

# On the Writing and the Interpretation of Contracts

Steven Shavell\*  
Harvard Law School

The major theme of this article is that the interpretation of contracts is in the interests of contracting parties. The general reasons are (a) that interpretation may improve on otherwise imperfect contracts; and (b) that the prospect of interpretation allows parties to write simpler contracts and thus to conserve on contracting effort. A method of interpretation is defined as a function whose argument is the written contract and whose value is another contract, the interpreted contract, which is what actually governs the parties' joint enterprise. It is shown that interpretation is superior to enforcement of contracts as written, and the optimal method of interpretation is analyzed.

## 1. Introduction

The major theme of this article is that the interpretation of contracts—their possible amplification, correction, and modification by adjudicators—is in the interests of contracting parties. The reasons are no doubt well-appreciated in at least a general sense: interpretation may improve on otherwise imperfect contracts; and the prospect of interpretation allows parties to write simpler contracts and thus to conserve on contracting effort.

As background, we know from common experience that parties may fail to provide for certain events in their contracts (suppose that they overlook the possibility of a leap year) and that they often employ broad terms that do not reflect their wishes in particular circumstances (suppose that they specify that material A should be used in construction but that they would really prefer substituting material B if an unusual problem arises with A). To explain why parties write such incomplete contracts, it is frequently suggested that many eventualities are hard to anticipate or describe in advance and that leaving out details saves time and effort.<sup>1</sup>

---

\*Samuel R. Rosenthal Professor of Law and Economics, Harvard Law School, Cambridge, MA 02138. Email: shavell@law.harvard.edu.

I wish to thank Ian Ayres, Pierpaolo Battigalli, Lucian Bebchuk, Richard Craswell, Gillian Hadfield, Oliver Hart, Louis Kaplow, Avery Katz, Eric Posner, Richard Posner, Alan Schwartz, Kathryn Spier, Jean Tirole, Joel Watson, and two anonymous referees for comments, and the John M. Olin Center for Law, Economics, and Business at Harvard Law School for research support.

1. This explanation for incomplete contracts was earlier emphasized by Williamson (1985). For discussion of it and of other reasons for incompleteness, see for example Hart and Holmström (1987), Hart and Moore (1999), and Tirole (1999).

We also observe that the courts actively engage in the interpretation of contracts. The courts fill gaps in contracts, resolve conflicts and ambiguities of language, and sometimes replace the parties' express terms with the courts' terms (such as to permit substitution of material B if a problem with A occurs) (see generally Farnsworth, 1999: chapter 7; and Section 6 below). Moreover, the interpretation of contracts is widely understood to influence how parties write contracts: the more closely the courts' interpreted contracts resemble the parties' true wishes, the more willing the parties are to leave gaps and to write fairly general terms, whereas parties are more willing to take extra pains to write more detailed contracts when courts refrain from interpreting terms or interpret terms in ways that run counter to their true desires.

Given this motivation, the writing of contracts and the courts' interpretation of them is examined here in a basic model of contracting, and the optimal method of interpretation is investigated.

In Section 2, the main assumptions of the model are stated. These include that parties are risk-neutral, that they have symmetric information, that they do not renegotiate contracts, that all variables are contractible, but that writing contracts involves costs that rise with the number of contractual terms.<sup>2</sup> A method of interpretation is defined as a function whose argument is the written contract and whose value is another contract, the interpreted contract. It is the interpreted contract, not the written contract, that actually governs the parties' joint enterprise.

In Section 3, the type of contract that parties choose to write is examined. In particular, it is asked when parties decide to stipulate specific terms, when they elect to write broad terms, and when they wish to leave gaps, given the courts' method of interpretation of contracts and given the cost of writing additional terms. Thus, note that how parties write contracts is influenced not only by writing costs, but also by the method of interpretation.

In Section 4, the courts' optimal method of interpretation of contracts is considered. For any method of interpretation, one can determine the contracts that parties will decide to write (as described in Section 3), their costs of doing so, the interpreted contract that will actually be employed, and consequently the expected payoff from the written contract. Of course, the contracts that are written will vary among contracting parties, depending on their underlying situation—the payoffs that they enjoy, given their acts and the contingencies that obtain. The underlying situation of a pair of contracting parties (the payoff function) is called their contractual type. It is assumed that the contractual type of contracting parties is not observable to the courts, so that the courts must use the same method of interpretation for all contracting parties. The optimal method of interpretation maximizes the expected value of contracts net of writing costs over the population of contractual types.

---

2. These assumptions are perhaps those that allow the most transparent development of the points of interest. However, as will be discussed in Section 6 and is noted later in the introduction, many of the qualitative conclusions hold independently of the contractual environment.

The conclusions about the optimal method of interpretation can be summarized as follows. First, some method of interpretation of contracts is always socially desirable: the optimal method of interpretation is strictly superior to literal enforcement of contracts as written. This is true because there exists, at the least, a way of filling gaps that allows some parties to reduce the number of terms in their contracts.

Second, the optimal method obeys a simple, fundamental necessary condition: the interpretation of a term in a contingency is that which maximizes the expected payoff for the subgroup of types who write the observed contract. (For instance, the optimal interpretation of a general term calling for use of material A is determined by what material is best in the contingency for the conditional distribution of types who write the observed contract.) In other words, to determine the optimal interpretation of a contractual term, the court behaves naively, as if its method of interpretation does not influence the set of contracts that are written (even though its method of interpretation generally does have this influence).

Although a necessary condition for optimal interpretation is that interpretation is best for those who write the observed contract, this is not a sufficient condition for optimal interpretation. For example, suppose that a gap might be filled in one of two ways: either as the majority wants or as the minority wants. If the gap is filled as the minority wants and this induces the majority to write a term reflecting their wishes, then the gap is filled in the optimal way for those (the minority) who leave the gap, so the necessary condition is satisfied. But this method of interpretation is not optimal, for it would be better to fill the gap as the majority wants, in order to reduce writing costs (then only the minority will bear the cost of writing a term reflecting their wishes). And if the gap is filled as the majority wants and this leads the minority to write a term reflecting their wishes, then the gap is filled in the optimal way for those (now the majority) who leave the gap, so the necessary condition is also satisfied.

Third, specific contractual terms are interpreted as they are written (so that if a term stipulates that material A should be used in a single, fully described contingency, this term will be respected by the courts)—in other words, it is not optimal for the courts to override specific terms—provided that the method of interpretation displays an independence property. Otherwise, it is possible that it is optimal for a specific term not to be interpreted as written, because a specific term may serve as a signal about the contracting parties' desires.

Fourth, it is sometimes desirable for the courts to override a general contractual term (such as a term requiring the use of material A in a broad set of circumstances). The reason that overriding a general term might be desirable is, on one hand, that the parties may prefer that a different action from that provided in the term be taken when certain problematic contingencies arise (the parties might be made better off if material B is substituted for material A if the use of A becomes difficult). On the other hand, the reason that the contracting parties might write the general term to begin with is, as indicated

at the outset, that they find this economical in order to save writing costs. Note too that bound up in the statement that the parties would often want to write a general term is that they are better off doing that than leaving gaps, for general terms frequently indicate their desires.<sup>3</sup>

The point of the foregoing paragraph bears emphasis: contractual terms that give outwardly clear, unambiguous instructions for contingencies (such as using material A in a broad range of circumstances) may sometimes be best for the courts not to enforce, because the parties do not really want the terms enforced as written in the particular contingency that occurred. Realistically, this point has substantial importance because of the omnipresence of general terms in contracts (on reflection, essentially any term is seen to cover multiple contingencies—a specific term is a theoretical ideal<sup>4</sup>). Thus, latent in a court's proper role with regard to virtually any contractual term is the possibility that the term should not be enforced as written.<sup>5</sup>

A fifth point follows from the fourth. Although it may be desirable for the courts not to enforce a contractual term as written, but rather to interpret it, because many contracting parties wish for that in certain circumstances, *some* contracting parties may not want the term overridden (some parties may want material A to be used even if a problematic contingency occurs—they do not want B to be substituted for A). This implies that it would be desirable for contracting parties to have a no-interpretation option (an “I really mean it” option) for each term. If exercised, this option would imply that the courts would enforce the term exactly as written.

In Section 5, the model is extended to allow for the possibility of presentation to courts of evidence beyond the contract (for instance, the contractual negotiation history, the parties' course of dealing, usual trade practice) at a cost. It is assumed for simplicity that this extrinsic-to-the-contract evidence is perfect, allowing the courts to determine the ideal contractual term. The effect of the ability to present such extrinsic evidence is determined, one of the main points being that specific terms are needed less often. It is also observed that the decision about the use of evidence ought to be made by the parties, not by the courts.

---

3. For instance, a person who is renovating his kitchen may specify that he wants an oak floor without qualification, even though he would not want oak three percent of the time, when certain unusual contingencies arise, such as that a shortage in oak develops and delivery is delayed by six months. In these contingencies, the person might want another hardwood (perhaps maple) substituted for oak, and the courts might well infer this preference for hardwood (rather than for tile or a composite material) from the fact that the term in question specifies oak.

4. Even a highly detailed term (such as a term specifying that oak should be used for a kitchen floor unless there is a delay in supply of over six months or a price increase of over 100%) typically omits explicit mention of a multitude of potentially relevant contingencies (such as that oak becomes subject to an insect pest or that a new kind of wood is discovered that is more durable than oak, essentially the same in appearance, and cheaper).

5. This conclusion should not be viewed by legal readers as an endorsement of an activist role for courts in the interpretation of contracts. The conclusions reached in the present article pertain to a stylized model and, as such, must be applied with due caution to the actual world of contracts and judicial practice.

