Econometrica, Vol. 70, No. 1 (January, 2002), 1–45

THE MIRRLEES APPROACH TO MECHANISM DESIGN WITH RENEGOTIATION (WITH APPLICATIONS TO HOLD-UP AND RISK SHARING)

BY ILYA SEGAL AND MICHAEL D. WHINSTON[1]

The paper studies the implementation problem, first analyzed by Maskin and Moore (1999), in which two agents observe an unverifiable state of nature and may renegotiate inefficient outcomes following play of the mechanism. We develop a first-order approach to characterizing the set of implementable utility mappings in this problem, paralleling Mirrlees's (1971) first-order analysis of standard mechanism design problems. We use this characterization to study optimal contracting in hold-up and risk-sharing models. In particular, we examine when the contracting parties can optimally restrict attention to simple contracts, such as noncontingent contracts and option contracts (where only one agent sends a message).

KEYWORDS: Implementation with renegotiation, first-order approach, option contracts, noncontingent contracts, relationship-specific investments, risk sharing.

1. INTRODUCTION

THIS PAPER STUDIES THE MECHANISM design problem in which two agents have complete information about each other's preferences, and renegotiate inefficient outcomes following play of the mechanism. Since efficiency is always achieved through renegotiation, the role of the mechanism is to influence the allocation of the available surplus between the two agents. This allocation is important when the parties want to create proper ex ante investment incentives, as in hold-up models, or to allocate ex ante risks optimally.

Maskin and Moore (1999) formulated the mechanism design problem for complete information environments with renegotiation, and characterized implementable social choice rules in such environments with a set of incentive-compatibility constraints. However, in environments with a continuum of states of the world, their approach yields a double continuum of incentive constraints,

[1] We are grateful to Sandeep Baliga, Aaron Edlin, Oliver Hart, Steve Matthews, Steve Tadelis, and seminar participants at Berkeley, Harvard, LSE, Michigan, MIT, Montreal, Northwestern, Penn, Yale, the 1998 Stanford Institute for Theoretical Economics, and the 1999 Summer Meetings of the North American Econometric Society for helpful discussions and comments. The comments of three anonymous referees and a co-editor were instrumental in improving the paper's exposition. We thank Federico Echenique and Luis Rayo for excellent research assistance. The first author gratefully acknowledges financial support and hospitality of the Hoover Institution for War, Revolution, and Peace, and financial support from the National Science Foundation. The second author gratefully acknowledges the financial support and hospitality of the George J. Stigler Center for the Study of the Economy and the State at the University of Chicago, and financial support from the National Science Foundation.

which is hard to analyze. This paper builds on the work of Maskin and Moore by developing a first-order approach to incentive-compatibility, which focuses on local incentive constraints, and provides a more convenient characterization of implementable social choice rules. Our analysis therefore extends Mirrlees's (1971) approach in the standard mechanism design setting to the problem of mechanism design in complete information environments with renegotiation.

Our development of this first-order characterization focuses on environments in which agents have quasilinear preferences and the state of the world is a one-dimensional variable $\theta \in [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$. The heuristics of our approach can be described as follows. Consider a direct revelation mechanism that prescribes an outcome as a function of the two parties' announcements of the state. Suppose that the prescribed outcome is renegotiated to an ex post efficient outcome in each state of the world. Let $\widetilde{U}_i(\theta_1, \theta_2, \theta)$ denote the equilibrium post-renegotiation utility of agent i in the mechanism given the two agents' announcements $(\theta_1, \theta_2) \in [\underline{\theta}, \bar{\theta}]^2$ in state θ. Since renegotiation always yields an efficient outcome, we must have

$$\widetilde{U}_1(\theta_1, \theta_2, \theta) + \widetilde{U}_2(\theta_1, \theta_2, \theta) = S(\theta)$$

for all $(\theta_1, \theta_2, \theta)$, where $S(\theta)$ is the maximum total surplus achievable in state θ. Thus, the direct revelation mechanism defines a constant-sum announcement game between the parties. Suppose that truth telling constitutes a Nash equilibrium of the game. If the functions $\widetilde{U}_i(\cdot)$ induced by the mechanism are differentiable, we can see how party i's equilibrium payoff $U_i(\theta) \equiv \widetilde{U}_i(\theta, \theta, \theta)$ depends on the state θ using an Envelope Theorem argument at each point $\theta \in (\underline{\theta}, \bar{\theta})$. Specifically, we can write

$$U_i'(\theta) = \left[\frac{\partial \widetilde{U}_i(\theta, \theta, \theta)}{\partial \theta_i}\right] + \left\{\frac{\partial \widetilde{U}_i(\theta, \theta, \theta)}{\partial \theta_{-i}}\right\} + \frac{\partial \widetilde{U}_i(\theta, \theta, \theta)}{\partial \theta}.$$

Just as in the standard Mirrlees approach, the term in square brackets is zero because agent i is maximizing his payoff by announcing truthfully. Moreover, in a setting with renegotiation, the term in curly brackets is *also* zero since the constant sum nature of the game implies that agent $-i$ *minimizes* agent i's payoff by announcing truthfully. Hence, we have

$$U_i'(\theta) = \frac{\partial \widetilde{U}_i(\theta, \theta, \theta)}{\partial \theta},$$

where the partial derivative is taken holding the agents' equilibrium announcements fixed.

In the setting we consider, the mechanism can prescribe an outcome $\langle x, t_1, t_2 \rangle$, where $x \in X$ is a decision, and t_i is the transfer to party i. The post-renegotiation utility of each party i takes the form $u_i(x, \theta) + t_i$. (In this setting, the function $\widetilde{U}_i(\cdot)$ is differentiable if $u_i(\cdot)$ is differentiable and the mechanism is differentiable, i.e., if $\langle x, t_1, t_2 \rangle$ is a differentiable function of agents' announcements.)

Thus, $\partial \widetilde{U}_i(\theta, \theta, \theta)/\partial\theta = \partial u_i(\hat{x}(\theta), \theta)/\partial\theta$, where $\hat{x}(\theta)$ is the decision prescribed by the mechanism when both parties announce that the state is θ (i.e., it is the "equilibrium decision rule"). The above display then implies that

$$U_i(\theta) = U_i(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{\partial u_i(\hat{x}(\tau), \tau)}{\partial\theta}\, d\tau,$$

which parallels the key condition in the Mirrlees approach to standard mechanism design problems.

Section 2 contains the formal development of our characterization of implementable utility mappings $\langle U_1(\cdot), U_2(\cdot)\rangle$ for this environment. One might wonder whether additional utility mappings might be implementable using mechanisms in which the agents' post-renegotiation utility functions $\widetilde{U}_i(\cdot)$ are not differentiable (e.g., discontinuous mechanisms that punish the agents for disagreements about the state of the world). In fact, we show that when the decision set X is connected, the utility mapping implemented by *any* mechanism must satisfy the above "Mirrlees condition," but with the equilibrium decision rule $\hat{x}(\theta)$ in general replaced with some other "generating" decision rule $x(\cdot)$. Intuitively, this generalized Mirrlees condition requires that an agent's equilibrium utility not vary faster with the state than his underlying payoff evaluated at *some* feasible decision. While this condition embodies only local incentive constraints, we identify a weak assumption on preferences under which the condition also implies global incentive-compatibility, and hence fully characterizes implementable utility mappings. We also show that this characterization offers a straightforward comparison between the utility mappings that are implementable with general two-sided message games and those implementable with relatively simple *option contracts* in which only one agent makes an announcement (for example, by electing whether to trade at a predetermined price), or with even simpler *noncontingent* contracts that prescribe a fixed outcome (requiring no messages at all).

While the generalized Mirrlees condition restricts the set of implementable utility mappings, it does not restrict the equilibrium decision rules in mechanisms implementing these mappings. Indeed, the equilibrium decision rule in a mechanism implementing a *given* utility mapping is indeterminate. To make predictions concerning equilibrium decision rules, we focus on a class of mechanisms that are "continuous" in a certain sense. Such continuous mechanisms may be attractive because they allow agents to approximate the outcome by transmitting only a limited amount of information. We show that continuous mechanisms can implement a wide class of utility mappings, so that agents will often incur no loss from using them. Using an envelope theorem for saddle-point problems formulated by Milgrom and Segal (2002), we extend to continuous mechanisms the above derivation for differentiable mechanisms. Thus, we establish that the equilibrium decision rule in a continuous mechanism must also generate, via the Mirrlees condition, the utility mapping implemented by the mechanism.

In the rest of the paper, we apply our characterization results to ex ante contracting problems. In Section 3, we study a general model of the classic hold-up problem, which includes as particular cases the models studied by Hart and

Moore (1990), Demski and Sappington (1991), Hermalin and Katz (1991), Edlin and Reichelstein (1996), Segal (1999), Che and Hausch (1999), and Edlin and Hermalin (2000). In the model, two parties make ex ante investments, which affect their valuations for ex post trade. These valuations are observed by both parties ex post, but are not verifiable. The parties sign an ex ante contract that specifies the mechanism to be played at the ex post stage. The trade prescribed by the contractual mechanism is always renegotiated to an ex post efficient trade, and so the role of the mechanism is to influence the allocation of ex post surplus in a manner that improves the parties' investment incentives.

In much of the literature on the hold-up problem, only a restricted class of the feasible contracts is considered (such as noncontingent or option contracts). In contrast, by characterizing the ways in which the parties' ex post payoffs can be made to depend on their ex ante investments, our implementation results allow us to determine the set of sustainable investments and the nature of the parties' optimal contracts quite generally. In Section 3, we first identify circumstances in which simple noncontingent contracts are in fact optimal, even though the parties may not be able to achieve the first-best. Using that result, we next study the nature of optimal contracts in an environment that includes as special cases the settings studied by Edlin and Reichelstein (1996) and Che and Hausch (1999), and generate predictions regarding when the quantity specified in an optimal contract is on average renegotiated upward or downward. Finally, we identify circumstances in which optimality requires more complex contracts, such as option contracts.

In Section 4, we apply our characterization results to the design of optimal risk-sharing arrangements (also studied by Hart and Moore (1988), Chung (1991), Green and Laffont (1992), and Aghion, Dewatripont, and Rey (1994)). Here we identify the ways in which the parties can feasibly share risks across random nonverifiable states of nature. We formulate conditions under which optimal risk sharing can be achieved with an option contract, and study how the direction of ex post renegotiation (upward or downward) depends upon the nature of ex ante uncertainty.

We conclude in Section 5 with some general comments about the Maskin-Moore approach to mechanism design and renegotiation. We contrast it with another approach, followed by some papers in the literature, in which renegotiation is possible only before play of the mechanism, and compare what is achievable in the two cases.

2. CHARACTERIZATION OF IMPLEMENTABLE UTILITY MAPPINGS

Consider an environment with two agents, labeled $i = 1, 2$. The set of possible states is the one-dimensional interval $\Theta \equiv [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$. Both agents observe the realization of state $\theta \in \Theta$, but no outsider does (this is mechanism design under "complete information"; see Moore (1990)). Upon observing the state, the agents play a mechanism, which prescribes a decision from a compact space X, as well as monetary transfers. A general mechanism is a pair of message sets (M_1, M_2)

and a collection of functions $\langle x(\cdot), t_1(\cdot), t_2(\cdot)\rangle$ mapping message pairs $(m_1, m_2) \in M_1 \times M_2$ into a decision and transfer payments, having the property $t_1(m_1, m_2) + t_2(m_1, m_2) = 0$.[2]

As in Maskin and Moore (1999), we assume that following the play of the mechanism, the agents renegotiate to an efficient outcome. The outcome of renegotiation can be described by a renegotiation function, which depends on the outcome $\langle x, t_1, t_2\rangle$ prescribed by the mechanism and on the true state of the world θ.[3,4] Although the agents have utility functions over the possible *post-renegotiation outcomes*, for the purposes of this section it will be convenient to focus instead on their induced utilities over the *pre-renegotiation outcome prescribed by the mechanism* (taking the renegotiation process into account). Specifically, we assume in this section that each agent *i*'s post-renegotiation payoff when the mechanism prescribes (pre-renegotiation) outcome $\langle x, t_1, t_2\rangle$ in state θ takes the quasilinear form $u_i(x, \theta) + t_i$. We assume that the function $u_i(x, \theta)$ is continuous in x and differentiable in θ, and that the partial derivative $\partial u_i(x, \theta)/\partial\theta$ is continuous in (x, θ). Finally, we assume that in each state θ, the two agents' post-renegotiation payoffs add up to some number $Z(\theta)$ that does not depend on the outcome prescribed by the mechanism:

$$(1) \qquad u_1(x, \theta) + u_2(x, \theta) = Z(\theta) \quad \text{for all} \quad x \in X, \quad \theta \in \Theta.$$

We will derive these properties in subsequent sections from more primitive assumptions about the agents' underlying preferences over post-renegotiation outcomes and the renegotiation process. In particular, condition (1) will follow from efficiency of renegotiation. For now, consider the following example:

EXAMPLE 1: Agents 1 and 2 are a buyer and a seller respectively, who can trade up to one unit of a good. The agents' underlying preferences are quasilinear in monetary transfers. The seller's cost is zero, while the buyer's marginal valuation for the good is $\theta \in [\underline{\theta}, \bar{\theta}]$. The mechanism prescribes a (message-contingent) outcome $\langle x, t_1, t_2\rangle$, where $x \in X \equiv [0, 1]$ is the prescribed trade between the agents, and the prescribed transfers satisfy $t_1 + t_2 = 0$. The agents then renegotiate to an efficient trade, splitting the renegotiation surplus equally. Suppose that $\underline{\theta} > 0$, so that the efficient trade is always $x = 1$, and the maximum achievable surplus is θ. Then the renegotiation surplus in state θ after the mechanism prescribes outcome $\langle x, t_1, t_2\rangle$ is $\theta(1 - x)$. The agents' post-renegotiation payoffs can

[2] The next sections consider settings with quasilinear preferences and a linear renegotiation function, in which this adding-up restriction on transfers is without loss of generality since any monetary waste would be renegotiated away in the same way in all states of the world. (See footnote 11.)

[3] Note that we need not constrain renegotiation to the choice of a decision from X. Thus, we can apply our framework to "incomplete contracting" models in which the agents negotiate ex post over some decisions that cannot be contracted upon ex ante. See Section 3 for examples.

[4] The decision set X could contain lotteries over a more primitive set of outcomes (such lotteries will be explicitly considered in some applications in Section 3). In this case, our framework assumes implicitly that renegotiation takes place after the play of the mechanism, but before the realization of the lottery (see Maskin and Moore (1999, Section 3) for a related discussion).

then be written as

$$\text{Buyer: } \theta x + t_1 + \frac{1}{2}\theta(1-x) = \frac{1}{2}\theta x + \frac{1}{2}\theta + t_1,$$

$$\text{Seller: } t_2 + \frac{1}{2}\theta(1-x) = -\frac{1}{2}\theta x + \frac{1}{2}\theta + t_2.$$

Thus, we have $u_1(x,\theta) = \theta x/2 + \theta/2$, $u_2(x,\theta) = -\theta x/2 + \theta/2$, and $Z(\theta) = \theta$.

If $(m_1(\theta), m_2(\theta))$ is a Nash equilibrium message pair of the mechanism in state θ, then the corresponding equilibrium outcome prescribed by the mechanism is given by

$$\begin{aligned}&\langle \hat{x}(\theta), \hat{t}_1(\theta), \hat{t}_2(\theta)\rangle \\ &\quad = \big\langle x(m_1(\theta), m_2(\theta)), t_1(m_1(\theta), m_2(\theta)), t_2(m_1(\theta), m_2(\theta))\big\rangle.\end{aligned}$$

In contrast to the standard implementation setting, this equilibrium outcome is always renegotiated to a surplus-maximizing outcome. Hence, our focus will be on each agent i's equilibrium post-renegotiation utility in the mechanism in state θ, given by $U_i(\theta) = u_i(\hat{x}(\theta), \theta) + \hat{t}_i(\theta)$.[5] The central objective in this setting is to identify the set of utility mappings $U_1, U_2 \colon \Theta \to \mathbb{R}$ that can be implemented through an appropriately designed mechanism. Condition (1) implies that any such mapping must satisfy

$$(2) \qquad U_1(\theta) + U_2(\theta) = Z(\theta) \quad \text{for all} \quad \theta \in \Theta.$$

In the remainder of this section, we characterize the other properties that implementable utility mappings must satisfy.

Appealing to the Revelation Principle, in characterizing implementability we can restrict attention to *direct revelation mechanisms*, in which the agents announce the state of the world, and truth telling constitutes a Nash equilibrium. Maskin and Moore (1999, Theorems 1, 2) describe the incentive constraints characterizing such direct revelation mechanisms. These constraints require that whenever the agents disagree on the state, i.e., announce $(m_1, m_2) = (\theta', \theta'') \in \Theta^2$ with $\theta' \neq \theta''$, and one agent's announcement is truthful, the other agent's lie does not make him better off. This is illustrated in Figure 1, where the outcome $\langle x(\theta', \theta''), t_1(\theta', \theta''), t_2(\theta', \theta'')\rangle$ prescribed by the mechanism for this disagreement (the northeast off-diagonal element in the table) must not give agent 1 a utility higher than $U_1(\theta'') = u_1(\hat{x}(\theta''), \theta'') + \hat{t}_1(\theta'')$ when the true state is θ'', and must not give agent 2 a utility higher than $U_2(\theta') = u_2(\hat{x}(\theta'), \theta') + \hat{t}_2(\theta')$ when the true state is θ'. Therefore, each possible disagreement (ordered pair of different states) gives rise to two incentive constraints, one for each agent.

[5] As noted by Maskin and Moore (1999), since the mechanism defines a constant-sum game between the agents, the Minimax Theorem implies that the equilibrium utilities $U_1(\theta), U_2(\theta)$ in state θ do not depend on which Nash equilibrium of this game is selected.

		Agent 2	
		θ'	θ''
Agent 1	θ'	$\widehat{x}(\theta'), \widehat{t}_1(\theta'), \widehat{t}_2(\theta')$	$x(\theta', \theta''), t_1(\theta', \theta''), t_2(\theta', \theta'')$
	θ''	$x(\theta'', \theta'), t_1(\theta'', \theta'), t_2(\theta'', \theta')$	$\widehat{x}(\theta''), \widehat{t}_1(\theta''), \widehat{t}_2(\theta'')$

FIGURE 1.—The direct revelation mechanism.

When the set of states Θ is finite, the Maskin-Moore characterization involves $2|\Theta| \cdot (|\Theta| - 1)$ constraints. When Θ is a continuum, however, it yields a double continuum of incentive constraints, which is difficult to analyze. The main result of this section shows that when Θ is an interval, the decision space X is connected, and the parties' post-renegotiation utilities are continuously differentiable and quasilinear, the Maskin-Moore incentive constraints can be reduced to local (first-order) incentive constraints, which provide a convenient characterization of implementable utility mappings.

