Contract Renegotiation in Agency Problems*

Aaron S. Edlin†and Benjamin E. Hermalin‡
University of California at Berkeley

First Version: November 1996
This Version: May 1997

Abstract

This paper studies the ability of an agent and a principal to achieve the first-best outcome when the agent invests in an asset that has greater value if owned by the principal than by the agent. When contracts can be renegotiated, a well-known danger is that the principal can holdup the agent, undermining the agent's investment incentives. We begin by identifying a countervailing effect: Investment by the agent can increase his value for the asset, thus improving his bargaining position in renegotiation. We show that option contracts will achieve the first best whenever this threat-point effect dominates the holdup effect. Otherwise, achieving the first best is difficult and, in many cases, impossible. In such cases, we show that if parties have an appropriate signal available, then the first best is still attainable for a wide class of bargaining procedures. A noisy signal, however, means that the optimal contract will involve terms that courts might view as punitive and so refuse to enforce.

Keywords: Contract renegotiation, bargaining, double moral hazard, relationship-specific investment.

JEL: C78, D23, K12, L23.

1 Introduction

We analyze agency problems with renegotiation, asking when is it possible to give an agent efficient incentives to work on an asset or project that a principal will use or market. A critical issue is whether the principal observes the agent's

---

action. Fudenberg and Tirole (1990) and Ma (1994) study the case where she does not. In this paper, we follow Hermalin and Katz (1991) by assuming that she observes his action, but cannot prove what it was in court. Unlike Hermalin and Katz, we focus much of our attention on situations in which the courts can verify only payments between the parties and possession (ownership) of the asset or project. We subsequently consider the same information structure as Hermalin and Katz, where we properly analyze the situation in which *both* parties have bargaining power simultaneously.[1]

As an example, consider the problem that Pixar and Disney faced when they collaborated to produce the animated feature *Toy Story*. Each party brought unique talents to the table: Pixar had the 3-D animation technology and Disney had the distribution and marketing expertise for animated films.[2] Prior to release, Disney could observe the quality of the film Pixar produced, but quality is sufficiently amorphous that it would be difficult, if not impossible, for it to be described in a contract or demonstrated in court unambiguously. Giving Pixar appropriate incentives would, then, seem problematic. If Disney commits to buy at a fixed price, then Pixar, acting opportunistically, has no incentive to work hard *for Disney*.[3] Alternatively, suppose that they enter an agreement that Disney will pay Pixar a fixed amount if Pixar makes what Disney judges to be a good film. Then Pixar will still have poor investment incentives, because it will worry about Disney acting opportunistically *ex post*; claiming the film is poor and bargaining to buy it at a lower price (see Williamson, 1985, for more on such holdups).[4] The parties may try to limit Disney's ability to holdup Pixar by guaranteeing Pixar a share of the film's profits, but this will undermine Disney's incentives to distribute and market the film. The parties' ability to contract around the incentive and holdup problems depends critically on the quality of the information that *is* verifiable and the nature of contract renegotiation.

Demski and Sappington (1991) and Bernheim and Whinston (1997) propose an easy solution to this problem, one that works even when the courts can verify only (1) payments between the parties and (2) ownership of the asset used in production: Specifically, the contract gives ownership of the film to Pixar, but gives Disney an option to purchase it at a price set equal to the film's value conditional on efficient effort by Pixar. Absent renegotiation, this contract is efficient, but, as we show, Disney has an incentive to let its option expire and renegotiate a lower price for the film.[5] Hence, this contract ultimately provides no protection from the holdup problem we observed above.

---

[1] Hermalin and Katz also consider the situation in which both parties have bargaining power simultaneously, but their analysis neglects to fold the results of renegotiation back into the incentive compatibility constraints.

[2] According to Pixar's CFO, the animated-film business is essentially a Disney monopoly. Source: "Lost Toys," *The Economist*, Vol. 342, No. 7999 (Jan. 11, 1997), p. 58.

[3] Pixar still has some incentive to develop the technology for other films, but not as much as if it had the additional incentive to work hard for Disney.

[4] Indeed, Disney captured around 90% of the profits (source: *ibid.*).

[5] That contracts in this industry are subject to renegotiation is evidenced by Disney and Pixar, who recently "scrapped" their original contract, signed in 1991 (source: "Disney in 10-Year, Five-Film Deal with Pixar," *New York Times*, February 25, 1997).

In the Demski-Sappington scheme, giving the agent initial ownership of the asset provides the agent with incentives to work efficiently, because otherwise the principal will let her option expire and leave the project with the agent. As previously noted, even if the agent takes the appropriate action, an opportunistic principal will let her option expire and bargain for a lower price in renegotiation. Typically, we would expect the principal to capture a share of the renegotiation surplus, so that the agent does not capture the full marginal contribution of his efforts to the principal's value. This holdup effect undermines the agent's incentives to work. Yet, there is a second effect. The agent's effort will strengthen his bargaining position, since if bargaining breaks down and he winds up with the project, his effort would give the project greater value.[6] We find that a suitably chosen option contract will provide first-best incentives when this second effect, the threat-point effect, dominates the holdup effect.[7] In this case, it is the threat of renegotiation that gives the agent incentives to work hard and, ironically, the role of an option is now to guard against *over*investment by the agent. When the holdup effect dominates the threat-point effect, on the other hand, we show that no option contract is efficient for any monotonic sharing rule.[8]

When the principal and agent's efforts are substitutes, the threat-point effect tends to dominate because the agent's effort has higher marginal contribution when he owns the project than when the principal owns the project and contributes her own effort. Conversely, the holdup effect dominates when their efforts are complements.[9] We also find that the threat-point effect will tend to be larger if effort reduces the risk premium (the difference between the expected and certainty equivalent value to the agent of keeping the asset), but smaller if

---

[6] This effect can be quite large. For example, according to "Woody and Buzz: The Untold Story," *New York Times*, February 24, 1997, Disney tripled its price to Pixar on future pictures out of concern that Pixar would break with Disney and take its now-proven technology to Warner Brothers or other studios.

[7] Nöldeke and Schmidt (1995, 1997) also consider models in which option contracts achieve the first best with renegotiation. Their results depend, however, on what will often be overly "optimistic" assumptions about the parties' ability to commit to their contracts in terms of the timing and scope of renegotiation. See our Conclusion for a further discussion of these issues and other differences between their models and ours.

[8] We say that A and B have a monotonic sharing rule if, when they bargain over a pot of money, their shares each increase with the amount in the pot. See Edlin and Reichelstein (1996).

[9] Complementarity and substitutability also play a role in Bernheim and Whinston (1997). In our paper, these concepts are defined with respect to the production process, a primitive. In contrast, Bernheim and Whinston speak of *strategic* complements and substitutes. Since the contract helps to fix the strategy spaces, whether actions are strategic complements or substitutes depends on the contract in place and may differ between principal and agent. For instance, if the contract allows the principal to choose the probability that she gets the asset ("quantity") and the agent to choose quality, then their actions are typically strategic complements for the principal; however, if the asset's value to agent is increasing in quality, then their actions will be strategic substitutes for the agent.

Complementarity and substitutability are also an issue in the control literature (see, e.g., Grossman and Hart, 1986; Chiu, 1996). Unlike here, where efficiency dictates that ownership (control) always be given to the principal in the end, this literature focuses on who should be given control.

effort increases the risk premium.

We can interpret the case of substitute efforts as the agent's investment being more general than relation-specific *on the margin*. The case of complementary efforts can be interpreted as the agent's investment being more specific than general *on the margin*. Since we assume the asset is more valuable to the principal than to the agent, investments are specific *on average*. These observations imply that what matters for efficiency is not whether investments are general or specific on average, but whether they are general or specific on the margin.

The possibility that there is no efficient option contract leads us, first, to consider general mechanisms and, then, to consider the Hermalin-Katz informational assumptions. In our analysis of general mechanisms we begin with the case of a risk-neutral agent. In this case, we show that the first best is implementable if and only if the threat-point effect dominates the holdup effect; and, thus, there is no loss of generality in restricting attention to option contracts. On the other hand, for a *risk-averse* agent, it is possible that a general mechanism could exploit his risk aversion to overcome the holdup effect even when the threat-point effect is weak; though this possibility is by no means guaranteed.

Perhaps a more plausible way to obtain the first best when the threat-point effect is weak is to link payments to some verifiable signal of effort. As Hermalin and Katz (1991) argued, if effort is observable, then a court can likely observe some imperfect signal of it. That is, for instance, in a dispute between Pixar and Disney about quality, the court could screen the film and render a judgment. Although the court's assessment of quality is likely a noisy signal of quality *vis-à-vis* true quality (effort), Hermalin and Katz show that in some circumstances it will nevertheless be sufficient for obtaining the first best when the parties can renegotiate in advance of the court's ruling.[10] However, just allowing for a verifiable signal turns out not to be sufficient for obtaining the first best: We give an example where despite having an informative signal, the first best is unattainable because the agent's bargaining power is unusually low when he expends first-best effort. On the other hand, we show the first best can be obtained provided the signal is sufficiently informative and the agent's bargaining share of any marginal increase in surplus is bounded away from zero.

In the Hermalin-Katz framework, the first-best contract we construct can be interpreted as demanding a low and sometimes negative payment if the signal suggests low effort. In many circumstances, courts would interpret this feature as a stipulated damage measure. When the signal is sufficiently informative, these stipulated damages will be small, and may be less than the expectation damage measure that a court would impose absent a contractual stipulation. If the signal is fairly uninformative, however, then the required damage payment must be quite high. (Loosely speaking, we might view this result as analogous to Becker (1968)'s classic observation that punishment must increase as the chance of detection decreases.) High damage payments risk being ruled punitive and therefore unenforceable. Hence, the penalty doctrine can get in the way of

---

[10] These circumstances are that the signal satisfy an informative condition they refer to as the "Convex-Hull Condition" and that one of the parties have all the bargaining power in renegotiation.

parties using their information to avoid the holdup problem. Our paper therefore adds a new voice to the chorus calling for an elimination of the penalty doctrine.

Our work is related to the large literature on trading mechanisms (see, e.g., Aghion *et al.*, 1994; Chung, 1991; Edlin and Reichelstein, 1996; Hermalin and Katz, 1993; Nöldeke and Schmidt, 1995; Rogerson, 1992; Segal and Whinston, 1997, among others). There are, however, important differences. First, we allow for a risk-averse agent, whereas this literature does not. Second, this literature models the agent's investment as affecting only his cost of production and trade, while we model the agent's investment as affecting the value of the asset. Hence, because the principal owns the asset in the end, the agent's investment affects the principal's valuation. This makes the investment "cooperative" in the sense of Che and Chung (1996) and Che and Hausch (1996).[11] The critical difference between our model and theirs is that we consider the possibility that the agent's effort has value even if the agent retains the project—the "no trade" case in their models. This possibility creates the threat-point effect we identified above, which can support efficient contracts when the parties' efforts are substitutes. On the other hand, our result that complementary efforts can cause inefficiency accords with Che and Hausch. Their model may be viewed as the limiting case in which efforts are maximally complementary so that the agent's effort yields no return without trade.

