Games **2013**, *4*, 347-366; doi:10.3390/g4030347

OPEN ACCESS

games

ISSN 2073-4336

www.mdpi.com/journal/games

Article

The Renegotiation-Proofness Principle and Costly Renegotiation

James R. Brennan and Joel Watson *

Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA

* Author to whom correspondence should be addressed; E-Mail: jwatson@ucsd.edu.

Received: 7 April 2013; in revised form: 9 June 2013 / Accepted: 9 July 2013 / Published: 25 July 2013

Abstract: We study contracting and costly renegotiation in settings of complete, but unverifiable information, using the mechanism-design approach. We show how renegotiation activity is best modeled in the fundamentals of the mechanism-design framework, so that noncontractibility of renegotiation amounts to a constraint on the problem. We formalize and clarify the Renegotiation-Proofness Principle (RPP), which states that any state-contingent payoff vector that is implementable in an environment with renegotiation can also be implemented by a mechanism in which renegotiation does not occur in equilibrium. We observe that the RPP is *not* valid in some settings. However, we prove a general monotonicity result that confirms the RPP's message about renegotiation opportunities having negative consequences. Our monotonicity theorem states that, as the costs of renegotiation increase, the set of implementable state-contingent payoffs becomes larger.

Keywords: contract theory; bargaining; negotiation; mechanism design

1. Introduction

In a real contractual relationship, the contracting parties may write a contract that directs an external enforcer (the court, for example) on how to interact with them later. However, the parties may not be jointly committed to their initial contract, to the extent that technology and the legal system allow them to renegotiate it before the external enforcer intervenes. For example, the initial contract may specify an externally-enforced outcome that is *ex post* inefficient in some contingency. If this contingency arose, the parties would then have the incentive to renegotiate the contractually-specified outcome. This "*ex post* renegotiation" can have important implications for the attainment of the parties' contractual goals.

On the theoretical side, researchers have studied renegotiation by finding ways of incorporating it into the standard mechanism-design framework, which has become an important tool for studying contract.[1]

[1] See [1] for the basic "mechanism design with *ex post* renegotiation" methodology. The mechanism-design approach is used in applied settings by, among others, [2–8].

The typical mechanism-design model specifies sets of *states* and *outcomes*, as well as the parties' preferences over the outcomes in each state. For contractual settings of *complete, but unverifiable information*, the state is interpreted as an event that the contracting parties jointly observe (but the external enforcer does not observe) and the outcome is interpreted as verifiable items that the external enforcer compels. A contract is then a game form (mechanism) that the external enforcer forces the parties to play after they observe the state. To represent *ex post* renegotiation, theorists commonly embed it in the specification of preferences; that is, they define payoffs in terms of what renegotiation would yield.

In the mechanism-design literature, the effect of renegotiation on contracting is represented by the Renegotiation-Proofness Principle (RPP), which states that any state-contingent outcome that can be implemented in an environment with renegotiation can also be implemented by a mechanism in which renegotiation does not occur in equilibrium. The RPP plays two roles in the literature. First, it helps simplify the analysis of implementation by allowing theorists to focus on so-called "renegotiation-proof" mechanisms. Second, the RPP captures the idea that the opportunity for parties to renegotiate imposes a constraint on the set of implementable state-contingent outcomes. In other words, the RPP conveys the message that the opportunity to renegotiate is bad for contracting. Unfortunately, the RPP is a sketchy "result." It is applied in various modeling exercises, but it has not been stated or validated in a general form.[2] Further, it is usually invoked without any formal modeling of renegotiation.

In this paper, we clarify the Renegotiation-Proofness Principle by explicitly modeling renegotiation in settings of complete, but unverifiable, information. We study general renegotiation costs, which includes the cases of free renegotiation and barred renegotiation (the two cases commonly studied in the literature) and everything between. In particular, we are motivated by the observation that, in reality, renegotiation can be moderately costly. For example, to alter a contract, parties may require the services of an attorney who charges them a fee.

Our modeling exercise has three components. First, we develop a method of incorporating costly renegotiation into the mechanism-design framework. In our model, renegotiation activity is defined as a component of the "outcome." That renegotiation cannot be controlled by the external enforcer is represented as a constraint on the outcomes that may be specified in a mechanism. In other words, renegotiation activity and its costs are designated in the fundamentals of the mechanism-design program, so that noncontractibility of renegotiation translates into a *constraint on the class of mechanisms*.[3]

Our formulation reveals that statements about "whether a mechanism can replicate what the players could achieve by renegotiating" may lack meaning. When renegotiation is properly incorporated into the mechanism-design framework, it is trivially true that a mechanism can replicate renegotiation; this is because renegotiation activity is specified in the outcome. Rather, the important issue is whether the effects of renegotiation can be achieved using some other available technology. For example, suppose the contracting parties would throw away resources in the renegotiation process. Then, one must ask whether they could arrange to throw away the same resources without renegotiating, perhaps motivated by an externally-enforced penalty.

Second, we provide a formal statement of the Renegotiation-Proofness Principle, and we develop the conditions under which it holds. In essence, the RPP involves a comparison between the renegotiation technology and other technologies. Most importantly, we show that the RPP is generally *invalid* in settings of moderate renegotiation costs. That is, in many settings, implementation necessarily involves renegotiation in equilibrium. We show that the RPP does hold when renegotiation is free, which is the common case that theorists have studied, but is not necessarily the most realistic case. These results highlight the intuition that underlies the literature's current understanding of the RPP.

[2] [9] and [10] utilize the RPP in models with incomplete information. The RPP fails in Reiche's [11] model.