Our analysis builds on the approach pioneered by Mirrlees (1971) for the standard mechanism design problem, in which only one agent observes the state of the world. To illustrate this connection, we first restrict attention to mechanisms in which only one agent, say agent i, is called upon to make an announcement (formally, the other agent has only one possible message: $|M_{-i}| = 1$). We shall call such mechanisms *agent i option mechanisms*. An argument identical to the standard mechanism design argument (for a setting without renegotiation) establishes that if the utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is implemented with an agent i option mechanism, it must satisfy

$$U_i(\theta) = U_i(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{\partial u_i(\hat{x}(\tau), \tau)}{\partial \theta}\, d\tau,$$

where $\hat{x}(\cdot)$ is the equilibrium decision rule prescribed by the mechanism. We shall refer to this condition as the *Mirrlees condition.* For differentiable mechanisms, it follows from the traditional Envelope Theorem, which establishes that $U_i'(\theta) = \partial u_i(\hat{x}(\theta), \theta)/\partial\theta$ for any $\theta \in (\underline{\theta}, \bar{\theta})$. A more general derivation, which does not restrict the class of allowed mechanisms, can be found in Milgrom and Segal (2002, Corollary 1). Note that by (1) and (2), the Mirrlees condition holds for agent i if and only if it holds for agent $-i$.

Since a generalized version of the Mirrlees condition will be repeatedly used in the paper, we introduce the following definition:

DEFINITION 1: A utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is *generated* by decision rule $x(\cdot)$ if for $i = 1, 2$,

$$U_i(\theta) = U_i(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{\partial u_i(x(\tau), \tau)}{\partial \theta}\, d\tau. \tag{3}$$

Note that any utility mapping generated by a decision rule is absolutely continuous (Kolmogorov and Fomin (1970)), and that all utility mappings generated by a given decision rule coincide up to a constant. Conversely, when $\partial u_i(x,\theta)/\partial\theta$ is a one-to-one function of x (as it will be under the single-crossing property defined below), any two decision rules generating the same utility mapping coincide almost everywhere.

Although any utility mapping that is implementable with an option mechanism is generated by some decision rule, the converse is not true. Nonetheless, a well-known sufficient condition for implementability with option mechanisms exists when agent i's post-renegotiation utility satisfies the following *single crossing property*:

SCP$_i$: *$X \subset \mathbb{R}$ and $\partial u_i(x,\theta)/\partial\theta$ is strictly increasing in $x \in X$ for all $\theta \in \Theta$.*

Under SCP$_i$, a standard mechanism design argument implies that the decision rule $\hat{x}(\cdot)$ can arise in an equilibrium of an agent i option mechanism if and only if $\hat{x}(\cdot)$ is nondecreasing. Also, observe that due to the adding-up condition (1), SCP$_i$ is equivalent to $u_{-i}(x,\theta)$ satisfying the single crossing property in $(-x,\theta)$. Therefore, the decision rule $\hat{x}(\cdot)$ can arise in an equilibrium of an agent $-i$ option mechanism if and only if $\hat{x}(\cdot)$ is nonincreasing. These arguments can be summarized as follows:

PROPOSITION 1: *If a utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is implementable with an agent i [agent $-i$] option mechanism having equilibrium decision rule $\hat{x}(\cdot)$, then it is generated by $\hat{x}(\cdot)$. Furthermore, under SCP$_i$, $\hat{x}(\cdot)$ must be nondecreasing [nonincreasing]. Conversely, under SCP$_i$, any utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ generated by a nondecreasing [nonincreasing] decision rule $x(\cdot)$ is implementable with an agent i [agent $-i$] option mechanism.*

The main result of this section extends this characterization to mechanisms in which both parties can send messages. We find that any utility mapping implementable with such a mechanism must still be generated by *some* decision rule, provided that the decision space X is connected. The converse is also true provided that the agents' payoffs satisfy the following property:

CONDITION $\pm$: *There exists a pair $(x^+, x^-) \in X^2$ such that for all $\theta \in \Theta$ and all $x \in X$,*

$$\frac{\partial u_1(x^-,\theta)}{\partial\theta} \le \frac{\partial u_1(x,\theta)}{\partial\theta} \le \frac{\partial u_1(x^+,\theta)}{\partial\theta}.$$

Note that by (1), Condition $\pm$ can be equivalently formulated for $u_2(\cdot)$. Note also that whenever SCP$_i$ holds, Condition $\pm$ is satisfied by choosing $(x^-, x^+) = (\min X, \max X)$ when $i = 1$, or $(x^-, x^+) = (\max X, \min X)$ when $i = 2$. Our main result can now be stated:

PROPOSITION 2: *Suppose the decision space X is connected. Then any implementable utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is generated by a decision rule $x(\cdot)$. Conversely, under Condition $\pm$, any utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ generated by a decision rule $x(\cdot)$ is implementable.*

PROOF: The proof builds on the following Lemma, which provides a more convenient characterization of implementability for connected decision spaces than that of Maskin and Moore (1999).[6]

LEMMA 1: *Suppose the decision space X is connected. Then a utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is implementable if and only if for any $\theta'', \theta' \in \Theta$ there exists $\tilde{x}(\theta'', \theta') \in X$ such that*

$$U_1(\theta'') - U_1(\theta') = u_1(\tilde{x}(\theta'', \theta'), \theta'') - u_1(\tilde{x}(\theta'', \theta'), \theta'). \tag{4}$$

PROOF OF LEMMA 1: The characterization result of Maskin and Moore (1999) implies that $\langle U_1(\cdot), U_2(\cdot)\rangle$ is implementable if and only if for any ordered pair of states (θ', θ'') there is an outcome $\langle x(\theta', \theta''), t_1(\theta', \theta''), t_2(\theta', \theta'')\rangle$ satisfying the following incentive constraints:

$$U_1(\theta'') \geq u_1(x(\theta', \theta''), \theta'') + t_1(\theta', \theta''), \tag{5}$$

$$U_2(\theta') \geq u_2(x(\theta', \theta''), \theta') + t_2(\theta', \theta''). \tag{6}$$

Using (1), (2), and the adding-up restriction on transfers, (6) can be rewritten as

$$U_1(\theta') \leq u_1(x(\theta', \theta''), \theta') + t_1(\theta', \theta''). \tag{7}$$

There is a $t_1(\theta', \theta'')$ for which (5) and (7) hold simultaneously if and only if

$$U_1(\theta'') - U_1(\theta') \geq u_1(x(\theta', \theta''), \theta'') - u_1(x(\theta', \theta''), \theta').$$

By reversing the roles of θ'' and θ' we see that a necessary and sufficient condition for implementing the mappings $\langle U_1(\cdot), U_2(\cdot)\rangle$ is that for every θ'' and θ' there is a pair $(x(\theta'', \theta'), x(\theta', \theta''))$ such that

$$\begin{aligned} u_1(x(\theta'', \theta'), \theta'') - u_1(x(\theta'', \theta'), \theta') &\geq U_1(\theta'') - U_1(\theta') \\ &\geq u_1(x(\theta', \theta''), \theta'') - u_1(x(\theta', \theta''), \theta'). \end{aligned}$$

Since X is connected and $u_1(\cdot)$ is continuous, the Intermediate Value Theorem implies that this condition can be satisfied if and only if the Lemma's statement holds. *Q.E.D.*

Now we establish both parts of the Proposition:

(i) *Necessity*. Suppose that $\langle U_1(\cdot), U_2(\cdot)\rangle$ are implementable. By Lemma 1, for any θ'' and θ' we have

$$|U_1(\theta'') - U_1(\theta')| \leq \left\{ \max_{(x,\theta)\in X\times\Theta} \left| \frac{\partial u_1(x, \theta)}{\partial \theta} \right| \right\} \cdot |\theta'' - \theta'|,$$

[6] Note that the Lemma holds for an arbitrary state space Θ. Thus, it provides an alternative characterization of Maskin-Moore incentive constraints in settings with connected outcome spaces and continuous preferences.

where the expression in curly brackets is well-defined given the fact that $u_1(\cdot)$ is continuously differentiable and X and Θ are compact. It follows that $U_1(\cdot)$ is absolutely continuous and from this we know that $U_1'(\cdot)$ exists almost everywhere and $U_1(\theta) = U_1(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} U_1'(\tau)\, d\tau$ (see Kolmogorov and Fomin (1970)).

We next argue that at any $\theta \in [\underline{\theta}, \bar{\theta})$ at which $U_1'(\theta)$ exists, there is a decision $x(\theta) \in X$ such that $U_1'(\theta) = \partial u_1(x(\theta), \theta)/\partial\theta$. To see this, consider a sequence $\{\theta^n\}_{n=1}^{\infty}$ such that $\theta^n > \theta$ for all n and $\theta^n \to \theta$ as $n \to \infty$. Then Lemma 1 implies that for each θ^n there is an $\tilde{x}(\theta^n, \theta)$ such that

$$\frac{U_1(\theta^n) - U_1(\theta)}{\theta^n - \theta} = \frac{u_1(\tilde{x}(\theta^n, \theta), \theta^n) - u_1(\tilde{x}(\theta^n, \theta), \theta)}{\theta^n - \theta}. \tag{8}$$

Moreover, using the Mean Value Theorem, (8) implies that

$$\frac{U_1(\theta^n) - U_1(\theta)}{\theta^n - \theta} = \frac{\partial u_1(\tilde{x}(\theta^n, \theta), \bar{\theta}^n)}{\partial\theta} \tag{9}$$

for some $\bar{\theta}^n \in [\theta, \theta^n]$. Given the compactness of X, there is a subsequence and a decision $x(\theta) \in X$ such that $\tilde{x}(\theta^n, \theta) \to x(\theta)$. Taking the limit of (9) along this subsequence we have

$$U_1'(\theta) = \frac{\partial u_1(x(\theta), \theta)}{\partial\theta}$$

wherever this derivative is well-defined. Hence, (3) must hold for $i = 1$. By (1) and (2), it must also hold for $i = 2$.

(ii) *Sufficiency*. Suppose that Condition $\pm$ holds and that the utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is generated by a decision rule $x(\cdot)$. Take any $\theta', \theta'' \in \Theta$, and suppose for definiteness that $\theta' \le \theta''$. By Condition $\pm$, we know that

$$\int_{\theta'}^{\theta''} \frac{\partial u_1(x^-, \tau)}{\partial\theta}\, d\tau \le \int_{\theta'}^{\theta''} \frac{\partial u_1(x(\tau), \tau)}{\partial\theta}\, d\tau \le \int_{\theta'}^{\theta''} \frac{\partial u_1(x^+, \tau)}{\partial\theta}\, d\tau.$$

This double inequality can be rewritten as

$$u_1(x^-, \theta'') - u_1(x^-, \theta') \le U_1(\theta'') - U_1(\theta') \le u_1(x^+, \theta'') - u_1(x^+, \theta').$$

Since $u_1(\cdot)$ is continuous and X is connected, by the Intermediate Value Theorem there exists $\tilde{x}(\theta'', \theta') \in X$ such that

$$U_1(\theta'') - U_1(\theta') = u_1(\tilde{x}(\theta'', \theta'), \theta'') - u_1(\tilde{x}(\theta'', \theta'), \theta').$$

Lemma 1 implies that $\langle U_1(\cdot), U_2(\cdot)\rangle$ is implementable. *Q.E.D.*

Proposition 2 offers a very convenient characterization of implementable utility mappings. In Sections 3 and 4, we will use this characterization to identify mechanisms that are optimal from the viewpoint of ex ante investment incentives or risk sharing. For example, we will identify settings in which the agents can restrict attention without loss to option mechanisms. A comparison of Propositions 1 and 2 shows that in environments satisfying a single-crossing property, the benefit of general two-agent announcement games relative to one-agent announcement games (option mechanisms) is *precisely* the ability to implement utility mappings generated by nonmonotonic decision rules. For example, for the payoffs described in Example 1 (which satisfy SCP_1), two-sided mechanisms can implement any absolutely continuous utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ that has $1/2 \leq U_1'(\theta) \leq 1$ almost everywhere, while agent 1 [agent 2] option mechanisms can implement only convex [concave] functions $U_1(\cdot)$. In some circumstances, the agents will be able to restrict attention to even simpler *noncontingent mechanisms*, in which no announcements are used (formally, $|M_1| = |M_2| = 1$), and a fixed outcome $(\hat{x}, \hat{t}_1, \hat{t}_2) \in X \times \mathbb{R}^2$ is prescribed. A utility mapping is implementable with a noncontingent mechanism if and only if it is generated by a constant decision rule. Thus, in Example 1, noncontingent mechanisms can be used to implement linear utility mappings in which the slope of $U_1(\cdot)$ is between 1/2 and 1.

Another difference between Propositions 1 and 2 lies in the fact that the former establishes that the equilibrium decision rule of any mechanism implementing a given utility mapping must generate the mapping, while the latter does not. To see why this is the case, note that starting with any incentive-compatible direct revelation mechanism implementing utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$, we can modify the equilibrium (on-diagonal) pre-renegotiation outcome of the mechanism (see Figure 1) to be any triple $\langle \hat{x}(\cdot), \hat{t}_1(\cdot), \hat{t}_2(\cdot)\rangle$ satisfying $U_i(\theta) = u_i(\hat{x}(\theta), \theta) + \hat{t}_i(\theta)$ for all θ. As long as the out-of-equilibrium (off-diagonal) prescriptions of the mechanism remain the same, the agents' incentive constraints will be preserved. Thus, the *equilibrium decision rule* in two-sided mechanisms implementing a given utility mapping is indeterminate. As for the *generating decision rule*, the proof of Proposition 2 constructs it as a limit of *out-of-equilibrium* decisions satisfying all local Maskin-Moore incentive constraints, i.e., constraints resolving disagreements in which one agent just slightly overstates or understates the true state.[7]

The indeterminacy of the equilibrium decision rule in two-sided message games can evoke different responses. One response is that we should be interested only in which utility mappings can be implemented, and not in how they are implemented, since there is a multitude of mechanisms that can do this (such an argument has been put forward by Maskin and Tirole (1999, Section 2.4)). In this light, Proposition 2 simply serves to bound the variation of utility across states of the world in implementable utility mappings. A second response is to restrict attention to a class of mechanisms with the property that the generating and equilibrium decision rules are linked. One motivation for doing so is that

[7] A related difference between the results is Proposition 2's assumption of connectedness of the decision set X. To see the need for this assumption, observe that, as the proof of Proposition 2 makes clear, under Condition $\pm$ *any* implementable utility mapping can be implemented with a mechanism that uses only decisions x^+ and x^- (both in and out of equilibrium). One thus needs to invoke the Intermediate Value Theorem to obtain a generating decision rule. In contrast, in the standard mechanism design approach used in Proposition 1, the equilibrium decision rule itself must generate the equilibrium utility mapping.

this will allow us to use the Mirrlees condition to generate potentially testable predictions for the observable equilibrium decisions in optimal mechanisms. The Envelope Theorem argument in the Introduction establishes this property for differentiable mechanisms. This result can be extended to a class of nonsmooth mechanisms by using a generalized Envelope Theorem formulated by Milgrom and Segal (2002) for *continuous* saddle-point problems. This generalized theorem can be applied to the following class of mechanisms:

DEFINITION 2: A mechanism with message sets (M_1, M_2) and outcome function $\langle x(\cdot), t_1(\cdot), t_2(\cdot)\rangle$ is *continuous* if the message sets are second-countable topological spaces,[8] and the outcome function is continuous in each of m_1 and m_2.

One argument in favor of continuous mechanisms is that they allow agents to approximate the outcome by transmitting only a limited amount of information. Indeed, in reality the agents' messages are finite strings of letters from a finite alphabet. The set of all such strings is countable. Second-countable topological spaces are those whose elements can be approximated arbitrarily closely with such finite strings. Continuity of the outcome function ensures that such approximate messages approximate the mechanism's outcome as well.

Note that a mechanism may be continuous even when the corresponding direct revelation mechanism is not (in the standard topology on the state space Θ). For example, any mechanism in which the agents' message spaces are finite or countable is continuous under the discrete topology on message spaces, even though the corresponding direct revelation mechanism is in general discontinuous. As another example, if the decision space X is second-countable, any agent i option mechanism can be thought of as a continuous mechanism, in which agent i's message space is a subset of the set $X \times \mathbb{R}^2$ of all possible outcomes, and the outcome function is the identity mapping. Thus, the class of continuous mechanisms is very wide.

Using this notion of a continuous mechanism, we have the following result:[9]

PROPOSITION 3: *If a utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is implemented by a continuous mechanism with equilibrium decision rule $\hat{x}(\cdot)$, then it is generated by $\hat{x}(\cdot)$. Conversely, if $X = [\underline{x}, \bar{x}] \subset \mathbb{R}$ and either SCP_1 or SCP_2 holds, any utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ generated by a decision rule $x(\cdot)$ that is continuous except on an (at most) countable set of points is implementable with a continuous mechanism.*

PROOF: See Appendix.

[8] For the definition of a second-countable topological space, see, e.g., Kolmogorov and Fomin (1970). Every separable metric space, and in particular every Euclidean space, is second-countable.

[9] In a previous version of this paper we showed that if one instead restricts attention to mechanisms with finite message spaces, then it is possible to implement any utility mapping generated by a step function. Moreover, this implies that any utility mapping can be approximately implemented with a finite mechanism.

To see the intuition for the first part of Proposition 3, recall that Proposition 2 constructs a generating decision rule as a limit of out-of-equilibrium decisions satisfying local Maskin-Moore incentive constraints. In a continuous mechanism, the equilibrium and out-of-equilibrium decisions are linked, which ensures that the equilibrium decision rule will also be a generating one.

The second part of Proposition 3 establishes that a wide set of implementable utility mappings can be implemented with continuous mechanisms. Given the robustness of continuous mechanisms to agents' small mistakes in their messages, we believe that the agents may adopt a continuous mechanism whenever its use involves no loss. In Sections 3 and 4, Proposition 3 will allow us to predict the equilibrium decision rules in mechanisms implementing optimal utility mappings in hold-up and risk-sharing contracting problems. Since these equilibrium decision rules will in general differ from the ex post efficient decisions, the restriction to continuous mechanisms will yield predictions regarding the direction of equilibrium renegotiation. (In particular, the Renegotiation-Proofness Principle will not hold with this restriction.)

We conclude this section with a discussion of the role of two simplifying assumptions: quasilinearity of payoffs and one-dimensionality of the state θ. While dramatically simplifying our analysis, quasilinearity is not crucial for at least some of our results. In its absence, the constant-sum condition (1) must be replaced with the condition that the agents' post-renegotiation payoffs are always on the (perhaps nonlinear) utility possibility frontier. Nevertheless, it is not hard to see that the necessity parts of Propositions 2 and 3 carry through in this case without modification.