2   Model

We begin with the same economic problem as Demski and Sappington (1991). A principal, who owns a transferable asset (e.g., patent, store, or movie idea) hires an agent. The agent undertakes an action $a \in [0, \infty)$. The principal observes this action, but it is not verifiable, so that contracts cannot be directly contingent on it. After observing the agent's action, the principal takes an action $b \in [0, \infty)$. These actions affect the asset's return, $r \in [\underline{r}, \infty)$, in a stochastic manner. Specifically, assume that the *density* over r conditional on a and b, $f(r|a, b)$, is increasing in a and b in the sense of first-degree stochastic dominance; that is, for all r, all $a \geq a'$ and $b \geq b'$

$$\int_{\underline{r}}^{r} f(z|a,b)\, dz \leq \int_{\underline{r}}^{r} f(z|a',b')\, dz.$$

Assume that $f(r|\cdot, b)$ and $f(r|a, \cdot)$ are both differentiable functions for all r, a, and b. We denote the expected return conditional on a and b as

$$R(a,b) = \int_{\underline{r}}^{\infty} r f(r|a,b)\, dr.$$

The principal is risk neutral and her utility equals $y - b$ where y is her income. Let her value of the asset, conditional on the agent's action a, be

$$V(a) \equiv \max_{b \in [0,\infty)} R(a,b) - b.$$

---

[11] We thank Bill Rogerson and Ilya Segal for pointing this connection out to us.

We assume that $V(a)$ is finite for all a and we observe, by the envelope theorem, that $V'(a) = \partial R / \partial a \geq 0$.

The agent's utility is $u(w) - a$, where w is his income. Assume that $u(\cdot)$ is strictly increasing, twice differentiable, and at least weakly concave (i.e., the agent is not risk loving).

The initial bargaining game between the principal and agent, which we do not model, yields the agent a level of utility that we normalize to 0. Since our focus is on achieving the first best, we need not worry about how this level of utility varies with the actual achievable outcome.

The first best is that the agent chooses a^* and the principal pays him p^* solving

$$\max_{p, a \in [0, \infty)} V(a) - p$$
$$\text{subject to } u(p) - a = 0. \tag{1}$$

We assume a unique solution to this program, with $a^* > 0$. The first-order condition with respect to a is

$$V'(a^*) = \frac{1}{u'[u^{-1}(a^*)]}. \tag{2}$$

To insure that the agent can be given sufficient incentives, we assume

Assumption 1: The domain of $u(\cdot)$ includes $\left(\underline{r} + u^{-1}(a^*) - V(a^*), \infty\right)$.

3   Option Contracts and Renegotiation

Demski and Sappington consider option contracts (p_1, p_2), where p_1 is an unconditional transfer from principal to agent (a negative p_1 is a payment to the principal). Upon acceptance of the contract, the principal transfers ownership of the asset to the agent together with p_1.[12]   After the agent has chosen his action, the principal has the option to buy back the improved asset at price p_2. If the principal declines to exercise her option, then the agent retains ownership. After the principal decides whether to exercise her option, she chooses her effort. The owner of the asset receives r.

Demski and Sappington show that if

$$\hat{p}_2 = V(a^*),$$

then it is an equilibrium for the agent to expend efficient effort a^*, and for the principal to exercise her option and choose b to maximize her value.

Demski and Sappington do not, however, consider the possibility that the agent and principal renegotiate if the principal does not exercise her option. If the principal lets her option expire, the agent will own the asset. If no sale is

---

[12] If the agent has initial ownership, he maintains ownership at this stage.

thereafter arranged, the principal will choose $b = 0$. The certainty equivalent payoff to the agent *for the asset* is, therefore,

$$CE\left(a, p_1\right) \equiv u^{-1}\left(\int_{\underline{r}}^{\infty} u\left(r + p_1\right) f\left(r|a, 0\right) dr\right) - p_1.$$

Observe that[13]

$$\lim_{p_1 \to \infty} p_1 + CE\left(a, p_1\right) = \infty \ \forall a. \tag{3}$$

We also assume that the agent is risk averse or the principal's effort is valuable or both. Consequently,[14]

$$V\left(a\right) > CE\left(a, p_1\right) \tag{4}$$

and so there are gains from renegotiating and assigning ownership to the principal. We presume, therefore, that the principal will end up buying the asset. In doing so, she and the agent must agree to a division of the *renegotiation surplus*

$$S(a, p_1) \equiv V\left(a\right) - CE\left(a, p_1\right).$$

We see no reason to assume that one of the parties captures all this surplus; more likely, it will be split between them. Following Edlin and Reichelstein (1996), we remain agnostic about the extensive form of this bargaining game. Rather, like them, we assume that the parties follow an *efficient monotonic sharing rule*:[15] that is, the agent and principal split the renegotiation surplus according to differentiable rules, $\sigma_A : \mathbb{R} \to \mathbb{R}_+$ and $\sigma_P : \mathbb{R} \to \mathbb{R}_+$ satisfying[16]

    A. (efficiency) $\sigma_A(S) + \sigma_P(S) \equiv S$; and either
    B. (weak monotonicity) $\sigma_A'\left(\cdot\right) \geq 0$, $\sigma_P'\left(\cdot\right) \geq 0$; or
    B'. (strict monotonicity) $\sigma_A'\left(\cdot\right) > 0$, $\sigma_P'\left(\cdot\right) > 0$.

Constant-shares bargaining (i.e., $\sigma_A\left(S\right) = \beta S$ and $\sigma_P\left(S\right) = \left(1 - \beta\right) S$) is a special case of a monotonic sharing rule.

In the case of renegotiation, the principal's payment to the agent will be

$$p\left(a, p_1\right) \equiv \sigma_A\left[S\left(a, p_1\right)\right] + CE\left(a, p_1\right).$$

---

[13]**Proof:** Note that

$$\int_{\underline{r}}^{\infty} u\left(r + p_1\right) f\left(r|a, 0\right) dr \geq u\left(\underline{r} + p_1\right)$$

for all p_1. Since $u^{-1}\left(\cdot\right)$ is increasing, we have

$$u^{-1}\left[\int_{\underline{r}}^{\infty} u\left(r + p_1\right) f\left(r|a, 0\right) dr\right] \geq \underline{r} + p_1$$

for all p_1. Since the left-hand side is $p_1 + CE\left(a, p_1\right)$, the result follows.

[14]**Proof:** $CE\left(a, p_1\right) \leq R\left(a, 0\right) \leq V\left(a\right)$, with one inequality strict if the agent is risk averse (certainty equivalent payoff less than expected payoff) or the principal's effort is productive ($V\left(a\right) > R\left(a, 0\right)$).

[15]Appendix A of Edlin and Reichelstein (1996) contains an explicit extensive-form bargaining game in which the equilibrium solution has surplus split monotonically.

[16]Monotonicity implies that $\sigma_A\left(\cdot\right)$ is differentiable almost everywhere. This is sufficient for the analysis that follows, but assuming differentiability everywhere simplifies the proofs.

The principal will, therefore, exercise her option if $p_2 \leq p(a, p_1)$. Otherwise, she will let it expire and renegotiate.

Demski and Sappington's approach is problematic when the agent does not have the power to make take-it-or-leave-it offers when renegotiating:

Proposition 1 *The Demski and Sappington equilibrium is* not *robust to renegotiation unless the agent has all the bargaining power when he chooses the first-best action.*

Proof: Consider the Demski and Sappington contract in which $\hat{p}_2 = V(a^*)$ and $\hat{p}_1$ divides the *ex ante* surplus. Suppose the agent chooses a^* as in the Demski and Sappington equilibrium. Observe

$$p(a^*, \hat{p}_1) = \sigma_A[S(a^*, \hat{p}_1)] + CE(a^*, \hat{p}_1) \leq V(a^*) = \hat{p}_2.$$

It follows that $p(a^*, \hat{p}_1) < \hat{p}_2$ unless $\sigma_A[S(a^*, \hat{p}_1)] = S(a^*, \hat{p}_1)$. If the agent does not have all the bargaining power, the principal will let her option expire rather than exercise it as required by the Demski and Sappington equilibrium. ∎

Remark 1 *Since $\sigma_A(0) = 0$,*

$$\sigma_A[S(a^*, \hat{p}_1)] = \int_0^{S(a^*, \hat{p}_1)} \sigma_A'(z)\, dz.$$

Consequently, Proposition 1 can be rewritten as: The Demski and Sappington equilibrium is not *robust to renegotiation unless the agent gets 100% of the marginal surplus for almost every level of surplus less than $S(a^*, \hat{p}_1)$ (i.e., unless $\sigma_A'(z) = 1$ for almost every z so $\int_0^{S(a^*, \hat{p}_1)} \sigma_A'(z)\, dz = S(a^*, \hat{p}_1)$).*

Remark 2 *Proposition 1 stands in stark contrast to Bernheim and Whinston (1997)'s model, in which the Demski-Sappington contract will achieve the first best (and won't be renegotiated along the equilibrium path) when the* principal *has all the bargaining power. Bernheim and Whinston's result depends on their assumption that renegotiation can only occur* before *the principal decides whether to exercise her option and* not *after.*[17]

In a world without renegotiation, the principal is indifferent between exercising her option and letting it expire when $p_2 = V(a)$. With renegotiation,

---

[17] Bernheim and Whinston consider option contracts in their Example 1. There are some differences between their model and ours—specifically, their principal makes no investment, quantities are continuous, and there is no uncertainty—but their analysis can be readily translated to our model. Their model of renegotiation (see their Section 4.4) assumes, unlike ours, that any renegotiation occurs prior to the principal's option-exercise decision and that this decision is irreversible—there is no subsequent chance for the principal to obtain ownership if she lets her option expire. Consequently, if $a = a^*$, then the principal cannot credibly threaten *not* to exercise an option with a strike price of $V(a^*)$. Renegotiation is, thus, irrelevant, so the first best is attainable using the Demski-Sappington contract.

she gets some of the surplus from renegotiating an expired option contract (provided $\sigma_P [S(a, p_1)] > 0$), so that she prefers renegotiating to exercising her option. Consequently, the principal can be expected to holdup the agent.

Holdup means that the agent cannot capture the full marginal contribution of his effort to the principal's value, $V'(a)$, and this reduces the agent's investment incentives. It turns out, however, that even though this holdup is unavoidable, it does not necessarily cause underinvestment. Offsetting the holdup effect is the threat-point effect: investment strengthens the agent's bargaining position by increasing the value of the asset if the agent retains it. If this threat-point effect is sufficiently large it will dominate the holdup effect, and an efficient option contract can exist with renegotiation. If the holdup effect dominates, then an efficient option contract does not exist. Mathematically, for monotonic sharing, the condition that the threat-point effect dominates can be expressed as

$$\frac{\partial CE(a^*, p_1)}{\partial a} \geq V'(a^*) \tag{5}$$

for p_1 solving[18]

$$u^{-1}(a^*) = p(a^*, p_1) + p_1 \tag{6}$$

That is, the threat-point effect dominates if the marginal impact of the agent's action increases his value for the asset (adjusted for risk) more than it increases the social (principal's) value for the asset. Since

$$\frac{\partial p(a, p_1)}{\partial a} = \sigma_A'[S(a, p_1)] V'(a^*) + (1 - \sigma_A'[S(a, p_1)]) \frac{\partial CE(a, p_1)}{\partial a}, \tag{7}$$

condition (5) implies

$$\frac{\partial p(a^*, p_1)}{\partial a} \in \left(V'(a^*), \frac{\partial CE(a^*, p_1)}{\partial a} \right) \tag{8}$$

for *strictly* monotonic sharing. Hence, when the threat-point effect dominates, the agent's action increases his payment more than it increases the principal's marginal value for the asset. So even though the principal is capturing some of the increased value on the margin, it is still possible to give the agent sufficient incentives. If (5) fails, then

$$\frac{\partial p(a^*, p_1)}{\partial a} \in \left(\frac{\partial CE(a^*, p_1)}{\partial a}, V'(a^*) \right) \tag{9}$$

---

[18] To see that a p_1 solving (6) exists, observe that the right-hand side of (6) is continuous in p_1 under the assumptions on $\sigma_A(\cdot)$ and $u(\cdot)$. Moreover, it lies between $CE(a^*, p_1) + p_1$ and $V(a^*) + p_1$. This upper bound equals $u^{-1}(a^*)$ if

$$p_1 = u^{-1}(a^*) - V(a^*)$$

(recall, from Assumption 1, that

$$r + u^{-1}(a^*) - V(a^*)$$

is in the domain of $u(\cdot)$ for all possible r). From (3), the lower bound goes to ∞ as $p_1 \to \infty$. Therefore a p_1 that solves (6) must exist.

for strictly monotonic sharing; consequently, sufficient incentives are no longer possible. The following proposition shows that (5) is necessary for an efficient option contract to exist under constant-shares bargaining.

