta)), s'(m(\theta)), \theta),$$

for $i = B, S$, where the summation is taken over $\Theta$. Given a contract $f$ and anticipating behavior at times 4 and 5, the seller chooses the investment level at time 2 that maximizes her payoff. This is the $x^f$ that maximizes

$$u_s(x \mid f) - \sigma(x).$$

Note that $x^f$ may differ from the first-best level of investment $x^*$, which maximizes

$$\sum q(\theta, x)[v(d^*(\theta), \theta) - c(d^*(\theta), \theta)] - \sigma(x),$$

where the summation is taken over $\Theta$. At time 1, the parties select the initial contract $f^*$ that maximizes the joint value of their relationship which, as a function of their contract $f$, is

$$u_B(x^f \mid f) + u_S(x^f \mid f) - \sigma(x^f) - \alpha(f).$$

3.2 Contracting costs: interpretation and assumptions.

Contracting and recontracting costs are represented by the function $\alpha$ and the variable $s$. The former gives the cost of writing an initial contract $f$, which generally comprises intrinsic
elements as well as elements that the law influences.\textsuperscript{21} The variable $s$ represents recontracting costs that partly occur naturally but also are a function of the parties’ contract and the legal rules. Complex contracts — those having a greater number of clauses or requiring a court to evaluate information from many different sources — are assumed to be more expensive to write than are simple contracts. To capture this idea, we adopt a formulation that is along the lines of Dye (1985), Anderlini and Felli (1994, 1999), MacLeod (2000), and Battigalli and Maggi (2001).\textsuperscript{22} In Battigalli and Maggi’s analysis, for example, a contract is a series of clauses linking combinations of various possible “inputs” (that they call \textit{elementary events}) to prescriptions of behavior (that they call \textit{elementary actions}). In our model, the inputs are message profiles; the prescriptions are the possible outcomes of the mechanism, $(d, p, s)$. For example, individual elementary events are: $m_{B} = \emptyset$ (“the buyer sends message $\emptyset$”); $m_{S} \neq \emptyset$ (“the seller does not send message $\emptyset$”); and $m_{D} = \emptyset$ (“the court verifies that the state is $\emptyset$”).

We need not focus on the technology of clause writing, but it is helpful to isolate certain components of contract creation costs on which the model depends.\textsuperscript{23} There is a cost $\alpha_{B}

\textsuperscript{21}Examples of ex ante contracting costs are (i) effort and time that parties spend determining possible contingencies, calculating optimal terms, and drafting language; (ii) payments to third parties, such as attorneys, who facilitate this activity; and (iii) technological investments that make messages or state verification possible. Examples of ex post costs (that we collapse into ex ante costs) are (iv) expenditures of time and money that the parties make during litigation or dispute resolution processes and (v) risk premia that risk-averse parties forfeit when enforcement has a random element.

\textsuperscript{22}See also Gray (1978).

\textsuperscript{23}Battagalli and Maggi (2001) associate a cost with each separate instance in which the contract refers to an elementary event or action. Further, they differentiate between the cost of the initial reference and the cost of later references. In our model, any contract $f$ with externally enforced components $d$, $p$, and $s$, can be analyzed by considering the cost of creating a series of clauses that represent $f$. Parties are assumed to use the most efficient language possible; that is,
associated with creating the ability to send message $m_B$; a cost $\alpha_S$ associated with creating the ability to send message $m_S$; and a cost $\alpha_D$ associated with creating the ability to send message $m_D$ that directly verifies the ex post state. A cost $\alpha_i$ is not paid if and only if the functions $d^i$, $p^i$, and $s^j$ are all constant in $m_i$ — that is, if the trading and pricing decisions and the renegotiation parameter do not depend on the message from channel $i$.\textsuperscript{24} Parties also incur a contracting cost $\gamma$ in order to specify a value of the renegotiation parameter $s$ that differs from the default parameter $S$.\textsuperscript{25}

The costs $\alpha_B$, $\alpha_S$, $\alpha_D$ and $\gamma$ relate to the “stark” aspects of contracts — whether the outcome is contingent on messages and whether the contract affects the renegotiation parameter. To see what is meant by “stark”, consider a contract that specifies trade of five units if and only if the buyer sends the message “The state is H; send five units;” otherwise, the contract specifies no trade. We let $\alpha_B$ be the cost of sending such a single buyer message; hence, this contract costs $\alpha_B$ to write. The parties could write a more complicated contract that also conditions only on buyer messages. Such a contract could recite: “The buyer takes twelve units if he announces that the state is H; the buyer takes five units if he announces that the state is L; there is no trade if the buyer sends any other message (or none)”. Parties would incur a cost greater than $\alpha_B$ to write this more complicated contract because the contract partitions the buyer’s message space more finely.

\textsuperscript{24}This is the formal reasoning underlying our assumption that it is costless to write a simple noncontingent contract, which does not require messages or verification.