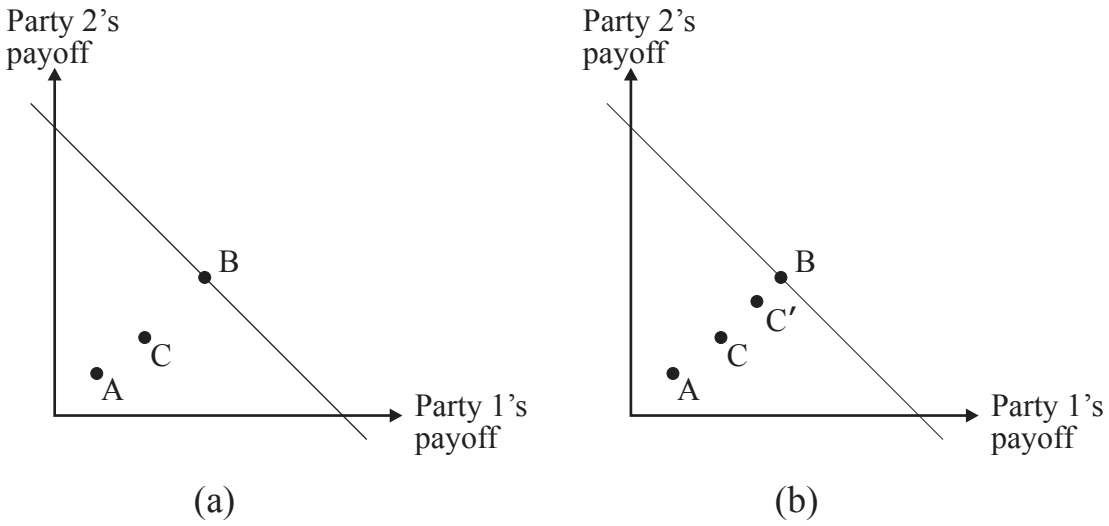
[3] In a related paper, [8] explains why models of mechanism design with *ex post* renegotiation can distort the scope of contracting if they do not explicitly account for the technology of trade. The points we make here are tangential to Watson's critique. While we do not explicitly model productive decisions, our model is consistent with Watson's methodology.

Third, we prove a general monotonicity result: that higher renegotiation costs imply a larger set of implementable state-contingent payoffs. This result elucidates the intuition about the negative effect of renegotiation opportunities that lies at the heart of the RPP. We thus argue that, although the RPP is not always valid, it does suggest a general and useful result about the effect of renegotiation costs.

Our analysis complements the work of [12], who study the implications of costly contracting and renegotiation and who identify real costs associated with the legal system. Further, it complements [8] and [13], who demonstrate the importance of explicitly modeling the technology of trade in a mechanism-design framework, and [14], who take a similar line in the analysis of evidence disclosure.

Before indulging in the details of our modeling exercise, we provide a graphical illustration of our results. For a given state, we can imagine that the set of feasible payoff vectors is the area below the diagonal line in Figure 1(a).

Figure 1. Renegotiation in payoff space for a given state.



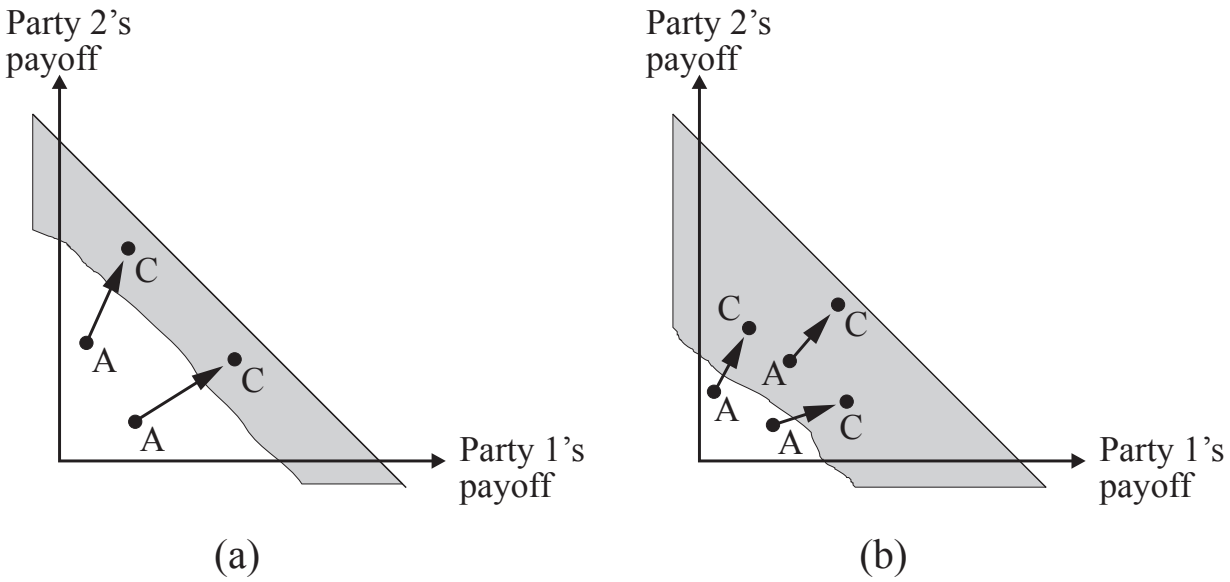
Suppose that, given the parties' initial contract, the external enforcer is poised to make a "public decision" that would, in the absence of renegotiation, lead to the payoff vector labeled A in Figure 1(a). Since this public decision is not *ex post* efficient, the parties would like to renegotiate their contract and specify a different decision for the external enforcer. If renegotiation is free, then they would presumably select a public decision that puts them on the efficient frontier at, say, point B. Anticipating this renegotiation, the parties could have designed their initial contract to achieve B without renegotiation. The RPP will be valid in such an environment, because, with the redesigned contract, the players will not renegotiate away from payoff vector B (otherwise, at least one player would fare worse and would, therefore, be unwilling to renegotiate).