Proposition 2 *Assume that (5) fails to hold for any p_1 solving (6) and that renegotiation follows a strictly monotonic sharing rule. Then no option contract (p_1, p_2) is efficient.*[19]

Proof: Suppose not. The agent's utility is

$$u\left(\min\left\{p\left(a, p_1\right), p_2\right\} + p_1\right) - a,$$

since the principal exercises her option if and only if $p_2 \leq p\left(a, p_1\right)$. Since $p\left(\cdot, p_1\right)$ is continuous, the agent will never choose a such that $p\left(a, p_1\right) > p_2$ (unless $a = 0$)—he could increase his utility by reducing a. Since the agent chooses $a^* \neq 0$, it follows that

$$u'\left(p\left(a^*, p_1\right) + p_1\right) \frac{\partial p\left(a^*, p_1\right)}{\partial a} \geq 1 \tag{10}$$

(where equality holds if $p\left(a^*, p_1\right) < p_2$). Since we've supposed that the first-best is achieved,

$$p\left(a^*, p_1\right) + p_1 = u^{-1}\left(a^*\right)$$

(i.e., (1) is satisfied). Using (2), the first-order condition (10) can be rewritten as

$$\frac{\partial p\left(a^*, p_1\right)}{\partial a} \geq V'\left(a^*\right),$$

which contradicts (9). ∎

In contrast, an efficient contract will exist for monotonic sharing rules if (5) holds and the first-order approach is valid.

Proposition 3 *Assume either that (5) holds for some p_1^* solving (6) and that sharing is monotonic; or that the agent receives 100% of the surplus on the margin when $a = a^*$ and $p_1 = p_1^*$. Then an efficient option contract exists provided that $u\left(p\left(a, p_1^*\right) + p_1^*\right) - a$ is quasi-concave in a and has a negative second derivative at any point at which it has a zero first derivative.*

Proof: Set $p_2^* = p\left(a^*, p_1^*\right)$. Then the agent chooses a to maximize

$$u\left(\min\left\{p\left(a, p_1^*\right), p_2^*\right\} + p_1^*\right) - a.$$

Since $p_2^* = p\left(a^*, p_1^*\right)$ and $p\left(\cdot, p_1^*\right)$ is increasing, it is never optimal for the agent to choose an $a > a^*$. Hence, the agent's problem is

$$\max_{a \leq a^*} u\left(p\left(a, p_1^*\right) + p_1^*\right) - a. \tag{11}$$

---

[19] This result extends to *weakly* monotonic sharing rules provided $\sigma'_A\left[S\left(a^*, p_1\right)\right] < 1$.

The conditions on $u\left(p\left(a, p_1^*\right) + p_1^*\right) - a$ ensure that

$$u'\left(p\left(\hat{a}, p_1^*\right) + p_1^*\right) \frac{\partial p\left(\hat{a}, p_1^*\right)}{\partial a} - 1 \geq 0$$

implies

$$u'\left(p\left(a, p_1^*\right) + p_1^*\right) \frac{\partial p\left(a, p_1^*\right)}{\partial a} - 1 > 0$$

for all $a < \hat{a}$. Hence a^* solves (11) if

$$u'\left(p\left(a^*, p_1^*\right) + p_1^*\right) \frac{\partial p\left(a^*, p_1^*\right)}{\partial a} - 1 \geq 0. \tag{12}$$

Using the definitions of p_1^*, p_2^*, and equation (2), we can rewrite inequality (12) as:

$$\frac{\partial p\left(a^*, p_1^*\right)}{\partial a} \geq V'\left(a^*\right).$$

This inequality holds if $\sigma_A'\left[S\left(a^*, p_1^*\right)\right] = 1$ or if sharing is monotonic and (5) holds. ∎

Note that the option price, $p\left(a^*, p_1^*\right)$, is less in Proposition 3 than in Demski and Sappington's paper (i.e., $p\left(a^*, p_1^*\right) \leq V\left(a^*\right)$). In our model, the principal must be indifferent between exercising her option and taking possession through renegotiation, while in Demski and Sappington she must be indifferent between exercising and foregoing possession. Unless $\sigma_A\left[S\left(a^*, p_1^*\right)\right] = S\left(a^*, p_1^*\right)$—which entails $\sigma_A'\left(z\right) = 1$ for almost every $z \in \left[0, S\left(a^*, p_1^*\right)\right]$ (see Remark 1 above)—taking possession through renegotiation is strictly preferable to foregoing possession altogether. Consequently, our price must be lower than theirs to account for renegotiation.

Observe, too, that once one introduces renegotiation, the purpose of giving the principal an option is *not* to induce effort, but to prevent overexertion. The case where an efficient contract exists is one where, without any contract, the threat-point effect is so strong that the agent would choose $a \geq a^*$. The principal's option keeps the agent from engaging in inefficient, rent-seeking behavior.

To explore the significance of condition (5) more readily, let the agent be risk neutral. Then $R\left(a, 0\right)$, the expected return to the asset if the principal supplies no effort, *equals* the agent's certainty equivalent value. It follows that condition (5) holds if

$$\frac{\partial^2 R}{\partial a \partial b} \leq 0;$$

that is, if the agent's and principal's actions are (weakly) substitutable. On the other hand, if the agent's and principal's actions are complementary, so that

$$\frac{\partial^2 R}{\partial a \partial b} > 0,$$

then condition (5) fails. Substitutability is a reasonable assumption in a variety of contexts—we argue later that it could hold for Pixar and Disney. On the

other hand, a large motivation for trade between principal and agent could be the complementary talents they bring to production. For a substantial class of agency problems, then, we should expect (5) to fail, so that we cannot achieve the first best using simple option contracts.

Observe that we can interpret trading situations in which the asset is value*less* to the agent (e.g., the situation considered in Che and Chung, 1996; Che and Hausch, 1996) to be ones in which investments are "maximally" complementary—action by the principal (including, possibly, just taking possession) is necessary for there to be any positive return to the agent's investment. Hence, it is not surprising that we—like Che and Hausch before us—find achieving the first best to be difficult in these situations (at least when there can be renegotiation). In particular, option contracts will *fail* to achieve the first best.

If the agent is risk averse, then this complements-substitutes dichotomy can be complicated by how the agent's wealth affects his attitudes toward risk and how his actions affect the gamble he faces if he keeps the asset. The dichotomy is maintained in its purest form, however, if there are no wealth effects and the riskiness of the return, r, is independent of a. For instance, if $u(w) = 1 - e^{-w}$ and r is distributed normally with mean $R(a, b)$ and variance τ^2, then

$$CE(a, p_1) = R(a, 0) - \frac{1}{2}\tau^2.$$

In this case, the risk premium, $\frac{1}{2}\tau^2$, is independent of the agent's action, a. In other settings, it is quite possible that the agent's action does affect the risk premium. If it reduces the risk premium, this will enhance the threat-point effect and, thereby, the possibility of an efficient outcome. Conversely, if the risk premium rises with a, then this reduces the threat-point effect, making the holdup effect dominate for a larger set of parameter values.

4   General Mechanisms

When the first best is unattainable with an option contract—i.e., when (5) fails—the obvious question is whether the first best could be achieved by a more general mechanism. We investigate that question in this section, where we show that the answer is "no" if the agent is risk neutral. If, however, the agent is risk averse, then the first best may be obtainable when (5) "just" fails.

4.1   The Risk-Neutral Case

Now assume that the agent is risk neutral. Initially, we assume that (5) fails.

Without loss of generality, assume the agent's utility is $w - a$. Consider a general announcement mechanism in which both agent and principal make announcements, $\hat{a}_A$ and $\hat{a}_P$ respectively, about the agent's actions. Conditional on these announcements, the agent is paid $p(\hat{a}_A, \hat{a}_P, \chi)$ and ownership of the asset is given to the principal ($\chi = 1$) with probability $x(\hat{a}_A, \hat{a}_P)$ and to the agent ($\chi = 0$) with probability $1 - x(\hat{a}_A, \hat{a}_P)$. We continue to assume that

renegotiation is possible after the mechanism is played out. Observe that if the principal receives ownership—the allocation is efficient—then there is no scope for renegotiation at this stage. We could also allow for renegotiation after announcements but *before* the mechanism is executed, but since the agent is risk neutral there are no gains from renegotiation at this point.

Define:

- $\tilde{v}(a) \equiv \int_{\underline{r}}^{\infty} r f(r|a,0)\, dr$ (agent's expected value of owning the asset at the end);

- $S(a) \equiv V(a) - \tilde{v}(a)$ (surplus from trading asset after the mechanism gives the agent ownership);

- $P(\hat{a}_A, \hat{a}_P, a) \equiv \sigma_A [S(a)] + \tilde{v}(a) + p(\hat{a}_A, \hat{a}_P, 0)$ (total payment to agent if the mechanism gives him initial ownership and the principal must bargain to acquire asset);

- $\tilde{w}(\hat{a}_A, \hat{a}_P, a) \equiv x(\hat{a}_A, \hat{a}_P)\, p(\hat{a}_A, \hat{a}_P, 1) + [1 - x(\hat{a}_A, \hat{a}_P)]\, P(\hat{a}_A, \hat{a}_P, a)$ (the agent's expected wage (utility) before playing the mechanism); and

- $w(a) \equiv \tilde{w}(a, a, a)$.

Observe that in the risk-neutral case, the failure of (5) is equivalent to

$$V'(a^*) > \tilde{v}'(a^*).$$

Observe, too, that the first-order condition (2) reduces to $V'(a^*) = 1$ in the risk-neutral case.