\textsuperscript{25}In some of what follows, we assume that parties can choose the renegotiation parameter freely, but this choice actually is subject to two constraints: (i) some recontracting costs may be exogenous; and (ii) the legal rules may restrict the parties’ freedom. Part 5 discusses the second constraint.
Contracting costs not captured by \( \alpha_B, \alpha_S, \alpha_D, \) and \( \gamma \) are denoted “complexity costs”. We do not analyze complexity costs in detail, but do make one simplifying assumption about the contracting cost structure:

**Assumption B**: It is costless to specify an outcome \((d, p, s)\).

We group the set of possible contracts that parties can write into four contract forms:

*Simple noncontingent*. Under this contract, the functions \( d', p', \) and \( s' \) are constant: The trade decision, price, and recontracting parameter do not depend on messages. A simple noncontingent contract costs \( \alpha(f) = 0 \) to write if \( s' = s \), and costs \( \alpha(f) = \gamma \) if \( s' \neq s \).

*Verified contingency*. A verified contingency contract prescribes a trading outcome that is conditioned only on verifiable evidence regarding the realized state, not on the parties’ messages. Parties must incur \( \alpha_D \) to create this contract form. Parties also incur \( \gamma \) if \( s' \neq s \) is specified in at least one contingency, and will incur complexity costs if they contract on several ex post states.

*Options*. The outcome under an option contract is a function of either the buyer’s message or the seller’s message, but not both. An option contract gives a party the option of trading at the specified prices or renegotiating. In the buyer-option case, contracting costs thus comprise \( \alpha_B, \alpha_D \) (if the contract requires the court to verify a datum directly), \( \gamma \) (if parties vary the default renegotiation parameter), and possibly complexity costs.

*Coordinated message*. The outcome depends on the messages of both the buyer and the seller. Contracting costs must include \( \alpha_B + \alpha_S \), and may also include \( \alpha_D, \gamma \), and complexity costs.

On our assumptions, simple noncontingent contracts are the least costly to write and coordinated message contracts are the most costly.
4. Results.

The standard “renegotiation-proofness principle” treats renegotiation as a constraint on contracting. Under the principle, parties can emulate the outcome of any ex post renegotiation with an appropriately designed mechanism (that specifies efficient outcomes in equilibrium). Because parties can achieve with a contractual mechanism everything they can achieve with ex post renegotiation, parties are assumed to prefer infinite renegotiation costs. The renegotiation-proofness principle does not necessarily hold under costly contracting, however, because emulating renegotiation may require a sophisticated mechanism that would be too expensive to construct and implement. Further, the principle does not hold when there are renegotiation costs, as Brennan and Watson (2001) show.

That renegotiation itself may be optimal in some contract scheme when contracting and recontracting costs are positive raises the question just how these costs affect the parties’ ability to achieve desired outcomes with particular contractual forms. In attempting to answer this question, we make the simplifying

**Assumption C:** Complexity costs are zero; that is, all contracting costs are summarized by the variables \( \alpha _B , \alpha _S , \alpha _D \) and \( \gamma \). The Appendix proves that versions of the results in the text hold for the more general setting in which Assumption C is relaxed.

Our first result shows that parties prefer very high renegotiation costs when they use coordinated message contracts (the most complicated form). These contracts must deter parties from dishonestly reporting the ex post state. This opportunity is heightened when parties can renegotiate. Verified contingency contracts also are adversely affected by renegotiation because...
they too yield efficient trade decisions and investment if courts enforce them as written. We summarize this logic in

**Proposition 1.** If it is optimal for parties to use either a coordinated message contract or a verified contingency contract and to specify \( s^*(m) \neq s \) for some message profile \( m \), then there is an optimal contract \( f^* \) (of the same form) that specifies \( s^*(m) = 0 \) for all \( m \). Further, \( d^*(\theta, \theta, \theta) = d^*(\theta) \) for each state \( \theta \).

The first sentence in Proposition 1 holds that parties to the specified contracts would prefer renegotiation to be infinitely costly. The second sentence says that this preference is held because \( f^* \) prescribes the ex post optimal trading decision for each state. Regarding notation, recall that the equilibrium message profile is \((\theta, \theta, \theta)\) in state \( \theta \).28

Our next result addresses the contractual form on the other side of the complexity spectrum: the simple noncontingent contract. Because this contract form is constant in the message profile, the contracted outcome can be described without the \( m \) argument.

**Proposition 2.** If the optimal contract \( f^* \) takes the simple noncontingent form, then the following conclusions hold generically.29 If \( s^r \neq s \) then \( s^r > 0 \). Further, the parties will adjust the renegotiation parameter (setting \( s^r = s \)) if \( \gamma \) is sufficiently small.