The story changes if renegotiation entails some moderate cost. The parties would like to renegotiate away from the public decision that would yield payoff A, but the cost of renegotiation may keep them from achieving payoff B. Instead, they obtain a payoff vector such as that labeled C in Figure 1(a). To be concrete, suppose that recontracting requires paying an attorney, who charges a fraction of the surplus that would be created by altering the public decision. Anticipating that their initial contract will be renegotiated, the parties expect payoff C in the given state. Interestingly, the RPP does not hold in this case, because the only way of achieving payoff vector C is with renegotiation. For example, consider what would happen if the parties wrote a contract that instructed the external enforcer to take the efficient public decision and also force the parties to throw away some resources, so that the final payoff vector is C. The parties would renegotiate this contract to avoid throwing away resources; but renegotiation

requires a payment to the attorney, and so, in the end, the parties realize payoff vector C′ shown in Figure 1(b).

Figure 2 portrays our monotonicity result.

Figure 2. Illustration of the monotonicity result.



The shaded region in Figure 2(a) depicts the set of possible payoff vectors that can be achieved in a given state. These payoffs include the effect of renegotiation. Specifically, the parties' initial contract may specify a public decision that would achieve a payoff labeled A; the parties renegotiate to obtain payoff C. This diagram represents, in payoff terms, a fixed renegotiation technology—which determines what the parties will do in the renegotiation phase (including paying an attorney) in each state. Figure 2(b) depicts a costlier renegotiation technology than that shown in Figure 2(a). For example, the attorney's fee in case (b) is a larger fraction of the surplus than it is in case (a). The higher cost is represented by shorter arrows from points A to points C. As illustrated, the set of achievable payoff vectors is larger in case (b), which implies a wider scope for implementing state-contingent payoff vectors.

The rest of this paper provides the details of our modeling exercise. Section 2 describes the contractual setting and develops our method of incorporating *ex post* renegotiation into a mechanism-design model. Section 3 states the Renegotiation-Proofness Principle and characterizes the conditions under which it is valid. Section 4 presents our monotonicity result. Section 5 contains a novel example that illustrates the failure of the RPP and shows that, to align investment-incentives, inefficient values may be required in some states.

2. The Contract Environment

We analyze a class of contractual relationships with external enforcement and complete, but unverifiable, information. Two players interact over four phases of time, as follows. In Phase 1, the players write a contract. The contract directs an external enforcer (a court, for example) on how to intervene in the fourth phase, as a function of verifiable information.

In Phase 2, the *state of the relationship* is realized and is commonly observed by the players. The state may be determined by actions that the players take and/or it may be influenced by random events;

we do not model how the state is determined. The state is not verifiable to the external enforcer. The set of possible states is denoted Θ.

In Phase 3, the players make decisions that are verifiable, but are not payoff relevant. For example, the parties may send messages to the external enforcer; the messages have no direct effect on the players' payoffs.

In Phase 4, productive decisions and external enforcement occur. Interaction in this phase defines the *physical outcome*, which is denoted d. Let D be the set of feasible physical outcomes; we assume D is independent of the realized state. Fourth-phase interaction is constrained by technology and the institutional environment, as discussed later in this section.

The player's payoffs from contractual interaction are defined by the function $u\colon D \times \Theta \to \mathbf{R}^2$, where $u(d, \theta)$ is the payoff vector of physical outcome d in state $\theta \in \Theta$. In forming their contract in the first phase, the players' goal is to *implement* a particular physical outcome—and, thus, a payoff vector—as a function of the state.

In much of our analysis, it will be convenient to work with physical outcomes in terms of their implied state-contingent payoffs. That is, instead of dealing with d directly, we deal with function $u(d, \cdot)$, which gives the payoff vector from d as a function of the state. Thus, we use the term *payoff outcome* (*outcome*, for short) for any mapping from the set of states to the set of payoff vectors. Let W be the set of payoff outcomes associated with the set of physical outcomes:

$$\begin{aligned} W \quad \equiv \quad & \{w\colon \Theta \to \mathbf{R}^2 \mid \text{ there exists } d \in D, \\ & \text{such that } w(\theta) = u(d, \theta) \text{ for all } \theta \in \Theta\}. \end{aligned}$$

2.1. Standard Mechanism Design Analysis

To this point, our description of contractual relationships has not indicated the precise structure of interaction in the third and fourth phases. A fully-specified model of any particular contractual relationship requires a more detailed account. In the standard mechanism-design approach to studying contract, theorists simplify the analysis by making (sometimes implicitly) three assumptions. First, theorists assume that all of the payoff relevant aspects of D are either directly verifiable to the external enforcer or are directly controlled by the external enforcer. Second, whenever they assume that an aspect of D is not directly controlled by the external enforcer, theorists assume that the external enforcer can compel fines or transfers that can be used to levy arbitrarily harsh punishments on individual players. For example, it is common in the literature to assume that, in Phase 4, the parties make verifiable decisions about (a) whether to trade an intermediate good and (b) a monetary transfer from one party to the other. After these decisions, the external enforcer can compel transfers or fines.[4]

These two assumptions motivate theorists to treat d as a "public decision"—that is, a decision made by the external enforcer. In other words, the players' verifiable actions are modeled, for all intents and purposes, as *alienable* (taken out of the players' hands). This assumption is commonly justified by noting that "forcing contracts" can be used to compel the players to take any specific action as a function of other verifiable events, in particular, the messages sent to the external enforcer in the Phase 3.