In Section 6, concluding comments are made about the generality of the conclusions, where it is explained that many (but not all) of the results hold regardless of assumptions made about the contractual environment, notably, concerning the contractibility of variables, renegotiation to avoid ex post inefficiency, and the information of the parties. Comments are also made about actual legal practice in light of the analysis.

Before proceeding, the relationship between this article and the economic literature on contracts should be noted. Articles of relevance include those concerning the costs of writing contracts; see Dye (1985) and, for example, Anderlini and Felli (1999), Battigalli and Maggi (2002), and Schwartz and Watson (2004). In these articles, the question considered is how parties ought to simplify their contracts in order to save writing costs, but it is presumed that the contracts will be enforced as they are written. Hence, the issue of the interpretation of contracts and its effects on the writing of contracts is not studied. More generally, the usual assumption in the incomplete contracting economic literature is that contracts are enforced as written, so that the interpretation of contracts is not examined.<sup>6</sup>

However, there are a number of articles (mostly in law reviews and law and economics journals) that do address, or touch upon, interpretation. See Anderlini, Felli, and Postlewaite (2001); Ayres and Gertner (1989, 1992); Bebchuk and Shavell (1991); Bernstein (1996); Goetz and Scott (1985); Hadfield (1994); Eric Posner (1998); Richard Posner (forthcoming); Schwartz (1992); and Schwartz and Scott (2003). While suggestive, these articles do not state and develop the general view of interpretation set out here as a function that transforms the written contract into the interpreted one.<sup>7</sup>

---

6. See for example the presentation of the contracting literature in Bolton and Dewatripont (2005). Some articles, however, mention informally that courts engage in interpretation; see for instance the survey by Hart and Holmström (1987:148).

7. Anderlini, Felli, and Postlewaite (2001) consider a model in which courts engage in a simple form of interpretation, voiding or not voiding a contract; the voiding of a contract can provide an implicit insurance benefit to the parties, who cannot contract in a set of indescribable contingencies. Ayres and Gertner (1989, 1992) and Bebchuk and Shavell (1991) focus on the specific point that courts can fill gaps so as to induce parties with private information to reveal it at the time of contracting. Goetz and Scott (1985) emphasize the difficulties courts face in interpreting non-standard express terms. Hadfield (1994) examines a model of the interpretation of best-effort clauses (she assumes that parties do not contract over effort levels but that courts observe effort levels and penalize parties for inadequate effort). Eric Posner (1998) focuses on the issue of whether courts should restrict attention to the written contract or consider evidence extrinsic to it. Richard Posner (2005) examines the relationship between contractual detail and reduced litigation-related interpretation costs. Schwartz (1992) suggests that in practice, courts often interpret contracts in ways that are not consistent with parties' wishes. Schwartz and Scott (2003) emphasize a similar theme in their discussion of interpretation, and Bernstein (1996) argues that evidence from contract dispute resolution in trade associations indicates that commercial parties generally want to avoid interpretation. Cohen (2000), Craswell (2000), and Katz (1998) survey contract interpretation from an economic viewpoint.

## 2. Assumptions

There is a population of risk-neutral contracting pairs. The joint monetary *payoffs* to a contracting pair depend on the *act* and the *contingency* that occurs. In particular, define

$a$  = an act, where  $a \in A$ , the universe of possible acts;

$\theta_i$  = a contingency, where  $\theta_i \in \Omega$ , the set of  $n$  possible contingencies;

$p_i$  = probability of  $\theta_i$ ;

$x(\theta_i, a)$  = joint payoff to a pair of parties if the contingency is  $\theta_i$  and the act is  $a$ .

An act may be interpreted as a vector with components corresponding to behaviors of both contracting parties. The occurrence of contingencies and acts is assumed to be verifiable by the courts. The joint payoff to the parties is the sum of the payoffs to each individually (for instance, the value of a good to the buyer minus the production costs incurred by the seller). The act(s) that maximize  $x(\theta_i, a)$  for a given contingency  $\theta_i$  will be called the *ideal act* given  $\theta_i$ ; this will be denoted as  $a^*(\theta_i)$ .<sup>8</sup>

A pair of contracting parties is identified by its *type*; let

$t$  = type of pair of contracting parties, where  $t \in T$ ,  
the universe of possible types.

The type  $t$  stands for the parties' contractual situation and is a parameter of the payoff function,  $x$ , which will sometimes be written  $x(a, \theta_i, t)$ . Also, the ideal act will sometimes be written  $a^*(\theta_i, t)$ . The courts cannot observe  $t$  but know its probability distribution; let

$F(t)$  = probability distribution of types  $t$ .

(The type  $t$ , however, will usually be suppressed in the notation.)

A *contract*  $K$  will be identified with a list of events and the act to be taken in each of the events. Let

$$K = \{(E_1, a_1), \dots, (E_m, a_m)\},$$

where the  $E_j$  are mutually exclusive (but not necessarily exhaustive) events in  $\Omega$ , and  $a_j$  is the act to be taken if  $E_j$  occurs. An event and the associated act, that is,  $(E_j, a_j)$ , is called a *term* of the contract.<sup>9</sup> A contract will also have a contract price, but our focus will be on its terms, and for ease,  $K$  will be

8. For simplicity, I will usually discuss the ideal act as if it is unique.

9. The description by parties of events and acts is implicitly assumed to be clear, so that no dispute can arise concerning what event or act was meant; thus, ambiguities of language are not taken into account in the model.

called a contract even though  $K$  leaves out the price.<sup>10</sup> If the event  $E_j$  in a term is a single contingency  $\theta_i$ , the term is called a *specific term*; otherwise, the term is called a *general term*, as it names an act to be done for more than one contingency. A contract is said to be a *fully detailed complete* contract if it has a specific term for each contingency. If it is not fully detailed but does provide for all contingencies through its general terms, it is *obligationally complete* (or simply *complete*); that is, an obligationally complete contract is such that the union of the events  $E_j$  in the contract is the universe  $\Omega$  of all possible contingencies. A contract has *gaps* if it is not obligationally complete, in other words, if the union of the  $E_j$  leaves out at least one contingency.<sup>11</sup>

Consider an example that will be amplified below. There are four alternative materials, A, B, C, and D, that the seller can use in making something,<sup>12</sup> and there are two possible contingencies, a normal contingency,  $\theta_1$ , and an unusual, problematic contingency,  $\theta_2$ . A contract that says to use A in contingency  $\theta_1$  and to use D in  $\theta_2$  is fully detailed; this contract has two terms,  $\{(\theta_1, A), (\theta_2, D)\}$ . A contract that says to use C no matter what is not fully detailed but is obligationally complete; the contract has one general term,  $\{(\theta_1 \text{ or } \theta_2, C)\}$ .<sup>13</sup> A contract that says only to use B in  $\theta_1$  has a gap because it does not provide for  $\theta_2$ ; the contract is  $\{(\theta_1, B)\}$ .

It is assumed that a contract involves a *writing cost*, which is a positive amount per term.<sup>14</sup> Let

$\alpha$  = cost of writing a term;  $\alpha > 0$ .

Thus, in the example, the contract that says to use A in  $\theta_1$  and D in  $\theta_2$  costs  $2\alpha$  because it has two terms, whereas the contract that says to use C no matter what costs  $\alpha$  because it has one term. Let

10. As will be seen, the assumption being made is that the method of interpretation does not depend on the contract price, only on its terms. This is a simplifying assumption because it means that when parties bargain over price, they are not thereby affecting the interpretation of the contract and thus its value. Were the assumption not made, it can be shown that much of what is to be said would continue to hold (for instance, Propositions 1, 2, 3, and 5), or that analogous conclusions would hold.

11. The term "obligationally complete" is first employed in Ayres and Gertner (1992). An obligationally complete contract might still be described as "incomplete" in the economics literature, since in that literature, incomplete contracts are less detailed than might be desirable for the parties but do not have gaps.

12. It is expositionally convenient to denote the acts in this example by A, B, C, and D rather than by  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$ . The example involves four acts because it turns out that this is the minimum necessary to illustrate the full range of possible outcomes that will be of interest.

13. Admittedly, in this simple example, with only two contingencies, the difference between writing the general term "use C whether  $\theta_1$  or  $\theta_2$  occurs" and writing the pair of specific terms "use C if  $\theta_1$  occurs" and "use C if  $\theta_2$  occurs" is not great. However, in realistic situations, the event  $E$  in a general contractual term will cover a vast multitude of contingencies, so that writing "use C if  $E$  occurs" is much easier than naming all the  $\theta$  in  $E$  and writing a separate term for each saying that C should be used.

14. If the cost of making contracts depends in a more complicated way on the description of acts and events, it will be evident that the qualitative nature of the major conclusions would not be altered.

$z(K)$  = number of terms in a contract  $K$ ,

so that the cost of writing a contract is  $\alpha z(K)$ .