Proposition 2 holds that parties to simple noncontingent contracts want the renegotiation surplus to exceed zero. As is illustrated in the example in Part 2, the investing party must anticipate

---

28 This and the following Propositions are proved in the Appendix.

29 By “generically,” we mean that the conclusions may fail to hold only in special knife-edge cases of the contracting environment. See the proof of the proposition for elaboration.
receiving sufficient surplus or it will not invest.\(^{30}\)

We denote a contractual relationship as having pure cooperative investment when \(c(d, 0)\) is constant in \(0\) (so that the seller’s investment only affects the buyer’s value of trade). We have for this case

Proposition 2'. In a setting of pure cooperative investment, there is a function \(B(x, d)\) with the following property: If parties use a simple noncontingent contract specifying \(d'\) and \(s'\), and the contract induces the seller to invest \(x'\), then it must be that \(s'\) is bounded from below by \(B(x', d')\).

Further, unless \(x'\) minimizes \(\sigma(x)\), \(B(x', d') > 0\). Finally, if \(x'\) is the highest cost investment (it maximizes \(\sigma(x)\)) and \(s' \neq s\), then it is optimal to have \(s' = 1\).

Proposition 2' holds that when investment is purely cooperative and parties use simple, noncontingent contracts, parties never prefer renegotiation to be infinitely costly; and sometimes prefer it to be costless. Regarding the intuition, cooperative investment directly benefits the buyer, so the seller must be directly motivated to invest. Since the investment outcome is stochastic, simple noncontingent contracts are renegotiated with positive probability, which implies that renegotiation serves the dual purpose of achieving ex post efficiency and ensuring the seller enough surplus to invest efficiently.\(^{31}\)

Propositions 1, 2 and 2' together show: Parties prefer moderate to low renegotiation costs when they use simple noncontingent contracts. In this event, parties would not impose high barriers to renegotiation if they could control the recontracting parameter. On the other hand, parties prefer very high renegotiation costs when they use the more sophisticated verified

---

\(^{30}\)This result is related to Huberman and Kahn’s (1988) conclusion that having the ability to renegotiate can allow parties to write simpler contracts.

\(^{31}\)Thus, Proposition 2' echos the themes of Che and Hausch (1999) and, less directly, Edlin and Reichelstein (1996).
contingency or coordinated message contracts. Parties to these contracts would ban renegotiation (set $s = 0$) if law and the technology permitted.

The parties’ preferences over renegotiation when they use one-sided option contracts depend on the nature of their investment. In the setting of pure self investment, where $v(d, \theta)$ is constant in $\theta$, a seller-option contract with $s = 0$ will induce the first-best level of investment $x^*$. However, with cooperative investment, the optimal option contract generally specifies $s > 0$, as the example in Part 2 demonstrated.

Turning to the contracting stage, Proposition 3 relates initial contracting costs to contractual form when these costs are sufficiently low to enable parties to use more sophisticated contractual forms.

**Proposition 3.** Suppose that the optimal investment $x^*$ cannot be supported using option contracts even when contracting is costless. (a) Fixing the other parameters at positive levels, if $\alpha_D$ is sufficiently small then parties optimally write verified contingency contracts. (b) If parameters $\alpha_M$, $\alpha_g$, and $\gamma$ are small relative to $\alpha_D$, then parties optimally write coordinated message contracts.

To summarize, high initial contracting costs lead parties to choose simple contracts, and in consequence to have a preference for moderate or low recontracting cost. Low initial contracting costs yield more sophisticated contractual forms and a party preference for high barriers to renegotiation. Parties always would prefer the State to set $\xi$ at the level that the parties themselves would choose because then they could avoid paying $\gamma$. This default rule approach to recontracting would be difficult to implement in practice, however, because the optimal $\xi$ varies with the contractual form that parties choose and the particular parameters of their deal.
5. Implications

5.1 Positive Implications

Contracting costs have been relatively neglected as a field of study. As a consequence, no papers we have found directly test the influence of these costs on contract form. This section sets out the empirical predictions that the Propositions above support, and some evidence relevant to them. Given how sketchy this evidence is, our predictions should be taken much more as invitations to do research than as confirmation.\textsuperscript{32}

1. Simple noncontingent Contracts: Contracts are more likely to take the simple noncontingent form when initial contracting costs are high relative to the gains the deal could create. More precisely,

A. Parties are more likely to use simple noncontingent contracts when their relationship is one shot. Regarding evidence, parties under a recent procurement practice write a detailed “master contract’ with a substantial number of terms. The buyer is expected to send a series of orders that specify only the items sought and a delivery time: All other aspects of each shipment are governed by the master contract, which is altered only when exogenous circumstances warrant. This practice suggests that complex contracts may become optimal when parties can spread fixed contracting costs over several deals, and is roughly consistent with the common observation that spot contracting is relatively simple.

B. The law encourages simple noncontingent contracting. As indicated in Part 1, Contract and Commercial Law create a one way rachet in favor of renegotiation. Courts

\textsuperscript{32}Predictions are put as declarative sentences. We set out relevant evidence where we have it.
discourage or do not enforce party efforts to make renegotiation more costly but permit party 
efforts to make renegotiation cheap. This discourages use of the sophisticated contract forms that 
disfavor renegotiation.