The third assumption theorists usually make is that the technology of interaction in Phase 3 is unrestricted. To be more precise, it is assumed that players have the opportunity to send and receive messages sequentially and simultaneously. It will be enough to assume that, in Phase 3, the players simultaneously and independently send messages to the external enforcer. Let $M \equiv M_1 \times M_2$ be the set of possible message profiles.

These assumptions justify treating the players' Phase 1 contracting problem as a standard mechanism-design problem, with fundamentals given by $\langle \Theta, D, u \rangle$. The contract formed by the players in Phase 1

[4] In the literature, the specific mechanics of trade and enforcement are usually not explicitly studied. See [8] regarding the appropriate modeling of trade and enforcement.

specifies a *game form* (M, g), which is defined by a message space, M, and an externally enforced mapping, $g : M \to D$. The game form can be equivalently written in terms of payoff outcomes, as (M, f), where $f : M \to W$ is defined by $f(m) \equiv u(g(m), \cdot)$ for every $m \in M$. Then, for every state θ, the game form defines an *induced game*, $\langle M, f(\cdot)(\theta)\rangle$, that is played in Phase 3. The game form is called a *mechanism*. Note that we focus on static mechanisms.

Behavior in the Phase 3 is modeled by Nash equilibrium, so the players' contracting problem is one of "Nash implementation" ([15]). A mechanism, along with a selection of equilibrium in each state, implies a *state-contingent value function*, $v : \Theta \to \mathbf{R}^2$, that gives the resulting payoff vector as a function of the state.[5] The revelation principle ([17,18] justifies constraining attention to (a) *direct revelation mechanisms*, where players send reports of the state (so $M \equiv \Theta^2$), and (b) truthful reporting equilibrium, where each player honestly reports the state.

For a direct revelation mechanism (Θ^2, f), we write $f(\theta_1, \theta_2) = w^{\theta_1\theta_2}$ for the payoff outcome of the message game when player 1 sends message θ_1 and player 2 reports θ_2. Thus, when the players send reports θ_1 and θ_2 in state θ, the payoff is $w^{\theta_1\theta_2}(\theta)$. Whether truthful reporting is a Nash equilibrium in every state is captured by:

Definition 1: *Mechanism* (Θ^2, f) *is called* **incentive-compatible** *if, for each* $\theta \in \Theta$ *and all* $\theta'_1, \theta'_2 \in \Theta$, $w_1^{\theta\theta}(\theta) \geq w_1^{\theta'_1\theta}(\theta)$ *and* $w_2^{\theta\theta}(\theta) \geq w_2^{\theta\theta'_2}(\theta)$.

Implementation of a state-contingent value function is defined by:

Definition 2: *Mechanism* (Θ^2, f) *is said to* **implement** *value function* v *if it is incentive-compatible and* $v(\theta) = w^{\theta\theta}(\theta)$ *for all* $\theta \in \Theta$.

Furthermore, we say that value function v is *implementable* if a mechanism exists that implements it. We let V denote the set of all implementable value functions.[6] That is:

$$V \equiv \{v : \Theta \to \mathbf{R}^2 \mid \text{ there is a mechanism } (\Theta^2, f) \text{ that implements } v\}.$$

2.2. The Standard Model of Mechanism Design with Ex Post Renegotiation

A key ingredient of the standard mechanism-design model is that the parties are committed to their chosen mechanism and, thus, to the public decision their messages prescribe. However, most real enforcement institutions do not allow such commitment. For example, it is not technologically feasible for a public court to administer an arbitrarily chosen mechanism. In addition, public courts do not enforce contracts verbatim, so, even if full enforcement of mechanisms were possible, institutional constraints limit the contracts that parties can choose.

Recognizing the real commitment problem, theorists have been led to study renegotiation of contracts. The following story illustrates the possibility of "*ex post* renegotiation."[7] Suppose the players agree to mechanism (Θ^2, g) in Phase 1; state θ is realized in Phase 2; and, in Phase 3, the parties send reports of the state, θ_1 and θ_2. Then, just after the reports are sent and assuming that the players each know the other's report, the players realize that the external enforcer is poised to select public decision $d = g(\theta_1, \theta_2)$ in Phase 4. However, if d is *inefficient* in the realized state—that is, there is another public decision, d', such that $u(d', \theta) > u(d, \theta)$—then the players have an incentive to alter what their mechanism prescribes, instructing the external enforcer to make a different public decision than d.[8] That

[5] Where there are multiple equilibria in an induced game, the players' contract specifies which equilibrium they will play. [16] discusses the issue of multiple equilibria and describes several general definitions of implementability.

[6] The set V is well-defined in any contractual environment, because any mechanism that prescribes a constant physical outcome, regardless of messages, will trivially implement some value function.

[7] Papers in the "mechanism design with *ex post* renegotiation" literature include [1] and [7].

[8] For vectors, "$\geq$" means greater than or equal in both dimensions, and "$>$" means a strict inequality in at least one dimension.

is, the players can substitute some d' for $g(\theta_1, \theta_2)$, just before the external enforcer makes the public decision. If the players can renegotiate in this way, then they cannot commit to their original mechanism; instead, the original mechanism sets the default point for possible renegotiation contingent on the state.[9]

To model *ex post* renegotiation, theorists have used the following clever trick (following [1]). They specify a "renegotiation function" $h : D \times \Theta \to D$, which gives the renegotiated public decision, d', as a function of the actual state, θ, and the decision, d, that the original mechanism prescribes. That is, $d' = h(d, \theta)$. Assuming that the players rationally anticipate renegotiation, this changes the induced game that they play in Phase 3. Rather than the message profile (θ_1, θ_2) yielding payoff vector

$$u(g(\theta_1, \theta_2), \theta)$$

in state θ, as would be the case without renegotiation, this message profile instead yields payoff vector

$$u(h(g(\theta_1, \theta_2), \theta), \theta).$$

Accordingly, one can redefine the utility function to incorporate the renegotiation activity, h. For every state θ and every public decision d, define:

$$\hat{u}(d, \theta) \equiv u(h(d, \theta), \theta).$$

Then, a setting of mechanism design with *ex post* renegotiation, given by $\langle \Theta, D, u, h \rangle$, is equivalent to the standard mechanism-design problem defined by $\langle \Theta, D, \hat{u} \rangle$.