Proposition 4 *Assume strict monotone sharing in renegotiation. Assume, too, that $V'(\cdot)$, $\tilde{v}'(\cdot)$, and $\sigma_A'[S(\cdot)]$ are continuous in a neighborhood of the first-best action, a^*. Then the first best is unattainable if (5) fails.*

Proof: We restrict attention to incentive compatible direct revelation mechanisms (without loss of generality according to the revelation principle), and we will show that no such mechanism can provide efficient investment incentives. Incentive compatibility requires that

$$
\begin{aligned}
\tilde{w}(a, a, a) &\geq \tilde{w}(a^*, a, a) \text{ and} \\
\tilde{w}(a^*, a^*, a^*) &\leq \tilde{w}(a^*, a, a^*);
\end{aligned}
$$

which, when $a < a^*$, can be interpreted, respectively, as saying that the agent can't improve his payoff by claiming to have worked hard (a^*) when he shirked, nor can the principal manage to pay a lower wage by claiming the agent shirked when he didn't. Combining these inequalities yields

$$
\begin{aligned}
w(a^*) - w(a) &\leq \tilde{w}(a^*, a, a^*) - \tilde{w}(a^*, a, a) \\
&= (1 - x(a^*, a))\, [P(a^*, a, a^*) - P(a^*, a, a)] \\
&= (1 - x(a^*, a))\, ([\sigma_A [S(a^*)] + \tilde{v}(a^*)] - [\sigma_A [S(a)] + \tilde{v}(a)]).
\end{aligned}
$$

Using the fact that $x(a^*, a) \geq 0$, together with the mean value theorem yields

$$w(a^*) - w(a) \leq (\sigma'_A [S(\tilde{a})] V'(\tilde{a}) + (1 - \sigma'_A [S(\tilde{a})]) \tilde{v}'(\tilde{a})) (a^* - a) \qquad (13)$$

for some $\tilde{a} \in (a, a^*)$. Then since $V'(a^*) = 1 > \tilde{v}'(a^*)$, $\sigma'_A(S) < 1$, and $V'(\cdot)$ and $\tilde{v}'(\cdot)$ are continuous, there exists an $a' < a^*$ such that

$$\sigma'_A [S(\tilde{a})] V'(\tilde{a}) + (1 - \sigma'_A [S(\tilde{a})]) \tilde{v}'(\tilde{a}) < 1$$

for all $\tilde{a} \in [a', a^*)$. It, then, follows from (13) that

$$\tilde{w}(a^*) - \tilde{w}(a') < a^* - a';$$

or, rewriting,

$$w(a^*) - a^* < w(a') - a'.$$

But this means that the agent prefers to choose action a' to action a^*, so that a^* is *not* implementable by any mechanism. ∎

The proof of this proposition shows that when (5) fails, any mechanism that tries to induce effort a^* would suffer from one of two problems: If the payments for announcements (a^*, a) are large, then the agent will prefer to shirk and claim he worked hard; or if payments are small, then even if the agent works hard, the principal will claim he shirked.

Suppose, instead that condition (5) is met. Observe that, if the agent is risk neutral,

$$p(a, p_1) = \sigma_A [S(a)] + \tilde{v}(a),$$

where $p(a, p_1)$ is as defined in Section 3. Hence, provided

$$\sigma_A [S(a)] + \tilde{v}(a) - a \qquad (14)$$

is quasi-concave, with a negative second derivative at any point at which it has a zero first derivative (as would be true, for instance, if $\sigma_A [S(a)] + \tilde{v}(a)$ were strictly concave), then the first best is attainable by Proposition 3. This observation and Proposition 4 yield:

Proposition 5 *Suppose (i) a risk-neutral agent; (ii) strictly monotone sharing in renegotiation; (iii) (14) is quasi-concave, with a negative second derivative at any point at which it has a zero first derivative; and (iv) $V'(\cdot)$ and $\tilde{v}'(\cdot)$ are continuous in a neighborhood of the first-best action, a^*. Then the following are all equivalent:*

- *The first best is attainable if and only if (5) holds;*[20]

---

[20] The results of Segal and Whinston (1997), independently established, can be applied in our framework to show that any message game can be replaced with a (possibly randomized) uncontingent contract that implements the same investments. By studying uncontingent contracts, one could then obtain our result that the first best is attainable with a risk-neutral agent if and only if (5) holds.

- *The first best is attainable if and only if it can be obtained by an option contract; and*

- *The first best is attainable if and only if the principal's and agent's actions are substitutes.*

These results are closely related to those of Che and Hausch (1996), who show in a trading model that the parties can never achieve the first best, and in fact do best to write no contract. The main difference between their analysis and ours is that they study the case of maximum complementarity—when the asset has no value to the agent.[21] Hence their model has no threat-point effect, which means (5) always fails. See the Conclusion for further discussion.

4.2   The Risk-Averse Case

When the agent is risk averse, efficiency dictates that the parties will wish to avoid exposing the agent to risk. Consequently, if $x\left(\hat{a}_A, \hat{a}_P\right) \in (0,1)$, the parties will renegotiate before χ is realized; that is, agree to a fixed wage in exchange for granting the principal ownership.[22] With this in mind, define

- $\widetilde{CE}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv \int_{\underline{r}}^{\infty} u\left[r + p\left(\hat{a}_A, \hat{a}_P, 0\right)\right] f\left(r|a, 0\right) dr - p\left(\hat{a}_A, \hat{a}_P, 0\right)$ (agent's certainty-equivalent value of owning the asset at the end);

- $\tilde{S}\left(a\right) \equiv V\left(a\right) - \widetilde{CE}\left(\hat{a}_A, \hat{a}_P, a\right)$ (surplus from trading asset after the mechanism gives the agent ownership);

- $\tilde{P}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv \sigma_A\left[\tilde{S}\left(a\right)\right] + \widetilde{CE}\left(\hat{a}_A, \hat{a}_P, a\right) + p\left(\hat{a}_A, \hat{a}_P, 0\right)$ (total payment to agent if the mechanism gives him ownership);

- $\tilde{w}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv x\left(\hat{a}_A, \hat{a}_P\right) p\left(\hat{a}_A, \hat{a}_P, 1\right) + \left[1 - x\left(\hat{a}_A, \hat{a}_P\right)\right] \tilde{P}\left(\hat{a}_A, \hat{a}_P, a\right)$ (the agent's expected wage after announcements but before the realization of χ);

- $\tilde{u}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv x\left(\hat{a}_A, \hat{a}_P\right) u\left[p\left(\hat{a}_A, \hat{a}_P, 1\right)\right] + \left[1 - x\left(\hat{a}_A, \hat{a}_P\right)\right] u\left[\tilde{P}\left(\hat{a}_A, \hat{a}_P, a\right)\right]$ (the agent's expected utility after announcements from the gamble over χ);

- $\widetilde{ce}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv u^{-1}\left[\tilde{u}\left(\hat{a}_A, \hat{a}_P, a\right)\right]$.

- $\tilde{s}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv \tilde{w}\left(\hat{a}_A, \hat{a}_P, a\right) - \widetilde{ce}\left(\hat{a}_A, \hat{a}_P, a\right)$ (surplus to be realized by renegotiating after announcements, but before χ is realized);

- $\tilde{W}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv \sigma_A\left[\tilde{s}\left(\hat{a}_A, \hat{a}_P, a\right)\right] + \widetilde{ce}\left(\hat{a}_A, \hat{a}_P, a\right)$ (total payment to agent);

---

[21] Another difference is that in their model χ represents quantity traded and is chosen from a continuum.

[22] For instance, if we viewed the court as a mechanism, the parties could settle their case after testifying but before the judge or jury renders a verdict—a common practice in litigation.

- $\tilde{U}\left(\hat{a}_A, \hat{a}_P, a\right) \equiv u\left[\tilde{W}\left(\hat{a}_A, \hat{a}_P, a\right)\right]$ (agent's utility); and

- $U\left(a\right) = \tilde{U}\left(a, a, a\right)$.

Observe, by construction, $\tilde{U}$ is monotonic in $\tilde{W}$.

Recall that by using an option, we can make the agent's incentive problem "one-sided." That is, by using an option, we need not worry about the agent over-exerting himself, only under-exerting himself. Specifically, assume that after the agent has chosen his action, but prior to announcements, the principal has the option to take possession of the asset in exchange for paying the agent $u^{-1}\left(a^*\right)$. Since, in a first-best equilibrium, the agent is paid $u^{-1}\left(a^*\right)$ with certainty if he chooses a^*, the agent won't choose $a > a^*$; so we can limit attention to $a \leq a^*$.

By the revelation principle, we can restrict attention to truth-telling equilibria. In such equilibria, the conditions for the first best to obtain are

$$U\left(a^*\right) - a^* \;=\; 0 \text{ and} \tag{IR}$$
$$U\left(a^*\right) - a^* \;\geq\; U\left(a\right) - a, \text{ for all } a < a^*. \tag{IC}$$

We can rewrite (IC) as

$$\frac{U\left(a^*\right) - U\left(a\right)}{a^* - a} \geq 1. \tag{IC$'$}$$

Truth-telling requires

$$\tilde{U}\left(a, a, a\right) \;\geq\; \tilde{U}\left(a^*, a, a\right) \text{ and}$$
$$\tilde{U}\left(a^*, a^*, a^*\right) \;\leq\; \tilde{U}\left(a^*, a, a^*\right);$$

that is, respectively, the agent can't improve his payoff by lying, nor can the principal improve hers by lying. Combining these two expressions yields

$$U\left(a^*\right) - U\left(a\right) \leq \tilde{U}\left(a^*, a, a^*\right) - \tilde{U}\left(a^*, a, a\right). \tag{15}$$

Using the mean-value theorem, the right-hand side equals

$$\tilde{U}_3\left(a^*, a, \tilde{a}\right)\left(a^* - a\right)$$

for some $\tilde{a} \in [a, a^*]$.[23] The inequality can be rewritten as

$$\frac{U\left(a^*\right) - U\left(a\right)}{a^* - a} \;\leq\; u'\left(\tilde{W}\right) \times \left[\sigma_A'\left(\tilde{s}\right) + [1 - \sigma_A'\left(\tilde{s}\right)]\frac{u'\left(\tilde{P}\right)}{u'\left(\widetilde{ce}\right)}\right] \tag{16}$$

$$\times [1 - x\left(a^*, a\right)] \times \left[\sigma_A'\left(\tilde{S}\right) V'\left(\tilde{a}\right) + \left[1 - \sigma_A'\left(\tilde{S}\right)\right]\frac{\partial \widetilde{CE}}{\partial \tilde{a}}\right],$$

where the functions $\tilde{W}$, $\tilde{s}$, $\widetilde{ce}$, $\tilde{S}$, $\tilde{P}$, and $\widetilde{CE}$ are all evaluated at $(a^*, a, \tilde{a})$.

From (IC$'$), a necessary condition for achieving the first best is that the right-hand side of (16) be at least one for all $a < a^*$. Some observations:

---

[23]Where $\tilde{U}_3$ is the partial derivative of $\tilde{U}$ with respect to its third argument.

1. If $x(a^*, a) = 1$ for all $a < a^*$, then the right-hand side of (16) must be zero, so the first best is not obtainable. This is intuitive: If trade is guaranteed, the agent has no incentives to expend effort.

2. If $x(a^*, a) = 0$ for all $a < a^*$—that is, we have an option contract like the ones in Section 3—then the right-hand side of (16) reduces to

$$u'\left[\tilde{P}(a^*, a, \tilde{a})\right] \times \left[\sigma'_A\left(\tilde{S}\right) V'(\tilde{a}) + \left[1 - \sigma'_A\left(\tilde{S}\right)\right] \frac{\partial \widetilde{CE}}{\partial \tilde{a}}\right]. \qquad (17)$$

Moreover, if x does not depend on announcements, then $p(\hat{a}_A, \hat{a}_P, 0)$ must also be independent of announcements for truth-telling. Consequently, (17) is continuous in $\tilde{a}$ and limits to

$$u'\left[u^{-1}(a^*)\right] \times \left[\sigma'_A\left(\tilde{S}\right) V'(a^*) + \left[1 - \sigma'_A\left(\tilde{S}\right)\right] \frac{\partial \widetilde{CE}}{\partial a^*}\right]$$

as $a \uparrow a^*$. Hence, the first best cannot be achieved if (5) fails to hold.

3. The most intriguing possibility occurs when $x \in (0, 1)$. In this case, with a risk-averse enough agent, it is possible that

$$\left[\sigma'_A(\tilde{s}) + [1 - \sigma'_A(\tilde{s})] \frac{u'\left(\tilde{P}\right)}{u'(\widetilde{ce})}\right] \times [1 - x(a^*, a)] > 1$$

(since P can be made less than $\widetilde{ce}$ and $u(\cdot)$ is concave). Hence, if (5) just fails to hold—that is, if

$$u'\left[u^{-1}(a^*)\right] \times \left[\sigma'_A\left(\tilde{S}\right) V'(a^*) + \left[1 - \sigma'_A\left(\tilde{S}\right)\right] \frac{\partial \widetilde{CE}}{\partial a^*}\right]$$

is not too much smaller than one—then the right-hand side of (16) can exceed one, at least for a in a neighborhood of a^*.