(C) The costs of writing state-contingent contracts are increasing in the number of 
relevant future states. This implies that, in periods of high volatility, parties write relatively 
simple contracts and rely on renegotiation to achieve good outcomes. There is some evidence 
relevant to this prediction. First, an index clause indicates that parties are using a verified 
contingency contract; under these clauses, the transaction price in any period is a function of 
verifiable aspects of the ex post state. Volatility increased substantially in the petroleum coke 
industry after 1973. A study of post-1973 contracts\(^33\) reported that the contract mix shifted from 
a primary reliance on contract index clauses to an even split “between those [contracts] relying 
on indexing and those relying on renegotiation”, but that “indexing ... functioned as part of the 
renegotiation process. The index was only expected to be in force for short periods.” Second, 
raw material prices are short-term volatile and commodity contracts seldom condition on future 
states.

2. Parties should prefer renegotiation to be cheap when it is costly to contract, and 
conversely. More precisely:

(A) Parties will attempt to reduce renegotiation costs when they use simple, 
noncontingent contracts or one sided option contracts. Data about renegotiation costs is hard to 
get, but there is a suggestive example. Fixed price contracts are common in raw materials 
markets though there is considerable price volatility. Parties thus anticipate frequent requests for 

\(^{33}\)Goldberg and Erickson (1987).
“adjustments” – i.e., for renegotiation. The cost of renegotiating simple contracts could be high were the decision maker to treat a willingness to make adjustments under certain market conditions as a willingness to make them under all market conditions. In response to this concern, the trade association rules that regulate disputes in many commodity markets commonly exclude evidence of prior accommodations under the current contract, or of accommodations under earlier contracts. This restriction facilitates renegotiation.

(B) Parties are more likely to use “no oral modification” terms, terms that restrict the authority of line agents to modify a deal, or other terms restricting renegotiation when they use more sophisticated contracts. As shown above, parties ex ante prefer not to renegotiate state-contingent and coordinated message contracts.

(C) Parties have an incentive explicitly to require renegotiation when they use the simpler contract forms and investment is cooperative. To understand this prediction, assume that the seller’s investment permits the buyer to use the product more efficiently, and that the seller has rivals. Then, when a simple contract specifies no trade in the ex post state that materializes, the buyer can credibly threaten to purchase the product more cheaply from a rival, even though renegotiation with the original seller would yield a positive gain. The buyer’s ability to make a credible exit threat may increase its bargaining power in renegotiation to the point where the seller would anticipate receiving too little surplus to invest efficiently (recall here Proposition 2’, holding that when investment is purely cooperative, the optimal renegotiation surplus is bounded from below). A possible contractual response to this possibility is to require the buyer to renegotiate in good faith. A good faith renegotiation requirement is difficult to police, and so

---

cannot reduce the buyer’s exit threat to zero. On the other hand, the requirement can increase the buyer’s exit costs by prohibiting such easily verifiable practices as buying elsewhere immediately after uncertainty is resolved or threatening to make a market contract during a renegotiation. Good faith renegotiation or price reopener terms sometimes are seen in long term contracts. Some commentators believe that variance in judicial outcomes is increasing in the size and scope of the evidentiary base. This is because the parties’ ability to know just what bit of evidence a court will later find dispositive falls as the evidentiary base expands. Risk neutral parties are unconcerned with variance as such, but variance can create “derivative” costs. For example, a seller may use the same contract with many buyers. An adverse construction of the contract could substantially reduce the seller’s profit or raise its costs in a period, thereby making it more difficult for the seller to pay creditors and suppliers. A party who believes that variance is increasing in the court’s interpretive base thus would prefer a merger clause. That such a belief is at least plausible supplies another reason to replace the current mandatory interpretive rule with a default.

3. Party efforts to reduce initial contracting costs should be increasing in the complexity of the deals they would like to make. More precisely:

(A) Merger clauses should be more likely in complex deals. A merger clause attempts to restrict an adjudicator’s interpretative base to the written words by excluding evidence of what was said and done during prior negotiations. Restricting the interpretive base is cost reducing because verification costs ($\alpha_D$ above) are reduced when parties need less evidence to verify each contractually relevant fact or obligation.36

(B) There should be a positive correlation between the use of the more complicated state-contingent or coordinated message contracts and the use of arbitrators, for two reasons:

(i) Arbitration proceedings are less costly than judicial proceedings, and specialist


36Some commentators believe that variance in judicial outcomes is increasing in the size and scope of the evidentiary base. This is because the parties’ ability to know just what bit of evidence a court will later find dispositive falls as the evidentiary base expands. Risk neutral parties are unconcerned with variance as such, but variance can create “derivative” costs. For example, a seller may use the same contract with many buyers. An adverse construction of the contract could substantially reduce the seller’s profit or raise its costs in a period, thereby making it more difficult for the seller to pay creditors and suppliers. A party who believes that variance is increasing in the court’s interpretive base thus would prefer a merger clause. That such a belief is at least plausible supplies another reason to replace the current mandatory interpretive rule with a default.
arbitrators are better than generalist courts at evaluating ex post states.