A contract  $K$  that the parties write is presumed to be treated by the courts according to a *method of interpretation*, denoted by  $M$ , which is known by contracting parties,<sup>15</sup> which the courts commit to employ,<sup>16</sup> and which is assumed to be costless to apply. This is a function whose argument is a contract  $K$  and whose value  $M(K)$  is a contract called the *interpreted contract*. Let

$$M(K) = M(\{(E_1, a_1), \dots, (E_m, a_m)\}) = \{(\theta_1, a(\theta_1)), \dots, (\theta_n, a(\theta_n))\},$$

where  $a(\theta_i)$  is the act undertaken in  $\theta_i$  and where the interpreted contract is assumed not to contain gaps. Note that  $M$  cannot be a function of the parties type  $t$ , since this is not observed by the courts.<sup>17</sup> An assumption that will be considered in one proposition (but not elsewhere employed) is that a method of interpretation displays *independence*, which is to say, it is such that  $a(\theta_i)$  is independent of contractual terms that do not cover  $\theta_i$ , and that how a gap is filled is independent of contractual terms.<sup>18</sup> (This independence assumption is simplifying (but unrealistic) because it implicitly rules out inference in the interpretation of one contractual term from the character of other terms.)

To illustrate a method of interpretation, consider the example mentioned above and the following method: specific terms are interpreted as written; gaps are filled with act A; general terms are interpreted as written, except that  $(\theta_1$  or  $\theta_2, C)$  is interpreted as  $\{(\theta_1, C), (\theta_2, B)\}$ , that is, the contract calling for C always is overridden in  $\theta_2$ .

It is assumed that contracts are not renegotiated (due to cost or inconvenience)<sup>19</sup> and that there is a substantial penalty for deviation from contracts,

15. In reality, the interpretation of contracts is guided by various doctrines and principles, so that knowing  $M$  involves learning these doctrines and principles, which may not be as difficult as learning an arbitrary function  $M$ .

16. The legal system is able to commit to employ rules by means of a variety of constraints on judges and juries, including the appeals process (meaning that decisions can be reversed if deviant) and rules of procedure and evidence. In reality, there may be uncertainty about which rules, here which  $M$ , courts would apply, but that will not be considered in this article.

17. If the courts do observe information about parties' type, then one can view  $F(t)$  as the distribution of  $t$  conditional on this information.

18. In other words, if  $\theta_i \in E_j$  in a term  $(E_j, a_j)$ , then  $a(\theta_i)$  does not depend on other terms; and if there is a gap at  $\theta_i$ , how the gap is filled does not depend on any terms.

19. There are often costs of renegotiation; problematic contingencies may occur at unforeseen times, and a decision must be quickly made by one party, making it difficult to bargain with the other contracting party. The reason for studying the assumption of no renegotiation is that it allows us to consider how a contractual term, if enforced, affects the joint value in the simplest possible setting. It would be distracting in one of the first formal investigations of contract interpretation to have to examine the more complex context in which there is renegotiation. For there, an important purpose of contractual terms is indirect, not to make "ex post" contractual performance efficient (for that would come about due to renegotiation, given symmetric information), but rather to improve the efficiency of ex ante investments by combating hold-up problems; hence, tracing the effect of a contractual term on joint value would be more complicated. However, it will be evident from the concluding comments that many of the results to be reached here would hold in the environment with renegotiation.



Table 1. Payoffs as a Function of Contingencies and Acts

	A	B	C	D
$\theta_1$	15	9	18	22
$\theta_2$	6	4	12	8

so that parties behave in the way set out in contracts<sup>20</sup> or, rather, in interpreted contracts. Hence, the (joint) expected value  $V$  of a contract  $K = \{(E_1, a_1), \dots, (E_m, a_m)\}$  if it were interpreted as it is written and has no gaps is

$$V(K) = \sum_i^n p_i x(\theta_i, a_i), \quad (1)$$

where the  $(\theta_i, a_i)$  are as determined by the contract.<sup>21</sup> However, because the interpreted contract  $M(K)$  is generally different from  $K$ , and because the contract involves a writing cost, the expected value (net of writing costs) of the contract  $K$  is given by<sup>22</sup>

$$V(M(K)) - \alpha z(K). \quad (2)$$

To illustrate the calculation of the expected value of a contract, let us continue with the example. Suppose that the method of interpretation is the one discussed above and that the payoffs to the parties are given in Table 1.

Assume as well that the probability of  $\theta_1$  is .8 and that of  $\theta_2$  is .2. Also assume that the writing cost  $\alpha$  per term is 2. Then several calculations of the expected value of contracts are these: If the contract is  $\{(\theta_1, D)\}$ , the interpreted contract,  $M(\{(\theta_1, D)\})$ , is  $\{(\theta_1, D), (\theta_2, A)\}$ , since the gap in  $\theta_2$  is filled with A; the number of terms in the written contract is one; hence, its expected value is  $.8 \times 22 + .2 \times 6 - 2 = 16.8$ . If the contract is  $\{(\theta_1, C), (\theta_2, A)\}$ , the interpreted contract is the same as the written one; the number of terms is two; and its expected value is  $.8 \times 18 + .2 \times 6 - 4 = 11.6$ . And if the contract is  $\{(\theta_1 \text{ or } \theta_2, C)\}$ , the interpreted contract is  $\{(\theta_1, C), (\theta_2, B)\}$ ; the number of terms is one; and its expected value is  $.8 \times 18 + .2 \times 4 - 2 = 13.2$ .

### 3. Choice of the Written Contract Given the Method of Interpretation

It is assumed that parties write the contract with highest expected value, that is, the contract maximizes the expected value of the interpreted contract minus the writing costs. (This assumption is consistent with the assumptions that

20. This simplifying assumption is natural to make, given the goal of studying the apparent meaning of contracts and whether interpretation is desirable. If the assumption were relaxed, and moderate damage measures for breach, such as the expectation measure, were allowed, the analysis would become more complicated (in part because courts would have to be able to observe variables in order to apply damage measures).

21. That is, for any  $\theta_i$ , find the event  $E_j$  containing  $\theta_i$  and define  $a_i$  to be the act in the term  $(E_j, a_j)$ .

22. It may be helpful to remind the reader that, given the assumptions that have been made of no renegotiation of contracts and that parties are aware that  $M(K)$  will be enforced by the courts, a party who is to take an action called for by  $M(K)$  simply does that (thus one can imagine that there are no actual proceedings in court).

parties are risk-neutral and that they each know the payoff function  $x(\theta_i, a)$ .) Thus, the parties choose a contract  $K$  to maximize (2). Their choice depends on the method of interpretation  $M$ ; let  $K(M)$  be their optimal contract (or, if the optimal contract is not unique, one of the optimal contracts).<sup>23</sup>

Note that parties would choose the ideal fully detailed contract, the contract with the ideal act  $a^*(\theta_i)$  for each contingency  $\theta_i$ , if contracts were enforced as written and there were no writing costs. This ideal contract (the terms of which will sometimes be referred to as what the parties want or desire) is a natural benchmark for comparison with written contracts and interpreted contracts.

Let us first consider the contract that the parties in the example would choose. It can be verified (by calculating the values of the possible contracts<sup>24</sup>) that the parties' best contract is  $\{(\theta_1 \text{ or } \theta_2, D)\}$ , a contract with a general term calling for use of D; since this contract is interpreted as written, its expected value is 17.2. The optimality of this contract for the parties given the method of interpretation can be explained roughly as follows. The ideal acts for the parties are D in  $\theta_1$  and C in  $\theta_2$ . Were the parties to write a fully detailed contract with two terms  $(\theta_1, D)$  and  $(\theta_2, C)$ , they would have to bear 4 in writing costs; their expected value would be 16. They can do better by writing a less detailed contract with one term, namely, the contract specifying D all of the time. It is true that under this contract, D rather than C is the act in  $\theta_2$ , but this reduces the payoff by only 4, and thus by an expected amount of only .8, which is less than the writing cost of 2 of an extra term. Among the contracts with one term, and thus involving the same writing cost, the parties could write a contract leaving a gap in  $\theta_2$ , but a gap would be filled with A, which is inferior for the parties to D.

The preceding example illustrates how parties trade off specificity of terms for savings in writing costs, but no simple characterization of the optimal written contract given the method of interpretation is apparent. However, it is worth noting that there are three ways in which each contingency  $\theta_i$  can be treated in a contract:  $\theta_i$  can be provided for in a specific term;  $\theta_i$  can be included in the event  $E_j$  of some general term  $(E_j, a_j)$ ; or  $\theta_i$  can be omitted, leaving a gap. For a contract to be optimal, the choice among the three possibilities just mentioned must be made correctly for each  $\theta_i$ .<sup>25</sup>

Let us next comment on a number of paradigmatic relationships that may hold among the terms of written contracts, interpreted contracts, and the parties' ideal contracts.

23. As is conventional, the parties are assumed to be able to carry out the maximization process without cost. This assumption is in some tension with the assumption that writing contracts involves cost. The results of the model thus depend on the view that reasonably good maximization involves lower cost than the writing of detailed contracts.

24. There are twenty-nine possible contracts: sixteen fully detailed contracts (such as  $\{(\theta_1, A), (\theta_2, B)\}$ ), eight contracts with one gap (such as  $\{(\theta_1, A)\}$  or  $\{(\theta_2, C)\}$ ), one contract with two gaps, and four contracts with general terms (such as  $\{(\theta_1 \text{ or } \theta_2, A)\}$ ).

25. This necessary condition was in effect the basis of the explanation of the best contract to write in the example just considered. For instance, that contingency  $\theta_2$  is part of a general term is due to the inferiority of  $\theta_2$  being provided for specifically (on account of the added writing cost) and to the inferiority of leaving a gap at  $\theta_2$  (on account of act A being worse than D).