Unfortunately, we have been unable to derive sufficiently general conditions under which the parties will be able to exploit the agent's risk aversion as outlined in Observation #3 to achieve the first best. The conditions under which risk aversion can be exploited in this way strike us as sufficiently limited as to be not worth pursuing here.[24] Instead, in the next section, we explore an alternative variant of the model with more general applications.

---

[24] Among the assumptions that are sufficient for achieving the first best are that the domain of $u(\cdot)$ be unbounded below, that the ratio

$$\frac{u'(w)}{u'\left[u^{-1}\left(k + \frac{1}{2}u(w)\right)\right]} \to \infty$$

as $w \downarrow -\infty$ for all constants $k > -\infty$, and $\partial R / \partial a \geq \nu > 0$ for all $a \in [0, a^*]$. Details available from the authors upon request.

5 Verifiable Information

So far, the only verifiable information is who owns the asset and what payments have been made. As the Introduction argued, however, in many settings, such as the Pixar-Disney example, the principal should be able to provide a court with at least imperfect evidence about a when she can observe a. For this reason, we now assume that a verifiable signal about a is available.

Let the verifiable signal be $\theta \in \Theta$ and let $G(\cdot|a)$ be a probability measure conditional on a that maps some σ-algebra of Θ into $\mathbb{R}$. We consider contracts in which the principal maintains ownership of the asset and commits to pay the agent a wage $w(\cdot) : \Theta \to \mathbb{R}$. Observe that if the agent chooses the first-best action, a^*, then the principal will choose her first-best action, b^*. Finally, we assume that between the time the principal observes a and the realization of θ that the principal and agent can renegotiate their contract.[25]

We begin by reviewing a result of Hermalin and Katz (1991) :

Proposition 6 *Assume there exists a contract $w(\cdot)$ that would induce the first-best action, a^*, from the agent* absent *renegotiation.[26] Then if either the principal or the agent has all the bargaining power in renegotiation (i.e., $\sigma_A(S) = 0$ or S), the first-best outcome is attainable.*

Proof: Follows from Hermalin and Katz's Corollary 3, their Proposition 3, and their discussion preceding their Proposition 3. ∎

Remark 3 *When the agent has all the bargaining power (i.e., $\sigma_A(S) = S$), then Demski and Sappington's option contract will also yield the first best (see our Proposition 1 above).*

Proposition 6 has the same drawback that Demski and Sappington does—it relies on one side being able to make a take-it-or-leave-it offer in renegotiation. Hermalin and Katz propose an extension of this result to a more general bargaining game (their Proposition 4). Unfortunately, their proof does not properly account for renegotiation (they show their contract is incentive compatible *absent* renegotiation, but not under renegotiation).

To appreciate what can go wrong in more general bargaining games, consider the following example. Assume $u(w) = \ln(w)$. There are two possible signals, θ_1 and θ_2. The probability of θ_1 conditional on a is given by

$$\Pr\{\theta_1|a\} = \left\{ \begin{array}{l} \frac{1}{2} \text{ if } a < 1 \\ \frac{3}{4} \text{ if } a \geq 1 \end{array} \right. .$$

Suppose the first-best action is $a = 1$. Observe that it is implementable *without* renegotiation by Proposition 2 of Hermalin and Katz (1991). Suppose that

---

[25] We assume that at the renegotiation stage the only information that can be used to forecast θ is a. See Hermalin and Katz (1991) for a discussion of what happens if additional information becomes available ("leaks") at renegotiation.

[26] Hermalin and Katz (1991) give mild conditions under which this is true if Θ is discrete. These conditions can be readily extended to cover a continuum of signals.

the principal's bargaining power in renegotiation increases with a. Specifically, assume she has all the bargaining power when $a = 1$, but no bargaining power when $a = 0$. The agent's utility if he chooses $a = 0$ is, then,

$$U_0 \left(w_1, w_2 \right) = \ln \left(\frac{1}{2} w_1 + \frac{1}{2} w_2 \right),$$

where $w_i = w \left(\theta_i \right)$. If the agent chooses $a = 1$, his utility is

$$U_1 \left(w_1, w_2 \right) = \frac{3}{4} \ln \left(w_1 \right) + \frac{1}{4} \ln \left(w_2 \right) - 1.$$

A necessary condition for $a = 1$ to be implementable *with* renegotiation is that there exist w_1 and w_2 such that

$$U_1 \left(w_1, w_2 \right) \geq U_0 \left(w_1, w_2 \right). \tag{18}$$

However, no such w_1 and w_2 exist.[27] Hence, the first best is *not* implementable under this bargaining game.

Admittedly, the bargaining game here is quite special; in particular, it is non-monotonic. Its purpose is solely to prove that the results of Proposition 6 are sensitive to the renegotiation process.

Proposition 7 *The fact that action $\tilde{a}$ is implementable* without *renegotiation does not entail that it is implementable for all renegotiation processes.*

To understand why it's not possible to implement the first best in the previous example, consider Figure 1.[28] Necessary conditions for the first best are that $U_1 \left(w_1, w_2 \right) = 0$ (the individual rationality constraint) and (18) hold. Given the individual rationality constraint, we can rewrite (18) as $U_0 \left(w_1, w_2 \right) < 0$. When the agent has all the bargaining power, his indifference curves in wage space are straight lines. When the principal has all the bargaining power, the agent's indifference curves are convex. U_0 is his indifference curve passing through $u^{-1} \left(0 \right)$ assuming he has the bargaining power. U_1 is his indifference curve passing through $u^{-1} \left(1 \right)$ assuming the principal has the bargaining power. For a contract to satisfy individual rationality, it must lie on U_1. For it to satisfy incentive compatibility (18), it must lie below U_0. Because, however, $u^{-1} \left(1 \right) > u^{-1} \left(0 \right)$ and the agent is sufficiently risk averse, U_0 is always below U_1; that is, no contract implements $a = 1$ given the bargaining game we assumed.

In light of this example, the obvious "remedy" is to restrict attention to bargaining games in which the agent's bargaining power is always sufficiently big. We capture the idea of "sufficiently big" by saying a monotonic sharing rule is *uniformly strict* if $\sigma'_A \left(\cdot \right) \in \left(\eta, 1 - \eta \right)$ for some $\eta > 0$. In this case, we can prove

---

[27] Maximizing $U_1 \left(w_1, w_2 \right) - U_0 \left(w_1, w_2 \right)$ with respect to w_1 and w_2 reveals that any w_1 and w_2 pair on the line $w_1 = 3 w_2$ is optimal. When $w_1 = 3 w_2$, $U_1 \left(w_1, w_2 \right) - U_0 \left(w_1, w_2 \right)$ reduces to $\ln \left(3^{3/4}/2 \right) - 1$, which is negative.

[28] In this figure and the next, we adopt the convention of labeling points on the 45° line by the common value of their coordinates (e.g., the point (w, w) would be labeled as w).
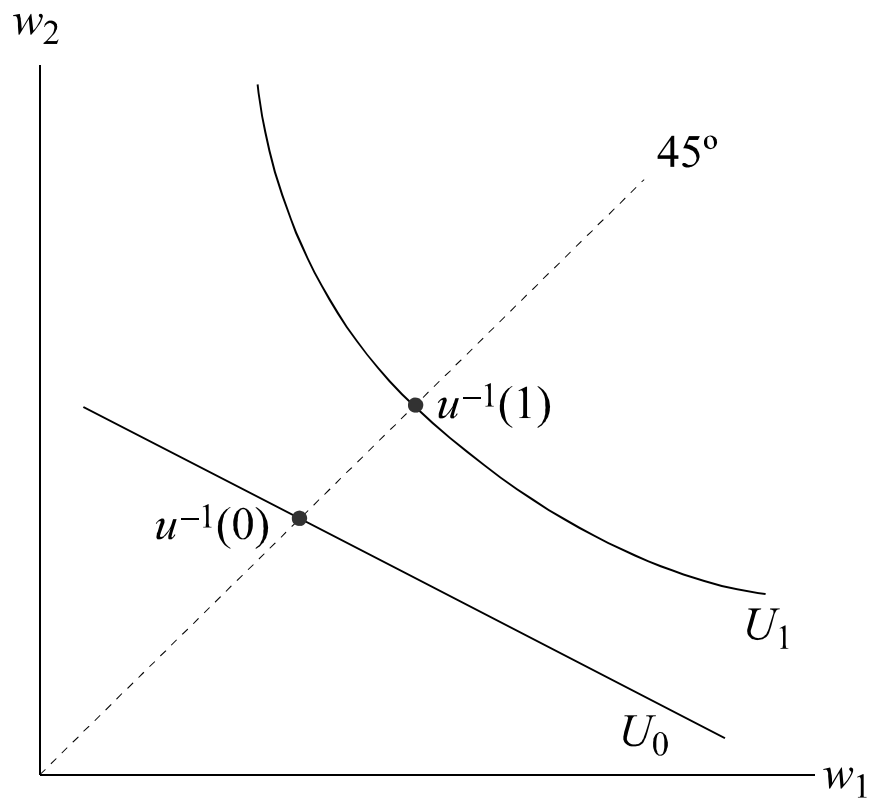
Figure 1: The first best, $a = 1$, is *not* attainable because U_1 can never be below U_0 when the principal has all the bargaining power at $a = 1$ and the agent has all of it at $a = 0$.

a slightly different, but correct version of Hermalin and Katz's Proposition 4. We show that even noisy signals will facilitate efficient contracts for uniformly strictly monotonic bargaining games.

Proposition 8 *Assume there exists a subset Θ^* of Θ such that $G(\Theta^*|a^*) > G(\Theta^*|a)$ for all $a \neq a^*$, that the range of $u(\cdot)$ is unbounded, and that the set of feasible actions, $\mathcal{A}$, is a* finite *subset of $[0, \infty)$. Then the first-best outcome is attainable for uniformly strictly monotonic sharing rules.*

Proof: Define $\pi(a) = G(\Theta^*|a)$ and consider the contract

$$w(\theta) = \begin{cases} w_1, \text{ if } \theta \in \Theta^* \\ w_2, \text{ if } \theta \notin \Theta^* \end{cases}.$$

After the agent chooses a, the parties will renegotiate. By assumption, the agent's resulting utility is

$$h(w_1, w_2, a) = u\left(\sigma_A\left[\hat{s}(w_1, w_2, \pi(a))\right] + \widehat{ce}(w_1, w_2, \pi(a))\right),$$

where

$$\begin{aligned} \widehat{ce}(w_1, w_2, \pi) &\equiv u^{-1}\left[\pi u(w_1) + (1-\pi) u(w_2)\right] \text{ and} \\ \hat{s}(w_1, w_2, \pi) &\equiv \pi w_1 + (1-\pi) w_2 - \widehat{ce}(w_1, w_2, \pi). \end{aligned}$$

A contract achieves the first best if and only if

$$\begin{aligned} h(w_1, w_2, a^*) - a^* &= 0; \text{ and} & (19) \\ h(w_1, w_2, a) - a &\leq h(w_1, w_2, a^*) - a^* \text{ for all } a \neq a^*. & (20) \end{aligned}$$

Define $w_2^*(w_1)$ so that $\langle w_1, w_2^*(w_1) \rangle$ satisfies (19). To see that $w_2^*(\cdot)$ is well-defined, observe that $\widehat{ce}(w_1, w_2, \pi)$ and $\pi w_1 + (1-\pi) w_2$ have positive first derivatives with respect to both w_1 and w_2. Hence, under a strictly monotonic sharing rule, h has positive first derivatives with respect to both w_1 and w_2. Together with the fact that

$$w_1 = w_2 = u^{-1}(a^*)$$

satisfy (19), these observations imply that $w_2^*(\cdot)$ indeed exists and, moreover, is decreasing and continuous. Note that $w_1 > u^{-1}(a^*)$ implies $w_2^*(w_1) < w_1$. For the remainder of the proof, we will consider only $w_1 > u^{-1}(a^*)$.