(ii) Arbitrators obey the parties’ interpretive instructions but courts commonly do not. This paper shows that efficiency is increasing in the ability of parties to affect initial and renegotiation costs. Thus, arbitration becomes attractive to parties for whom it may be particularly important to affect these costs – that is, to parties who want to give interpretive instructions to the adjudicator, such as not to consider certain forms of evidence (i.e., prior negotiations) or to enforce the original contract rather than a renegotiated contract. There is some evidence that parties who use arbitration routinely do give interpretive instructions. See Bernstein (1996, 2001). Further, such instructions seem more important in connection with sophisticated contracts, so the use of arbitration may be increasing in contract complexity.

(C) Parties should restrict the use of custom to determine the meaning of contract terms. Parties litigate because one of them contests the existence of a custom, or its applicability to the instant case. Courts resolve these disputes by making independent assessments of an asserted custom’s normative desirability, in general or as applied. That is, courts treat customs much as they treat precedents from other jurisdictions, that courts are free to follow, alter or reject. See Craswell (2000). Thus, litigation costs are increasing in the ease with which parties can introduce evidence of custom. Parties thus have an incentive to preclude resort to custom in adjudication. And commercial parties often do attempt, in contracts and trade association rules, to restrict an arbitrator’s recourse to custom as an interpretive resource (recall that arbitrators obey interpretive instructions).

That so little data exists relating contract costs to contract form implies the need for serious empirical research. Nevertheless, the theory seems plausible and there apparently is little
contradictory evidence. This suggests that it is appropriate to consider the normative relevance of positive contracting costs. Part 5.2 next considers briefly how Contract Law should change.

5.2 Normative Implications

(1) The Parol Evidence Rule: This rule provides that when parties intend a writing to contain all of their rights and duties, evidence of prior or contemporaneous negotiations is inadmissible to show what the writing does. Two questions arise in litigation under this rule: Supposing that a contract can have several parts, (a) Did parties intend the writing fully to memorialize at least some aspects of what their agreement covered? (b) If so, does the writing contain only some or all of the parties’ agreement? A party disadvantaged by a literal interpretation of the words thus has an incentive to introduce evidence that some or all of the writing is incomplete when read in context. Courts encourage this incentive because they permit extensive recourse to prior and contemporaneous negotiations to resolve interpretive disputes. Consequently, the parol evidence rule is less effective in practice than its formal statement might suggest.

(2) The Merger Clause Rule: Parties may respond to this concern by adding a “merger clause” to the writing that recites, in essence: “This contract contains the entire agreement of the parties”. This response may be ineffective. A leading authority claims: “there has been a

37 The courts’ interpretative stance regarding question (a) is summarized in Restatement (Second) of Contracts §209(3), which provides that when “parties reduce an agreement to a writing which in view of its completeness and specificity appears to be a complete agreement, it is taken to be an integrated [that is, complete] agreement unless it is established by other evidence that the writing did not constitute a final expression.” (Emphasis added) The courts’ interpretative stance regarding question (b) is summarized in the Official Comment to §2-209(3) that “a writing cannot of itself prove its own completeness, and wide latitude must be allowed for inquiry into circumstances bearing on the intention of the parties”. (Emphasis added)
tendency to deny such [merger] clauses conclusive effect.38 The current contextualist interpretive legal regime thus is largely mandatory. The example in Part 2 above illustrated the consequence: parties sometimes will shift from the more sophisticated but more efficient contract forms to the more simple but less efficient contract forms.

(3) **Course of Performance, Course of Dealing and Usage of Trade:** The parole evidence rule bars courts from using evidence of prior or contemporaneous negotiations in connection with the current contract, but the rule does not bar the introduction of evidence regarding the parties’ practice under other agreements, the parties’ behavior under the current contract, or the customary meaning of the contract language. Section 2-208 of the UCC (and the Common Law) clarify the effect of this gap in the rule by providing that practice under prior contracts or under the current contract, and “usage of trade” [i.e., custom] “shall be relevant to determine the meaning of the [current] agreement.” The UCC does say that an “express” term shall control if one exists, but goes on to recite that a “course of performance shall be relevant [in a litigation] to show a waiver or modification of any term inconsistent with such course of performance.” As argued above, these rules raise the cost of renegotiation, and thus seem out of place when parties use simple noncontingent contracts or option contracts.

(4) **The No Modification Rules:** Parties prefer to restrict renegotiation when they use state-contingent or coordinated message contracts. The Common Law held that any contract could be modified by a later contract. Courts therefore would not enforce contract clauses banning renegotiation, and also would not enforce clauses requiring modifications to be in writing.39 The

---

38Farnsworth (1999) at 436.