The analysis for a multidimensional state $\theta = (\theta_1, \ldots, \theta_K)$ with $K > 1$ is more involved. While we cannot extend our full characterization of implementable utility mappings to this case, we can provide a partial characterization by using the constraints that the agents do not wish to misrepresent any given dimension θ_k for a fixed value of θ_{-k}. Specifically, using the necessity parts of Propositions 2 and 3, we can provide a *necessary* condition for how the agents' utilities can change with changes in θ_k:[10]

COROLLARY 1: *Suppose $\Theta = \prod_{k=1}^{K} [\underline{\theta}_k, \bar{\theta}_k] \subset \mathbb{R}^K$, the decision space X is connected, $u_i(\cdot, \cdot)$ is continuous, and $\partial u_i(x, \theta)/\partial \theta_k$ exists and is continuous. Then for any implementable utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ there exists a decision rule $x^k(\cdot)$ such that for all $\theta = (\theta_k, \theta_{-k}) \in \Theta$,*

$$U_i(\theta_k, \theta_{-k}) = U_i(\underline{\theta}_k, \theta_{-k}) + \int_{\underline{\theta}_k}^{\theta_k} \frac{\partial u_i(x^k(\tau_k, \theta_{-k}), \tau_k, \theta_{-k})}{\partial \theta_k} \, d\tau_k.$$

[10] Full characterization of implementable utility mappings with multidimensional type spaces is difficult even in standard mechanism design. Here the problem becomes even more difficult if discontinuous mechanisms are allowed, since we cannot ensure the existence of a single decision rule that generates a given utility mapping in all dimensions at once.

Moreover, if the utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ *is implemented by a continuous mechanism with equilibrium decision rule* $\hat{x}(\cdot)$, *then for all* $\theta = (\theta_k, \theta_{-k}) \in \Theta$,

$$U_i(\theta_k, \theta_{-k}) = U_i(\underline{\theta}_k, \theta_{-k}) + \int_{\underline{\theta}_k}^{\theta_k} \frac{\partial u_i(\hat{x}(\tau_k, \theta_{-k}), \tau_k, \theta_{-k})}{\partial \theta_k}\, d\tau_k.$$

We will use Corollary 1 in the analysis of hold-up in Section 3, where we will care only about the dependence of ex post utilities on ex ante investments, which we will assume can be aggregated into one dimension (θ_k), and not about their dependence on exogenous uncertainty (θ_{-k}).

3. APPLICATION TO HOLD-UP

In this section we apply the implementation results of the previous section to the study of contracting in hold-up models. The model we study consists of four stages, which are depicted in Figure 2.

In the first stage, the parties can write a contract governing, to some degree, their future trading relations. Formally, this contract specifies a mechanism that the parties will play in stage 3.

In the second stage, the two parties simultaneously choose their investments $a_1 \in A_1$ and $a_2 \in A_2$ respectively. At the same time, the random state of nature $\varepsilon \in \mathfrak{E}$ is realized. We assume that A_1, A_2 are compact connected sets in Euclidean spaces, and $\mathfrak{E}$ is a probability space. Both investments $a = (a_1, a_2)$ and the random state ε are assumed to be observed by the two parties, but not verifiable. The cost of investment a_i in state ε is given by the function $\psi_i(a_i, \varepsilon)$.

In the third stage, the parties play the contractually specified mechanism. The mechanism prescribes an outcome $\langle x, t_1, t_2\rangle$, where $x \in X$ is a nonmonetary decision (see examples below for specific interpretations) and $t_i \in \mathbb{R}$ is a monetary transfer to agent i. As in the previous section, we assume that X is a connected compact space, and that $t_1 + t_2 = 0$.

Finally, in the fourth stage the parties engage in bargaining, with the disagreement point given by the outcome $\langle x, t_1, t_2\rangle$ prescribed in stage 3. Party i's utility from this disagreement outcome is

$$v_i(x, a, \varepsilon) + t_i - \psi_i(a_i, \varepsilon).$$

We assume that bargaining results in an efficient outcome, generating a total surplus (excluding investment costs) of $S(a, \varepsilon)$ in state ε following investments a.

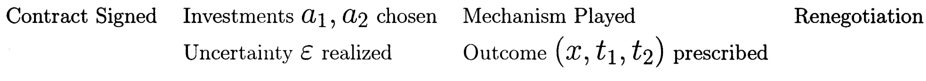

FIGURE 2.—Timing.

Moreover, we assume that each party i receives a fixed share λ_i of the renegotiation surplus $[S(a,\varepsilon)-v_i(x,a,\varepsilon)-v_{-i}(x,a,\varepsilon)]$, with $\lambda_1+\lambda_2=1$. Then party i's post-renegotiation utility as a function of the disagreement outcome $\langle x,t_1,t_2\rangle$ can be written as $w_i(x,a,\varepsilon)+t_i$, where[11]

$$\begin{aligned}(10)\qquad w_i(x,a,\varepsilon)&=v_i(x,a,\varepsilon)-\psi_i(a_i,\varepsilon)+\lambda_i[S(a,\varepsilon)-v_i(x,a,\varepsilon)-v_{-i}(x,a,\varepsilon)]\\&=[\lambda_{-i}v_i(x,a,\varepsilon)-\lambda_i v_{-i}(x,a,\varepsilon)]+\lambda_i S(a,\varepsilon)-\psi_i(a_i,\varepsilon).\end{aligned}$$

In general, the form of $S(a,\varepsilon)$ depends on the specifics of the application. In some applications, ex ante contracts are "complete," in the sense that the parties' ex post renegotiation concerns only the level of the same decision x that can be prescribed in the ex ante contract (as well as transfers). In such cases, the ex post surplus is given by

$$(11)\qquad S(a,\varepsilon)=\max_{x\in X}\sum_{i=1,2}v_i(x,a,\varepsilon).$$

In other cases, ex ante contracts are "incomplete" in the sense that there are variables that are contractible ex post but are not contractible ex ante (see Example 5 below). Some of our results will rely on the "complete contracting" condition (11).

Given that the parties always renegotiate to realize the aggregate ex post surplus $S(a,\varepsilon)$, the goal of an ex ante contract is to sustain the ex ante investments that achieve the highest possible expected surplus net of investment costs, $E_\varepsilon[S(a,\varepsilon)-\sum_{i=1,2}\psi_i(a_i,\varepsilon)]$.

Our model includes as special cases a number of different versions of the hold-up model that appear in the literature:

EXAMPLE 2—*The hold-up models of Edlin and Reichelstein (1996) and Che and Hausch (1999)*: In these models, $X\subseteq\mathbb{R}_+$ is the set of possible quantities sold by party 1 (the seller) to party 2 (the buyer), the quantity traded is the only ex post decision variable and it can be specified in the ex ante contract (hence (11) holds), $v_2(x,a,\varepsilon)$ represents the buyer's valuation for x units of the good, and $-v_1(x,a,\varepsilon)$ represents the seller's cost of producing x units. The two papers also assume that the parties' investments are one-dimensional, i.e., $a_1,a_2\in\mathbb{R}$. In addition, Edlin and Reichelstein consider only "self-investments," for which each party i's utility is directly affected only by its own investment a_i, while Che and Hausch focus on cases of "cooperative" investments, for which the investment a_i of each party i directly affects only party $j\neq i$'s utility.

EXAMPLE 3—*The hold-up model of Segal (1999)*: This model has the same form as the previous example, except that n possible goods can be traded, and so $X\subset\mathbb{R}^n_+$.

[11] Although up to this point we have simply assumed that the transfers sum to zero, this is without loss of generality given our renegotiation process. In particular, any prescribed outcome $\langle x,\tilde t_1,\tilde t_2\rangle$ such that $\tilde t_1+\tilde t_2<0$ leads to the same post-renegotiation utilities for the two parties as does the outcome $\langle x,t_1,t_2\rangle$ with $t_i=\tilde t_i-\lambda_i(\tilde t_1+\tilde t_2)$ for $i=1,2$, for which $t_1+t_2=0$.

EXAMPLE 4—*The asset ownership models of Demski and Sappington (1991) and Edlin and Hermalin (2000)*: There is one asset, and two risk-neutral parties.[12] The asset is initially owned by party 1 and managed by party 2. In stage 2 the two parties can make investments that increase the value of the asset to whomever owns it ex post (in Demski and Sappington (1991) only the owner makes ex ante investments at stage 2). In these models, $x \in [0, 1]$ denotes the manager's probability of owning the asset ex post, and it is the only ex post contractible decision variable (hence (5) holds). Letting $\hat{v}_i(a, \varepsilon)$ denote the asset's value when owned by party i (which may depend on i, for example, because of ex post moral hazard), the parties' disagreement utilities can be written as $v_i(x, a, \varepsilon) + t_i$, where[13]

$$v_1(x, a, \varepsilon) = (1 - x)\hat{v}_1(a, \varepsilon),$$
$$v_2(x, a, \varepsilon) = x\hat{v}_2(a, \varepsilon).$$

EXAMPLE 5—*The incomplete contracting setting of Grossman and Hart (1986), Hart and Moore (1990), and Hart (1995)*: In the simplest version of this model (see Chapter 2 of Hart (1995)), there is a single asset that can be owned ex post by one of two risk-neutral parties. The initial contract can specify the probability $x \in [0, 1]$ that party 1, rather than party 2, owns the asset. Ex post, however, there is a decision regarding the utilization of the asset. Absent an ex post agreement between the parties, the owner makes the utilization decision to maximize his ex post payoff; the resulting return to party i when party j owns the asset is $\hat{v}_{ij}(a, \varepsilon)$. (Hart and Moore's (1990) Assumption 3 restricts attention to the case of self-investments, in which $\hat{v}_{ij}(\cdot)$ depends only on a_i.) Party i's payoff from the disagreement outcome $\langle x, t_1, t_2 \rangle$ is therefore $v_i(x, a, \varepsilon) = x\hat{v}_{i1}(a, \varepsilon) + (1 - x)\hat{v}_{i2}(a, \varepsilon)$. Unlike in the previous examples, the fact that asset utilization can be specified contractually ex post, but not ex ante, means that $S(a, \varepsilon) > \max_{x \in X} \sum_{i=1,2} v_i(x, a, \varepsilon)$ whenever the efficient ex post outcome given (a, ε) involves joint utilization of the asset.

EXAMPLE 6—*The moral hazard model of Hermalin and Katz (1991)*: Party 1 is an agent and party 2 is a principal. The parties' investments a (efforts) determine the distribution of a verifiable output $q \in \{q_1, \ldots, q_K\}$ where $q_k \in \mathbb{R}$. The probability of q_k is given by $\Pr\{q_k | a\}$. Let X denote the set of output-contingent

[12] Demski and Sappington (1991) and Edlin and Hermalin (2000) allow the parties to be risk-averse.

[13] As noted in Section 2, when the set X contains lotteries over more primitive outcomes, we assume that renegotiation takes place after play of the mechanism, but prior to the realization of the lottery. Note, however, that when the parties are risk neutral (as in this example), we could equally well imagine that renegotiation occurs only after the realization of the lottery: the expected post-renegotiation utilities of the parties are the same for both timings. When the parties are risk-averse, however, our framework requires that renegotiation be possible between the play of the mechanism and any random realization of the mechanism's prescription; intuitively, if this were not the case, random prescriptions could be used to dissipate surplus in a way that could not be avoided through renegotiation.

compensation schemes $(x_1, \ldots, x_K)$ in which $x_k \in [\underline{x}_k, \bar{x}_k]$.[14] Each party i is a (weakly) risk-averse expected utility maximizer with a Bernoulli utility function over ex post income. To fit this application into our framework, we need to assume that each party i has constant absolute risk aversion, so that the parties' ex post certainty equivalents can be written as

$$\text{(12)}\quad \text{the agent:}\quad v_1(x,a)+t_1=-\frac{1}{r_1}\ln\left[\sum_k \Pr\{q_k|a\}\cdot\exp\{-r_1x_k\}\right]+t_1,$$

$$\text{the principal:}\quad v_2(x,a)+t_2=-\frac{1}{r_2}\ln\left[\sum_k \Pr\{q_k|a\}\cdot\exp\{-r_2(q_k-x_k)\}\right]+t_2,$$

where $r_i \geq 0$ is party i's coefficient of absolute risk aversion. After a is chosen, the parties can renegotiate the incentive scheme x. For (10) to describe the parties' post-renegotiation payoffs, we assume that bargaining splits the available certainty equivalent renegotiation surplus in fixed proportions. The optimal ex post compensation scheme maximizes the sum of the parties' certainty equivalents given a, which gives rise to a total (certainty equivalent) surplus $S(a)$ given by (11). This compensation scheme is determined by optimal risk sharing considerations; for example, when the agent (party 1) is risk-averse and the principal (party 2) is risk-neutral, the scheme fully insures the agent, e.g., by setting $x_k = 0$ for all k. In this setting, the parties may optimally specify a different compensation scheme ex ante, or use a more complicated mechanism in which the prescribed compensation scheme depends on the parties' announcements, in order to create adequate investment incentives.

In the remainder of this section, we investigate some of the implications of the implementation results of Section 2 for optimal contracting in this class of hold-up models. Many of the applications in the hold-up literature restrict attention to a simple class of contracts, such as option contracts or noncontingent contracts. (Sometimes this restriction is justified by the fact that, in the specific environments studied, such simple mechanisms achieve the first-best.) Here we shall be interested in what can be achieved when we allow for fully general contracts.[15] Let $W_i(a, \varepsilon)$ define the equilibrium value of agent i's post-renegotiation utility $w_i(x, a, \varepsilon) + t_i$ under the contract, given investments a and the realization of uncertainty ε. Our implementation results will tell us what utility mappings $\langle W_1(a, \varepsilon), W_2(a, \varepsilon)\rangle$ are implementable with general contracts, which will in turn determine the set of investments a that can be sustained.

An important complication arises from the fact that the state (a, ε) is in general multidimensional, while the model of Section 2 was developed for a one-dimensional state variable θ. To apply the results of Section 2 here, we shall need

[14] We impose bounds on compensation payments in order to ensure that X is compact. We also assume that the parties receive no information about the conditional distribution of q given a before the mechanism is played. Thus, there is no uncertainty ε in this example.

[15] For an early attempt at such an analysis, see Green and Laffont (1989).

to make some form of aggregation assumption. In what follows, we shall use two distinct aggregation assumptions. The first (and weaker) aggregation condition is the following:[16]

CONDITION A: *Post-renegotiation utility functions take the form*

$$w_i(x, a, \varepsilon) = u_i(x, \phi(a), \varepsilon) + g_i(a, \varepsilon) \quad \textit{for} \quad i = 1, 2,$$

where $\phi(\cdot)$, $g_i(\cdot,\cdot)$, *and* $u_i(\cdot,\cdot,\cdot)$ *are real-valued functions,* $\phi(\cdot)$ *and* $E_\varepsilon[g_i(\cdot,\cdot)]$ *are differentiable in* a, $u_i(\cdot,\cdot,\cdot)$ *is continuous in x, and* $\partial u_i(\cdot,\cdot,\cdot)/\partial\phi$ *exists and is continuous in* (x, ϕ).

Condition A says that the part of party i's post-renegotiation payoff that depends on the decision x prescribed by the mechanism depends on a only through a one-dimensional aggregate measure $\phi(a)$. The function $u_i(x, \phi(a), \varepsilon)$ contains the part of $[\lambda_{-i}v_i(\cdot) - \lambda_i v_{-i}(\cdot)]$ in (10) that depends upon x, while $g_i(\cdot)$ contains $\lambda_i S(\cdot), \psi_i(\cdot)$, and any parts of $[\lambda_{-i}v_i(\cdot) - \lambda_i v_{-i}(\cdot)]$ that do not depend upon x. Note that if Condition A is satisfied for one party, say party i, it is also satisfied for the other party: since $w_{-i}(x, a, \varepsilon) = S(a, \varepsilon) - w_i(x, a, \varepsilon) = -u_i(x, \phi, \varepsilon) + S(a, \varepsilon) - g_i(a, \varepsilon)$, we can take $u_{-i}(x, \phi, \varepsilon) = -u_i(x, \phi, \varepsilon)$ and $g_{-i}(a, \varepsilon) = S(a, \varepsilon) - g_i(a, \varepsilon)$. This also implies that we can assume without loss of generality that $\sum_{i=1,2} u_i(x, \phi, \varepsilon) = 0$ for all (x, ϕ, ε).

Condition A is satisfied whenever only one party i makes a one-dimensional investment choice $a_i \in \mathbb{R}$. However, it is far less restrictive than this. For example, it will be satisfied when one party i makes a multi-dimensional investment choice, as long as only one dimension affects i's post-renegotiation preferences over x (in particular, investment effects need not aggregate in other parts of i's payoff function, such as $\psi_i(\cdot)$ and $S(\cdot)$). It can also be satisfied in cases in which both parties make investments. For example, it is satisfied whenever $E_\varepsilon[\lambda_i S(a, \varepsilon) - \psi_i(a_i, \varepsilon)]$ is differentiable in a for $i = 1, 2$ and the parties' utilities take the following separable form:

CONDITION S: *The decision space* $X = [\underline{x}, \bar{x}] \subset \mathbb{R}$, *and*

$$v_i(x, a, \varepsilon) = \bar{v}_i(a)x + \hat{v}_i(x, \varepsilon) + \tilde{v}_i(a, \varepsilon),$$

where $\bar{v}_i(\cdot), \hat{v}_i(\cdot), \tilde{v}_i(\cdot)$ *are real-valued functions,* $\bar{v}_i(\cdot)$ *is differentiable,* $\hat{v}_i(\cdot)$ *is continuous in x, and* $E_\varepsilon[\tilde{v}_i(\cdot, \varepsilon)]$ *is differentiable in a.*

[16] For simplicity, we include in Condition A the assumptions on differentiability with respect to investments a. These assumptions are not needed to apply the implementation results of Section 2, but are necessary for our focus later on the agents' first-order conditions. A similar point applies with regard to Condition AA below.

Indeed, with these utilities, Condition A is satisfied by taking

$$\phi(a) = \lambda_2 \bar{v}_1(a) - \lambda_1 \bar{v}_2(a),$$
$$u_i(x, \phi(a), \varepsilon) = \delta_i \phi(a) x + \lambda_{-i} \hat{v}_i(x, \varepsilon) - \lambda_i \hat{v}_{-i}(x, \varepsilon),$$
$$g_i(a, \varepsilon) = \lambda_{-i} \tilde{v}_i(a, \varepsilon) - \lambda_i \tilde{v}_{-i}(a, \varepsilon) + \lambda_i S(a, \varepsilon) - \psi_i(a_i, \varepsilon),$$

where $\delta_1 = 1$ and $\delta_2 = -1$. Condition *S*, in turn, includes as a special case the separability condition (A3) specified by Edlin and Reichelstein (1996, p. 492), which assumes in addition that $v_i(\cdot)$ does not depend on a_{-i}.[17]

The second (and stronger) aggregation condition, which we label Condition AA, assumes in addition that the part of each party's post-renegotiation payoff that is affected by x depends on the investment aggregate $\phi(a)$ and uncertainty ε only through the one-dimensional aggregate measure $\theta(\phi(a), \varepsilon)$. Formally, we have the following:

CONDITION AA: *Post-renegotiation utility functions take the form*

$$w_i(x, a, \varepsilon) = u_i(x, \theta(\phi(a), \varepsilon)) + g_i(a, \varepsilon) \quad \textit{for} \quad i = 1, 2,$$

where $\phi(\cdot), \theta(\cdot,\cdot), u_i(\cdot,\cdot)$*, and* $g_i(\cdot,\cdot)$ *are real-valued functions,* $\phi(\cdot)$ *and* $E_\varepsilon[g_i(\cdot,\cdot)]$ *are differentiable in* a*,* $\theta(\cdot,\cdot)$ *is differentiable in* ϕ*,* $u_i(\cdot,\cdot)$ *is continuous in* x*, and* $\partial u_i(\cdot,\cdot)/\partial\theta$ *exists and is continuous in* (x, θ)*.*

For example, when Condition S holds, Condition AA implies that $\hat{v}_i(x, \varepsilon) = \gamma_i(\varepsilon) x + \sigma_i(x)$, in which case $\theta(\phi, \varepsilon) = \phi + [\lambda_2 \gamma_1(\varepsilon) - \lambda_1 \gamma_2(\varepsilon)]$. As with Condition A, we can without loss of generality take $\sum_{i=1,2} u_i(x, \theta) = 0$ for all (x, θ).