Observe that, for $\langle w_1, w_2^*(w_1) \rangle$, (20) can be rewritten as

$$h(w_1, w_2^*(w_1), a) - a \leq 0 \text{ for all } a \neq a^*. \quad (21)$$

Let

$$Q(w_1, w_2, \pi) \equiv \sigma_A\left[\hat{s}(w_1, w_2, \pi)\right] + \widehat{ce}(w_1, w_2, \pi),$$

then we can rewrite (20) using (19) and (21) as

$$Q(w_1, w_2^*(w_1), \pi(a^*)) - Q(w_1, w_2^*(w_1), \pi(a)) \geq u^{-1}(a^*) - u^{-1}(a) \quad (22)$$

for all $a \neq a^*$. Observe that

$$
\begin{aligned}
\frac{\partial Q\left(w_1, w_2^*\left(w_1\right), \pi\right)}{d\pi} &= \sigma_A'\left[\hat{s}\left(w_1, w_2^*\left(w_1\right), \pi\right)\right]\left(w_1 - w_2^*\left(w_1\right)\right) \\
&\quad + \left(1 - \sigma_A'\left[\hat{s}\left(w_1, w_2^*\left(w_1\right), \pi\right)\right]\right)\frac{\partial \widehat{ce}}{\partial \pi}.
\end{aligned}
$$

Note $\partial \widehat{ce}/\partial \pi > 0$ because $w_1 > w_2^*\left(w_1\right)$. Since the sharing rule is uniformly strictly monotonic, there exists an $\eta > 0$ such that

$$
\frac{\partial Q}{d\pi} \geq \eta\left(w_1 - w_2^*\left(w_1\right)\right). \tag{23}
$$

Now, employ the mean value theorem and, *then*, (23), to arrive at

$$
Q\left(w_1, w_2^*\left(w_1\right), \pi\left(a^*\right)\right) - Q\left(w_1, w_2^*\left(w_1\right), \pi\left(a\right)\right) \geq \left[\pi\left(a^*\right) - \pi\left(a\right)\right]\eta\left(w_1 - w_2^*\left(w_1\right)\right).
$$

Since $\mathcal{A}$ is finite, we can choose $\hat{a} \in \mathcal{A}\backslash\{a^*\}$ to minimize $\pi\left(a^*\right) - \pi\left(a\right)$.

$$
Q\left(w_1, w_2^*\left(w_1\right), \pi\left(a^*\right)\right) - Q\left(w_1, w_2^*\left(w_1\right), \pi\left(a\right)\right) \geq \left[\pi\left(a^*\right) - \pi\left(\hat{a}\right)\right]\eta\left(w_1 - w_2^*\left(w_1\right)\right). \tag{24}
$$

Since $w_2^*\left(\cdot\right)$ is a continuous and decreasing function and $\pi\left(a^*\right) > \pi\left(\hat{a}\right)$, we can find a w_1^* such that the right-hand side of (24) exceeds $u^{-1}\left(a^*\right) - u^{-1}\left(0\right)$. Transitivity then yields

$$
Q\left(w_1^*, w_2^*\left(w_1^*\right), \pi\left(a^*\right)\right) - Q\left(w_1^*, w_2^*\left(w_1^*\right), \pi\left(a\right)\right) \geq u^{-1}\left(a^*\right) - u^{-1}\left(a\right)
$$

for all $a \neq a^*$. Hence, (22) and, thus, (20) holds. By construction, (19) holds. Therefore, the contract $\langle w_1^*, w_2^*\left(w_1^*\right)\rangle$ is efficient. ∎

Intuitively, as the agent switches from a to a^* he raises the probability π that $\theta \in \Theta^*$. Raising π increases what the principal is willing to pay the agent to buy out the contract at a rate $w_1 - w_2$. The agent's bargaining position also improves with increases in π, but even ignoring this, the agent receives at least a share η of the principal's extra willingness to pay under uniformly monotonic sharing. Hence, by driving the wedge $w_1 - w_2$ sufficiently large (maintaining the (IR) constraint), we can induce an arbitrarily large final pay difference between actions a^* and a. For a finite set of actions, this is sufficient to find a first-best contract.

One "strike" against Proposition 8 is that it changes the action space from the non-negative real numbers to a finite (although arbitrarily large) subset of them. As a matter of economics this is not terrible—an argument can be made that the world is discrete and continuous models are the greater abstraction. Nevertheless, for completeness, we "extend" the analysis to the continuous case.

Proposition 9 *Same assumptions as Proposition 8, except let $\mathcal{A} = [0, \infty)$. Assume, in addition, that $G\left(\Theta^*|\cdot\right)$ is a continuous function. Then for any $\varepsilon > 0$, there is a contract that induces an action within ε of the first-best action, a^*.*

Proof: The notation is the same as in the previous proof. Since $\pi(\cdot)$ is continuous, let $\hat{a} \in \arg\max_{[0, a^* - \varepsilon]} \pi(a)$, and choose w_1^* sufficiently high that

$$Q\left(w_1^*, w_2^*\left(w_1^*\right), \pi\left(a^*\right)\right) - Q\left(w_1^*, w_2^*\left(w_1^*\right), \pi\left(\hat{a}\right)\right) \geq u^{-1}\left(a^*\right) - u^{-1}\left(0\right).$$

Then the agent prefers a^* to any $a \in [0, a^* - \varepsilon]$ given the contract $\langle w_1^*, w_2^*\left(w_1^*\right)\rangle$. Moreover, since $w_1^* > w_2^*\left(w_1^*\right)$ and a^* maximizes $\pi(a)$, choosing a^* dominates choosing any $a > a^*$. Hence, the contract $\langle w_1^*, w_2^*\left(w_1^*\right)\rangle$ must implement some $a \in (a^* - \varepsilon, a^*]$. ∎

In other words, under appropriate conditions on the distribution of the signal conditional on the agent's action, we can be sure of getting arbitrarily close to the first best when the agent's action space is continuous.

The condition that there exist some set $\Theta^* \subset \Theta$ such that a^* maximizes $G\left(\Theta^*|a\right)$ was important for establishing these results. This condition would be satisfied, for example, if $\theta = a + \nu$, where ν is a mean-zero random variable whose density is quasi-concave with a mode of zero; that is, loosely, if the signal is a noisy observation of the action.[29] This condition allows us to reduce the problem to a two-state problem. In this two-state world, an agent's indifference curve lies between the straight line that would be his indifference curve if he had all the bargaining power (e.g., U_0 in Figure 1) and the curve that would be his indifference curve if the principal had all the bargaining power (e.g., U_1 in Figure 1).

Consider Figure 2. Let $\mathcal{A} = \{a_1, a_2, a^*\}$ and let U^* be the indifference curve passing through $u^{-1}\left(a^*\right)$ conditional on $a = a^*$. To satisfy individual rationality, a first-best contract must lie on U^*. Let $U\left(\tilde{a}\right)$ be the indifference curve passing through $u^{-1}\left(\tilde{a}\right)$ conditional on $a = \tilde{a}$. Observe that since $G\left(\Theta^*|a^*\right) > G\left(\Theta^*|\tilde{a}\right)$, U^* is steeper when crossing the $45°$ line than any $U\left(\tilde{a}\right)$. To satisfy incentive compatibility, a first-best contract must lie below $U\left(\tilde{a}\right)$ for all $\tilde{a} \neq a^*$. Consider, e.g., $\mathbf{w}^*$ in Figure 2. Since it lies on U^* but below $U\left(\tilde{a}\right)$ for all $\tilde{a} \neq a^*$, it is a first-best contract. Proposition 8 establishes that we can draw Figure 2; that is, it shows that the indifference curves under a uniformly strictly monotonic sharing rule are not too flat for low a nor too convex for a^*, either of which could prevent us from finding a $\mathbf{w}^*$. As $\tilde{a} \to a^*$, the intersection of $U\left(\tilde{a}\right)$ and U^* may approach $(+\infty, -\infty)$. This is why Proposition 8 restricts attention to a finite set of feasible actions and why, with a continuum of feasible actions, we are only assured of getting within ε of the first best.

As Figure 2 illustrates, a first-best contract could require choosing a point far from the full-insurance ($45°$) line. In principle, this could potentially create two difficulties for attaining the first best that are not contemplated by our model. First, since renegotiation entails going from a risky contract (e.g., $\mathbf{w}^*$) to full insurance, the farther away the risky contract is from full insurance, the more the parties have at stake in their bargaining. If bargaining costs—so far assumed to be zero—are increasing in the amount at stake, then the bargaining

---

[29] See Hermalin and Katz (1991) for further discussion of this assumption and its relationship to their less restrictive "convex-hull condition."
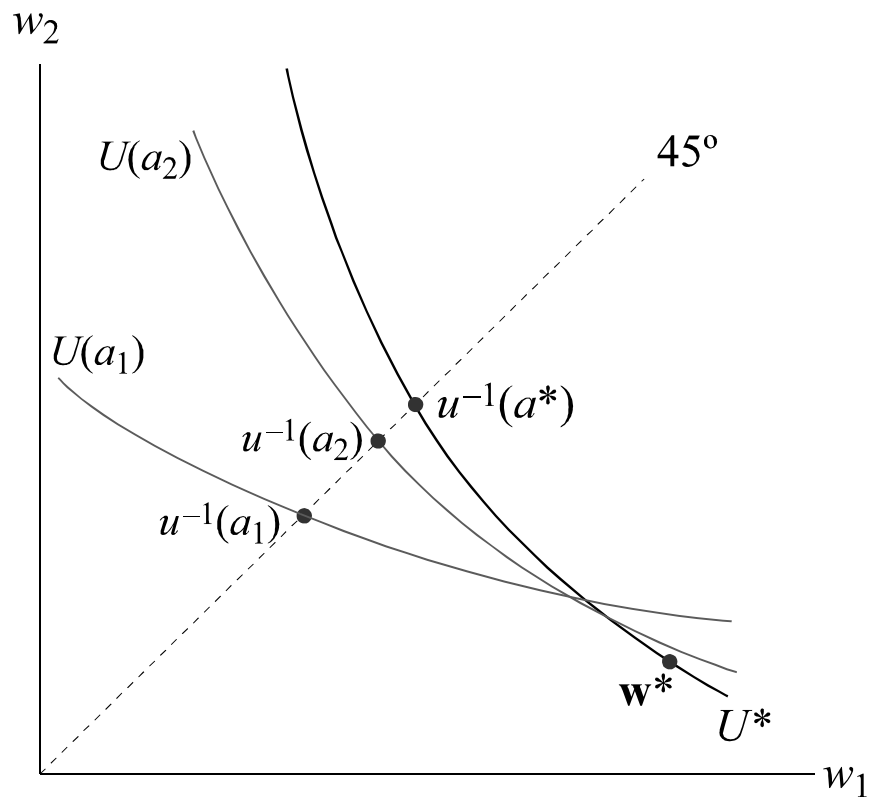
Figure 2: The first best is attainable under uniformly strict monotonic bargaining.

costs associated with a contract too far from full insurance could lead the parties to choose an initial contract closer to full insurance even though it would not have optimal incentive properties (i.e., would induce an action other than the first-best action). Second, it is possible that a first-best initial contract could have a very low w_2 (including possibly a negative value depending on $u(\cdot)$). Consequently, the courts might interpret w_2 as a penalty (especially if negative) and refuse to enforce it, citing the *penalty doctrine* or the law's abhorrence of forfeiture. Anticipation of such an eventuality would undermine the agent's incentives. It is likely, therefore, that the results in Propositions 8 and 9 depend upon a court's willingness to enforce punitive liquidated damage clauses; or, viewed otherwise, to impose specific performance.