39See Blum (2001)
UCC, in §2-209, reversed the latter rule for sale of goods contracts, but then erected procedural and substantive barriers to the enforcement of no-oral-modification terms. Regarding procedure, such a term must be separately signed by the party that did not propose it. Regarding substance, “an attempt at modification ... can operate as a waiver.” This rule means in practice that if a party takes a costly action in reliance on an oral modification promise, the no-oral-modification term becomes unenforceable. These no modification rules are inefficient; rather, parties should be permitted to choose the renegotiation parameter.40

There is a folk theorem genre of result in the contract theory literature holding that parties already can do this by involving a third entity. The theorem has A and B contracting with each other that if they later renegotiate they must pay $v > 0 to C. The required payment will deter renegotiation. Such three party schemes actually raise the same issues as two party contracts in which parties agree not to renegotiate. Parties to a two party scheme have an incentive to ignore a no renegotiation clause in order to achieve ex post efficiency. Thus the clause, recall, would be effective only if the law permitted parties later to reinstate the transfers that the original contract specified.41

The law also is needed for three party schemes. Were A and B to perform the original contract, though it is ex post inefficient, then C, the third party, would get nothing. C thus has an incentive to negotiate for a portion of the renegotiation surplus in return for waiving his right to receive $v$. If a third party scheme were legally enforceable, however, then C could sue for $v$ after

---

40 Jolls (1997) makes a similar recommendation but she argues, consistent with the prior literature, that parties generally would prefer to set $s = 0$ (that is, ban renegotiation altogether). As we show, parties hold this preference for some contract forms but not others.

41 See note 6, supra.
agreeing to waive this right, just as a party to a contract with a no renegotiation term could renege on his no renegotiation promise. As a consequence, C’s promise to permit A and B to renegotiate in return for a payment that is less than $v$ would not be credible: A and B would realize that renegotiation actually would cost them $v$, and so would prefer to perform.

No modification clauses are absent from current contracts because they are unenforceable (today, parties are held to the promises in the renegotiated contract). Three party schemes seem not to be seen, apparently because they also cannot work without legal enforcement. Current courts are no more likely to enforce three party schemes than two party schemes.  

(5) Agreements to Agree: Simple noncontingent contracts and one sided option contracts may achieve efficiency by specifying performance in some ex post states but no trade in others. Gains from trade were assumed always to exist in the model, however, so parties were expected to renegotiate in the no trade states. Renegotiation ensured the seller enough surplus to motivate her choice of the efficient investment level. As indicated above, this happy outcome may not occur when a buyer can use the threat to purchase from the seller’s rival to capture most of the ex post surplus for himself. Parties sometimes respond to the buyer’s incentive to behave strategically with terms requiring the parties to renegotiate in good faith in specified circumstances. American courts are split on the enforceability of these “agreements to agree.” Some courts think it is too difficult to give content to the obligation (what is “good faith”?), and

---

42The contracting parties perhaps could make a three party scheme robust to collusion by choosing a third party who cannot accept money. For example, if the parties directed that $v$ be paid to the state as a fine and designated a local prosecutor or attorney general to play the role of C, then for A and B to offer C a share of the renegotiation surplus in return for nonenforcement would be an illegal bribe. Public officials, however, seldom lawfully could participate in such schemes.
so do not enforce the clauses; while other courts think they can effectively police the bargaining process and so do enforce. The analysis here suggests that the latter practice is best: Efficiency would be increased if courts attempted always to enforce renegotiation-in-good-faith terms in the contexts modeled here.

The Contract Law rules questioned here seem attractive when the parties are individual persons. In these cases, perhaps the best normative justification for using the state’s power to coerce performance is that the recalcitrant party actually consented to the deal. An effective judicial search for true consent seems to some to require consideration of all relevant evidence, while many of the reforms proposed here would permit parties substantially to restrict a court’s interpretive base. The rhetoric of courts and many scholars regarding interpretation commonly does presuppose a picture of natural persons attempting to contract. The model here, in contrast, applies to two firms with linear utility functions who are attempting to maximize the size of the pie when information is asymmetric, and who are repeat market players. When this is the real picture, efficiency is an attractive normative goal, and it implies changing the current mandatory rules of the game to defaults.

6. Conclusion

This paper embeds positive initial contracting and renegotiation costs in an otherwise standard mechanism design model. The extension yields several interesting implications about party preferences over these costs and over the relation between them. Thus, parties generally prefer low initial contracting costs because this maximizes party freedom to choose the contractual form that is optimal in their circumstances. When parties choose forms that
themselves ensure efficient investment and trade (such as a complete mechanism), they strongly prefer that these contracts not be renegotiated. Initial contracting costs can be high in relation to contractual gains, however, and then parties choose more simple contractual forms that require renegotiation to ensure efficient investment and trade. Our conclusions regarding contracting costs imply the existence of contracting practices that actually are seen, such as the explicit contractual requirement that parties renegotiate in good faith, and party efforts to facilitate renegotiation when they use simple noncontingent contracts.