(a) A specific term is interpreted as written and results in the ideal act for the parties: This occurs when a specific term  $(\theta_i, a)$  is interpreted as written and when the expected benefit of writing a specific term exceeds its writing cost  $\alpha$ . The expected benefit is determined by the probability of the contingency  $\theta_i$  and the benefit of the specific term, that is, the difference between the payoff under the ideal act  $a^*(\theta_i)$  and the payoff under the best alternative (either a gap or the best general term).<sup>26</sup>

(b) A general term is interpreted as written and always results in the ideal act for the parties: This occurs when a general term  $(E_j, a_j)$  is interpreted as written and  $a_j$  is the ideal act for all contingencies in  $E_j$ .<sup>27</sup>

(c) A general term is interpreted as written but does not always result in the ideal act for the parties: This occurs when a general term  $(E_j, a_j)$  is interpreted as written, the parties do not want  $a_j$  to be performed in at least some contingency  $\theta_i$  in  $E_j$ , but the expected gain from an alternative contract (writing an explicit term for  $\theta_i$ , including  $\theta_i$  in another general term, leaving a gap) is lower.<sup>28</sup>

(d) A general term is overridden in a contingency and results in a superior act for the parties: This occurs when a general term  $(E_j, a_j)$  is overridden in  $\theta_i$  and the act  $a_j$  is replaced with a better act, such as  $a^*(\theta_i)$ , saving the parties the cost of writing a separate term or of failing to obtain what they want in another general term or from a gap.<sup>29</sup>

(e) A general term is overridden in a contingency, but the parties would be better off if the term were enforced as written: This occurs when a general term  $(E_j, a_j)$  is overridden in  $\theta_i$ ,  $a_j$  is a better act in  $\theta_i$ , and the parties would be worse off with another term.<sup>30</sup>

(f) A gap is filled with the ideal act for the parties: This occurs when a gap for  $\theta_i$  is filled with  $a^*(\theta_i)$ . In this case, it is optimal for the parties to leave a gap, for they then obtain what they want without a writing cost.<sup>31</sup>

26. To illustrate in the example, if the writing cost  $\alpha$  is .2 (rather than 2), the contract selected by the parties is the fully detailed ideal contract,  $\{(\theta_1, D), (\theta_2, C)\}$ , for only by writing this contract will the ideal acts be taken, and the writing cost is low enough to justify that. In particular, the expected value of  $\{(\theta_1, D), (\theta_2, C)\}$  is 19.6, whereas the expected value of  $\{(\theta_1 \text{ or } \theta_2, D)\}$  is 19; the latter contract is no longer superior because the writing cost is so low.

27. Suppose that the payoffs in the table are modified: the payoff from D in  $\theta_2$  changes from 8 to 18, so that the ideal act is D in both  $\theta_1$  and  $\theta_2$ . Then the chosen contract is  $\{(\theta_1 \text{ or } \theta_2, D)\}$ , which is a general term that is interpreted as written, and which always results in the parties' ideal act.

28. This was already shown in the example, for the chosen contract was  $\{(\theta_1 \text{ or } \theta_2, D)\}$  but the parties want C rather than D in  $\theta_2$ .

29. Suppose in the example that the payoff from B is 19 in  $\theta_1$  and 14 in  $\theta_2$ , and the payoff from C is 28 in  $\theta_1$ . Then the ideal acts are C in  $\theta_1$  and B in  $\theta_2$ , the chosen contract is  $\{(\theta_1 \text{ or } \theta_2, C)\}$ , and this is interpreted as  $\{(\theta_1, C), (\theta_2, B)\}$ .

30. Suppose in the example that the payoff from C in  $\theta_1$  is 28. Then the ideal act is C in both  $\theta_1$  and  $\theta_2$ , and the contract that would be written is  $\{(\theta_1 \text{ or } \theta_2, C)\}$ , which would be interpreted as  $\{(\theta_1, C), (\theta_2, B)\}$  even though the parties want C in  $\theta_2$ . Note here that the parties could obtain what they want with the fully detailed contract  $\{(\theta_1, C), (\theta_2, C)\}$ , but this would not be worth the added writing costs. In particular, the expected value of  $\{(\theta_1 \text{ or } \theta_2, C)\}$  is 21.2 (because it is interpreted as  $\{(\theta_1, C), (\theta_2, B)\}$ ), and the expected value of  $\{(\theta_1, C), (\theta_2, C)\}$  is only 20.8.

31. Suppose in the example that the payoff from A in  $\theta_2$  is 14. Then the ideal acts are D in  $\theta_1$  and A in  $\theta_2$ , the contract that would be written is  $\{(\theta_1, D)\}$ , and the gap in  $\theta_2$  would be filled with A.

(g) A gap is filled with an act different from the ideal act for the parties: This occurs when a gap for  $\theta_i$  is filled with an act that the parties do not want but that is still superior to them to alternatives.<sup>32</sup>

#### 4. The Optimal Method of Interpretation

Given a method of interpretation  $M$ , we can calculate the expected value of the contract chosen by contracting parties of any type  $t$ . To be explicit, note that the value of any enforced contract, given by (1), depends on  $t$  because the  $x(\theta_i, a)$  are in fact  $x(\theta_i, a, t)$ ; hence, (1) may be written as  $V(K, t)$ . Also, the chosen contract  $K(M)$  described in the last section depends on  $t$  as well as on  $M$ , so we may write  $K(M, t)$  rather than  $K(M)$ . Hence, the expected value  $W$  of the contract chosen by type  $t$  given  $M$  is

$$W(M, t) = V(M(K(M, t)), t) - \alpha z(K(M, t)). \quad (3)$$

(The right side reflects the two effects of  $M$ :  $M$  influences the choice of the written contract, and  $M$  then determines how the written contract is interpreted.) Hence, social welfare, that is, the expected value of contracts over the population of different types  $t$  is

$$S = \int_T W(M, t) dF(t). \quad (4)$$

Before continuing, an issue needs to be noted (even though it is somewhat distracting): for any method of interpretation, there is a family of equivalent methods of interpretation resulting in the same interpreted contracts, writing costs, and level of social welfare  $S$ . In particular, let  $\pi$  denote a permutation of the set of acts  $A$  (that is,  $\pi$  is a 1:1 mapping from  $A$  to  $A$ ). Given any method of interpretation  $M$ , define the  $\pi$ -permuted version  $M_\pi$  of  $M$  by  $M_\pi(\{(E_1, a_1), \dots, (E_m, a_m)\}) = M(\{(E_1, \pi(a_1)), \dots, (E_m, \pi(a_m))\})$ . It is evident that the parties can obtain any interpreted contract under  $M_\pi$ , and with the same number of terms, as they can under  $M$  (where they write  $(E_i, a_i)$  under  $M$ , let them substitute  $(E_i, \pi^{-1}(a_i))$  under  $M_\pi$ , and conversely. Accordingly, for all parties, the menu of opportunities under  $M_\pi$  and under  $M$  are identical; they will choose contracts resulting in the same interpreted contracts, and thus  $W(M, t) = W(M_\pi, t)$ . In other words, we have

*Remark 1.* Let  $M_\pi$  be any  $\pi$ -permuted version of a method  $M$  of interpretation. Then the interpreted contracts for parties of each type  $t$  will be identical under  $M_\pi$  and  $M$ , and social welfare  $S$  will be the same under  $M_\pi$  as under  $M$ .

32. Suppose in the example that the payoff from B in  $\theta_2$  is 9 and from both C and D in  $\theta_2$  is 2. Then the ideal acts are D in  $\theta_1$  and B in  $\theta_2$ , and the contract that would be written is  $\{(\theta_1, D)\}$ , which would be filled with A in  $\theta_2$ . Note that in this case,  $\{(\theta_1 \text{ or } \theta_2, D)\}$  is not desired by the parties since act D in  $\theta_2$  is inferior to act A in  $\theta_2$ , and the fully detailed ideal contract  $\{(\theta_1, D), (\theta_2, B)\}$  is not worth the writing costs.

The optimal method of interpretation maximizes the expected value of contracts over the population of different types  $t$ ;<sup>33</sup> that is, the optimal method maximizes (4).<sup>34</sup> Because of Remark 1, we know that the optimal method is not unique, but for ease, let us speak of “the” optimal method in any case.<sup>35</sup>

Let us first show that some method of contract interpretation different from literal enforcement of contracts as written is optimal. Under literal enforcement, whatever contract parties write is the interpreted contract and it is assumed that their contract leaves no gaps (gap filling is not in the spirit of courts that merely do what parties want, that do not take an active role in stating the meaning of a contract). As noted in the introduction, literal enforcement is generally presumed in the literature on writing costs. We have

*Proposition 1.* The optimal method of contract interpretation is strictly superior to literal enforcement of contracts as written.

*Note.* The reason that this is true is that it is always possible to do better than literal enforcement by interpreting some gaps in ways that will allow at least some contracting parties to save writing costs.<sup>36</sup> (Of course, in general, the optimal method of interpretation involves more than the mere filling of gaps.)

*Proof.* Assume that contracts are literally enforced. Consider any type  $t$  of contracting parties and any term  $(E_j, a_j)$  in the contract they write. Now consider a method of interpretation  $M$  in which all terms are interpreted as written and gaps for  $\theta_i$  in  $E_j$  are filled with act  $a_j$ . Under this method, the type  $t$  can be made better off by writing the same contract as before, except leaving gaps for all  $\theta_i$  in  $E_j$ : under this contract, the parties of type  $t$  will obtain the same interpreted contract as before, but their writing cost will be reduced by  $\alpha$ . All other types will be at least as well off as before, as they can write the same contract they had before and be equally well off. ■

33. The optimal method can be viewed as one that minimizes the costs of judicial errors in interpretation (plus writing costs), for such errors correspond to interpretations of a term that do not maximize the value of the contract for parties of type  $t$ .