2.3. A More General Approach

The analytical method described in the previous subsection is shorthand for explicitly modeling renegotiation activity. Although it has been useful in the literature, this shorthand method is not well-suited for studying settings in which renegotiation entails a cost. For example, suppose renegotiation requires payments to an attorney. To study this setting using the literature's trick, one would have to define utility function $\hat{u}$ to embody these payments. However, then we have a payoff-relevant aspect of strategic interaction (transfers made to a third party) that is not specified in the fundamentals of the mechanism-design problem. This contradicts the mechanism-design ideal—that all payoff-relevant aspects of interaction are included in the "outcome." As a result, we cannot represent formally, for example, whether the external enforcer can take an action that achieves the same payoffs that could be reached with renegotiation. This is critical, because we need to compare, in payoff terms, the renegotiation technology with other technologies. Note as well that the renegotiation activity cannot be put in terms of the literature's "h" function without including its payoff-relevant aspects in the definition of the "outcome." Thus, it is not obvious even how to define whether renegotiation occurs.

The proper way of analyzing renegotiation while adhering to the mechanism-design framework is to (i) represent renegotiation activity in the fundamentals of the mechanism-design problem and (ii) represent noncontractibility of renegotiation behavior as a constraint on mechanism design. To make sense of this exercise, let us start from scratch by reviewing the building blocks of mechanism design.

Broadly speaking, in any design setting, the players are to interact in a "grand game." Some aspects of the grand game are open to design, whereas other aspects are exogenously determined. In the standard mechanism-design model, the set of "public decisions" (outcomes) and their payoff consequences are exogenously given, but it is assumed that one can design a game form that arbitrarily maps to this set. Thus, to the extent that the public decisions represent individual actions of the players, these actions are assumed to be verifiable.

[9] When trade decisions are explicitly modeled, it is more appropriate to call this "interim renegotiation;" see [8]. There may also be "ex ante renegotiation"—occurring before players send messages—which we discuss in the Conclusion, but do not study here.

Renegotiation activity must be specified in the grand game as well. Because it generally is payoff relevant, we adhere to convention by including it in the specification of the outcome. In principle, an outcome could specify both a bargaining protocol (the manner in which renegotiation takes place) and the players' actual behavior in this protocol. The key assumption is that renegotiation activity is non-contractible, so that the players cannot commit to how they will renegotiate.[10] In other words, actual renegotiation behavior must constitute an equilibrium, and this will depend on what occurred earlier in the game.[11]

To represent the renegotiation opportunity, we suppose that interaction in Phase 4 of the contractual relationship consists of *renegotiation activity* followed by a *public action* that the external enforcer compels. Let us denote the public action as x, and let X be the set of possible public actions. By way of interpretation, the external authority either directly controls x, or x is verifiable and can be compelled by the threat of external punishment.

At the beginning of Phase 4, there is a *default public action*, $\underline{x}$, that the players know will be chosen by the external enforcer if they do not renegotiate it. That is, $\underline{x}$ is the public action specified by the players' contract for the given message profile from Phase 3. We summarize renegotiation activity in terms of (i) the renegotiated public action, $x \in X$, and (ii) the vector of transfers and expenditures made by the players during the renegotiation process, which we denote by t. Assume that $t \in \mathbf{R}^2_-$, where

$$\mathbf{R}^2_- \equiv \{(t_1, t_2) \in \mathbf{R}^2 \mid t_1 + t_2 \leq 0\}.$$

Here, t_i is the transfer to player i. Note that the total transfer is nonpositive; if it is negative, then the players' jointly have a renegotiation expenditure.

As an example, an outcome of renegotiation may be "Each player pays \$200 to an attorney, who modifies their contract so that 'trade nothing' is put in place of 'trade 68 bushels of wheat at \$15 per bushel'." This specification of renegotiation activity is represented by (x, t), where $x =$ 'trade nothing' and $t = (-200, -200)$.

We use the term *minimal renegotiation activity* to mean that the default public action is not renegotiated. Specifically, if $\underline{x}$ is the default public action, then minimal renegotiation activity is the case of $(x, t) = (\underline{x}, \underline{t})$, where $\underline{t} \equiv (0, 0)$. This concept will be important for stating the Renegotiation-Proofness Principle in Section 3.

Renegotiation activity is non-contractible, due to either technological or institutional limitations, and therefore, the external enforcer does not control it. However, at the end of Phase 4, the enforcer compels x. Failure to renegotiate implies $x = \underline{x}$, so that the enforcer selects the default public action. Since the players know the state when they renegotiate, the renegotiation activity is described by some function, $\gamma : \Theta \to X \times \mathbf{R}^2_-$, that represents how (x, t) is conditioned on the state, θ. Let Γ be the set of all such functions. Furthermore, we write $\gamma(\theta) = (\gamma_x(\theta), \gamma_t(\theta))$, where $\gamma_x : \Theta \to X$ and $\gamma_t : \Theta \to \mathbf{R}^2_-$.