Contract Law encourages courts to search thoroughly for the parties’ actual intentions in creating the contract and in renegotiating it. We show that this search has yielded mandatory legal rules that make it extremely difficult for parties to restrict renegotiation, and that can increase greatly the cost of creating sophisticated contracts. As a consequence, parties now have legal incentives to use the more simple contract forms, though these may be the least efficient in a world of more cooperative courts. The search for actual intent rather than the intent that is most consistent with the parties’ writing, we argue, is largely misplaced when sophisticated firms prefer to tie courts to the written words. Thus, Contract Law should change materially (in ways detailed above) to reflect the fact that efficiency is the appropriate normative objective for business contracts, and that efficiency is best served by rules that minimize initial contracting costs and, more broadly, that permit parties to choose the interpretative rules that govern their relationship.

Contracting costs seldom are treated endogenously in the theoretical and empirical literature on contracting. Nor has much attention been paid to the law’s effect on parties’ choice of contract form. This paper thus is an early formal cut at a difficult subject. That we are able to
generate a fairly large set of positive and normative implications with a relatively simple model suggests that the law and economics of costly contracting is an important subject.
Appendix: Generalization and Proofs.

This appendix analyzes contracting environments for a weaker version of Assumption C; and it also provides proofs of the Propositions in the text. We start with technical definitions.

Let \( M = \Theta^3 \) denote the message space. Given a contract \( f \), we call a subset \( K \) of \( M \) a contract event if \( K \) represents exactly the set of message profiles that the mechanism maps to a single outcome — that is, for some \( m \in M \), we have

\[
(d^f(m), p^f(m), s^f(m)) = (d^f(m'), p^f(m'), s^f(m'))
\]

if and only if \( m' \in M \).

Any contract can be written as a list of events and their associated outcomes. More precisely, a contract defines a partition of the message space and it specifies an outcome for each element of the partition. Because we assume that it is costless for parties to specify an outcome (Assumption B), contracting cost is treated here as a function of the partition of the message space. This cost is composed of \( \alpha_B, \alpha_S, \alpha_D, \) and complexity costs relating to the fineness or coarseness of the partition. In place of Assumption C, we make the following weaker Assumption C': Contracting costs are weakly increasing in the size of the implied partition of the message space. That is, if contract \( f \) implies a partition that is a refinement of the partition implied by contract \( f' \), then \( \alpha(f) \geq \alpha(f') \).

We call a contract event \( K \) a null event if

\[
K \cap \{(\theta, \theta, \theta) | \theta \in \Theta \}.
\]

Finally, we call \( K \) a state 0 event if \( (\theta, \theta, \theta) \in K \) and either

\[
K \subset \{(\theta, \theta', \theta) | \theta' \in \Theta \}
\]

or

\[
K = \{(\theta', \theta'', \theta) | \theta', \theta'' \in \Theta \}.
\]

If \( K \) is a null event, then it is a set of message profiles that would not occur in equilibrium. If \( K \) is a state 0 event, then \( K \) is either a set of message profiles where the buyer and the seller both report 0, or it is the set of message profiles where \( \theta \) is directly verified.

Proposition 4. There is an optimal contract \( f^* \) with the following properties. Given \( f^* \), every null event and every state 0 event for which \( f^* \) specifies \( s \neq \underline{s} \) turns out to have \( s = 0 \). Further, if \( f^* \) admits a state 0 event, then \( d^{f^*}(0, 0, \theta) = d^*(0) \).

In less formal language, the first conclusion of Proposition 4 is that, for all null and state events of \( f^* \), whenever \( f^* \) prescribes a different renegotiation parameter than the default \( \underline{s} \), the contract bars renegotiation. The second conclusion is that the contract prescribes the ex post optimal trading decision for all state events.

Proof of Proposition 4: Suppose \( f \) is an optimal contract. Represent \( f \) as a partition \( P \) of \( M \) and a list of outcomes, one for each element of the partition. Let contract \( f^* \) specify the same partition \( P \). For each element \( K \) of partition \( P \), we define the outcome specified by \( f^* \) in the following way.
1. If \( K \) is a null event and if \( f \) specifies \( s \neq \xi \) for this event, then let \( f^* \) prescribe the same outcome as specified by \( f \) except with \( s = 0 \).

2. If \( K \) is a state \( \theta \) event and if \( f \) specifies \( s \neq \xi \) for this event, then let \( f^* \) prescribe decision \( d^*(\theta) \) and renegotiation parameter \( s = 0 \) for this event; the price \( p \) is set so that the seller obtains the same payoff under \( f^* \) as she does under \( f \), for \( K \).

3. If \( K \) is a state \( \theta \) event and if \( f \) specifies \( s = \xi \) for this event, then let \( f^* \) prescribe decision \( d^*(\theta) \) and renegotiation parameter \( s = \xi \) for this event; the price \( p \) is set so that the seller obtains the same payoff under \( f^* \) as he does under \( f \), for \( K \).