34. Note that types are presumed not to be able to send messages to the court declaring their type—the method of interpretation is assumed to depend only on the contractual terms. Were parties able to declare their type at no cost, then the ideal contract would be costlessly achieved trivially: let the mechanism be that if contracting parties declare they are of type  $t$ , the contract is always interpreted to be the ideal contract for type  $t$ . Then parties would always announce their true type, not spend anything on contract terms, and obtain the ideal contract. The justification for the assumption that parties cannot costlessly declare their type is that this is tantamount to declaring the entire function  $x(\theta_i, a, t)$ , which should be more costly than naming the ideal contract, but doing that has been assumed to be expensive. In other words, the whole point of the present article is that it is expensive to convey information about contractual desires, so in keeping with that assumption, it is presumed that parties cannot costlessly convey their type.

35. In particular, when under an optimal method, a term is interpreted as written; we will focus on that method rather than a permuted version under which terms are not interpreted as written.

36. The optimal method of interpretation is not necessarily superior to literal enforcement if, contrary to the assumption of this article, interpretation involves a cost. Proposition 1 might thus be better expressed by the statement that the optimal method of interpretation has positive gross value.

It is useful to illustrate the optimal method of interpretation with the often-discussed case of filling a single gap where there are two possible ways to do so. Specifically, assume that there are two contingencies  $\theta_1$  and  $\theta_2$ , that all parties want the same act C in  $\theta_1$ , that a fraction  $q$  of parties want act A in  $\theta_2$  and would obtain a positive payoff  $r$  from A and 0 from the other act B, and that the remaining fraction  $1 - q$  of parties wants act B in  $\theta_2$  and would obtain a positive payoff  $y$  from B and 0 from act A. In this case, the optimal method of interpretation is to interpret a gap in  $\theta_2$  as either A or B.<sup>37</sup> If a gap is interpreted as A, then the A-types will leave a gap in  $\theta_2$  and the B-types will leave a gap if  $p_2 y < \alpha$  and otherwise will write a term, spending  $\alpha$ , and obtain  $p_2 y - \alpha$ . Thus, the loss relative to the first-best will be  $p_2 y$  if this is less than the writing cost, and otherwise the loss will be the writing cost, so the expected loss will be  $(1 - q)\min(p_2 y, \alpha)$ . Similarly, if a gap is filled with B, A-types will leave a gap if  $p_2 r < \alpha$ , and the expected loss will be  $q\min(p_2 r, \alpha)$ . Thus, a gap should be filled with A if and only if

$$q\min(p_2 r, \alpha) > (1 - q)\min(p_2 y, \alpha). \quad (5)$$

One case is where the writing cost is low enough that both groups would write their preferred term if the gap is not filled as they want. In that case, (5) reduces to  $q > (1 - q)$ , so the gap is filled with the term preferred by the majority, in order to minimize writing costs. Another case is where the writing cost is high enough that neither group would write its preferred term. In that case, (5) reduces to  $qr > (1 - q)y$ , so that both the population proportion and the loss from the wrong term matter and the optimal decision is not necessarily to fill the gap with what the majority wants. In summary, we have<sup>38</sup>

*Remark 2.* In the simple situation where there are two ways to fill a gap, condition (5) determines how the gap is optimally filled. Hence, the gap should be filled as the majority desires if the writing cost  $\alpha$  is sufficiently low—in which case, each group would write its preferred term when the gap would not be filled as it wants; otherwise, filling the gap as the majority desires might not be optimal.

In the situation just examined, suppose that the writing cost  $\alpha$  is sufficiently low that each group will write its preferred term if the gap in  $\theta_2$  is not filled as it wants, and suppose that  $q > .5$ , so that A-types are in the majority and the gap should thus be filled with A. Observe then that if the gap is filled with A, all parties who leave a gap will be A-types, so the interpretation of the gap will be optimal given the contracts that are written. And observe too that if the gap is—suboptimally—filled with B, all parties who leave a gap will be B-types, so the interpretation of the gap will also be optimal given how contracts are

37. It is obviously best to interpret a gap in  $\theta_1$  as C since all parties want that.

38. Ayres and Gertner state a similar conclusion (1989:114).

written (now differently). These observations show that interpreting a contractual term optimally given the way that contracts are written is not *sufficient* for interpretation to be optimal (as was noted in the introduction). The reason is that although the method of interpretation may be optimal given how contracts are written, the method also influences how contracts are written, and might not affect that in a desirable way.

The observations also raise the question of whether interpreting a term optimally, taking as given the way contracts are written, is a *necessary* condition for the method of interpretation to be optimal. The answer is yes. As is now shown, a fundamental property of the optimal method of interpretation is that interpretation must be “naively” socially optimal—the interpretation of a term must be best for the types of parties who actually write the observed contract, which is to say, *ignoring* any effects of the method of interpretation on how contracts are written.

*Proposition 2.* (a) Under the optimal method of interpretation, the interpretation of a contract must be optimal *given* the constraint that the contracts that are written by parties are fixed: if  $M^*$  is the optimal method of interpretation—so that  $K(M^*, t)$  is the contract written by parties of type  $t$ —then  $M^*$  must maximize  $\int_T V(M(K(M^*, t)), t) dF(t)$  over  $M$ .

(b) Hence, in any contingency  $\theta_i$ , the act called for under the optimal method of interpretation is the act that maximizes the expected payoff to contracting parties in  $\theta_i$  over the conditional distribution of types who write the observed contract  $K$ .

*Notes.* The reasoning establishing part (a) is essentially as follows. Suppose that, under the optimal method of interpretation  $M^*$ , social welfare is not maximized given the assumption that the contracts that parties write will be fixed as those that they do write under  $M^*$ . Then there exists another method  $M'$  producing higher welfare given the contracts written under  $M^*$ . Now, in fact, the contracts written under  $M'$  will generally be different (since the contracts that are written depend on the method of interpretation), but this can only raise social welfare. Hence,  $M^*$  could not have been optimal.

Part (b), a corollary of part (a), is a substantial aid in determining the optimal method of interpretation because it means that attention can be restricted to a limited class of methods of interpretation.

*Proof.* To prove part (a), we want to show that if  $M^*$  is the optimal method of interpretation, then  $M^*$  maximizes  $\int_T V(M(K(M^*, t)), t) dF(t)$  over  $M$ . Assume otherwise, that there exists an  $M'$  such that  $\int_T V(M(K(M^*, t)), t) dF(t)$  is higher, or equivalently, that

$$\begin{aligned} & \int_T [V(M'(K(M^*, t)), t) - \alpha z(K(M^*, t))] dF(t) \\ & > \int_T [V(M^*(K(M^*, t)), t) - \alpha z(K(M^*, t))] dF(t). \end{aligned} \quad (6)$$

However, we know that

$$\begin{aligned} & \int_T [V(M'(K(M', t)), t) - \alpha z(K(M', t))] dF(t) \\ & \geq \int_T [V(M'(K(M^*, t)), t) - \alpha z(K(M^*, t))] dF(t), \end{aligned} \quad (7)$$

since, for each  $t$ ,  $V(M'(K(M', t)), t) - \alpha z(K(M', t)) \geq V(M'(K(M^*, t)), t) - \alpha z(K(M^*, t))$ , because the parties choose their contracts optimally given  $M'$ . Combining (6) and (7), we obtain

$$\begin{aligned} & \int_T [V(M'(K(M', t)), t) - \alpha z(K(M', t))] dF(t) \\ & > \int_T [V(M^*(K(M^*, t)), t) - \alpha z(K(M^*, t))] dF(t), \end{aligned} \quad (8)$$

which contradicts the supposed optimality of  $M^*$ .

Part (b) follows from part (a): since  $M$  maximizes  $\int_T V(M(K(M^*, t)), t) dF(t)$ ,  $M$  must maximize the expected payoff for each  $\theta_i$ . Hence, for any observed contract and any  $\theta_i$ , the interpreted act  $a$  maximizes  $\int_{T(K)} x(\theta_i, a, t) dF(t | K = K(M^*, t))$ , where  $T(K) = \{t | K = K(M^*, t)\}$ . ■

Let us next consider the optimal interpretation of a specific contractual term.

*Proposition 3.* (a) Specific terms are interpreted as written under the optimal independent method of interpretation. (b) However, in general, a specific term might not be interpreted as written under the optimal method of interpretation.

*Note.* If the parties go to the expense of writing a term covering only a single contingency  $\theta_i$ , it might be thought that it is optimal for the term be interpreted as written, for then they can obtain their ideal act  $a^*(\theta_i)$  in  $\theta_i$  by writing it. But this logic overlooks the point that if the method of interpretation does not display independence, courts may make inferences from one term about other terms in the contract. In particular, it is possible that by writing a specific term that calls for an action different from the action  $a^*(\theta_i)$ , the parties can signal their type. If so, this signal can be usefully employed in the interpretation of the other terms in the contract, and the court can then interpret the specific term in the way the parties really want, which is to say, not as written, but as  $a^*(\theta_i)$ .<sup>39</sup> Although this explains why a specific term might not be interpreted as written under the optimal method of interpretation, it follows from Proposition 2 that, if all types  $t$  who write the observed contract would want a specific term to be interpreted as written, then that must be done under the optimal method of interpretation.