The physical outcome in Phase 4 is $d = (\underline{x}, \gamma)$, the default public action paired with the renegotiation activity. Thus, we have $D \equiv X \times \Gamma$. We assume that payoffs are additive in renegotiation expenditures. That is, there is a function $\tilde{u} : X \times \Theta \to \mathbf{R}^2$, such that, in every state $\theta \in \Theta$ and for each outcome $d = (\underline{x}, \gamma)$, the payoff vector is

$$u(d, \theta) \equiv \tilde{u}(\gamma_x(\theta), \theta) + \gamma_t(\theta).$$

Thus, the players' payoffs depend only on the state, the public action and any transfers and expenditures made during renegotiation.

Here is a condition on public actions and payoffs that will play a role in whether the Renegotiation-Proofness Principle holds:

[10] Noncontractibility may be due to the external enforcer refusing to condition on the renegotiation behavior or it could be that the renegotiation activity is unverifiable apart from whatever it is that the players agree on.

[11] Standard models of renegotiation in the literature typically assume that the bargaining protocol is fixed. Some, such as [19], suppose that parameters of the bargaining protocol can be designed. Our model allows for both settings. We elaborate in footnote 14 below.

Definition 3: *The contractual setting is said to be* **comprehensive** *if, for any state, $\theta \in \Theta$, and any $(x,t) \in X \times \mathbf{R}^2_-$, there exists a public action, $\underline{x}' \in X$, such that $\tilde{u}(\underline{x}',\theta) = \tilde{u}(x,\theta) + t$.*

This condition means that, for every state, any payoff vector that can be achieved through renegotiation can also be achieved with minimal renegotiation activity, with a suitably chosen default public action.[12]

As an example, suppose the default public action $\underline{x}$ is "trade 50 units of the intermediate good at \$20 per unit" and the decision x^* is "trade 62 units of the intermediate good at \$19 per unit." Further suppose that x^* is an efficient public action in state θ, meaning that x^* maximizes $\tilde{u}_1(x,\theta) + \tilde{u}_2(x,\theta)$ by choice of x. Let $\tilde{u}(\underline{x},\theta) = (240, 350)$ and $\tilde{u}(x^*,\theta) = (300, 410)$. Imagine that $\underline{x}$ is specified as the default public action. In state θ, the players may renegotiate to select x^*, with each player making an expenditure of 30, yielding the payoff vector $(270, 380)$. The expenditures may be money paid to an attorney whose services are required to alter a contract. Comprehensiveness would require the existence of another public action, $\underline{x}'$, such that $\tilde{u}(\underline{x}',\theta) = (270, 380)$, a payoff which then could be achieved with minimal renegotiation activity $(\underline{x}', \underline{t})$. For instance, $\underline{x}'$ might be "trade 62 units of the intermediate good at \$19 per unit and each donate 30 to charity." Thus, a fairly broad range of contractible items is necessary for comprehensiveness.

Standard mechanism-design analysis can be employed for setting $\langle \Theta, D, u \rangle$, except that there are now constraints on γ. These constraints fall into two categories. First, there may be institutional constraints. We represent these as feasibility restrictions on the renegotiation activity, as a function of the state. Specifically, in state θ and with default public action $\underline{x}$, the players' renegotiation activity is restricted to some set $Y(\underline{x},\theta) \subset X \times \Gamma$. We are especially interested in how Y represents restrictions that are due to intrinsic costs of renegotiation—time spent bargaining and modifying the contract, payments made to attorneys, and so on. For example, renegotiating the default decision "sell 600 bushels of wheat at \$10 per bushel" to another public action, "sell 500 bushels at \$11 each", may require a nonneglgble expenditure.

The second constraint on γ is behavioral: it must be consistent with an appropriate theory of bargaining behavior. In other words, the selection of an element in Y depends on one's theory of negotiation. At this point, we do not adopt any particular bargaining theory, but we assume that the bargaining theory identifies a single element of $Y(\underline{x},\theta)$ for every state θ and every default public action $\underline{x}$. Importantly, note that this selection is a function of both θ and $\underline{x}$, which are given at the beginning of Phase 4. Thus, there exists a function $y\colon X \times \Theta \to X \times \mathbf{R}^2_-$, such that

$$y(\underline{x},\theta) = (y_x(\underline{x},\theta), y_t(\underline{x},\theta)) = (x,t) \in Y(\underline{x},\theta)$$

is the outcome of renegotiation in state θ, given default public action $\underline{x}$. We call y the *renegotiation function*.[13]

Our analysis hereinafter takes function y as fundamental. Keep in mind that this function represents both the constraints of Y and the theory of negotiation. For notational ease, we sometimes write $y^{\underline{x}} \equiv y(\underline{x},\cdot)$. For a specific default public action $\underline{x}$ and a state θ, we speak of $\tilde{u}(\underline{x},\theta)$ as the *disagreement payoff*, and we call

$$u((\underline{x}, y^{\underline{x}}),\theta) = \tilde{u}(y_x(\underline{x},\theta),\theta) + y_t(\underline{x},\theta)$$

the *renegotiated payoff*.

In summary, the institutional and behavioral restrictions imply that the physical outcome in Phase 4 must be an element of the subset of D given by

$$D^* \equiv \{(\underline{x},\gamma) \in D \mid \gamma(\theta) = y(\underline{x},\theta) \text{ for every } \theta \in \Theta\}.$$

[12] It may be natural to require that any default public action may be renegotiated to any other public action, for a sufficient renegotiation expenditure. Though intuitively appealing, this is not necessary for our analysis.