To consider the legal issues in a more concrete fashion, let's return to the Pixar-Disney example. Suppose that Disney promises to pay Pixar $\bar{w}$ for a good film. In exchange, Pixar promises to make a good film ($a = 1$). Assume the value of a good film is V_G, while the value of a bad one is $V_B < V_G$. Assume it is efficient (first best) for Pixar to work hard (i.e., choose $a = 1$); i.e.,

$$V_G - u^{-1}(1) \geq V_B - u^{-1}(0).$$

Should a dispute arise between Disney and Pixar, the court would find itself unable to determine whether the film is good. It can, however, hear the testimony of Siskel and Ebert, two reputable and prominent film critics. Their appraisal, $\theta \in \{V_B, V_G\}$, is distributed

$$\Pr\{\theta = V_G | a\} = \left\{ \begin{array}{l} z_0 \text{ if } a < 1 \\ z \text{ if } a \geq 1 \end{array} \right. ,$$

where $z > z_0$.

Suppose the courts refuse to enforce stipulated damages and instead impose *expectation damages*, which in this context would be a penalty $V_G - V_B$, if $\theta = V_B$.[30] Pixar's payment is, then,

$$w(\hat{s}) = \left\{ \begin{array}{l} \bar{w} \text{ if } \theta = V_G \\ \bar{w} + V_B - V_G \text{ if } \theta = V_B \end{array} \right. .$$

Clearly, all a except $a = 0$ and $a = 1$ are strictly dominated strategies for Pixar. For convenience, assume $\sigma'_A(S) = \frac{1}{2}$. From the proof of Proposition 8, a sufficient condition for the first best to be implementable is

$$V_G - V_B \geq \frac{u^{-1}(1) - u^{-1}(0)}{\frac{1}{2}(z - z_0)}. \tag{25}$$

Suppose that a good film generates substantial surplus; specifically, suppose

$$V_G - V_B > 2\left[u^{-1}(1) - u^{-1}(0)\right]$$

(the difference in benefits exceeds twice the difference in costs). Hence, if $z \gg z_0$, so that $z - z_0 \approx 1$, then the payment difference $(V_G - V_B)$ is big enough to induce

---

[30] See Restatement (Second) of Contracts, Sections 356, 1981; and UCC Sections 2-718 (1).

Pixar to choose $a = 1$. A value for $\bar{w}$ can then be found such that the Pixar's individual rationality constraint is just satisfied. However, as $z - z_0 \downarrow 0$, (25) will fail, meaning that the first best is not implementable with expectation damages. To implement $a = 1$ would, then, require allowing contracts in which $w(V_G) - w(V_B) > V_G - V_B$; i.e., allowing stipulated damages in excess of expectation damages.

When $z - z_0 \downarrow 0$, Siskel and Ebert's appraisal becomes an increasingly poor signal, so that true damages are increasingly difficult to ascertain. Since one criterion for determining whether a stipulated damages clause is valid is that damages be difficult to ascertain, it is tempting to see our findings as consistent with legal rules.[31] However, courts typically require that the stipulated damages be a reasonable *ex ante* anticipation of damages, or a reasonable *ex post* approximation, or both.[32] In this case, the maximum anticipated damage is $V_G - V_B$, so stipulated damages that substantially exceeded this amount would be viewed as an invalid penalty. Yet, as we've seen, such damages may be necessary to provide adequate incentives if there is no reliable signal to determine when they should be assessed. The situation is quite similar to the general deterrence argument of Becker (1968) that penalties must be many times the social cost of an illicit activity if the chances of apprehension are slim. The legal rules for determining the validity of liquidated damage clauses do not seem to account for this fact adequately.

6   Conclusion

It is problematic to provide an agent with efficient incentives if (1) contracts will be renegotiated when they prescribe inefficient outcomes and (2) the parties share the gains from such renegotiation. This problem is particularly acute when no verifiable information beyond ownership and monetary transfers is available, the problem we considered in Section 3.

To summarize our analysis of this problem and to relate it to the trading literature, consider Figure 3 (this discussion is influenced by ongoing work with Ilya Segal and Michael Whinston). The trading literature considers the efficiency of contracting when the agent's investment affects the principal's value for trade or the agent's cost of trade. In our context, trade is equivalent to the principal taking possession at the end of the game. Consequently, we can interpret the principal's value as $V(\cdot)$ and the agent's production cost as $CE(\cdot, p_1)$ (since $CE(\cdot, p_1)$ is the agent's opportunity cost of transferring ownership).[33] The ver-

---

[31] See Restatement (Second) of Contracts, Section 356, Comment B, factor 2.

[32] See Restatement (Second) of Contracts, Section 356, Comment B, factor 1, and California Civil Code, Section 1671 (b).

[33] Observe that for a *given* a, it doesn't matter whether we interpret $CE(a, p_1)$ as a production cost (a trading model) or as the foregone benefit of retaining ownership (an agency model). The agent's investment incentives do, however, depend on the interpretation. In particular, going from an agency model to a trading model we have to subtract $CE(a, p_1)$ from his payoffs in both the trade and no-trade outcome. Since CE is a function of a, this will change the agent's investment incentives. Consequently, some caution needs to be exercised
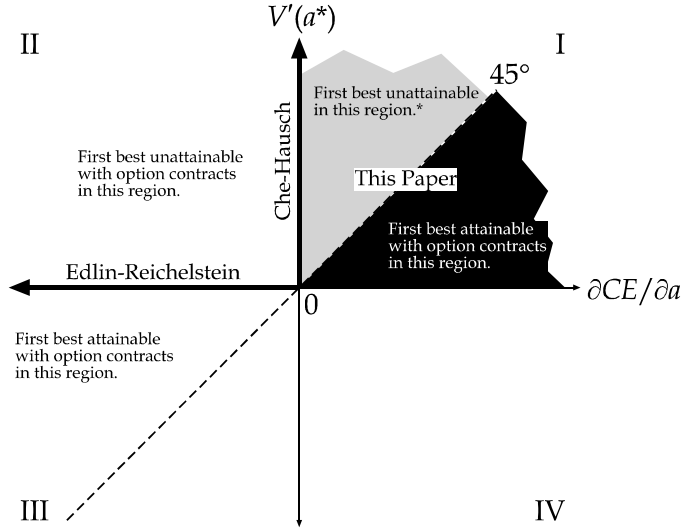
$V'(a^*)$

II   I

                                                          45°

First best unattainable
in this region.*

First best unattainable   This Paper
with option contracts
in this region.   First best attainable
                                   with option contracts
                                   in this region.

Che-Hausch

Edlin-Reichelstein $\partial CE/\partial a$

0

First best attainable
with option contracts
in this region.

III   IV

Figure 3: The Relationship Among Recent Papers in the Trading Literature with Renegotiation (* comparing risk-neutral agents; for risk-averse agents the first best is unattainable in this region when attention is restricted to option contracts).

tical axis shows the marginal return to the principal from the agent's action, $V'(a^*)$, while the horizontal axis shows the marginal increase to the agent's production cost, $\partial CE(a^*, p_1)/\partial a$. Edlin and Reichelstein (1996) consider the bold (negative) portion of the horizontal axis, where the agent's investment lowers his production costs (i.e., $\partial CE/\partial a < 0$) but does not affect the principal's value (i.e., $V'(a^*) = 0$).[34] In contrast, Che and Hausch (1996) consider the bold (positive) portion of the vertical axis, where the agent's investment is in quality and, so, raises the principal's value (i.e., $V'(a^*) > 0$), but does not affect the cost of production (i.e., $\partial CE/\partial a = 0$). Our paper has analyzed the positive quadrant (I), where the agent's investment raises both the principal's value and the cost of production.

Because surplus is divided in renegotiation, the agent cannot capture all of the marginal contributions of his effort to the principal's value ($V'(a^*)$). Although this undermines the agent's investment incentives, we identified a countervailing effect: Investment by the agent increases his opportunity cost of trade ($\partial CE/\partial a > 0$), which improves his bargaining position in renegotiation and thereby increases his final payoff. Unlike Che and Hausch, this means that there will not generically be underinvestment. In particular, below the 45°

when interpreting our analysis in a trading-model context.

[34] Aghion *et al.* (1994), Chung (1991), and Nöldeke and Schmidt (1995) also consider the bold portion of the horizontal axis. Unlike our paper and Edlin and Reichelstein, these papers effectively assign all the bargaining power to one party.

line—the region where the threat-point effect dominates the hold-up effect—there will actually be *over*investment if the parties don't sign a contract *ex ante*. In this region, an option guards against such overinvestment: Giving the principal an option to purchase the asset limits the agent's investment incentives because he gets no return *on the margin* when the principal exercises her option. If the option price is set appropriately, then at low levels of investment, the agent's return is high because the principal won't exercise her option. These considerations mean that the first best can be obtained in the darkly shaded region below the 45° line.

Above the 45° line, the agent will underinvest if he owns the asset initially and there is no *ex ante* contract: The agent's marginal return is a convex combination of $V'(a)$ and $\partial CE(a, p_1)/\partial a$, and so must be less than the social marginal return, $V'(a)$. Option contracts cannot improve this situation. Ownership provides the agent maximum incentives, so the principal's threat to take ownership—exercise the option—can only serve to lessen incentives. The degree to which an option lessens incentives depends on its strike price. As just noted, an infinite strike price—unencumbered ownership—provides maximum incentives. In contrast, a zero strike price—equivalently, purchasing the agent's investment in advance—provides no incentives. That is, incentives are increasing (weakly) in the strike price.

The fact that option contracts cannot provide stronger incentives than would prevail absent any contract stands in apparent contradiction to Demski and Sappington (1991), where incentives are increasing as the strike price tends to $V(a^*)$ from either direction. The difference has to do with renegotiation; we allow for it, but Demski and Sappington implicitly rule it out.

Figure 3 leads us to extend our agency model to other quadrants. Propositions 2 and 4 would apply equally well to Quadrant II if we allowed a to have a negative marginal contribution when $b = 0$: For this case, the first-best can't be achieved because the threat-point effect actually intensifies the holdup effect—investment lowers the value of the project to the agent on the margin. Quadrants III and IV, including the entire horizontal axis, are of little interest in an agency model: The principal's marginal value for the asset can be nonpositive at the efficient level of investment, a^*, only if $a^* = 0$; but the least-cost action (i.e., 0) can always be implemented at first-best cost.