Observe that under Conditions A and AA the component $g_i(a, \varepsilon)$ of party i's post-renegotiation utility is unaffected by the outcome of the mechanism. Therefore, the equilibria in any contractual mechanism can depend only on the pair (ϕ, ε) (and only on the aggregate value $\theta(\phi, \varepsilon)$ under Condition AA), and can affect payoffs only through the portion $u_i(\cdot) + t_i$ of party i's post-renegotiation payoff.

Consider, for a moment, the case in which the stronger Condition AA holds. Because of the above observation, party i's equilibrium utility in a contract can be written in the form $W_i(a, \varepsilon) = U_i(\theta(a, \varepsilon)) + g_i(a, \varepsilon)$, where $U_i(\theta)$ denotes the equilibrium value of $u_i(x, \theta) + t_i$ under the contract. For the purpose of characterizing which utility mappings $\langle W_1(\cdot), W_2(\cdot)\rangle$ are achievable using contracts, it suffices to focus our study on which utility mappings $\langle U_1(\cdot), U_2(\cdot)\rangle$ can be implemented. Since we have $\sum_{i=1,2} u_i(x, \theta) = 0$ for all (x, θ), this problem fits directly into the implementation framework of Section 2, where condition (1) holds taking $Z(\theta) = 0$ for all θ. We can therefore use Proposition 2 to characterize implementable utility mappings.

[17] In fact, Condition A holds when each party i's payoff takes the more general form $v_i(x, a, \varepsilon) = \bar{v}_i(a)\mu(x, \varepsilon) + \hat{v}_i(x, \varepsilon) + \tilde{v}_i(a, \varepsilon)$. We restrict attention to the stronger Condition S to simplify exposition.

Likewise, when only Condition A holds, a contract can affect only how the equilibrium value of $u_i(x, \phi, \varepsilon) + t_i$, which we denote by $U_i(\phi, \varepsilon)$, depends on (ϕ, ε). Although the state (ϕ, ε) is now multidimensional, from the viewpoint of investment incentives we need only be concerned with how $U_i(\phi, \varepsilon)$ can depend on the investment aggregate ϕ for each given ε, which we can characterize using Corollary 1.

Given the mapping $U_i(\phi, \varepsilon)$ induced under a given contract, an investment vector $a^0 = (a_1^0, a_2^0) \in A_1 \times A_2$ is a (pure strategy) investment Nash equilibrium under the contract if and only if[18]

$$a_i^0 \in \arg\max_{a_i \in A_i} E_\varepsilon[U_i(\phi(a_i, a_{-i}^0), \varepsilon) + g_i(a_i, a_{-i}^0, \varepsilon)] \quad \text{for} \quad i = 1, 2. \tag{13}$$

In this case, we shall say that the contract *sustains* investments a^0. For later reference, note that if $E_\varepsilon U_i(\phi(a), \varepsilon)$ is differentiable in ϕ, and if the contract sustains Nash equilibrium investment vector a^0 in the interior of A (denoted by $\operatorname{int} A$), then the following first-order conditions must hold:

$$\nabla_{a_i} E_\varepsilon U_i(\phi(a^0), \varepsilon) + \nabla_{a_i} E_\varepsilon g_i(a^0, \varepsilon) = 0 \quad \text{for} \quad i = 1, 2. \tag{14}$$

3.1. *When a Noncontingent Contract is Optimal*

In this section, we investigate when the agents can restrict their attention, without loss, to noncontingent contracts, i.e., contracts that prescribe a fixed outcome $(\hat{x}, \hat{t}_1, \hat{t}_2)$, independent of any messages. To answer this question, we establish conditions under which it is guaranteed that any sustainable investment vector a^0 can be sustained using an appropriately designed noncontingent contract.[19] Our main results establish conditions under which, for any sustainable investment vector $a^0 \in \operatorname{int} A$, there exists an appropriately designed noncontingent contract in which a^0 satisfies the necessary first-order conditions (14) for being a Nash equilibrium.[20] We address the issue of second-order conditions after establishing this result.

The logic of our results is most easily seen in the case in which there is no uncertainty, only party 1 has an investment choice $a_1 \in A_1 \subset \mathbb{R}$, and the initial contract sustaining a_1^0 induces an equilibrium utility function $U_1(a_1)$ that is differentiable at $a_1 = a_1^0$. Then, provided that a_1^0 is in the interior of A_1, the first-order condition (14) can be written as

$$U_1'(a_1^0) + g_1'(a_1^0) = 0.$$

[18] In principle, we could allow for mixed-strategy (and correlated) equilibria by interpreting the investment sets A_i to include investment decisions that are contingent on the realization of the random variable ε (hence, we would still be able to write disagreement payoffs as functions $v_i(x, a, \varepsilon)$). The difficulty, however, is that Condition A will fail to hold except in very special cases (e.g., where only one agent has an investment choice, and he randomizes over two possible investment levels).

[19] We do not rule out the possibility that the noncontingent contract we construct has Nash equilibrium investment vectors other than a^0.

[20] Note that any noncontingent contract prescribing a decision $\langle \hat{x}, \hat{t}_1, \hat{t}_2 \rangle$ gives rise to a utility mapping that is differentiable in a, since under such a contract we have $U_i(\phi(a), \varepsilon) = u_i(\hat{x}, \phi(a), \varepsilon) + \hat{t}_i$.

Since $\langle U_1(\cdot), U_2(\cdot)\rangle$ is implementable, by Proposition 2 it is generated by a decision rule x: $A_1 \to X$. Hence, letting $\hat{x} = x(a_1^0)$, we know that $U_1'(a_1^0) = \partial u_1(\hat{x}, a_1^0)/\partial a_1$, and so

$$\frac{\partial u_1(\hat{x}, a_1^0)}{\partial a_1} + g_1'(a_1^0) = 0. \tag{15}$$

But (15) is also the first-order condition for party 1's investment choice under a noncontingent contract that specifies $\hat{x}$ as the default trade. Hence, as long as the noncontingent contract satisfies appropriate second-order conditions, it also sustains investment level a_1^0.

Proposition 4 extends this logic to cover cases in which both agents may invest, uncertainty is present, and the equilibrium utility functions $U_i(\phi, \varepsilon)$ may not be differentiable.[21] The result assumes only the weaker aggregation property Condition A, so that uncertainty need not aggregate with investments. In this case, we use Corollary 1 to characterize how party i's payoff $U_i(\phi, \varepsilon)$ can depend on the investment aggregate ϕ for any given ε, and therefore how party i's expected payoff $E_\varepsilon U_i(\phi, \varepsilon)$ can depend on ϕ. For Proposition 4, we assume that the following condition holds:

CONDITION H$^\pm$: *For each ϕ, there exist $x^+(\phi), x^-(\phi) \in X$ such that*

$$\frac{\partial u_1(x^-(\phi), \phi, \varepsilon)}{\partial \phi} \leq \frac{\partial u_1(x, \phi, \varepsilon)}{\partial \phi} \leq \frac{\partial u_1(x^+(\phi), \phi, \varepsilon)}{\partial \phi}$$

for all $x \in X$ and all $\varepsilon \in \mathfrak{E}$.

Note that since $u_1(\cdot) = -u_2(\cdot)$, Condition H$^\pm$ can be equivalently formulated for $u_2(\cdot)$. In words, Condition H$^\pm$ says that the trades $x^+(\phi)$ and $x^-(\phi)$ that make the parties' payoffs maximally and minimally responsive to the investment aggregate ϕ can be chosen independently of the realization of the uncertainty ε. In the absence of uncertainty (i.e., when $u_i(\cdot)$ does not depend on ε), Condition H$^\pm$ holds trivially. When uncertainty is present, however, Condition H$^\pm$ does impose restrictions. For example, consider the case in which Condition AA holds. In this case, we have $\partial u_1(x, \phi, \varepsilon)/\partial\phi = \partial u_1(x, \theta(\phi, \varepsilon))/\partial\theta \cdot \partial\theta(\phi, \varepsilon)/\partial\phi$, and so Condition H$^\pm$ requires that for each ϕ there exist $x^+(\phi)$ and $x^-(\phi)$ such that

$$\begin{aligned}\frac{\partial u_1(x^-(\phi), \theta(\phi, \varepsilon))}{\partial \theta}\frac{\partial\theta(\phi, \varepsilon)}{\partial\phi} &\leq \frac{\partial u_1(x, \theta(\phi, \varepsilon))}{\partial \theta}\frac{\partial\theta(\phi, \varepsilon)}{\partial\phi}\\ &\leq \frac{\partial u_1(x^+(\phi), \theta(\phi, \varepsilon))}{\partial \theta}\frac{\partial\theta(\phi, \varepsilon)}{\partial\phi}\end{aligned}$$

[21] For an example of such nondifferentiability, consider the option contract giving the buyer the right to buy one unit at some price p. In the absence of uncertainty, the buyer's equilibrium utility is typically nondifferentiable in investment at any point at which he is indifferent about exercising his option.

for all $x \in X$ and all $\varepsilon \in \mathfrak{E}$. Therefore, under Condition AA, with $\partial\theta(\phi,\varepsilon)/\partial\phi > 0$ and θ having full support on $[\underline{\theta}, \bar{\theta}]$ for every ϕ, Condition $H^{\pm}$ holds if and only if Condition $\pm$ is satisfied. Likewise, if Condition $\pm$ holds, then Condition $H^{\pm}$ holds if and only if the sign of $\partial\theta(\phi,\varepsilon)/\partial\phi$ does not depend on ε.[22]

PROPOSITION 4: *Under Conditions A and* $H^{\pm}$, *if there exists a contract sustaining* $a^0 \in \operatorname{int} A$, *then there exists a noncontingent contract that implements a utility mapping* $\langle U_1(\cdot), U_2(\cdot)\rangle$ *satisfying the first-order conditions* (14).

PROOF: See Appendix.

We also establish a stronger result for the environment in which the stronger aggregation property Condition AA holds and the aggregate $\theta(\phi,\varepsilon)$ has a conditional distribution function $F(\theta|\phi)$ with a uniformly bounded derivative in ϕ. These stronger assumptions allow us to derive the optimality of a noncontingent contract under weaker conditions than Condition $H^{\pm}$. Another reason for our interest in this environment is that, as the following lemma shows, the added structure allows us to rewrite the first-order conditions (14) for investment choice in terms of the decision rule $x(\cdot)$ generating the contract's utility mapping:

LEMMA 2: *Suppose that Condition AA holds and* $\theta(\phi,\varepsilon)$ *has a conditional distribution function* $F(\theta|\phi)$ *with support* $[\underline{\theta}, \bar{\theta}]$ *and a uniformly bounded derivative in* ϕ *in a neighborhood of* $\phi^0 \equiv \phi(a^0)$. *Then a utility mapping generated by decision rule* $x(\cdot)$ *satisfies the first-order conditions* (14) *if and only if*

$$(16) \qquad -\left[\int \frac{\partial u_i(x(\theta),\theta)}{\partial\theta} F_\phi(\theta|\phi^0)\,d\theta\right]\nabla_{a_i}\phi(a^0) + \nabla_{a_i} E_\varepsilon g_i(a^0,\varepsilon) = 0, \quad \textit{for} \quad i = 1, 2.$$

PROOF: Suppose that the contract implements utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$. Using integration by parts, party i's expected equilibrium utility can be written as

$$E_\varepsilon U_i(\theta(\phi,\varepsilon)) = \int U_i(\theta)\,dF(\theta|\phi) = U_i(\bar{\theta}) - \int U_i'(\theta) F(\theta|\phi) d\theta.$$

Since the utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ is generated by decision rule $x(\cdot)$, we can write

$$\begin{aligned} E_\varepsilon U_i(\theta(\phi,\varepsilon)) &= U_i(\bar{\theta}) - \int U_i'(\theta) F(\theta|\phi)\,d\theta \\ &= U_i(\bar{\theta}) - \int \frac{\partial u_i(x(\theta),\theta)}{\partial\theta} F(\theta|\phi)\,d\theta. \end{aligned}$$

[22] Segal (1999) provides an example in which Condition $H^{\pm}$ is not satisfied. In that paper, investments affect the value and cost of only one "real" good out of n goods, and the real good is determined randomly (i.e., is determined by ε). Thus, the trades that maximize and minimize responsiveness to investment depend on ε. As a result, the conclusion of Proposition 4 does not hold: general message-contingent contracts perform better than noncontingent contracts.

Hence, differentiating with respect to ϕ (Bartle (1995, Corollary 5.9)), we have

$$\frac{\partial}{\partial \phi} E_\varepsilon U_i(\theta(\phi, \varepsilon)) = -\int \frac{\partial u_i(x(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi)\, d\theta. \qquad \textit{Q.E.D.}$$

Observe that if the parties use continuous contracts, in which, by Proposition 3, the implemented utility mapping is generated by the *equilibrium* decision rule, then (16) makes a potentially testable prediction regarding equilibrium decisions in any contract sustaining a given interior investment vector a^0. We shall have more to say about this later in the paper.

We can now establish our second result on when noncontingent contracts are optimal:

PROPOSITION 5: *Suppose that Condition AA holds and that $\theta(\phi, \varepsilon)$ has a conditional distribution $F(\theta|\phi)$ with support $[\underline{\theta}, \bar{\theta}]$ and a uniformly bounded derivative in ϕ in a neighborhood of $\phi^0 \equiv \phi(a^0)$. Suppose in addition that Condition $\pm$ holds and that $F_\phi(\theta|\phi)$ has a constant sign across all θ and ϕ.*[23] *If there exists a contract sustaining $a^0 \in \operatorname{int} A$, then there exists a noncontingent contract that implements a utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$ satisfying the first-order conditions* (14).

PROOF: For definiteness, let $F_\phi(\theta|\phi) \le 0$ for all θ and ϕ. Let $x(\cdot)$ denote the decision rule generating the utility mapping of the original contract. Under the assumptions of the proposition, we can write

$$\begin{aligned} -\int \frac{\partial u_i(x^-, \theta)}{\partial \theta} F_\phi(\theta|\phi^0)\, d\theta &\le -\int \frac{\partial u_i(x(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi^0)\, d\theta \\ &\le -\int \frac{\partial u_i(x^+, \theta)}{\partial \theta} F_\phi(\theta|\phi^0)\, d\theta. \end{aligned}$$

Using connectedness of X and continuity, there must exist $\hat{x} \in X$ such that

$$-\int \frac{\partial u_i(\hat{x}, \theta)}{\partial \theta} F_\phi(\theta|\phi^0)\, d\theta = -\int \frac{\partial u_i(x(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi^0)\, d\theta.$$

Since the first-order condition for the original contract can be rewritten as (16) according to Lemma 2, the above display implies that

$$-\left[\int \frac{\partial u_i(\hat{x}, \theta)}{\partial \theta} F_\phi(\theta|\phi^0)\, d\theta\right] \nabla_{a_i} \phi(a^0) + \nabla_{a_i} E_\varepsilon g_i(a^0, \varepsilon) = 0, \quad \text{for} \quad i = 1, 2.$$

[23] In fact, the result follows under the weaker assumption that for each ϕ there exist $x^+(\phi)$, $x^-(\phi) \in X$ such that

$$\frac{\partial u_1(x^-(\phi), \theta)}{\partial \theta} F_\phi(\theta|\phi) \le \frac{\partial u_1(x, \theta)}{\partial \theta} F_\phi(\theta|\phi) \le \frac{\partial u_1(x^+(\phi), \theta)}{\partial \theta} F_\phi(\theta|\phi)$$

for all $x \in X$ and all $\theta \in [\underline{\theta}, \bar{\theta}]$.

But this is exactly the first-order condition (16) for the investment vector a^0 under the noncontingent contract that prescribes decision $\hat{x}$ (and whose utility mapping is generated by this constant decision). *Q.E.D.*

Recall that under Conditions AA and $\pm$, Condition $H^{\pm}$ requires that $\partial\theta(\phi,\varepsilon)/\partial\phi$ be of constant sign for all (ϕ,ε), which is stronger than the requirement in Proposition 5 that $F_\phi(\theta|\phi)$ not change sign. Only this weaker condition is needed in the proposition because under Condition AA we need only be concerned with finding trades that minimize and maximize investment incentives at each possible value of θ, rather than at all realizations of ε.

Propositions 4 and 5 do not insure that the investment vector a^0 satisfies the appropriate second-order conditions under the constructed noncontingent contract $\hat{x}$. The problem arises even when each party i's underlying payoff $v_i(\hat{x}, a, \varepsilon)$ and the total surplus $S(a,\varepsilon)$ are concave in investments a, and each party i's investment cost $\psi_i(a,\varepsilon)$ is convex in investments. Indeed, when $\lambda_i > 0$, party i's post-renegotiation payoff (10) is negatively affected by the other party's disagreement utility $v_{-i}(\cdot)$; hence this payoff need not be concave in a_i when $v_{-i}(\cdot)$ depends on a_i (as in Che and Hausch (1999) and Demski and Sappington (1991)).[24] Thus, the noncontingent contract constructed in Proposition 4, while satisfying first-order conditions at $a = a^0$, may not sustain a^0. One way to rule out this possibility is by assuming that the parties' payoffs under all noncontingent contracts are concave in their own investments:

Condition C: For each $i = 1, 2$, all $\hat{x} \in X$, and all $a_{-i} \in A_{-i}$,

$$E_\varepsilon[u_i(\hat{x}, \phi(a_i, a_{-i}), \varepsilon) + g_i(a_i, \varepsilon)] \quad \text{is strictly concave in} \quad a_i \in A_i.$$

This condition would be satisfied, for example, whenever the cost functions $\psi_i(a_i, \varepsilon)$ are sufficiently convex.

3.2. *Optimal Noncontingent Contracts*

Recent work on hold-up models has identified cases in which the first-best can be sustained despite ex post nonverifiability, as well as cases in which nonverifiability constrains the parties to the point at which the second-best contract is effectively no contract (i.e., specifies no trade at all). In particular, Edlin and Reichelstein (1996) show that, in a model of self-investments, first-best investments can be sustained under their separability condition with a noncontingent contract that specifies the average efficient trade level given efficient investments. Che and Hausch (1999), on the other hand, show that when one party's investment has a large effect on the other party's utility, the "second-best" contract will specify no trade at all (note that this is also a noncontingent contract).