4. Otherwise, have \( f^* \) prescribe the same outcome as does \( f \) for event \( K \).

Finally, suppose \( f^* \) prescribes the same (truthful) behavior at time 4 as \( f \) prescribes. Contract \( f^* \) has the same cost as does contract \( f \). It also has all of the properties described in Proposition 4. Furthermore, the parties have the same incentives at time 4 — to report truthfully — with contract \( f^* \) as they do with contract \( f \). Finally, by the construction of \( f^* \) (in particular, the way the prices are set), we have \( u_S(x \mid f^*) = u_S(x \mid f) \) for every investment level \( x \); hence, the seller has the same investment incentive. We also have \( u_B(x \mid f^*) = u_B(x \mid f) \). Thus, \( f^* \) and \( f \) have the same cost, \( f^* \) and \( f \) induce the same investment, and \( f^* \) has state-contingent payoffs that are at least as large as the ones under \( f \). This proves that \( f^* \) is optimal. \( Q.E.D. \)

**Proof of Proposition 1:** We use Proposition 4 to prove Proposition 1. Suppose that, under Assumption C, it is optimal for the parties to use a coordinated message contract \( f' \) that specifies \( s \neq \xi \) in some contingency. Since complexity costs are assumed to be zero, this contract will cost \( c_S + c_B + \gamma \). Note that, at the same cost, the parties could write a coordinated message contract \( f \) that has the finest possible partition of the message space and specifies the same outcome for each message profile as does contract \( f' \). Contract \( f \) thus partitions the message space into separate contract events for each of the messages sent by the parties — where every set

\[
\{(\theta', \theta'' \in \Theta) \mid \theta \in \Theta\}
\]

is a separate event, for each \( \theta' \) and \( \theta'' \). Contract \( f \) is obviously optimal. Note, further, that every event in the partition implied by \( f \) is either a null event or a state event. Proposition 4 then implies the existence of an optimal contract \( f^* \) that specifies \( d^*(\theta, \theta, \theta) = d^*(\theta) \) for each state and \( s = 0 \) whenever \( f' \) sets \( s \neq \xi \). In fact, we can assume that \( f^* \) specifies \( s(m) = 0 \) for every message profile \( m \). The same method can be used for the case in which \( f' \) is a verified contingency contract. \( Q.E.D. \)

**Proof of Proposition 2:** Suppose \( f' \) is an optimal simple noncontingent contract specifying \( s' \neq \xi \). Let the seller choose investment level \( x' \) under contract \( f' \). Because \( x' \) solves the seller’s optimization problem at time 2, it is the case that

\[
 u_S(x' \mid f') - \sigma(x') \geq u_S(x \mid f') - \sigma(x)
\]
for every \( x \in X \). From Assumption A, it must be that \( d' \) (the decision prescribed by \( f' \)) is not ex post optimal in some state \( \theta' \) that occurs with positive probability following investment \( x' \). In state \( \theta' \), the parties’ strictly prefer to renegotiate ex post. If the parties’ contract bars renegotiation, however, \( (s' = 0) \), then allowing the parties to share in the renegotiation surplus would disrupt the seller’s incentive to select \( x' \). In other words, the seller’s incentive constraint is binding:

\[
 u_s(x' | f') - \sigma(x') = u_s(x | f') - \sigma(x)
\]

for some \( x \neq x' \). However, this equality occurs only in knife-edge cases. To see this, observe that if, holding all other aspects of the technology fixed, \( \sigma(x') \) were lowered, then the seller’s incentive constraint would hold with slack when \( s = 0 \). The optimal contract would then specify a higher value of \( s \) (so the parties could realize some renegotiation surplus). Further, fixing the other aspects of the contracting environment, parties generally will not prefer the default parameter \( s \) for any investment level \( x \) become only a finite number of values of \( s \) would be optimal. This implies that if \( \gamma \) is low enough, parties will set \( s \neq s \). Q.E.D.

Note that Proposition 2 does not require Assumption C.

**Proof of Proposition 2**: Because \( d(d, \Theta) \) is constant in \( \Theta \), the seller has an incentive to choose investment level \( x' \) only if

\[
 \Sigma q(\theta, x') s' \pi_s r(d, \theta) - \sigma(x') \geq \Sigma q(\theta, x) s' \pi_s r(d, \theta) - \sigma(x)
\]

for all \( x \), where the summation is taken over \( \Theta \). Rearranging this expression yields

\[
 s' \pi_s \Sigma r(d, \theta) [q(\theta, x') - q(\theta, x)] \geq \sigma(x') - \sigma(x)
\]

The bound \( B(x', d) \) can be defined as the maximum of

\[
 [\sigma(x') - \sigma(x)] / \pi_s \Sigma r(d, \theta) [q(\theta, x') - q(\theta, x)],
\]

over all \( x \) for which \( \sigma(x') > \sigma(x) \). The conclusion about \( s' = 1 \) obviously holds when \( \sigma(x') \geq \sigma(x) \) for all \( x \).

**Proof of Proposition 3**: Obvious.
References


Bull, Jesse, “Costly Evidence Production and the Limits of Verifiability,” unpublished manuscript.


De Fraja, Gianni, “After You Sir. Hold-Up, Direct Externalities, and Sequential Investment,” 26