*Proof.* We show part (a) here and part (b) in the appendix. Hence, let us prove that any independent method of interpretation  $M$  in which specific terms

39. Nevertheless, my intuition is that the possibility that it would be optimal not to interpret a specific term as written is not likely; for there should be a rich set of opportunities in the contract for parties to signal their type apart from use of a specific term.



are not interpreted as written is (weakly) dominated by another such method  $M'$  defined to be the same as  $M$ , except that under  $M'$  specific terms are interpreted as written. Consider the contract  $K = K(M)$  that parties of some type  $t$  choose under  $M$ . If  $K$  does not contain specific provisions, then  $M'(K) = M(K)$  by definition of  $M'$ , so the parties are as well off choosing  $K$  under  $M'$  as under  $M$ . And since the parties can choose  $K$  under  $M'$  and might choose a different contract, they must be at least as well off under  $M'$  as under  $M$ . If  $K$  contains specific provisions, let the parties replace  $K$  with  $K'$ , which is the same as  $K$  except that in place of each specific provision  $(\theta_i, a)$  in  $K$  is the specific provision  $(\theta_i, a^*(\theta_i))$ . The parties will be at least as well off under  $M'(K')$  as under  $M(K)$ :  $K$  and  $K'$  have the same number of terms and thus involve the same writing cost; any term that is not specific in  $K$  will also be a term in  $K'$  and will be interpreted in the same way under  $M'$  as under  $M$  (this step makes implicit use of the independence assumption about  $M$ , for the interpretation of a term that is not specific is assumed here not to depend on specific terms); and any term that is specific in  $K$  will be replaced with  $(\theta_i, a^*(\theta_i))$  and will result in the ideal act  $a^*(\theta_i)$  for  $\theta_i$  rather than  $a(\theta_i)$ . Since, then, for any type  $t$ , parties will be at least as well off under  $M'$  as under  $M$ , the expected value of  $M'$  is at least that of  $M$ . ■

The next proposition states that the paradigmatic types of outcomes discussed earlier in Section 3 can all occur under the optimal method of interpretation.

*Proposition 4.* Under the optimal method of interpretation, when contracting parties write the contracts that are best for them, the following different types of outcome are possible for a pair of contracting parties:

- (a) a gap is filled in a way that is ideal for the parties;
- (b) a gap is filled in a way that is not ideal for the parties;
- (c) a general term is interpreted as written, and this is ideal for the parties;
- (d) a general term is interpreted as written, but this is not ideal for the parties;
- (e) a general term is overridden in a contingency, and this is better for the parties;
- (f) a general term is overridden in a contingency, but interpreting the term as written would be better for the parties.

This is shown in the appendix. ■

Now let us consider an *opt-out rule*, under which contracting parties can specify that any term will not be interpreted but rather enforced as written.

*Proposition 5.* The opt-out rule, which allows contracting parties to specify that any term that they write not be interpreted by courts, is socially desirable.

*Note.* The opt-out rule raises social welfare because it allows contracting parties to avoid outcome (f) of Proposition 4, that a general term would be

overridden when that would lower the parties' well-being.<sup>40</sup> (In the example given in the last paragraph of the appendix, a BB type could guarantee that their contract  $\{(\theta_1 \text{ or } \theta_2, B)\}$  would not be overridden by opting out of contract interpretation, whereas under the optimal method of interpretation, B would be overridden with C in  $\theta_1$ .)

*Proof.* Let  $M$  be the optimal method of interpretation in the absence of the opt-out possibility for contracting parties. Any pair of contracting parties will be at least as well off if they have the opt-out provision under  $M$  than if not, and some parties may be better off, given possibility (f) of Proposition 4. Hence, social welfare might be higher under  $M$  with the opt-out rule, and thus under whatever is the optimal method of interpretation under the opt-out rule. ■

## 5. Evidence Beyond the Contract

Let us now consider briefly the possibility that the court considers not only the contract, but also evidence going beyond the contract. In fact, the courts often do consider such evidence, including the parties' contract negotiating history, their business dealings with each other during the life of the contract, other contracts they have made, and customs and norms in their industry. Let

$\beta$  = cost of presenting evidence,

where the cost is borne by the parties. If evidence is presented, assume for simplicity that it is perfect, allowing the court to determine the parties' ideal act  $a^*(\theta_i, t)$  in the contingency  $\theta_i$ . Assume also that the cost of presenting evidence exceeds the cost of writing a specific term,

$$\beta > \alpha, \quad (9)$$

the motivation being that at least as much information usually has to be presented to convince a court that some action  $a$  is optimal as merely to name that action in a contract, and further that parties are likely to be contesting each other's evidence in court. Assume also that the parties can costlessly specify in their contract the set of contingencies for which evidence will be presented.<sup>41</sup> Let

40. It is implicitly assumed here that it costs the parties nothing to add an opt-out clause to a contractual term; the motivation is that the marginal cost of adding such a simple clause to a term that they have already considered and written should be negligible. In any event, if there is a marginal cost to adding an opt-out clause, the social value of the opt-out rule would fall but could never be negative.

41. This assumption of costlessness is motivated by the fact that the parties do not specify an action for any contingency in  $R$ , but it is an inessential assumption. If it were supposed that there is a cost  $r$  per contingency that is included in  $R$ , then in Proposition 6,  $p^*$  would be  $(\alpha - r)/\beta$ , and (10) would be replaced by  $\beta < x(a^*(\theta_i, t), \theta_i, t) - x(a(\theta_i), \theta_i, t) - r/p_i$ .

$R$  = set of contingencies in which evidence will be presented.

Assume then that if  $\theta_i \in R$ , evidence will be presented and the act that will be enforced is  $a^*(\theta_i, t)$  for the parties of type  $t$ . Finally, suppose for ease that the method of interpretation  $M$  does not depend on  $R$ .

Let us examine how parties will write contracts, assuming that specific terms are interpreted as written.<sup>42</sup> We have

*Proposition 6.* Suppose that contracting parties stipulate in their contract in what set of contingencies  $R$  they will present evidence to the court. Then

- (a) a specific term will not be written for any contingency with a probability less than or equal to the threshold  $p^* = \alpha/\beta$ ;
- (b) evidence will not be presented for any contingency with a probability exceeding  $p^*$ , that is,  $R$  does not contain any  $\theta_i$  for which  $p_i > p^*$ ; and
- (c) evidence will be presented if and only if the cost is less than the joint loss from not doing so, namely,

$$\beta < x(a^*(\theta_i, t), \theta_i, t) - x(a(\theta_i), \theta_i, t). \quad (10)$$

That is, given the terms of the contract, (10) determines  $R$ .

*Note.* Because it is more expensive to present evidence to determine  $a^*(\theta_i, t)$  than to provide for it in the contract, it makes sense that parties will provide explicitly for it rather than present evidence if the likelihood of the contingency is sufficiently high; but if the likelihood is low, they save by avoiding the ex ante cost of specific provision. This explains parts (a) and (b). Part (c) is clear; if the loss from a less-than-ideal act in  $\theta_i$  under a general contractual term would exceed the cost  $\beta$  of presenting evidence, then the parties will arrange to present evidence in order to eliminate the loss.

*Proof.* Writing a specific term for a contingency  $\theta_i$  and stipulating that if  $\theta_i$  occurs, evidence be presented, are substitutes, in that each results in the ideal act  $a^*(\theta_i, t)$  if  $\theta_i$  occurs. The cost of a specific term (rather than a gap or including  $\theta_i$  in a general term) is  $\alpha$ , and the expected cost of having evidence presented if  $\theta_i$  occurs (by including  $\theta_i$  in  $R$ ) is  $p_i\beta$ . Accordingly, if  $p_i < p^*$ , then  $p_i\beta < \alpha$ ; it is cheaper to have the court consider evidence ex post than to bear the certain cost  $\alpha$  of a specific term, demonstrating (a) (if  $p_i = p^*$ , the parties are indifferent, and we assume for convenience that they would not write a specific term). Conversely, if  $p_i > p^*$ , then  $p_i\beta > \alpha$ ; it is more expensive to have the court consider evidence ex post than to bear the certain cost  $\alpha$  of a specific term, demonstrating (b). Part (c) is self-explanatory. ■

42. This is a feature of the optimal method of interpretation, assuming that it displays independence; essentially the same proof as that of Proposition 3(a) applies.

Parts (a) and (b) do not hold if the assumption of independence about  $M$  is relaxed, for then writing a specific term generally affects the interpretation of all other terms, not just the outcome in  $\theta_i$ .

Two observations may be added. First, the parties' choice of the set  $R$  in which to present evidence depends on their type  $t$ —notably, to know what their loss would be if they did not present evidence and relied on the court's interpretation depends on their type (see (10)). Hence, the court is not itself able to determine when it would be best for evidence to be presented; if the court decides when evidence is presented, social welfare will fall. Second, as a general matter, the optimal method of interpretation will change from what it is in the absence of the possibility of presentation of evidence, for, among other factors, the danger of large losses from errors in interpretation is bounded by the cost  $\beta$  of presentation of evidence.

## 6. Concluding Comments

### 6.1 Generality of the Analysis

The basic structure of the model of optimal interpretation examined here applies independently of the assumptions made about the contractual environment, in particular, relating to the contractibility of variables, renegotiation, and the information of the parties. That is, in any contractual environment, one can consider a method of interpretation  $M$  to be a function that transforms a written contract  $K$  into the interpreted contract  $M(K)$ ; contracting parties of type  $t$  will, given  $M$  and the assumed bargaining process, select a contract  $K(M, t)$ , and the actual contract that is employed will thus be  $M(K(M, t))$ ; the problem of the court, supposing that it does not observe  $t$ , will be to choose  $M$  to maximize  $\int W(M, t) dF(t)$ , where  $W(M, t)$  is the addition to social welfare when parties of type  $t$  choose  $K(M, t)$  and the contract that governs is  $M(K(M, t))$ .