[24] This problem does not arise with self-investments (i.e., when $v_{-i}(\cdot)$ does not depend on a_i), as in Edlin and Reichelstein (1996) and Hart and Moore (1990).

Given these results, one might wonder what contracts are optimal for more general settings, in which investments have both self- and cooperative effects. In particular, we would like to know how the nature of investments determines the level of trade that the parties specify in an optimal contract. (For example, how general is the finding of Che and Hausch that with cooperative investment effects, the optimal ex ante contract specifies a low level of trade, which is renegotiated upward ex post?)

In this subsection, we address these issues by characterizing optimal noncontingent contracts in situations in which Conditions A, $H^{\pm}$, and C hold, and therefore a noncontingent contract is optimal by Proposition 4. (At the end of the section we formulate a property that is shared by all optimal continuous contracts.) Due to the complexity of the general problem, we restrict attention to a far more specialized situation. Specifically, we assume that Condition S holds, which implies Conditions A and $H^{\pm}$ (for the latter, take $x^+(\phi) \equiv \bar{x}$ and $x^-(\phi) \equiv \underline{x}$). In addition, we restrict attention to cases in which only one party (say, party 1) invests, choosing a one-dimensional investment $a_1 \in A_1 \subset \mathbb{R}$, and in which $\tilde{v}_2(a_1, \varepsilon)$ does not depend on a_1. Finally, as in Edlin and Reichelstein (1996) and Che and Hausch (1999), we assume that there is complete contracting, so that (11) holds.

To begin, we define the first-best investment a_1^*, which for simplicity we assume is unique. In our quasilinear environment, this is the investment that maximizes expected net surplus:

$$(17) \qquad a_1^* = \arg\max_{a_1 \in A_1} E_\varepsilon[S(a_1, \varepsilon) - \psi_1(a_1, \varepsilon)].$$

When this investment is interior ($a_1^* \in \operatorname{int} A_1$), it must satisfy the first-order condition

$$(18) \qquad \frac{\partial}{\partial a_1} E_\varepsilon[S(a_1^*, \varepsilon)] = \frac{\partial}{\partial a_1} E_\varepsilon[\psi_1(a_1^*, \varepsilon)].$$

For later reference, note also that by the Envelope Theorem (see, e.g., Milgrom and Segal (2002, Theorem 1)) we have

$$(19) \qquad \frac{\partial}{\partial a_1} E_\varepsilon[S(a_1^*, \varepsilon)] = \frac{\partial}{\partial a_1} E_\varepsilon[v_1(x^*(a_1^*, \varepsilon), a_1^*, \varepsilon)] + \frac{\partial}{\partial a_1} E_\varepsilon[v_2(x^*(a_1^*, \varepsilon), a_1^*, \varepsilon)],$$

where $x^*(a_1^*, \varepsilon)$ is an ex post optimal trading rule.

Now, consider a noncontingent contract that prescribes trade level $\hat{x}$. If this contract sustains investment choice a_1^* by party 1, the following first-order condition must hold:

$$(20) \qquad \frac{\partial}{\partial a_1} E_\varepsilon[\lambda_2 v_1(\hat{x}, a_1^*, \varepsilon) - \lambda_1 v_2(\hat{x}, a_1^*, \varepsilon) + \lambda_1 S(a_1^*, \varepsilon)] = \frac{\partial}{\partial a_1} E_\varepsilon[\psi_1(a_1^*, \varepsilon)].$$

Substituting from (18) for $\partial E_\varepsilon[\psi_1(a_1^*, \varepsilon)]/\partial a_1$ in (20), we see that if the noncontingent contract $\hat{x}$ sustains the first-best investment a_1^*, then

$$\frac{\partial}{\partial a_1} E_\varepsilon[\lambda_2 v_1(\hat{x}, a_1^*, \varepsilon) - \lambda_1 v_2(\hat{x}, a_1^*, \varepsilon) - \lambda_2 S(a_1^*, \varepsilon)] = 0.$$

Substituting for $\partial E_\varepsilon[S(a_1^*, \varepsilon)]/\partial a_1$ from (19), we obtain

$$\lambda_2 \frac{\partial}{\partial a_1} E_\varepsilon[v_1(\hat{x}, a_1^*, \varepsilon) - v_1(x^*(a_1^*, \varepsilon), a_1^*, \varepsilon)] = \lambda_1 \frac{\partial}{\partial a_1} E_\varepsilon[v_2(\hat{x}, a_1^*, \varepsilon)] + \lambda_2 \frac{\partial}{\partial a_1} E_\varepsilon[v_2(x^*(a_1^*, \varepsilon), a_1^*, \varepsilon)].$$

Using Condition S and the assumption that $\tilde{v}_2(a_1, \varepsilon)$ does not depend on a_1, this equation can be rewritten as

$$\hat{x} = \left[1 + \frac{\bar{v}_2'(a_1^*)}{\lambda_2 \bar{v}_1'(a_1^*) - \lambda_1 \bar{v}_2'(a_1^*)}\right] E_\varepsilon[x^*(a_1^*, \varepsilon)]. \tag{21}$$

Thus, under our assumptions, Proposition 4 implies that the first-best investment a_1^* is sustainable if and only if there is an $\hat{x} \in X$ that satisfies (21). This is formally stated in the following proposition, which also characterizes a second-best optimal contract when a_1^* cannot be sustained:

PROPOSITION 6: *Suppose that Conditions S and C hold, we have complete contracting (i.e., (11) holds), only party* 1 *invests,* $A_1 \subset \mathbb{R}$, $\tilde{v}_2(a_1, \varepsilon)$ *is independent of* a_1, *and the efficient investment is* $a_1^* \in \operatorname{int} A_1$. *Define the real-valued quantity*

$$x^{**} = \left[1 + \frac{\bar{v}_2'(a_1^*)}{\lambda_2 \bar{v}_1'(a_1^*) - \lambda_1 \bar{v}_2'(a_1^*)}\right] E_\varepsilon[x^*(a_1^*, \varepsilon)].$$

If $x^{**} \in X$, *then the noncontingent contract that prescribes the trade level* $\hat{x} = x^{**}$ *sustains the first-best investment* a_1^*. *Otherwise,* a_1^* *cannot be sustained by any contract. Moreover, if the expected ex ante surplus* $E_\varepsilon[S(a_1, \varepsilon) - \psi_1(a_1, \varepsilon)]$ *is quasiconcave in* a_1, *and* $\operatorname{sign}[\lambda_2 \bar{v}_1'(a_1) - \lambda_1 \bar{v}_2'(a_1)] = $ *const, then the noncontingent contract that prescribes trade level*

$$\hat{x} = \begin{cases} \underline{x} & \text{if } x^{**} < \underline{x}, \\ \bar{x} & \text{if } x^{**} > \bar{x}, \end{cases}$$

is a (second-best) optimal contract.

PROOF: See Appendix.

Proposition 6 allows us to compare the level of trade $\hat{x}$ specified in the optimal noncontingent contract with the average level of actual (post-renegotiation) trade arising under this contract. When $\hat{x} \in (\underline{x}, \bar{x})$, the first-best investment level is

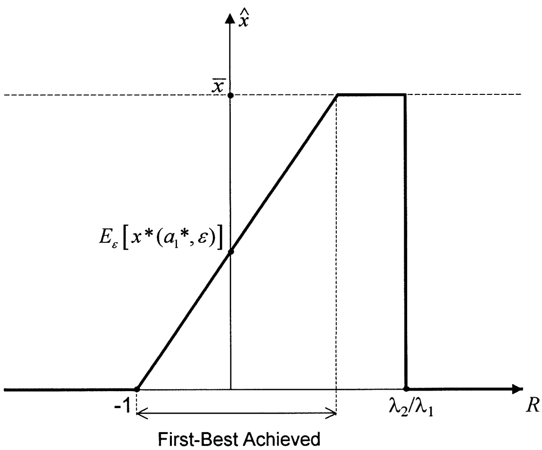

FIGURE 3.—Optimal noncontingent contract.

sustainable, and the comparison hinges on the sign of the fraction in (21). The numerator represents the externality that party 1's investment has on party 2. In the absence of such externality, the optimal noncontingent contract specifies the average ex post efficient trade level, as we already know from Edlin and Reichelstein. When the externality is positive (resp. negative), the contractually-specified trade should be modified from the average ex post efficient level so as to encourage (resp. discourage) investment. Whether an increase in $\hat{x}$ encourages or discourages investment depends, in turn, on the denominator in (21). When the denominator is positive, an increase in $\hat{x}$ encourages investment; conversely, when the denominator is negative, an increase in $\hat{x}$ discourages investment.[25]

The result is illustrated in Figure 3, which plots the optimal noncontingent contract's quantity as a function of $R = \bar{v}_2'(a_1^*)/\bar{v}_1'(a_1^*)$, which can be thought of as a measure of the investment's external effect on party 2.[26] Note that the self-investment case of Edlin and Reichelstein corresponds to $R = 0$, while the purely cooperative investment case of Che and Hausch corresponds to $R = +\infty$ or $R = -\infty$. For the purpose of illustration, we assume that $X = [0, \bar{x}]$.

When $R = 0$, we have $\hat{x}(0) = E_\varepsilon x^*(a_1^*, \varepsilon)$, which is in accordance with the result of Edlin and Reichelstein (1996): when investment is purely self-investment, the optimal noncontingent contract specifies the expected ex post efficient trade given the first-best investment a_1^*. When $R > 0$, party 1's investment has a posi-

[25] To see this, note that $\partial^2 u_1(x, \phi(a_1), \varepsilon)/\partial x \partial a_1 = \lambda_2 \bar{v}_1'(a_1) - \lambda_1 \bar{v}_2'(a_1)$.

[26] Note that in some settings we may have $\bar{v}_1'(a_1^*) < 0$. For example, in the asset model of Demski and Sappington (1991) and Edlin and Hermalin (2000), investment of party 1 (the agent) increases the ex post value of the asset in either party's hands. If $x \in [0, 1]$ is the probability that party 2 receives the asset, then $\bar{v}_1(\cdot)$ is the negative of the asset's value in party 1's hands, while $\bar{v}_2(\cdot)$ is its value in party 2's hands; hence we have $\bar{v}_1'(\cdot) < 0$ and $\bar{v}_2'(\cdot) > 0$.

tive externality on party 2, and therefore should be encouraged. As long as the denominator in (21) is positive, this encouragement is achieved by increasing $\hat{x}(R)$ above the expected ex post efficient level. Contrast this effect to the setting of Che and Hausch, where investment is encouraged by *reducing* contractual trade.

At some $R > 0$, $\hat{x}(R)$ hits the maximum possible trade $\bar{x}$. The function remains at this upper bound up to $R = \lambda_2/\lambda_1$. Beyond this point, the investment effect of the prescribed trade level is reversed, and the effect identified by Che and Hausch (1999) now prevails. Since party 1 underinvests under any contract, it is now optimal to set $\hat{x}(R) = 0$.

When $R < 0$, investment has a negative externality, and therefore should be discouraged. For this reason, $\hat{x}(R)$ decreases as R decreases until $R = -1$, at which point $\hat{x}(R)$ hits zero and stays there all the way to $R = -\infty$ (which corresponds to purely cooperative investment).

According to Proposition 6, we can also use the graph to see when ex post nonverifiability reduces ex ante surplus (when the bounds of $X = [0, \bar{x}]$ bind), and when it does not (when the bounds do not bind).

Proposition 6 characterizes the level of trade in the optimal noncontingent contract in the present setting. However, when the aggregation Condition AA holds and $F(\theta|\phi)$ has a uniformly bounded derivative in ϕ, we can also say something about the expected level of trade prescribed in the equilibrium of *any* optimal continuous contract. Recall from Proposition 3 that the equilibrium decision rule $\hat{x}(\cdot)$ in a continuous contract generates the implemented utility mapping. From Lemma 2 we know that if this continuous contract sustains $a_1^0 \in \operatorname{int} A$, then

$$\int \frac{\partial u_i(\hat{x}(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi(a_1^0))d\theta = \int \frac{\partial u_i(\hat{x}, \theta)}{\partial \theta} F_\phi(\theta|\phi(a_1^0))d\theta,$$

where $\hat{x}$ is the level of trade prescribed by the noncontingent contract sustaining a_1^0. In the present setting, however, $\partial u_1(x, \theta)/\partial \theta = x$. Moreover, recall that when Conditions S and AA both hold, we must have $\theta(\phi, \varepsilon) = \phi + [\lambda_2\gamma_1(\varepsilon) - \lambda_1\gamma_2(\varepsilon)]$, in which case $F_\phi(\theta|\phi) = -f(\theta|\phi)$, the density of θ conditional on ϕ.[27] The above equation can then be rewritten as

$$\int \hat{x}(\theta) f(\theta|\phi(a_1^0))d\theta = \hat{x}.$$

Hence, the expected level of trade prescribed in any optimal continuous contract must be equal to the level of trade in the optimal noncontingent contract described in Proposition 6.

[27] To see this, let $H(\cdot)$ and $h(\cdot)$ denote, respectively, the probability distribution and density functions of $[\lambda_2\gamma_1(\varepsilon) - \lambda_1\gamma_2(\varepsilon)]$, and observe that $F_\phi(\theta|\phi) = \partial H(\theta - \phi)/\partial\phi = -H'(\theta - \phi) = -h(\theta - \phi) = -f(\theta|\phi)$.

3.3. *When Options or More Complex Contracts are Optimal*

In subsection 3.1 we derived conditions under which the two agents can restrict themselves to the use of a noncontingent contract without loss. In this section we consider what can be said about the nature of optimal contracts when the assumptions leading to this result do not hold. In particular, we are interested in when the parties can nonetheless restrict themselves without loss to the use of *option contracts*, in which only one party needs to make an announcement. Given Proposition 4, noncontingent contracts may not be optimal when one of the three conditions—A, $H^{\pm}$, or C—is not satisfied. We consider each of these possibilities in turn.

3.3.1. *Condition A is not satisfied*

EXAMPLE 7: Let $x \in X = [0, 1]$ be the probability of trading an indivisible good between party 2 (the seller) and party 1 (the buyer). Let party 1 be the only investing party, and suppose that party 1 makes a two-dimensional investment $a_1 = (a_{11}, a_{12}) \in A_1 = [0, \frac{1}{2}]^2$. Party 1's investment cost function is $\psi_1(a_{11}, a_{12}) = \frac{1}{2}(a_{11})^2 + \frac{1}{2}(a_{12})^2$. The setting is one of complete contracting (condition (11) holds) and party 1's payoff function $v_1(\cdot)$ is

$$v_1(x, a_1, \varepsilon) = [\varepsilon \cdot 4a_{11} + (1-\varepsilon)a_{12}]x,$$

where $\mathfrak{E} = \{0, 1\}$ and each value of ε has probability 1/2. Party 2's payoff function is $v_2(x, a_1, \varepsilon) = -x$. The bargaining shares are $(\lambda_1, \lambda_2) = (0, 1)$ (party 1 has no bargaining power), so that $w_1(x, a_1, \varepsilon) = v_1(x, a_1, \varepsilon) - \psi_1(a_1)$.

The ex post surplus-maximizing decision rule is to trade if and only if $\varepsilon = 1$ and $a_{11} \geq \frac{1}{4}$. The first-best investment levels can then be calculated as $(a_{11}^*, a_{12}^*) = (\frac{1}{2}, 0)$. However, a noncontingent contract that specifies trade level $\hat{x}$ results in investment levels $(a_{11}, a_{12}) = (2\hat{x}, \frac{1}{2}\hat{x})$; hence, it cannot sustain the first-best investments. In contrast, as Proposition 7 below shows, the party 1 option contract in which party 1 chooses x and the resulting transfers are $t_2(x) = -t_1(x) = x$ does sustain (a_{11}^*, a_{12}^*).

Intuitively, when Condition A does not hold, a noncontingent contract may be too blunt an instrument to control all investment levels at the same time. Option contracts may do better in such situations. In particular, an option contract can sustain the first-best in complete contracting settings in which only one party invests and the other party's valuation is not affected by investments or uncertainty (note that the above example satisfies these assumptions):

PROPOSITION 7: *Suppose that in a setting of complete contracting* (*condition* (11) *holds*) *only party* 1 *invests and* $v_2(\cdot) = v_2(x)$. *Then any first-best investment* a_1^* *can be sustained with a party* 1 *option contract in which party* 1 *chooses* x *and the corresponding transfers are* $t_1(x) = -t_2(x) = v_2(x)$.

PROOF: Note that party 2's post-renegotiation payoff under this contract is $\lambda_2\{\max_{\tilde{x}\in X}[v_1(\tilde{x}, a_1, \varepsilon)+v_2(\tilde{x})]-[v_1(x, a_1, \varepsilon)+v_2(x)]\}$, which is nonnegative for all investment and quantity choices of party 1. This implies that following investment choice a_1 and realization of uncertainty ε, party 1 optimally chooses an ex post optimal quantity $x^*(a_1, \varepsilon) \in \arg\max_{\tilde{x}\in X}[v_1(\tilde{x}, a_1, \varepsilon) + v_2(\tilde{x})]$, since by doing so he reduces party 2's payoff to zero. Given this choice, party 1's ex post payoff given a_1 and ε is equal to the ex post social surplus $S(a, \varepsilon) = v_1(x^*(a_1, \varepsilon), a_1, \varepsilon)+v_2(x^*(a_1, \varepsilon))$. Hence, party 1 will invest efficiently ex ante. *Q.E.D.*

3.3.2. *Condition $H^{\pm}$ is not satisfied*

EXAMPLE 8: Let $x \in [0, 1] = X$ be the probability of trading an indivisible good between party 2 (the seller) and party 1 (the buyer). Let party 1 be the only investing party, with $A_1 = [0, 1]$, and investment cost $\psi_1(a_1) = \frac{1}{2}(a_1)^2$. The parties' valuations are $v_1(x, a_1, \varepsilon) = (1 + \varepsilon a_1)x$ and $v_2(x, a_1, \varepsilon) = -(1 + \frac{\varepsilon a_1}{2})x$. Assume again that party 1 has no bargaining power (i.e., the bargaining shares are $(\lambda_1, \lambda_2) = (0, 1)$), so that $w_1(x, a_1, \varepsilon) = v_1(x, a_1, \varepsilon) - \psi_1(a_1)$. Note that this setting fails to satisfy Condition $H^{\pm}$: party 1's investment increases its valuation for trade when $\varepsilon > 0$, but reduces it when $\varepsilon < 0$.

Given an investment a_1 and realization of uncertainty ε, the ex post efficient trade probability is equal to 1 if $\varepsilon > 0$ and to 0 if $\varepsilon < 0$. Thus, the maximum ex post surplus is $S(a_1, \varepsilon) = a_1 \max\{\varepsilon/2, 0\}$. Given this, the first-best level of ex ante investment is $a_1^* = E[\max\{\varepsilon/2, 0\}]$, which is positive as long as ε has a nondegenerate distribution. However, if the distribution of ε is symmetric around zero, any noncontingent contract yields investment level $a_1 = 0$. On the other hand, the option contract in which party 1 can choose between the outcomes $(x, t_1, t_2) = (0, 0, 0)$ and $(x, t_1, t_2) = (\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$ sustains the first-best investment a_1^*.