[13] One can think of y as specifying the equilibrium outcome of some non-cooperative bargaining game or the outcome prescribed by a cooperative-game theory of negotiation. See [20] for discussion of related modeling issues.

The set D^* is precisely the set of contractible physical outcomes for the setting of mechanism design with *ex post* renegotiation. Thus, the setting, $\langle \Theta, D, u, y \rangle$, is equivalent to the standard mechanism-design problem defined by $\langle \Theta, D^*, u \rangle$. In words, because some aspects of interaction in Phase 4 are not controlled by the external enforcer, implementation is constrained by the theory of how these aspects are resolved. The physical outcomes that are consistent with the renegotiation theory are simply a subset of the set of all possible physical outcomes.[14]

As with the basic model of mechanism design, we can define the set of payoff outcomes as follows:

$$\begin{aligned} W \equiv \{ & w \colon \Theta \to \mathbf{R}^2 \mid \text{ there exists } \underline{x} \in X \text{ such that} \\ & w(\theta) = u((\underline{x}, y^{\underline{x}}), \theta) \text{ for all } \theta \in \Theta \}. \end{aligned}$$

We let $\hat{W}^\theta$ be the set of payoff outcomes that have minimal renegotiation activity in state θ:

$$\begin{aligned} \hat{W}^\theta \equiv \{ & w \colon \Theta \to \mathbf{R}^2 \mid \text{ there exists } \underline{x} \in X, \text{ such that} \\ & w(\theta') = u((\underline{x}, y^{\underline{x}}), \theta') \text{ for all } \theta' \in \Theta, \text{ and } y(\underline{x}, \theta) = (\underline{x}, \underline{t}) \}. \end{aligned}$$

2.4. A Class of Costly Renegotiation Functions

In this subsection, we describe a parameterized class of renegotiation functions to represent the idea that the players cannot extract all of the potential surplus from changing the contractually-specified public action. For example, suppose that public action $\underline{x}$ is about to be enforced in state θ. If there is another public action, $x \in X$, for which

$$\tilde{u}_1(x, \theta) + \tilde{u}_2(x, \theta) > \tilde{u}_1(\underline{x}, \theta) + \tilde{u}_2(\underline{x}, \theta),$$

then the players would like to renegotiate the default public action. However, this may require an expenditure. We suppose that the players must jointly pay α in order to re-specify the public action. Further, they must pay fraction β of the surplus created by changing the public action. A transfer between the players is unrestricted. The costs impose constraints on the feasible renegotiation activity, which are given by:

$$Y(\underline{x}, \theta) = \left\{ (x, t) \;\middle|\; t_1 + t_2 \leq -\beta \min \left\{ 0, \sum_{i=1,2} [\tilde{u}_i(x, \theta) - \tilde{u}_i(\underline{x}, \theta)] \right\} - \alpha I_{x \neq \underline{x}} \right\}, \tag{1}$$

Here, $I_{x \neq \underline{x}}$ is the indicator function that equals zero when $x = \underline{x}$ and equals one otherwise.

Constrained to Y in state θ and with default public action $\underline{x}$, the players' available renegotiation surplus is:

$$r(\underline{x}, \theta) = \max_{(x,t) \in Y(\underline{x}, \theta)} \sum_{i=1,2} \left[\tilde{u}_i(x, \theta) + t_i - \tilde{u}_i(\underline{x}, \theta) \right]. \tag{2}$$

We suppose that the players choose x to achieve this surplus and that they select transfers to split the surplus according to fixed bargaining weights π_1 and π_2, where $\pi_1, \pi_2 \geq 0$ and $\pi_1 + \pi_2 = 1$. That is, $y(\underline{x}, \theta)$ solves (2) and gives player i the payoff $\tilde{u}_i(\underline{x}, \theta) + \pi_i r(\underline{x}, \theta)$.

[14] We noted above that this model can incorporate "renegotiation design." There are two ways of doing this. First, one can consider a parameterized class of renegotiation functions and evaluate how the predictions of the model vary with this parameter. This is suitable for settings in which the parameter (such as bargaining weights) is set at the beginning and is independent of the state and messages, as in [19]. One can also look at renegotiation functions that represent different bargaining protocols/weights in different states. Second, one can assume that the default public action is multidimensional and includes a bargaining parameter, which then allows for the bargaining protocol/weights to be conditioned on the message profile.

The renegotiation theory embodied in Equations 1 and 2 is quite flexible. For example, we obtain the standard "free-renegotiation" case (equivalent to the setting described in Subsection 2.2) by specifying $\alpha = \beta = 0$. We also use the term "costless renegotiation" to describe this case. If $\beta = 0$, but $\alpha > 0$, so that

$$Y(\underline{x}, \theta) = \{(x, t) \in Y \mid x = \underline{x} \text{ or } t_1 + t_2 \leq -\alpha\},$$

then we have the case in which the players must pay only a lump sum, α, to make any change in the specified public action. In the extreme, we could have $\alpha = \infty$, which means renegotiation is not possible. Finally, when $\alpha = 0$, but $\beta > 0$, then there are only proportional costs of renegotiation activity. In this case, the players renegotiate to the *ex post* efficient public action, but they lose some fraction of the surplus to transaction costs.

In practical terms, α and β may represent transaction costs that are inherent in the process of negotiation or expenditures that must be paid to third parties. Here is a simple and realistic foundation: for a positive integer Λ, let the state be a vector $\theta = (\theta^1, \theta^2, \ldots, \theta^\Lambda)$, where $\theta^\lambda \in \{0, 1\}$ for $\lambda = 1, 2, \ldots, \Lambda$. Let the public action be a similar vector $x = (x^1, x^2, \ldots, x^\Lambda)$, where $x^\lambda \in \{0, 1\}$ for $\lambda = 1, 2, \ldots, \Lambda$. In this setting, productive interaction is multidimensional. Each dimension may be interpreted as an individual component of a good or service that one player provides to the other, where $x^\lambda = 1$ means that component λ is included ("on"). The state determines the set of desired components, meaning that it is optimal to set $x^\lambda = \theta^\lambda$. Let the players' joint value of the public action be $\sum_{\lambda \in \Lambda}(1 - \left|\theta^\lambda - x^\lambda\right|)$.