Moreover, many of the conclusions reached about optimal interpretation hold in fairly broad circumstances, because the arguments for the conclusions often did not depend on the particulars of the model studied here. Proposition 1, that some method of interpretation is desirable, is true whenever contracting parties can save writing costs or contracting effort by leaving a gap.<sup>43</sup> Proposition 2, that the optimal method  $M^*$  is naively optimal—is best for the conditional distribution of types  $t$  who write the observed contract—is valid as long as contracting parties maximize their joint

---

43. Recall that the essence of the argument for the proposition was that if contracts are enforced as written (there is no interpretation), one can select any term in the contract written by any type  $t$  and employ the method of interpretation  $M$  under which gaps are filled with the act  $a_j$  that had been specified in this term.  $M$  will then allow the type  $t$  to leave a gap in the term and obtain the same contract as before, and other types can always write the contracts that they had before. This argument that  $M$  will lead to an increase in social welfare applies whatever the nature of contractual terms (for instance, if a term names damages for breach), as long as contracting parties save writing cost or contracting effort by leaving a gap.

welfare  $W(M, t)$  given  $M$ .<sup>44</sup> Likewise, Propositions 3 and 4 should generally be valid.

Let me now comment on two factors that were not taken into account in the analysis and how doing so would affect the conclusions. One factor that was not considered in the model is renegotiation, since in the model it was assumed that contractual terms directly determine outcomes, even though they might be inefficient. Suppose instead that one makes assumptions as in much of the literature on contracting: renegotiation always leads to efficient outcomes ex post, and the purpose of contracts is to improve the choice of ex ante investment, notably by implicitly combating the problem of hold-up (see for example Hart and Holmström, 1987; and Tirole, 1999). In this type of model, although the value of interpreted contracts would be determined by how well the contracts improve investment decisions, the main qualitative conclusions about optimal interpretation would hold, as indicated above. However, Proposition 2(a), that  $M^*$  is naively optimal, must be carefully construed. It does *not* imply that the court interprets a term in the way that is ex post efficient for those who write the observed contract, but rather that the court interprets the term so as to foster ex ante investment incentives for the types who write the observed contract. Suppose, for instance, that a contract specifies specific performance, and it is obvious to the court that performance is inefficient because the cost of performance far exceeds the value to the buyer. This does not imply that the court should interpret the contract as excusing performance. Since the parties may well have chosen specific performance to enhance the buyer's

---

44. The proof is essentially that given above. In particular, we want to show that  $M^*$  maximizes  $\int_T V(M(K(M^*, t)), t) dF(t)$  over  $M$ . Then if the claim is not true, there exists an  $M'$  such that

$$\begin{aligned} & \int_T [V(M'(K(M^*, t)), t) - \alpha(K(M^*, t))] dF(t) \\ & > \int_T [V(M^*(K(M^*, t)), t) - \alpha(K(M^*, t))] dF(t), \end{aligned} \quad (6')$$

where  $\alpha(K)$  is some writing or effort cost associated with a contract  $K$  (not necessarily the simple one assumed in the analysis). However, we know that

$$\begin{aligned} & \int_T [V(M'(K(M', t)), t) - \alpha(K(M', t))] dF(t) \\ & \geq \int_T [V(M'(K(M^*, t)), t) - \alpha(K(M^*, t))] dF(t), \end{aligned} \quad (7')$$

since, for each  $t$ ,  $V(M'(K(M', t)), t) - \alpha(K(M', t)) \geq V(M'(K(M^*, t)), t) - \alpha(K(M^*, t))$ , because of the assumption that the parties choose  $K$  to maximize their joint welfare  $W(M', t) = V(M'(K), t) - \alpha(K)$  given  $M'$ . Combining (6') and (7'), we obtain

$$\begin{aligned} & \int_T [V(M'(K(M', t)), t) - \alpha(K(M', t))] dF(t) \\ & > \int_T [V(M^*(K(M^*, t)), t) - \alpha(K(M^*, t))] dF(t), \end{aligned} \quad (8')$$

which contradicts the supposed optimality of  $M^*$ .

incentives to invest,<sup>45</sup> the best thing for the court to do may well be to enforce specific performance (that is, the court may know that the purpose of specific performance is to improve investment incentives and that ex post efficiency, here nonperformance, will come about through renegotiation).

Another factor of interest is asymmetry of information between the contracting parties. In the model, the parties were assumed to have symmetric information. If they do not, then two of the conclusions reached do not hold. First, the conclusion that the optimal method of interpretation  $M^*$  is naively optimal is not true. A prototypical example is where a contracting party chooses an inefficient term in order to masquerade as another kind of party and thereby to gain a price advantage (see also Spier, 1992). The optimal interpretation of the term might be to override it to obtain the efficient action, even though that interpretation is not best on average for the set of individuals who write the observed contract, thus contradicting Proposition 2(a).<sup>46</sup> The second major difference in conclusions when information is asymmetric is that the optimality of allowing parties to opt out of interpretation might not hold. For instance, some individuals who obtain mortgage loans with rates that fluctuate in an uncapped way with the market rate of interest might underestimate the likelihood of a large increase in the market rate. By signing an opt-out clause, they prevent courts from interpreting their mortgage contracts so as to protect them from extreme variation in the market rate; disallowing opt-out might therefore be socially beneficial.

## 6.2 Legal Practice

Several points about the interpretation of contracts in practice seem worth making in the light of the foregoing analysis. First, as stated at the outset, contractual interpretation is an important function of the courts (one commentator cites 25% as the fraction of contract cases concerned with interpretation; see Farnsworth, 1999:426), and interpretation is much more often concerned with the overriding of terms and related matters than with the filling of gaps in the sense of unprovided for contingencies.<sup>47</sup>

Second, the ability of parties to control interpretation by explicitly opting out of interpretation of a term is circumscribed,<sup>48</sup> even though a legal policy

45. For example, in Rogerson (1984), an early model of breach remedies with renegotiation, specific performance leads to superior buyer investment decisions than does the expectation measure or the reliance measure of damages for breach.

46. Also, it is evident why the proof of Proposition 2(a) does not hold when parties have asymmetric information. Inequality (7') in note 44 does not necessarily apply, for the contract that is chosen does not necessarily maximize the contracting parties' joint welfare  $V(M'(K), t) - \alpha(K)$  given  $M'$ .

47. Perusal of chapter 7 of Farnsworth (1999), or discussion with practicing lawyers, will reveal that interpretation usually involves overriding terms, resolving ambiguities of language, or settling internal contradictions in terms, and only unusually involves outright failure to provide instructions for a contingency. (Nevertheless, commentators often describe interpretation as the filling of gaps.)

48. Chapter 7 of Farnsworth (1999); personal communications with teachers of contracts courses.

of allowing such opting out often is socially desirable (Proposition 6). (Still, the reluctance of courts to allow opting out could be justified by a concern that a party to the contract is opting out of interpretation because, as just noted above, and unlike in the model, he has less information than the other party and fails to understand that opting out disadvantages him or because of ambiguity in the meaning of words.)

Third, the issue of the use of evidence extrinsic to contracts in their interpretation is of significance in actual practice and is also much debated. Whereas it was assumed in Section 5 that extrinsic evidence was perfect, this evidence is highly imperfect in reality and is very costly to consider (especially because of the tendency of parties to contest negotiating history, oral statements, course of dealing). Thus, the question of whether the value of extrinsic evidence in contractual interpretation exceeds its cost to the parties is a real one. Accordingly, the ability of the parties to control whether the courts will examine extrinsic evidence would seem to be of substantial importance to them. However, their power to limit interpretation to the written contract is restricted<sup>49</sup> (which is of a piece with their difficulty in opting out of interpretation altogether). Possibly this is due to the courts' belief that the parties' true desires should be discovered at trial, without due regard to the cost to the parties of the necessary inquiry. A factor working in the opposite direction, though, is that since the parties do not bear the costs to the court of engaging in interpretation, the parties might specify socially excessive interpretation to the degree that they can control the amount of interpretation.

## Appendix

*Proof of Proposition 3(b).* The following example demonstrates the possibility that a specific term is not enforced as written under the optimal method of interpretation, and thus demonstrates part (b) of Proposition 3. There are two contingencies  $\theta_1$  and  $\theta_2$ , each occurring with probability .5; five acts, A, B, C, D, E; and a writing cost  $\alpha$  of 1. A contractual type will obtain a positive payoff of 4 under one preferred act in each contingency and 0 otherwise. For instance, the type AB obtains a payoff of 4 if A is the act in  $\theta_1$  but obtains 0 otherwise in that contingency, and the type obtains a payoff of 4 if B is the act in  $\theta_2$  but obtains 0 otherwise in that contingency. The types in the population are these: AA, CB, CC, DD, EE, AB, AE, CD, DE, EA, DA, EB, AC, DB, CE, and AD, and the fraction of each type is .01, except that the fraction of type AD is .85. Now consider a method  $M$  of interpretation and the contracts chosen by the different types as described in Table 2.

Note that the method  $M$  is described for all contracts with one term and for the double gap, but not for fully detailed two-term contracts. (However, it will

49. Notably, the parol evidence rule, requiring a court to focus on the written contract and to bar extrinsic evidence if the contract was intended to be an "integrated" expression of the parties' desires, is subject to many exceptions. See generally chapter 7 of Farnsworth (1999).