Intuitively, when Condition $H^{\pm}$ fails to hold, a noncontingent contract may be a poor vehicle for providing investment incentives because the effect of specifying a high trade level may be positive for some realizations of ε, but negative for others.

We now study conditions under which option contracts are optimal in situations in which Condition $H^{\pm}$ is not satisfied. Proposition 7 already provides one such result. In what follows we provide another for cases in which Condition AA holds, $X = [\underline{x}, \bar{x}] \subset \mathbb{R}$, and the distribution function of θ conditional on $\phi(a)$, $F(\theta|\phi)$, has a uniformly bounded derivative in ϕ. (Example 8 is such a case as long as ε has bounded support.)

Recall from Proposition 5 that if $a^0 \in \text{int}A$ can be sustained by a contract, then there is a noncontingent contract that implements a utility mapping satisfying the first-order conditions (14), as long as Condition $\pm$ holds and $F_\phi(\theta|\phi)$ is of constant sign. Intuitively, the constant sign of $F_\phi(\theta|\phi)$ ensures that the effect of increasing x on the incentive for investment does not change sign across states, as in Example 8. The following result shows that when the sign of $F_\phi(\theta|\phi)$ changes across states in a well-ordered way, the first-order condition for a^0 can be satisfied with an option contract.

PROPOSITION 8: *Suppose that there exists a contract sustaining $a^0 \in \operatorname{int} A$. Suppose also that Condition AA holds, that $X = [\underline{x}, \bar{x}] \subset \mathbb{R}$, and that the distribution function of θ conditional on ϕ, $F(\theta|\phi)$, has a uniformly bounded derivative with respect to ϕ in a neighborhood of $\phi^0 \equiv \phi(a^0)$. If SCP_i holds for $i = 1$ or 2, and if $F_\phi(\theta|\phi^0)$ changes sign at most once on $\theta \in \Theta$, then there exists an option contract that implements utility mappings satisfying the first-order conditions* (14).

PROOF: See Appendix.

The condition that $F_\phi(\theta|\phi)$ changes sign at most once can be interpreted as saying that small changes in ϕ result in either elementary increases or decreases in risk in the distribution of θ (not necessarily mean-preserving). In Example 8, for instance, we have $\phi(a) = a_1$, $\theta(\phi, \varepsilon) = 1 + \varepsilon\phi$, and $F(\theta|\phi) = H((\theta - 1)/\phi)$, where $H(\cdot)$ is the distribution function of ε. Thus, letting $h(\cdot)$ denote the density function of ε, we can write

$$F_\phi(\theta|\phi) = -\left(\frac{\theta - 1}{\phi^2}\right) h\left(\frac{\theta - 1}{\phi}\right),$$

which changes sign once at $\theta = 1$ when $\phi > 0$. Proposition 8 therefore implies that in Example 8 there exists an option contract satisfying the first-order conditions for sustaining investment level a_1^*. The option contract described in the example satisfies the appropriate second-order conditions as well.

3.3.3. *Condition C is not satisfied*

EXAMPLE 9: Suppose that only party 1 invests with $A_1 = [0, 1.5]$, there is no uncertainty, $X = [0, 20]$, and the functions $v_i(\cdot)$ take the form $v_1(x, a_1) = a_1 x - x^2/2$ and $v_2(x, a_1) = \sqrt{a_1}x$. Note that both utility functions are concave in a_1. Assume also that there is complete contracting, so that condition (11) holds. Then the ex post optimal trade is $a_1 + \sqrt{a_1}$, and the available ex post surplus is $S(a_1) = (a_1 + \sqrt{a_1})^2/2$. Let party 1's investment cost be $\psi_1(a_1) = 3(a_1)^2/4 + a_1\sqrt{a_1}$ (note that it is convex in a_1). Then the socially optimal investment is $a_1^* = 1$. The unique noncontingent contract $\hat{x}$ under which a_1^* satisfies party 1's first-order condition is $\hat{x} = 3[1 - (\lambda_1/2\lambda_2)]^{-1}$.[28] Suppose now that the bargaining shares are $\lambda_1 = 0.6$ and $\lambda_2 = 0.4$, so that $\hat{x} = 12$. The ex ante payoff of party 1 in this contract is

$$\pi_1(a_1) = [\lambda_2(a_1\hat{x} - \hat{x}^2/2) - \lambda_1\sqrt{a_1}\hat{x}] + \lambda_1 S(a_1) - \psi_1(a_1).$$

Simple calculation shows that while $\pi_1'(a_1^*) = 0$, we have $\pi_1''(a_1^*) = 0.6 > 0$, so that party 1's second-order condition is violated. Hence, party 1 will not optimally choose a_1^* in this noncontingent contract.

[28] See the discussion leading to condition (21).

In cases in which noncontingent contracts fail to satisfy second-order conditions, it is interesting to ask under what conditions option contracts are optimal. While we do not have a general answer to this question, we find that this is indeed true when only one party invests and the parties' valuations for trade are certain (which includes the above example):[29]

PROPOSITION 9: *Suppose that only party* 1 *invests and the functions* $v_i(\cdot)$ *are continuous and do not depend on* ε. *Then if investment* a_1^0 *is sustained by a contract, it is sustained by a party* 2 *option contract*.

PROOF: Suppose a_1^0 is sustained by a contract with message spaces M_1, M_2 and outcome function $\langle x(\cdot,\cdot), t_1(\cdot,\cdot), t_2(\cdot,\cdot)\rangle$, with $t_1(m_1, m_2) = -t_2(m_1, m_2)$. Let $u_i(x, a_1) = \lambda_{-i} v_i(x, a_1) - \lambda_i v_{-i}(x, a_1)$ and let $\langle U_1(a_1), U_2(a_1)\rangle$ be the equilibrium ex post utility mapping. Consider now the "tariff" contract giving party 2 the option to choose any $x \in X$ with the accompanying transfer

$$\begin{aligned} T_2(x) &= \limsup_{\tilde{x}\to x} \sup\left\{t_2(m_1^*(a_1^0), m_2)\colon m_2 \in M_2, x(m_1^*(a_1^0), m_2) = \tilde{x}\right\} \\ &= -T_1(x). \end{aligned}$$

(Note that $T_2(x(m_1^*(a_1^0), m_2^*(a_1^0)) > -\infty$, and $T_2(\cdot)$ is bounded above because $U_2(a_1^0)$ is finite.) By construction, $T_2(\cdot)$ is an upper semicontinuous function, and so party 2 has an optimal choice of $x \in X$ for any a_1. Let $\widetilde{U}_i(a_1)$ denote party i's equilibrium utility in the tariff contract.

By the zero-sum nature of the message game, for all $a_1 \in A_1$ we can write

$$\begin{aligned} \widetilde{U}_1(a_1) &= \min_{x\in X}[u_1(x, a_1) + T_1(x)] \\ &= \inf_{m_2\in M_2}[u_1(x(m_1^*(a_1^0), m_2), a_1) + t_1(m_1^*(a_1^0), m_2)] \end{aligned}$$

(the second equality obtains by compactness of X and continuity of $u_1(\cdot, a_1)$).

By the Minimax Theorem applied to the message game induced by the original contract,

$$\widetilde{U}_1(a_1) \le \inf_{m_2\in M_2} \sup_{m_1\in M_1} \left[u_1(x(m_1, m_2), a_1) + t_1(m_1, m_2)\right] = U_1(a_1).$$

On the other hand, since $\langle m_1^*(a_1^0), m_2^*(a_1^0)\rangle$ is by assumption a Nash equilibrium of the message game induced by the original contract following $a_1 = a_1^0$,

$$\widetilde{U}_1(a_1^0) = u_1(x(m_1^*(a_1^0), m_2^*(a_1^0)), a_1^0) + t_1(m_1^*(a_1^0), m_2^*(a_1^0)) = U_1(a_1^0).$$

Therefore, if a_1^0 was sustained by the original contract, it is also sustained by the new tariff contract. *Q.E.D.*

[29] We thank Aaron Edlin for suggesting this result and Steve Matthews for alerting us to an error in an earlier version of its proof.

4. APPLICATION TO RISK SHARING

In this section, we use our implementation results to study the form of optimal contracts when the objective of contracting is risk sharing. We assume that there is a one-dimensional ex post state of the world $\theta \in [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$ that is unknown to the parties ex ante and is observable, but not verifiable, ex post. An ex post outcome is a vector $\langle x, t_1, t_2 \rangle$, where $x \in X$ is the nonmonetary component of the outcome (X is a connected compact space), and $t_i \in \mathbb{R}$ is the monetary transfer to party i. The ex post monetary payoff to each party i is $v_i(x, \theta) + t_i$, and we assume that the maximum achievable ex post total surplus is $S(\theta) = \max_{x \in X}\{v_1(x, \theta) + v_2(x, \theta)\}$. (This rules out cases of "incomplete contracting," as discussed in Section 3.) We assume that the parties are (weakly) risk-averse. The ex ante expected utility of party i is $E_\theta W_i(U_i(\theta))$, where $W_i(\cdot)$ is a smooth, increasing, and concave Bernoulli utility function, and $U_i(\theta)$ is the ex post payoff of party i in state θ measured in monetary terms. Given that renegotiation always selects an ex post efficient outcome, the objective of an ex ante contract is to allocate risks optimally.[30]

The timing is exactly as in Section 3, except that here the parties have no investment decisions at stage 2. As in Section 3, we assume that in stage 4 the parties split the renegotiation surplus in fixed proportion. We denote the post-renegotiation monetary payoff to party i (exclusive of the monetary transfer t_i) by

$$\begin{aligned} (22) \qquad u_i(x, \theta) &= v_i(x, \theta) + \lambda_i[S(\theta) - v_i(x, \theta) - v_{-i}(x, \theta)] \\ &= [\lambda_{-i} v_i(x, \theta) - \lambda_i v_{-i}(x, \theta)] + \lambda_i S(\theta). \end{aligned}$$

Note that in contrast to Section 3, here we find it most convenient to let $u_i(\cdot)$ stand for *all* of party i's post-renegotiation payoff, and not just the part affected by the contract; hence, here we have $Z(\theta) = S(\theta)$. Sufficient conditions for applying our implementation results in Section 2 are then that $v_i(\cdot, \cdot)$ is continuous in x, that $\partial v_i(\cdot, \cdot)/\partial\theta$ exists and is continuous in (x, θ), and that $S(\cdot)$ is continuously differentiable.

4.1. *Properties of First-Best Contracts*

In this subsection we study the properties of "first-best" contracts, i.e., contracts implementing Pareto optimal risk-sharing allocations. (Later we will discuss how to derive second-best contracts in situations in which first-best risk sharing is not implementable.) As shown by Borch (1962), for any Pareto optimal risk-sharing utility mapping $\langle U_1(\cdot), U_2(\cdot) \rangle$ there exist nondecreasing functions $f_1(\cdot), f_2(\cdot)$, with $\sum_{i=1,2} f_i(S) = S$ for all S, such that $U_i(\theta) = f_i(S(\theta))$. The

[30] We restrict attention to deterministic mechanisms (i.e., those that assign to each pair of messages a nonrandom outcome). Our analysis could be extended to randomized mechanisms (as in Example 5 in Section 3) if the parties have constant absolute risk aversion and if renegotiation after play of the mechanism takes place prior to the randomization.

surplus sharing functions $f_i(\cdot)$ depend on the parties' Bernoulli utility functions $W_i(\cdot)$, as well as on the selected point on the ex ante utility possibility frontier. Differentiating the Borch condition with respect to θ, and letting $\rho_i(S) = f_i'(S)$, we can write

$$U_i'(\theta) = \rho_i(S(\theta)) \cdot S'(\theta),$$

where $\rho_i(S) \geq 0$ for $i = 1, 2$, $\sum_i \rho_i(S) = 1$ for all $S \in \mathbb{R}$.

As an example, consider the case in which the parties' risk preferences satisfy constant absolute risk aversion (CARA), i.e., $W_i(U) = -(1/r_i)e^{-r_i U}$, with $r_i \geq 0$. In this case, by equalizing marginal rates of substitution across states, it can be seen that *all* Pareto optimal utility mappings satisfy the above condition with $\rho_i(S) = r_i^{-1}/(r_1^{-1} + r_2^{-1}) \equiv \rho_i$. This number can be interpreted as the "relative risk tolerance" of party i.

Proposition 2 establishes that for $\langle U_1(\cdot), U_2(\cdot)\rangle$ to be implementable, there must be a decision rule $x(\cdot)$ that generates it, and hence we must have $U_i'(\theta) = \partial u_i(x(\theta), \theta)/\partial\theta$ for almost all θ. Substituting in the above equation, we obtain that for almost all θ,

$$\frac{\partial u_i(x(\theta), \theta)}{\partial \theta} = \rho_i(S(\theta)) \cdot S'(\theta). \tag{23}$$

Assume that $X = [\underline{x}, \bar{x}] \subset \mathbb{R}$ and a single-crossing property (either SCP_1 or SCP_2) holds. Then, by the Implicit Function Theorem, the above condition fully pins down the generating decision rule $x(\theta)$ for almost all θ. Furthermore, Proposition 3 says that under these assumptions, any implementable utility mapping can be implemented with a continuous mechanism (see footnote 31). In such a mechanism the equilibrium decision rule is also a generating decision rule, and therefore must satisfy condition (23) almost everywhere. Also, note that in the special case of CARA risk preferences, $\rho_i(S) \equiv \rho_i$ for all Pareto optimal risk-sharing utility mappings, and so all such utility mappings are generated by the same decision rule, regardless of the selected point on the ex ante utility possibility frontier.

In what follows, we restrict attention to situations in which only party 1's utility is uncertain. By the Envelope Theorem, $S'(\theta) = \partial v_1(x^*(\theta), \theta)/\partial\theta$, where $x^*(\theta)$ is an ex post optimal trade level in state θ. Substituting this expression into the risk-sharing condition (23) for $i = 1$, and substituting $u_1(x, \theta)$ from (22), the condition can be rewritten as

$$\frac{\partial v_1(x(\theta), \theta)}{\partial \theta} = \left[1 - \frac{\rho_2(S(\theta))}{\lambda_2}\right] \frac{\partial v_1(x^*(\theta), \theta)}{\partial \theta}. \tag{24}$$

We will use this condition to study the properties of first-best risk-sharing contracts.

Note first that when party 1 is risk-neutral, we have $\rho_2(S) \equiv 0$ and condition (24) is satisfied by taking $x(\theta) = x^*(\theta)$. That is, any first-best contract implements the utility mapping that is generated by the ex post efficient decision rule $x^*(\cdot)$. To understand this intuitively, note that the decision rule $x(\theta) = x^*(\theta)$ can be implemented with an option contract, in which party 1 can choose any x at the price $T_2(x) = -T_1(x) = -v_2(x) + F$. This contract fully insures party 2, giving it a constant utility of F. Therefore, it is Pareto optimal.

Now consider the case in which both parties are risk averse. To compare $x(\theta)$ to $x^*(\theta)$, use the Mean Value Theorem to rewrite condition (24) as follows:

$$\text{(25)} \qquad \frac{\partial^2 v_1(\tilde{x}(\theta), \theta)}{\partial x\, \partial\theta} \cdot [x(\theta) - x^*(\theta)] = -\frac{\rho_2(S(\theta))}{\lambda_2} \frac{\partial v_1(x^*(\theta), \theta)}{\partial\theta}$$

for some $\tilde{x}(\theta)$ between $x(\theta)$ and $x^*(\theta)$. Recall that under SCP$_i$ the sign of the cross-partial derivative $\partial^2 v_1(x, \theta)/\partial x\, \partial\theta$ is constant. If, in addition, the sign of the partial $\partial v_1(x, \theta)/\partial\theta$ is known, we can predict the sign of $[x(\theta) - x^*(\theta)]$:

PROPOSITION 10: *Suppose that uncertainty θ affects only party 1's utility, and either SCP$_1$ or SCP$_2$ holds. Then*:

(i) *if*

$$\frac{\partial v_1(x, \theta)}{\partial\theta} \bigg/ \frac{\partial^2 v_1(x, \theta)}{\partial x\, \partial\theta} \geq 0$$

for any $(x, \theta) \in X \times \Theta$, then any decision rule $x(\cdot)$ generating a Pareto optimal utility mapping must satisfy $x(\theta) \leq x^*(\theta)$ for almost all θ;

(ii) *if*

$$\frac{\partial v_1(x, \theta)}{\partial\theta} \bigg/ \frac{\partial^2 v_1(x, \theta)}{\partial x \partial\theta} \leq 0$$

for all $(x, \theta) \in X \times \Theta$, then any decision rule $x(\cdot)$ generating a Pareto optimal utility mapping must satisfy $x(\theta) \geq x^*(\theta)$ for almost all θ.

The assumptions in cases (i) and (ii) mean that the state θ moves party 1's marginal utility $\partial v_1(x, \theta)/\partial x$ and its total utility $v_1(x, \theta)$ in the same or opposite directions, respectively. To see the intuition behind this result, suppose for definiteness that $\partial v_1(x, \theta)/\partial\theta > 0$, i.e., θ increases party 1's utility (and, therefore, total surplus). As observed before, when $x(\theta) = x^*(\theta)$, party 2 is fully insured, and party 1 bears all the risk. Unless party 1 is risk neutral, this allocation of risk is not optimal—it is optimal to shift at least some risk from party 1 to party 2. When $\partial v_1(x, \theta)/\partial\theta$ is increasing in x, we reduce party 1's exposure to risk by reducing x below $x^*(\theta)$. Conversely, when $\partial v_1(x, \theta)/\partial\theta$ is decreasing in x, we reduce party 1's exposure to risk by increasing x above $x^*(\theta)$.

To have a concrete example, suppose that party 1's utility takes the following form:

$$\text{(26)} \qquad v_1(x, \theta) = \theta\beta(x).$$

In this case, condition (24) can be rewritten as

$$\text{(27)} \qquad \beta(x(\theta)) = \left[1 - \frac{\rho_2(S(\theta))}{\lambda_2}\right]\beta(x^*(\theta)),$$

or using the Mean Value Theorem,

$$\beta'(\tilde{x}(\theta)) \cdot [x(\theta) - x^*(\theta)] = -\frac{\rho_2(S(\theta))}{\lambda_2}\beta(x^*(\theta))$$

for some $\tilde{x}(\theta) \in X$. Therefore, $x(\theta)$ is below or above $x^*(\theta)$ depending on whether $\beta(x^*(\theta))$ and $\beta'(\tilde{x}(\theta))$ have the same or opposite sign. Consider two examples:

1. Party 1 is a buyer whose valuation for $x \geq 0$ units of the good in state θ is $v_1(x, \theta) = \theta x$. Since θ increases both the buyer's utility and his marginal utility, any decision rule $x(\cdot)$ generating a Pareto optimal utility mapping has $x(\theta) \leq x^*(\theta)$ for almost all θ.