Critical to these conclusions are our assumptions that exactly one unit can be traded and its return is unverifiable. Hence, unlike Edlin and Reichelstein, we cannot set a higher quantity in the *initial* contract to create the appropriate threat points for renegotiation. Moreover, unlike Hermalin and Katz, we cannot make the *initial* contract contingent on returns. Another critical assumption concerns the parties inability to avoid renegotiation. Unlike Nöldeke and Schmidt (1995) and Bernheim and Whinston (1997), who limit renegotiation to before the exercise date of the option, we recognize that the parties could wish to renegotiate after the exercise date of the option.[35] Hence, in our model,

---

[35] The Nöldeke and Schmidt model also differs from ours on other dimensions: In particular, they assume investments are privately and inalienably beneficial, whereas we assume invest-

the principal's threat not to exercise her option is credible, which creates the holdup effect.

Lastly, consider our assumption that the principal always values ownership of the asset more than the agent *at the exercise date*. This assumption makes renegotiation when the agent owns the asset—and thus the holdup of the agent—unavoidable. If this were not true; in particular, if they valued the asset equally at that time, then renegotiation after the exercise date would be pointless because there would be no efficiency to be gained. There would, thus, be no danger of holdup at this stage. Since it was this holdup that prevented the Demski-Sappington contract from being efficient, the consequence of equal valuation at the exercise date is that the Demski-Sappington contract implements the first best throughout Quadrant I. This also explains why Nöldeke and Schmidt (1997) find that the Demski-Sappington option price is renegotiation proof and implements the first best throughout Quadrant I: Nöldeke and Schmidt assume equal valuation by assuming that the there is no risk and that the exercise date occurs *after* the principal invests.

When the parties' valuations are unequal at the exercise date, then the agent must internalize the principal's value when investing and an option contract will not provide him with sufficient incentives to do so unless the threat-point effect dominates the now unavoidable holdup effect. The assumption of equal values strikes us as unrealistic in many settings. After all, the principal's greater value could create the motive for trade in the first place.[36] Second, in an agency setting, the agent's greater risk aversion would make the principal value the asset more.[37] Third, it may be technically difficult to ensure that the exercise date falls after the principal's investment; particularly since the principal has an incentive to increase her bargaining power either by postponing her investment or by moving up the exercise decision. For instance, consider the Nöldeke and Schmidt (1997) model in which options are exercised *after* the principal's investment (and the parties' valuations are the same), but imagine the following

---

ments enhance a tradeable asset; and they follow the bargaining game put forward by Hart and Moore (1988) for renegotiation, which essentially gives all the bargaining power to one party, whereas we allow both parties to have bargaining power.

[36] That is, ignoring risk but interpreting CE as the agent's valuation function, our analysis shows that the first best is not always attainable when $CE < V$; in particular, it won't be attainable above the $45°$ line in Quadrant I. Pixar and Disney offer an example in which $CE < V$: The view in the film industry is that an animated film is more valuable if owned by Disney than if owned by another studio (see note 2 *supra*).

[37] To begin, suppose, that the principal has no investment to make. Then $CE < V$ and whether the first best is attainable depends on whether one is in the top half or bottom half of Quadrant I (in particular, since the principal makes no investment, this conclusion does not depend on whether the option's exercise date falls before the principal's investment, as in our model, or after, as in Nöldeke and Schmidt).

Even if the principal does invest and this investment occurs *prior* to the exercise date, the Demski-Sappington option contract will fail to achieve the first best if the agent is risk averse: Under this contract, the principal's surplus if she invests efficiently and exercises her option is zero. If she lets her option expire and renegotiates, she likely captures some of the difference between V and CE in renegotiation (i.e., a *positive* fraction of the surplus). We would, therefore, expect meaningful renegotiation with a risk-averse agent, which re-introduces the holdup problem we identified earlier.

change: Suppose that prior to her investing, but after the agent invested, the principal can burn her copy of the option contract in front of the agent. Then the principal has a credible threat not to invest: Any investment she makes would be subject to expropriation by the agent, because she can no longer prove she has the right to purchase the asset at an agreed price.[38] Given that her threat not to invest is credible,[39] the parties risk an inefficient outcome unless a new contract is signed that restores investment incentives to the principal. But, then, the agent is again exposed to holdup, so whether efficiency can be realized will again depend on whether the threat-point effect dominates the holdup effect.

Returning to Quadrant I and an agency interpretation, whether the threat-point effect or the holdup effect dominates depends in large part on whether the agent and principal's actions are productive substitutes or complements. Indeed, if the agent is risk neutral—or in the CARA case we considered—complements implies being above the 45° line, where the holdup effect dominates; while substitutes implies being below it, where the threat-point effect dominates.

The dichotomy between complementary and substitutable actions could offer insights into what activities take place within a firm and which are done through the market: If, as is often assumed, control is greater within the firm than outside, then it might be efficient for a firm to encompass complementary activities because they are difficult to procure contractually.[40] For example, it may have made sense for General Motors to buy Fisher Body, because General Motors' efforts at designing and marketing automobiles were complementary to Fisher Body's efforts to make auto bodies.[41] In contrast, Disney and Pixar may be efficient as distinct firms if the better Pixar makes its animation technology, the smaller the marginal return to Disney's marketing activities is for attracting movie-goers.[42]

---

[38] Note, at this stage, the agent also has an incentive to deny a pre-existing option contract, since that way he guarantees he gets full value for the asset (which, recall, is the same for both parties).

[39] We also note that even if burning the option contract is impossible, the principal's threat *not* to invest is *not* incredible: In the Nöldeke and Schmidt equilibrium she is just indifferent between investing efficiently and not investing at all. Admittedly this indifference can be broken by having the agent invest $a^* + \varepsilon$, which is why, as a matter of neo-classical game theory, there is not a subgame-perfect equilibrium in which the principal doesn't invest. On the other hand, the experimental literature on the ultimatum game and the theoretical literature on fairness (see Rabin, 1993, on both these literatures) suggests that if the principal were not promised a significant (i.e., greater than ε) share of the surplus generated by her investment, then she would actually prefer not to invest to hurt the agent.

[40] Although coming from a rather different model, Grossman and Hart (1986) reached a similar conclusion. It is worth noting, however, that more recently, Chiu (1996), working in a similar vein to Grossman and Hart, has challenged their conclusion as being overly sensitive to the assumed bargaining game between the parties.

[41] General Motors efforts expanded the market for their cars which made any efforts by Fisher more valuable.

[42] That is, assume revenue is $\rho(q, m)$, where q is "quality," m is marketing activity and $\frac{\partial^2 \rho}{\partial q \partial m} < 0$. For example, suppose a customer goes to the movies if either a friend recommends it or he is impressed by the marketing. Let $\mu(q)$ be the probability that a friend does *not* recommend it ($\mu' < 0$) and let $\lambda(m)$ be the probability that he is *not* impressed by the marketing ($\lambda' < 0$). Assuming independence, the probability he goes to the movie is $1 -$

Providing an agent with efficient incentives is less problematic when there is additional verifiable information about the agent's action, as discussed in Section 5. Not surprisingly, more verifiable information improves the situation. In particular, for uniformly strictly monotonic sharing rules, a sufficient condition for achieving the first best (or an outcome within ε of the first best) is that there be a set of verifiable signals that is most likely to occur if the first-best action is chosen.

Achieving the first-best will be most difficult when the principal and agent must take complementary actions, but only a noisy signal is available. Complementarity implies that an option contract won't work. Moreover, although the signal means that the contracts we constructed for Propositions 8 and 9 should work, the wage given a "bad" realization of the signal could be far below the wage given a "good" signal if the signal is noisy. Courts will very likely view this difference as punitive, and refuse to enforce the contract; thereby precluding the first best. The law's hostility to penalties is revealed by the Restatement (Second) of Contracts, 1981, American Law Institute Publishers, St. Paul, Section 356, comment (a):

> The central objective behind the system of contract remedies is compensatory, not punitive. *Punishment of a promisor for having broken his promise has no justification on either economic or other grounds* and a term providing such a penalty is unenforceable on grounds of public policy (emphasis added).[43]

Our analysis rejects the premise that supra-compensatory damages have "no justification on ... economic ... grounds": There are situations in which the parties do better with supra-compensatory remedies.[44] The essential intuition is simple. When it is difficult to observe effort, then a party may often get away with breach, and so must be punished when breach is actually "proven"; this intuition is similar to Becker's (1968) observation that if a criminal is only apprehended occasionally, the penalty must exceed the social cost to induce optimal deterrence. Hence, our analysis suggests replacing the penalty doctrine with a more flexible policy.

References

Aghion, P., Dewatripont, M., and Rey, P. (1994). Renegotiation design with unverifiable information. *Econometrica*, **62**, 257–282.

$\mu(q)\lambda(m)$, which has the required negative cross partial derivative.

[43] See also Schwartz (1990), Restatement (Second) of Contracts, Sections 356, 359 (1), 1981; and UCC Sections 2-718 (1), 2-716 (1)).

[44] Schwartz (1990) has argued that the courts should not strike what appear *to them* to be supra-compensatory remedies, because what the law views as compensatory may, in fact, be undercompensatory. Our analysis pushes the objection one step farther by rejecting the Restatement's premise that supra-compensatory remedies are undesirable.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, **76**, 169–217.

Bernheim, B. D. and Whinston, M. D. (1997). Incomplete contracts and strategic ambiguity. *American Economic Review*. forthcoming.

Che, Y.-K. and Chung, T.-Y. (1996). Contract damages and cooperative investments. University of Wisconsin Working Paper.

Che, Y.-K. and Hausch, D. B. (1996). Cooperative investments and the value of contracting: Coase vs. Williamson. University of Wisconsin Working Paper.

Chiu, Y. S. (1996). Noncooperative bargaining, hostages, and optimal asset ownership. Working paper, Dept. of Decision Sciences and Managerial Economics, Chinese University of Hong Kong.

Chung, T.-Y. (1991). Incomplete contracts, specific investments, and risk sharing. *Review of Economic Studies*, **58**, 1031–1042.

Demski, J. S. and Sappington, D. E. M. (1991). Resolving double moral hazard problems with buyout agreements. *RAND Journal of Economics*, **22**, 232–240.

Edlin, A. S. and Reichelstein, S. (1996). Holdups, standard breach remedies, and optimal investment. *American Economic Review*, **86**, 478–501.

Fudenberg, D. and Tirole, J. (1990). Moral hazard and renegotiation in agency contracts. *Econometrica*, **58**(6), 1279–1320. November.

Grossman, S. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, **94**(4), 691–719. August.

Hart, O. D. and Moore, J. (1988). Incomplete contracts and renegotiation. *Econometrica*, **56**, 755–785.

Hermalin, B. E. and Katz, M. L. (1991). Moral hazard and verifiability: The effects of renegotiation in agency. *Econometrica*, **59**, 1735–1753.

Hermalin, B. E. and Katz, M. L. (1993). Judicial modification of contracts between sophisticated parties: A more complete view of incomplete contracts and their breach. *Journal of Law, Economics, & Organization*, **9**, 230–255.

Ma, C.-t. A. (1994). Renegotiation and optimality in agency contracts. *Review of Economic Studies*, **61**(1), 109–129. January.

Nöldeke, G. and Schmidt, K. M. (1995). Option contracts and renegotiation: A solution to the hold-up problem. *RAND Journal of Economics*, **26**(2), 163–169. Spring.

Nöldeke, G. and Schmidt, K. M. (1997). Sequential investments and options to own. Working paper, University of Bonn.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, **83**(5), 1281–1302. December.

Rogerson, W. P. (1992). Contractual solutions to the hold-up problem. *Review of Economic Studies*, **59**, 777–793.

Schwartz, A. (1990). The myth that promisees prefer supracompensatory remedies: An analysis of contracting for damage measures. *Yale Law Journal*, **100**(2), 369–407. November.

Segal, I. and Whinston, M. (1997). Implementation with renegotiation: A note. Mimeo.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. The Free Press, New York.