Table 2. Written and Interpreted Contracts Chosen by Different Types

Type	Written Contract	Interpreted Contract
AA	$\{(\theta_1 \text{ or } \theta_2, A)\}$	$\{(\theta_1, A), (\theta_2, A)\}$
CB	$\{(\theta_1 \text{ or } \theta_2, B)\}$	$\{(\theta_1, C), (\theta_2, B)\}$
CC	$\{(\theta_1 \text{ or } \theta_2, C)\}$	$\{(\theta_1, C), (\theta_2, C)\}$
DD	$\{(\theta_1 \text{ or } \theta_2, D)\}$	$\{(\theta_1, D), (\theta_2, D)\}$
EE	$\{(\theta_1 \text{ or } \theta_2, E)\}$	$\{(\theta_1, E), (\theta_2, E)\}$
AB	$\{(\theta_1, A)\}$	$\{(\theta_1, A), (\theta_2, B)\}$
AE	$\{(\theta_1, B)\}$	$\{(\theta_1, A), (\theta_2, E)\}$
CD	$\{(\theta_1, C)\}$	$\{(\theta_1, C), (\theta_2, D)\}$
DE	$\{(\theta_1, D)\}$	$\{(\theta_1, D), (\theta_2, E)\}$
EA	$\{(\theta_1, E)\}$	$\{(\theta_1, E), (\theta_2, A)\}$
DA	$\{(\theta_2, A)\}$	$\{(\theta_1, D), (\theta_2, A)\}$
EB	$\{(\theta_2, B)\}$	$\{(\theta_1, E), (\theta_2, B)\}$
AC	$\{(\theta_2, C)\}$	$\{(\theta_1, A), (\theta_2, C)\}$
DB	$\{(\theta_2, D)\}$	$\{(\theta_1, D), (\theta_2, B)\}$
CE	$\{(\theta_2, E)\}$	$\{(\theta_1, C), (\theta_2, E)\}$
AD	$\{.\}^*$	$\{(\theta_1, A), (\theta_2, D)\}$

\*The contract has no terms—gaps for  $\theta_1$  and  $\theta_2$ .

not matter how  $M$  is defined for such contracts.) Note as well that  $M$  does not obey independence, for the interpretation of gaps depends on the accompanying specific term.

It is evident why each type chooses the contract shown in the table. It is obvious that type AD will choose the double gap, as that is costless. Each other type obtains its ideal contract at a cost of 1 by writing the indicated one-term contract, whereas if it spends nothing and leaves a double gap, its expected payoff would fall by at least 2.

That  $M$  is optimal follows from two observations. All types obtain their ideal contracts with contracts with one term or the double gap contract. The type that writes the double gap contract is AD, the most numerous (.85 vs. .01).

Observe that type AE writes the specific term  $(\theta_1, B)$ , yet B is interpreted as A, so the specific term is not interpreted as written, which is what we claimed. (One can view the explanation as follows. Type AE implicitly signals its type by writing B as the act in the specific term for  $\theta_1$ , but having done that, it is best for A to be the actual interpreted act in  $\theta_1$ .)

Moreover, although there are other optimal methods, under *any* optimal method, the contract  $\{(\theta_1, B)\}$  will be interpreted differently from B. In particular, observe that any permutation of the first 15 rows of the written contract column will also be optimal: for any permutation is just another way of allowing each of the 15 types with fraction .01 to obtain their ideal contracts with a one-term contract; the double gap contract must be for the AD types due to their high fraction. But there is *no* type for which B is the ideal act in  $\theta_1$ . Thus, since the interpreted contract must be the ideal contract for some type, it cannot be interpreted as B. ■



*Proof of Proposition 4.* We want to demonstrate that the outcomes listed in Proposition 4 are possible under an optimal method of interpretation. To do this, we will use the example given in the proof of Proposition 3(b) in the appendix and variations of it. The example involves a general term that is interpreted as written, where this is ideal for the parties, for instance  $\{(\theta_1 \text{ or } \theta_2, A)\}$ ; it also has a general term that is not interpreted as written, where this is ideal for the parties, for instance  $\{(\theta_1 \text{ or } \theta_2, B)\}$ ; it involves as well a contract with a gap that is filled in an ideal way, for instance  $\{(\theta_2, E)\}$ .

To show that a gap may be filled in a way that is not ideal for the parties, let there be a group of BB, for which there is not only a payoff of 4 if B is the act in either contingency, but also a payoff of 3 if E is the act in  $\theta_1$ . Assume that this group is small enough that the optimal  $M$  does not change. The BB type will choose  $\{(\theta_2, B)\}$ , so the gap will be interpreted as E but they would prefer that B be the act in  $\theta_1$ .

To show that a general term may be interpreted as written but that this is not what is desired by parties, let there be a group of CA, for which there is not only a payoff of 4 if C is the act in  $\theta_1$  and if A is the act in  $\theta_2$ , but also a payoff of 3 if A is the act in  $\theta_1$ . Assume that this group is small enough that the optimal  $M$  does not change. The CA type will choose  $\{(\theta_1 \text{ or } \theta_2, A)\}$ , which will be interpreted as written, but they would prefer that C be the act in  $\theta_1$ .

Last, to show that a general term may not be interpreted as written but that the parties would prefer that it be interpreted as written, let there be a group of BB, for which there is not only a payoff of 4 if B is the act in either contingency, but also a payoff of 3 if C is the act in  $\theta_1$ . Assume that this group is small enough that the optimal  $M$  does not change. The BB type will choose  $\{(\theta_1 \text{ or } \theta_2, B)\}$ , which will be interpreted as C in  $\theta_1$ , but they would prefer that B be the act in  $\theta_1$ . ■

## References

- Anderlini, Luca, and Leonardo Felli. 1999. "Incomplete Contracts and Complexity Costs," 46 *Theory and Decision* 23–50.
- Anderlini, Luca, Leonardo Felli, and Andrew Postlewaite. 2001. "Courts of Law and Unforeseen Contingencies." University of Pennsylvania Law School, Institute for Law and Economics, Research Paper No. 01-05.
- Ayres, Ian, and Robert Gertner. 1989. "Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules," 99 *Yale Law Journal* 87–130.
- . 1992. "Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules," 101 *Yale Law Journal* 729–73.
- Battigalli, Pierpaolo, and Giovanni Maggi. 2002. "Rigidity, Discretion, and the Costs of Writing Contracts," 92 *American Economic Review* 798–817.
- Bebchuk, Lucian, and Steven Shavell. 1991. "Information and the Scope of Liability for Breach of Contract: The Rule of *Hadley v. Baxendale*," 7 *Journal of Law, Economics, and Organization* 284–312.
- Bernstein, Lisa. 1996. "Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms," 144 *University of Pennsylvania Law Review* 1765–821.
- Bolton, Patrick, and Mathias Dewatripont. 2005. *Contract Theory*. Cambridge: MIT Press.
- Cohen, George M. 2000. "Implied Terms and Interpretation in Contract Law," in Boudewijn Bouckaert and Gerrit de Geest, eds., *Encyclopedia of Law and Economics*, Vol. 3: 78–99. Cheltenham, UK: Edward Elgar.

- Craswell, Richard. 2000. "Contract Law: General Theories," in Boudewijn Bouckaert and Gerrit de Geest, eds., *Encyclopedia of Law and Economics*, Vol. 3: 1–24. Cheltenham, UK: Edward Elgar.
- Dye, Ronald A. 1985. "Costly Contract Contingencies," 26 *International Economic Review* 233–50.
- Farnsworth, E. Allan. 1999. *Contracts*. 3rd edition. New York: Aspen.
- Goetz, Charles J., and Robert E. Scott. 1985. "The Limits of Expanded Choice: An Analysis of the Interactions between Express and Implied Contract Terms," 73 *California Law Review* 261–322.
- Hadfield, Gillian K. 1994. "Judicial Competence and the Interpretation of Incomplete Contracts," 23 *Journal of Legal Studies* 159–84.
- Hart, Oliver, and Bengt Holmström. 1987. "The Theory of Contracts," in Truman Bewley, ed., *Advances in Economic Theory*, Fifth World Congress, chapter 3. Cambridge, UK: Cambridge University Press.
- Hart, Oliver, and John Moore. 1999. "Foundations of Incomplete Contracts," 66 *Review of Economic Studies* 115–38.
- Katz, Avery. 1998. "Contract Formation and Interpretation," in Peter Newman, ed. *The New Palgrave Dictionary of Economics and the Law*, Vol. 1: 425–32. London: Macmillan.
- Posner, Eric A. 1998. "The Parol Evidence Rule, The Plain Meaning Rule, and the Principles of Contractual Interpretation," 146 *University of Pennsylvania Law Review* 533–77.
- Posner, Richard A. Forthcoming. "The Law and Economics of Contract Interpretation." *Texas Law Review*.
- Rogerson, William P. 1984. "Efficient Reliance and Damage Measures for Breach of Contract," 15 *Rand Journal of Economics* 39–53.
- Schwartz, Alan. 1992. "Relational Contracts in the Courts: An Analysis of Incomplete Agreements and Judicial Strategies," 21 *Journal of Legal Studies* 271–318.
- Schwartz, Alan, and Robert E. Scott. 2003. "Contract Theory and the Limits of Contract Law," 113 *Yale Law Journal* 541–619.
- Schwartz, Alan, and Joel Watson. 2004. "The Law and Economics of Costly Contracting," 20 *Journal of Law, Economics & Organization* 2–31.
- Spier, Kathryn E. 1992. "Incomplete Contracts and Signalling," 23 *RAND Journal of Economics* 432–43.
- Tirole, Jean. 1999. "Incomplete Contracts: Where Do We Stand?" 67 *Econometrica* 741–81.
- Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.