2. Party 1 is a buyer who always needs to consume exactly $\bar{x}$ units of a good, and can either procure it from party 2 (the seller) or produce it himself at unit cost θ. If the buyer purchases $x \leq \bar{x}$ units from the seller, he needs to produce $(\bar{x} - x)$ units, and his utility can be written as $v_1(x, \theta) = -\theta(\bar{x} - x)$. Since θ reduces the buyer's utility but increases his marginal utility, any decision rule $x(\cdot)$ generating a Pareto optimal utility mapping has $x(\theta) \geq x^*(\theta)$ for almost all θ.

Recall from Proposition 3 that when the parties use a continuous mechanism, the equilibrium decision rule $\hat{x}(\cdot)$ and generating decision rule $x(\cdot)$ coincide.[31] Thus, Proposition 10 predicts the direction of renegotiation in any optimal continuous contract. Moreover, when party 1's payoff is described by (26), expression (27) demonstrates that the extent of equilibrium renegotiation is increasing in party 2's relative risk-tolerance and decreasing in party 2's relative bargaining power.

Finally, we investigate when the parties can achieve the first-best using an option contract, which by Proposition 1 can be done by checking when the decision rule generating the first-best utility mapping is monotonic. For simplicity, we suppose first that the parties' Bernoulli utilities have the CARA form described above, with $\rho_2(S(\theta)) \equiv \rho_2 < \lambda_2$. Also, suppose that $\beta(\cdot)$ is strictly monotonic, e.g., for definiteness, $\beta'(x) > 0$ for all $x \in X$. Then expression (27) implicitly defines $x(\theta)$ as an increasing function of $x^*(\theta)$. Furthermore, the Monotone Selection Theorem (Milgrom and Shannon (1994)) establishes that $x^*(\theta)$ is nondecreasing in θ, and so $x(\theta)$ must also be nondecreasing in θ. Since SCP_1 is satisfied $(\partial^2 u_1(x, \theta)/\partial x\, \partial\theta = (1-\lambda_1)\partial^2 v_1(x, \theta)/\partial x \partial\theta = (1-\lambda_1)\beta'(x) > 0)$, we can conclude that optimal risk sharing can be implemented with a party 1 option contract. This conclusion can be extended to more general utility functions:

[31] The sufficiency part of Proposition 3 implies that in this section's setting, any implementable continuously differentiable utility mapping can be implemented with a continuous mechanism. When both parties have sufficiently smooth and strictly concave Bernoulli utility functions $W_i(\cdot)$, application of the Implicit Function Theorem to the first-order conditions for optimal risk sharing establishes that the optimal utility mapping is indeed continuously differentiable; thus the parties can restrict themselves to continuous mechanisms without loss.

PROPOSITION 11: *Suppose that the parties' Bernoulli utilities satisfy Constant Absolute Risk Aversion, that uncertainty θ affects only party* 1*'s utility, that either SCP_1 or SCP_2 holds, and that*

$$\frac{\partial^2 v_1(x,\theta)}{\partial\theta^2} \leq 0 \quad \text{and} \quad \frac{\partial^3 v_1(x,\theta)}{\partial x\,\partial\theta^2}\frac{\partial v_1(x',\theta)}{\partial\theta} \Big/ \frac{\partial^2 v_1(x,\theta)}{\partial x\,\partial\theta} \leq 0$$

for all $(x, x', \theta) \in X^2 \times \Theta$. Then any Pareto optimal risk sharing utility mapping can be implemented with a party 1 *option contract*.

PROOF: See Appendix.

4.2. *Comparison to the Implicit Labor Contracts Literature*

Our results are closely related to the literature on implicit labor contracts (see, for example, the 1983 symposium in the *Quarterly Journal of Economics*). That literature is concerned with optimal risk sharing between a worker and a firm, when uncertainty in the firm's revenue function is privately observed by the firm. Our model is different in two important respects. First, unlike the labor contract literature, our model has no trade-off between ex ante risk sharing and ex post distortion (such as unemployment), because the parties always renegotiate ex post distortions under complete information. Therefore, in our model first-best risk sharing may be achievable at no ex post cost. The second difference is that, since the realization of uncertainty is now observed by both parties, two-sided message games may now be useful, rather than option contracts in which only the firm sends a message.

Despite these differences, the two models' predictions are similar. First, Proposition 11 demonstrates that under some conditions, there is no need to use two-sided message games: optimal risk sharing can be implemented by giving an option to the party whose preferences are uncertain (the firm). If the firm is risk-neutral, then the first-best can be implemented with the firm's option in which it always chooses the ex post efficient trade $x^*(\theta)$. Observe that this contract can also be implemented when there is no ex post renegotiation, and when the worker does not observe the signal θ. Thus, in this simple case the two models predict exactly the same contract.

When the firm is risk-averse, both in our model and in the standard model, the optimal contractual decision $x(\theta)$ will differ from the ex post efficient decision $x^*(\theta)$. In our model, in contrast to the standard one, this deviation is always renegotiated ex post; hence it is costless. Yet the optimal direction of deviation from $x^*(\theta)$ is the same in both models. To see this intuitively, observe that in the standard model, a small deviation away from $x^*(\theta)$ has a first-order risk sharing effect, but only a second-order loss from ex post distortion. Therefore, regardless of whether an ex post distortion is present, the parties will want to move away from $x^*(\theta)$ at least slightly in the direction dictated by risk sharing considerations.

4.3. *When the First-Best is not Implementable*

To have an example where first-best risk sharing cannot be achieved, suppose that the parties' Bernoulli utilities take the CARA form, with $\rho_2(\theta) \equiv \rho_2 > \lambda_2$, and that $\partial v_1(x, \theta)/\partial\theta > 0$ for all $x \in X$. Then the first-best condition (24) cannot be satisfied. Intuitively, in this case the second-best will involve making the two sides of (24) as close as possible; for example, under SCP_1 (i.e., $\partial^2 v_1(x, \theta)/\partial x\,\partial\theta > 0$), we should optimally set $x(\theta) = \min X$. This intuition can be formalized by setting up the second-best problem as an optimal control problem, where θ is time, U_1 is the state variable, and the generating decision x is the control variable. The motion equation is then given by the generalized Mirrlees condition. With CARA Bernoulli utilities, the parties' certainty equivalents are quasilinear in lump-sum transfers; thus every Pareto optimal contract should maximize the "total surplus"—the sum of the two parties' certainty equivalents. This maximization problem can be solved using Pontryagin's Maximum Principle. A solution to such a problem will generally have some regions of Θ on which the constraint $x \in [\underline{x}, \bar{x}]$ does not bind, and the Euler equation (24) holds. On other regions of Θ, the constraint will bind, and the generating decision will be either $\underline{x}$ or $\bar{x}$.[32]

5. CONCLUSION

In this paper, we have developed a first-order approach to the problem of mechanism design with renegotiation under complete information, and have examined its implications in hold-up and risk sharing models. This first-order approach, which extends Mirrlees's (1971) analysis of standard mechanism design problems, offers a convenient characterization of the set of implementable utility mappings. In the hold-up model, for example, this characterization allowed us to identify the properties of fully optimal second-best contracts, and to ask when these can take some of the simple forms assumed in the existing literature (e.g., noncontingent and option contracts). In risk-sharing problems, this characterization allowed us to identify conditions under which first-best risk sharing could be achieved, and the nature of the contracts that accomplish this.

It should be noted that our view of renegotiation, which follows Maskin and Moore (1999), differs from some other views adopted in the contracting literature. For example, Hart and Moore (1988), Noldeke and Schmidt (1995), and Spier and Whinston (1995) allow renegotiation only before the contractual mechanism, but not after it. Under this alternative assumption, the mechanism is renegotiated prior to its play in those states of the world θ in which it would yield an ex post inefficient equilibrium outcome. Thus, the parties' utility mappings under any contract must still add up to the available ex post surplus, i.e., satisfy condition (2). On the other hand, since outcomes of the mechanism cannot be renegotiated after play commences, the mechanism no longer defines a

[32] Green and Laffont (1992) develop this optimal control technique to analyze a risk-sharing model with complete information and renegotiation, restricting attention to party 1 option contracts and to the case where party 2 has all bargaining power, i.e., $\lambda_2 = 1$.

constant-sum game, and our Mirrlees characterization of the parties' utility mappings under the mechanism no longer applies. Rather, the results of Moore and Repullo (1988, Section 5) imply that as long as θ affects the parties' underlying preferences,[33] they can design a multi-stage mechanism implementing *any* utility mapping that satisfies the adding-up condition (2).[34]

The adding-up condition, by itself, may preclude implementation of efficient investments in bilateral hold-up models due to the "moral hazard in teams" problem identified by Holmstrom (1982). For example, if both parties' investments affect the parties' ex post payoffs only through a one-dimensional state of the world θ, then the adding-up constraint prevents giving each party its marginal contribution to ex post surplus. As a result, with sufficient smoothness of payoff functions, first-best investments cannot be sustained. On the other hand, when only one party invests and its investment affects ex post preferences, the first-best can be implemented by making the investing party the residual claimant of ex post surplus.[35,36]

Our analysis shows that when renegotiation is always possible after play of the mechanism, the extra constraint imposed on the parties' contract design problem (relative to the case in which renegotiation occurs only before play of the mechanism) is given precisely by the Mirrlees condition (3).[37] Furthermore, this constraint sometimes binds: for example, subsection 3.2 shows that under this constraint, the first-best investment level may not be sustainable even in one-sided investment problems.

Which view of renegotiation is closer to reality depends on the technological commitments available to the parties. For example, an inefficient ownership structure prescribed by a contract may be renegotiated without loss of efficiency. On the other hand, the production of an inefficient widget prescribed by a contract may entail an irreversible loss of surplus. Thus, a careful analysis of

[33] In our quasilinear environment, this means that for any two different states of the world there is an agent whose utility functions in the two states do not coincide up to an additive constant.

[34] This assumes that the courts can enforce contractual trade, as in Noldeke and Schmidt (1995) and Spier and Whinston (1995). In contrast, Hart and Moore (1988) assume that the courts cannot verify which party has refused to trade, which further restricts the set of implementable utility mappings.

[35] In trading models, in which the parties' no-trade ex post utilities are zero, this can be achieved with a two-stage game in which the investing party makes a take-it-or-leave-it trading offer to the other party, and no trade is implemented following a rejection.

[36] First-best investments can also be implemented when both parties invest if their investments can be inferred from the one-dimensional ex post state. See Legros and Matsushima (1991) and Legros and Matthews (1993) for characterizations of implementable investments in team moral hazard problems with and without uncertainty respectively.

[37] Another view of renegotiation is taken by Aghion, Dewatripont, and Rey (1994) and Chung (1991). One can view these papers as specifying that renegotiation takes place after play of the mechanism, but also as allowing the mechanism to specify parameters that affect the parties' ex post bargaining powers. In particular, these papers consider contracts that specify, along with the pre-renegotiation outcome, party 1's bargaining share $\lambda_1 \in [0, 1]$. Our Mirrlees characterization of implementable utility mappings can be applied to such contracts by expanding the "decision set" to $X \times [0, 1]$.

technological and legal structure is important for understanding the role of renegotiation. At the same time, it may be possible to test between different theories of renegotiation indirectly, by comparing their predictions for the form of optimal contracts. Therefore, by providing a characterization of optimal contracts under the Maskin-Moore theory of renegotiation, our paper takes a step towards a better understanding of the role of renegotiation in real-world contracting.

Dept. of Economics, Stanford University, Stanford, CA 94305, U.S.A.; ilya.segal@stanford.edu

and

Dept. of Economics, Northwestern University, 2003 Sheridan Rd., Evanston, IL 60208-2600, U.S.A.; mwhinston@northwestern.edu

Manuscript received August, 1998; final revision received May, 2001.

APPENDIX: PROOFS

PROOF OF PROPOSITION 3: Consider a continuous mechanism with message sets (M_1, M_2) and the outcome function $\langle x(\cdot), t_1(\cdot), t_2(\cdot)\rangle$. The mechanism induces a zero-sum game between the agents, in which agent 1's payoff is

$$f_1(m_1, m_2, \theta) = u_1(x(m_1, m_2), \theta) + t_1(m_1, m_2).$$

If $(m_1(\theta), m_2(\theta))$ is a Nash equilibrium message pair in state θ, then agent 1's equilibrium utility is $U_1(\theta) = f_1(m_1(\theta), m_2(\theta), \theta)$. Theorem 4 of Milgrom and Segal (2002) offers a Mirrlees representation for the saddle value function of a zero-sum game. We will argue that the assumptions of this Theorem are satisfied in our setting. First, the partial derivative $\partial f_1(m_1, m_2, \theta)/\partial\theta = \partial u_1(x(m_1, m_2), \theta)/\partial\theta$ is continuous in each of m_1 and m_2, since the mechanism's decision component $x(m_1, m_2)$ of the outcome function is continuous in each of m_1 and m_2, and $\partial u_1(x, \theta)/\partial\theta$ is continuous by assumption. Furthermore, the integrable bound and equidifferentiability assumptions are implied by the continuity of $\partial u_1(x, \theta)/\partial\theta$ on the compact set $X \times \Theta$. The Theorem then establishes that the saddle value $U_1(\cdot)$ is differentiable almost everywhere, and

$$U_1(\theta) = U_1(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} \frac{\partial f_1(m_1(\tau), m_2(\tau), \tau)}{\partial\theta}\, d\tau.$$

Since $\partial f_1(m_1(\theta), m_2(\theta), \theta)/\partial\theta = \partial u_1(x(m_1(\theta), m_2(\theta)), \theta)/\partial\theta$, we see that the equilibrium decision rule $\hat{x}(\theta) \equiv x(m_1(\theta), m_2(\theta))$ generates the utility mapping $\langle U_1(\cdot), U_2(\cdot)\rangle$.

We prove the second statement of the Proposition by constructing a continuous mechanism implementing $\langle U_1(\cdot), U_2(\cdot)\rangle$. This will be a direct revelation mechanism, i.e., the message spaces are $M_i = \Theta$ for each agent $i = 1, 2$. The outcome function is given by

$$x(\theta_1, \theta_2) = \begin{cases} \tilde{x}(\theta_1, \theta_2) & \text{if } \theta_1 \neq \theta_2, \\ x(\theta_1) & \text{otherwise,} \end{cases}$$

$$t_i(\theta_1, \theta_2) = U_i(\theta_i) - u_i(x(\theta_1, \theta_2), \theta_i) \quad \text{for } i = 1, 2,$$

where $\tilde{x}(\theta_1, \theta_2)$ satisfies (4), and exists due to Lemma 1. Note that under SCP$_i$, $\tilde{x}(\theta_1, \theta_2)$ is uniquely defined for $\theta_1 \neq \theta_2$. Note also that by construction, using zero-sum conditions (1) and (2), for $\theta_1 \neq \theta_2$ we have

$$\begin{aligned} t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2) &= U_1(\theta_1) - u_1(\tilde{x}(\theta_1, \theta_2), \theta_1) + U_2(\theta_2) - u_2(\tilde{x}(\theta_1, \theta_2), \theta_2) \\ &= [U_1(\theta_1) - U_1(\theta_2)] - [u_1(\tilde{x}(\theta_1, \theta_2), \theta_1) - u_1(\tilde{x}(\theta_1, \theta_2), \theta_2)] = 0. \end{aligned}$$

It is also easy to see that $t_1(\theta, \theta) + t_2(\theta, \theta) = 0$ for all θ.

It is straightforward to verify that if agent $-i$ announces the true state θ, then agent i obtains utility $U_i(\theta)$ regardless of his own announcement. Therefore, truth telling constitutes a Nash equilibrium of the mechanism, and the equilibrium utility mapping is $\langle U_1(\cdot), U_2(\cdot)\rangle$.

While the constructed mechanism may not be continuous in the standard (real-line) topology on the message space Θ, it is continuous if the topology on Θ is coarsened in the following way. Let $C \subset \Theta$ be the set of discontinuity points of the generating decision rule $x(\cdot)$, and take the topology on Θ that is generated by a basis that includes a basis in the standard topology on Θ and all singleton subsets of C. By construction, the new topology on Θ is second-countable.

We now show that the decision rule $x(\theta_i, \theta_{-i})$ is continuous in $\theta_i \in \Theta$ under the new topology (continuity of the transfer rules follows easily). Continuity at any point $\theta_i \in C$ is trivial, and we can concentrate on points $\theta_i \in \Theta \backslash C$. First observe that by the Implicit Function Theorem $\tilde{x}(\theta_i, \theta_{-i})$ is continuous (and even differentiable) in θ_i at any $\theta_i \neq \theta_{-i}$. Since $x(\theta_i', \theta_{-i}) = \tilde{x}(\theta_i', \theta_{-i})$ for θ_i' in a neighborhood of such θ_i, the mechanism is continuous at θ_i. Thus, it remains to check the continuity of the decision rule $x(\theta_i, \theta_{-i})$ in θ_i at $\theta_i = \theta_{-i} \in \Theta\backslash C$. Since by construction the generating decision rule $x(\cdot)$ is continuous at θ_{-i}, the generated utility mapping $U_i(\cdot)$ must be differentiable at θ_{-i}, with $U_i'(\theta_{-i}) = \partial u_i(x(\theta_{-i}), \theta_{-i})/\partial\theta$. By (4), this implies that

$$\begin{aligned}\frac{u_i(\tilde{x}(\theta_i, \theta_{-i}), \theta_i) - u_i(\tilde{x}(\theta_i, \theta_{-i}), \theta_{-i})}{\theta_i - \theta_{-i}} &= \frac{U_i(\theta_i) - U_i(\theta_{-i})}{\theta_i - \theta_{-i}} \to U_i'(\theta_{-i}) \\ &= \frac{\partial u_i(x(\theta_{-i}), \theta_{-i})}{\partial\theta} \quad \text{as} \quad \theta_i \to \theta_{-i}.\end{aligned}$$

By the Mean Value Theorem, the first fraction in the above display equals $\partial u_i(\tilde{x}(\theta_i, \theta_{-i}), \tilde{\theta}_i(\theta_i))/\partial\theta$ for some $\tilde{\theta}_{-i}(\theta_i)$ between θ_i and θ_{-i}. Thus, we must have

$$\frac{\partial u_i(\tilde{x}(\theta_i, \theta_{-i}), \tilde{\theta}_i(\theta_i))}{\partial\theta} \to \frac{\partial u_i(x(\theta_{-i}), \theta_{-i})}{\partial\theta} \quad \text{as} \quad \theta_i \to \theta_{-i}.$$

But this implies, by SCP$_i$, that $\tilde{x}(\theta_i, \theta_{-i}) \to x(\theta_{-i})$ as $\theta_i \to \theta_{-i}$; thus $x(\theta_i, \theta_{-i})$ is also continuous in θ_i at $\theta_i = \theta_{-i}$. *Q.E.D.*

PROOF OF PROPOSITION 4: Observe first that for each party i we must have

$$\nabla_{a_i} E_\varepsilon g_i(a^0, \varepsilon) = \lambda_i \nabla_{a_i} \phi(a^0) \tag{28}$$

for some $\lambda_i \in \mathbb{R}$. Indeed, otherwise there would exist a_i close to a_i^0 such that $\phi(a_i, a_{-i}^0) = \phi(a^0)$ while $E_\varepsilon g_i(a_i, a_{-i}^0, \varepsilon) > E_\varepsilon g_i(a^0, \varepsilon)$, thus violating the equilibrium condition (13).

If $\nabla_{a_i}\phi(a^0) = 0$, then by (28) we must have $\nabla_{a_i} E_\varepsilon g_i(a^0, \varepsilon) = 0$, and therefore *any* noncontingent contract satisfies party i's first-order condition (14). Consider therefore a party i such that $\nabla_{a_i}\phi_i(a^0) \neq 0$, namely, $\partial\phi_i(a^0)/\partial a_{ik} \neq 0$ for some investment dimension k. Define the function $a_{ik}(\phi)$ implicitly by

$$\phi(a_{ik}(\phi), a_{i,-k}^0, a_{-i}^0) = \phi.$$

Under our assumptions, the Implicit Function Theorem establishes that $a_{ik}(\phi)$ is uniquely defined and differentiable in a neighborhood of $\phi^0 \equiv \phi(a^0)$.

Define

$$G_{ik}(\phi) = E_\varepsilon g_i(a_{ik}(\phi), a_{i,-k}^0, a_{-i}^0, \varepsilon).$$

Using the chain rule, the Implicit Function Theorem, and (28),

$$G_{ik}'(\phi^0) = [\partial E_\varepsilon g_i(a^0, \varepsilon)/\partial a_{ik}] \cdot a_{ik}'(\phi^0) = \frac{\partial E_\varepsilon g_i(a^0, \varepsilon)/\partial a_{ik}}{\partial\phi(a^0)/\partial a_{ik}} = \lambda_i. \tag{29}$$

Letting $V_i(\phi) = E_\varepsilon[U_i(\phi, \varepsilon)]$, the equilibrium condition (13) implies that in a neighborhood of ϕ^0,

$$V_i(\phi) - V_i(\phi^0) + [G_{ik}(\phi) - G_{ik}(\phi^0)] \le 0 \tag{30}$$

for all k.

Suppose that, say, $\partial\phi(a^0)/\partial a_{1k} \neq 0$ for some dimension k. Consider the properties of party 1's equilibrium utility function $U_1(\phi, \varepsilon)$ under the original contract. From Corollary 1 we know that there exists a decision rule $x(\cdot)$ such that for all $\phi'' \ge \phi'$ and all ε,

$$U_1(\phi'', \varepsilon) - U_1(\phi', \varepsilon) = \int_{\phi'}^{\phi''} \frac{\partial u_1(x(\tau, \varepsilon), \tau, \varepsilon)}{\partial \phi} d\tau.$$

Using Condition H$^\pm$, this difference can be bounded as follows:

$$\int_{\phi'}^{\phi''} \frac{\partial u_1(x^-(\tau), \tau, \varepsilon)}{\partial \phi} d\tau \le U_1(\phi'', \varepsilon) - U_1(\phi', \varepsilon) \le \int_{\phi'}^{\phi''} \frac{\partial u_1(x^+(\tau), \tau, \varepsilon)}{\partial \phi} d\tau.$$

Taking the expectation over ε, we have

$$\int_{\phi'}^{\phi''} E_\varepsilon\left[\frac{\partial u_1(x^-(\tau), \tau, \varepsilon)}{\partial \phi}\right] d\tau \le V_1(\phi'') - V_1(\phi') \le \int_{\phi'}^{\phi''} E_\varepsilon\left[\frac{\partial u_1(x^+(\tau), \tau, \varepsilon)}{\partial \phi}\right] d\tau.$$

Note that $\partial u_1(x^-(\tau), \tau, \varepsilon)/\partial\phi = \min_{x\in X} \partial u_1(x, \tau, \varepsilon)/\partial\phi$ is a continuous function of τ by the Theorem of the Maximum. Therefore, the left integrand, which is the expectation of this expression over ε, is also continuous in τ. By the same logic, the right integrand is also continuous in τ. Therefore, by the Mean Value Theorem, for all $\phi'' > \phi'$ there exist $\phi^-(\phi', \phi''), \phi^+(\phi', \phi'') \in [\phi', \phi'']$ such that

$$\begin{aligned} E_\varepsilon\left[\frac{\partial u_1(x^-(\phi^-(\phi', \phi'')), \phi^-(\phi', \phi''), \varepsilon)}{\partial \phi}\right] &\le \frac{V_1(\phi'') - V_1(\phi')}{\phi'' - \phi'} \\ &\le E_\varepsilon\left[\frac{\partial u_1(x^+(\phi^+(\phi', \phi'')), \phi^+(\phi', \phi''), \varepsilon)}{\partial \phi}\right]. \end{aligned} \tag{31}$$

Take a sequence $\phi^n \searrow \phi^0$. Using (30) and the first inequality in (31) for $(\phi', \phi'') = (\phi^0, \phi^n)$, we obtain

$$-\frac{G_{1k}(\phi^n) - G_{1k}(\phi^0)}{\phi^n - \phi^0} \ge E_\varepsilon\left[\frac{\partial u_1(x^-(\phi^-(\phi^0, \phi^n)), \phi^-(\phi^0, \phi^n), \varepsilon)}{\partial \phi}\right].$$

Note that $\phi^-(\phi^0, \phi^n) \to \phi^0$, and moreover, since X is compact, the sequence can be chosen so that $x^-(\phi^-(\phi^0, \phi^n)) \to x^r$ for some $x^r \in X$. Then, taking the limit (and recalling that $\partial u_1(\cdot)/\partial\phi$ is continuous in x and ϕ), we obtain

$$-G'_{1k}(\phi^0) = -\lambda_1 \ge E_\varepsilon\left[\frac{\partial u_1(x^r, \phi^0, \varepsilon)}{\partial \phi}\right]. \tag{32}$$

Now take instead a sequence $\phi^n \nearrow \phi^0$. Using (30), and the second inequality in (31) for $(\phi', \phi'') = (\phi^n, \phi^0)$, we obtain

$$-\frac{G_{1k}(\phi^0) - G_{1k}(\phi^n)}{\phi^0 - \phi^n} \le E_\varepsilon\left[\frac{\partial u_1(x^+(\phi^+(\phi^n, \phi^0)), \phi^+(\phi^n, \phi^0), \varepsilon)}{\partial \phi}\right].$$

Note that $\phi^+(\phi^n, \phi^0) \to \phi^0$, and moreover, since X is compact, the sequence can be chosen so that $x^+(\phi^+(\phi^n, \phi^0)) \to x^l$ for some $x^l \in X$. Then, taking the limit, we obtain

$$-G'_{1k}(\phi^0) = -\lambda_1 \le E_\varepsilon\left[\frac{\partial u_1(x^l, \phi^0, \varepsilon)}{\partial \phi}\right]. \tag{33}$$

Putting (32) and (33) together, we see that by continuity and connectedness of X, there must exist $\hat{x} \in X$ such that

$$G'_{1k}(\phi^0) = -E_\varepsilon\left[\frac{\partial u_1(\hat{x}, \phi^0, \varepsilon)}{\partial \phi}\right] = -\frac{\partial}{\partial \phi}[E_\varepsilon u_1(\hat{x}, \phi^0, \varepsilon)].$$

Hence, by (29),

$$\lambda_1 = -\frac{\partial}{\partial \phi}[E_\varepsilon u_1(\hat{x}, \phi^0, \varepsilon)]. \tag{34}$$

Substituting (34) into (28), we see that a noncontingent contract specifying decision $\hat{x}$ satisfies the first-order condition (14) for party 1.

Now, suppose also that $\partial \phi_2(a^0)/\partial a_{2m} \neq 0$ for some dimension m. Since by construction $V_1(\cdot) = -V_2(\cdot)$, (30) for $i = 1$ and $i = 2$ together imply that in a neighborhood of ϕ^0,

$$\pi(\phi) = [G_{1k}(\phi) - G_{1k}(\phi^0)] + [G_{2m}(\phi) - G_{2m}(\phi^0)] \leq 0.$$

Since $\pi(\phi^0) = 0$, ϕ^0 must be a local maximum of $\pi(\cdot)$, and thus the following first-order condition must hold:

$$\pi'(\phi^0) = G'_{1k}(\phi^0) + G'_{2m}(\phi^0) = 0,$$

and therefore, by (29), $\lambda_2 = -\lambda_1$. Using (34) and the fact that $u_1(\cdot) = -u_2(\cdot)$, we have

$$\lambda_2 = -\lambda_1 = \frac{\partial}{\partial \phi}[E_\varepsilon u_1(\hat{x}, \phi^0, \varepsilon)] = -\frac{\partial}{\partial \phi}[E_\varepsilon u_2(\hat{x}, \phi^0, \varepsilon)].$$

Substituting into (28), we see that a noncontingent contract specifying decision $\hat{x}$ satisfies the first-order condition (14) for party 2. *Q.E.D.*

PROOF OF PROPOSITION 6: The derivation in the text shows that the noncontingent contract $\hat{x} = x^{**}$ satisfies the first-order condition (14) with $a_1^0 = a_1^*$, and under Condition C it also satisfies the appropriate second-order condition. This establishes the first statement. The derivation in the text also implies that if $x^{**} \notin X$, then a_1^* cannot be sustained by a noncontingent contract, hence by Proposition 4, it cannot be sustained by any contract. This establishes the second statement. To establish the third statement in this case, let $a_1^0(\hat{x}) \in A_1$ denote the investment implemented by the noncontingent contract $\hat{x}$ (which is uniquely defined due to Condition C). Then by Proposition 4, the set of investments that are sustainable with a contract can be described as $A_1^0 = \{a_1^0(\hat{x})\colon \hat{x} \in X\}$. $a_1^0(\cdot)$ is a continuous function by Berge's Theorem of the Maximum, and the set $X = [\underline{x}, \tilde{x}]$ is compact and connected; therefore A_1^0 must also be compact and connected, i.e., it is a closed interval. Since $a_1^* \notin A_1^0$ and the ex ante welfare function is quasiconcave, the second-best optimal investment is the point in A_1^0 that is closest to a_1^*. Since by assumption $\partial^2 u_1(\hat{x}, \phi(a_1), \varepsilon)/\partial \hat{x}\, \partial a_1 = \lambda_2 \bar{v}_1'(a_1) - \lambda_1 \bar{v}_2'(a_1)$ has a constant sign, the Monotone Selection Theorem (Milgrom and Shannon (1994)) establishes that the function $a_1^0(\hat{x})$ is monotone, which implies that the second-best optimal $\hat{x} \in X$ is the point in X that is closest to x^{**}. *Q.E.D.*

PROOF OF PROPOSITION 8: Suppose that the utility mapping implemented by the original contract is generated by the decision rule $x(\cdot)$. Let $\Theta_+ = \{\theta \in \Theta\colon F_\phi(\theta|\phi^0) > 0\}$ and $\Theta_- = \{\theta \in \Theta\colon F_\phi(\theta|\phi^0) \leq 0\}$. Under SCP_i, we can write

$$\int_{\Theta_+} \frac{\partial u_i(\underline{x}, \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta \leq \int_{\Theta_+} \frac{\partial u_i(x(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta \leq \int_{\Theta_+} \frac{\partial u_i(\bar{x}, \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta,$$

$$\int_{\Theta_-} \frac{\partial u_i(\bar{x}, \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta \leq \int_{\Theta_-} \frac{\partial u_i(x(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta \leq \int_{\Theta_-} \frac{\partial u_i(\underline{x}, \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta.$$

By connectedness of X and continuity, there exist x_+ and x_- such that

$$\int_{\Theta_+} \frac{\partial u_i(x_+, \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta = \int_{\Theta_+} \frac{\partial u_i(x(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta,$$
$$\int_{\Theta_-} \frac{\partial u_i(x_-, \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta = \int_{\Theta_-} \frac{\partial u_i(x(\theta), \theta)}{\partial \theta} F_\phi(\theta|\phi^0) d\theta.$$

Therefore, since the original contract satisfies first-order conditions (16), the utility mapping generated by the decision rule

$$\tilde{x}(\theta) = \begin{cases} x_+, & \text{for } \theta \in \Theta_+, \\ x_-, & \text{for } \theta \in \Theta_-, \end{cases}$$

will also satisfy (16). Moreover, since by assumption $F_\phi(\theta|\phi)$ changes sign at most once on $\theta \in \Theta$, Θ_- must lie either wholly below or wholly above Θ_+. Therefore, $\tilde{x}(\cdot)$ is monotonic, and since SCP_i is satisfied, Proposition 1 implies that the utility mapping generated by $\tilde{x}(\cdot)$ can be implemented with an option contract. (Which party has the option depends on the comparison between x_+ and x_-, and on which of the sets Θ_- and Θ_+ lies above the other.) *Q.E.D.*

PROOF OF PROPOSITION 11: Differentiating condition (24) with $\rho_2(S(\theta)) \equiv \rho_2 = const$, we obtain

$$\frac{\partial^2 v_1(x(\theta), \theta)}{\partial x\, \partial \theta} x'(\theta) = \left[1 - \frac{\rho_2}{\lambda_2}\right] \frac{\partial^2 v_1(x^*(\theta), \theta)}{\partial x\, \partial \theta} x^{*\prime}(\theta) - \frac{\rho_2}{\lambda_2} \frac{\partial^2 v_1(x^*(\theta), \theta)}{\partial \theta^2} + \left[\frac{\partial^2 v_1(x^*(\theta), \theta)}{\partial \theta^2} - \frac{\partial^2 v_1(x(\theta), \theta)}{\partial \theta^2}\right]. \tag{35}$$

Suppose for definiteness that we have SCP_1, i.e., $\partial^2 v_1(x, \theta)/\partial x\, \partial \theta > 0$. Using the Mean Value Theorem and (25), the last term can then be written as

$$\frac{\partial^2 v_1(x^*(\theta), \theta)}{\partial \theta^2} - \frac{\partial^2 v_1(x(\theta), \theta)}{\partial \theta^2} = \frac{\partial^3 v_1(\hat{x}(\theta), \theta)}{\partial x\, \partial \theta^2} [x^*(\theta) - x(\theta)]$$
$$= -\frac{\rho_2}{\lambda_2} \frac{\partial v_1(x^*(\theta), \theta)}{\partial \theta} \frac{\partial^3 v_1(\hat{x}(\theta), \theta)}{\partial x\, \partial \theta^2} \Big/ \frac{\partial^2 v_1(\tilde{x}(\theta), \theta)}{\partial x\, \partial \theta}$$

for some $\hat{x}(\theta)$, $\tilde{x}(\theta)$ between $x^*(\theta)$ and $x(\theta)$. Our assumptions ensure that this expression is nonnegative. Also, by SCP_1, the Monotone Selection Theorem (Milgrom and Shannon (1994)) implies that $x^{*\prime}(\theta) \geq 0$. Finally, since the first-best can be implemented, we must have $\rho_2 \leq \lambda_2$. Putting all these facts together, we see that the right-hand side in (35) is nonnegative. Therefore, we must have $x'(\theta) \geq 0$. By Proposition 1, this implies that $x(\theta)$ can be implemented with an option contract. The proof for the case where SCP_2 holds is similar. *Q.E.D.*

REFERENCES

AGHION, P., M. DEWATRIPONT, AND P. REY (1994): "Renegotiation Design with Unverifiable Information," *Econometrica*, 62, 257–282.

BARTLE, R. G. (1995): *The Elements of Integration and Lebesgue Measure*. New York: John Wiley & Sons.

BORCH, K. (1962): "Equilibrium in a Reinsurance Market," *Econometrica*, 30, 424–444.

CHE, Y.-K., AND D. HAUSCH (1999): "Cooperative Investments and the Value of Contracting," *American Economic Review*, 89, 125–147.

CHUNG, T. Y. (1991): "Incomplete Contracts, Specific Investments, and Risk-Sharing," *Review of Economic Studies*, 58, 1031–1042.

DEMSKI, J., AND D. SAPPINGTON (1991): "Resolving Double Moral Hazard Problems with Buyout Agreements," *RAND Journal of Economics*, 22, 232–240.

EDLIN, A., AND B. HERMALIN (2000): "Contract Renegotiation and Options in Agency Problems," *Journal of Law, Economics, and Organization*, 16, 395–423.

EDLIN, A., AND S. REICHELSTEIN (1996): "Holdups, Standard Breach Remedies, and Optimal Investment," *American Economic Review*, 86, 478–501.

GREEN, J., AND J.-J. LAFFONT (1989): "Contract Renegotiation and the Underinvestment Effect," mimeo, Harvard University.

——— (1992): "Renegotiation and the Form of Efficient Contracts," *Annales d'Economie et de Statistique*, 25/26, 123–150.

GROSSMAN, S., AND O. HART (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691–719.

HART, O. (1995): *Firms, Contracts, and Financial Structure*. Oxford: Clarendon Press.

HART, O., AND J. MOORE (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755–785.

——— (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119–1158.

HERMALIN, B., AND M. KATZ (1991): "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency," *Econometrica*, 59, 1735–1753.

HOLMSTROM, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340.

KOLMOGOROV, A. N., AND S. V. FOMIN (1970): *Introductory Real Analysis*. Englewood Cliffs, N. J.: Prentice-Hall.

LEGROS, P., AND H. MATSUSHIMA (1991): "Efficiency in Partnerships," *Journal of Economic Theory*, 55, 296–322.

LEGROS, P., AND S. MATTHEWS (1993): "Efficient and Nearly-Efficient Partnerships," *Review of Economic Studies*, 60, 599–611.

MASKIN, E., AND J. MOORE (1999): "Implementation and Renegotiation," *Review of Economic Studies*, 66, 39–56.

MASKIN, E., AND J. TIROLE (1999): "Two Remarks on the Property-Rights Literature," *Review of Economic Studies*, 66, 139–150.

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," forthcoming in *Econometrica*.

MILGROM, P., AND C. SHANNON (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

MIRRLEES, J. (1971): "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38, 175–208.

MOORE, J. (1990): "Implementation in Environments with Complete Information," in *Advances in Economic Theory*, ed. by J.-J. Laffont. Cambridge: Cambridge University Press.

MOORE, J., AND R. REPULLO (1988): "Subgame Perfect Implementation," *Econometrica*, 56, 1191–1220.

NOLDEKE, G., AND K. SCHMIDT (1995): "Option Contracts and Renegotiation: A Solution to the Hold-Up Problem," *RAND Journal of Economics*, 26, 163–179.

SEGAL, I. (1999): "Complexity and Renegotiation: A Foundation for Incomplete Contracts," *Review of Economic Studies*, 66, 57–82.

SPIER, K., AND M. D. WHINSTON (1995): "On the Efficiency of Privately Stipulated Damages for Breach of Contract: Entry Barriers, Reliance, and Renegotiation," *RAND Journal of Economics*, 26, 180–202.