Suppose that the legal institution requires contract modifications to be drafted and recorded in proper legal format, so that the players must employ the services of an attorney. It takes time and effort for an attorney to prepare the document, and this cost is increasing in the number of components that the attorney has to specify. Specifically, the attorney's cost of providing the service consists of a fixed cost, α, and a cost of $\beta \in (0, 1)$ for each modified component. Assume that the market for attorneys is competitive, so the players jointly pay $\alpha + \delta\beta$, where δ is the number of components modified. Clearly, the players want to modify every mismatched component (where $x^\lambda \neq \theta^\lambda$) and will obtain exactly the payoffs described earlier in this subsection.

3. The Renegotiation-Proofness Principle

In this section, we use the concept of minimal renegotiation to formally evaluate the Renegotiation-Proofness Principle. The first subsection provides a simple, general characterization of the RPP. The second subsection reproduces the characterization in terms of payoff vectors before and after renegotiation.

3.1. Characterization

The following is our formal definition of "renegotiation-proof." Recall that, for any mechanism (Θ^2, f), we let $w^{\theta_1\theta_2} = f(\theta_1, \theta_2)$ denote the payoff outcome when the players send reports θ_1 and θ_2.

Definition 4: *We say that a mechanism* (Θ^2, f) **does not necessitate renegotiation** *if* $w^{\theta\theta} \in \hat{W}^\theta$ *for every* $\theta \in \Theta$.

A mechanism necessitates renegotiation if there is a state, θ, in which truthful reports lead to a payoff outcome that requires non-minimal renegotiation activity in state θ.

Recall that V denotes the set of implementable value functions. Let V^{NR} be the set value functions that are implemented by mechanisms that do not necessitate renegotiation. We express the Renegotiation-Proofness Principle in terms of V and V^{NR}.

Definition 5: *The Renegotiation-Proofness Principle is said to be* **valid** *if* $V = V^{\mathrm{NR}}$.

This notion of renegotiation-proofness requires only that minimal renegotiation activity follow truthful reports.[15]

Our first theorem is a straightforward characterization of the conditions under which the RPP is valid. First, we have another definition.

Definition 6: *Given a comprehensive contractual setting, renegotiation function* y *is said to have the* **terminal property** *if for every* $\underline{x} \in X$ *and every* $\theta \in \Theta$*, there exists a public action* $\underline{x}' \in X$*, such that* $u((\underline{x}, y^{\underline{x}}), \theta) = \tilde{u}(\underline{x}', \theta)$ *and* $y(\underline{x}', \theta) = (\underline{x}', \underline{t})$.

In other words, the terminal property holds if, for every state θ and every default public action $\underline{x}$, one can find a public action $\underline{x}'$ that (i) yields the same payoff vector as would follow from renegotiation of $\underline{x}$ and (ii) would itself not be renegotiated. Note that a failure of comprehensiveness would make condition (i) infeasible.

Theorem 1: *The Renegotiation-Proofness Principle is valid if and only if the contractual setting is comprehensive and the renegotiation function has the terminal property.*

The proof of Theorem 1 is straightforward and makes use of the following result.

Lemma 1: *Suppose the contractual setting is comprehensive and the renegotiation function has the terminal property. Then, for every* $w \in W$ *and every* $\theta \in \Theta$*, there is a payoff outcome* $w^{\theta\theta} \in \hat{W}^{\theta}$*, such that* $w^{\theta\theta}(\theta) = w(\theta)$.

Proof: Because $w \in W$, there exists some $\underline{x} \in X$, such that $w(\theta) = u((\underline{x}, y^{\underline{x}}), \theta)$. Comprehensiveness and the terminal property imply the existence of $\underline{x}' \in X$, such that $u((\underline{x}, y^{\underline{x}}), \theta) = \tilde{u}(\underline{x}', \theta)$ and $y(\underline{x}', \theta) = (\underline{x}', \underline{t})$. Define payoff outcome $w^{\theta\theta}$ in association with $\underline{x}'$, so that $w^{\theta\theta}(\theta') \equiv u((\underline{x}', y^{\underline{x}'}), \theta')$ for every $\theta' \in \Theta$. Clearly, $w^{\theta\theta}(\theta) = w(\theta)$ and $w^{\theta\theta} \in \hat{W}^{\theta}$. ***Q.E.D.***

Proof of Theorem 1: First, we prove that comprehensiveness and the terminal property imply $V = V^{\mathrm{NR}}$. Clearly, $V^{\mathrm{NR}} \subset V$. Take any $v \in V$. We must show that $v \in V^{\mathrm{NR}}$. Let (Θ^2, f) be a mechanism that implements v and define $w^{\theta_1\theta_2} \equiv f(\theta_1, \theta_2)$ for all $\theta_1, \theta_2 \in \Theta$. For each $\theta \in \Theta$, Lemma 1 implies the existence of $\hat{w}^{\theta} \in \hat{W}^{\theta}$, such that $\hat{w}^{\theta}(\theta) = w^{\theta\theta}(\theta)$. Define a function $f' : \Theta^2 \to W$ so that, for every $\theta \in \Theta$, $f'(\theta, \theta) = \hat{w}^{\theta}(\theta)$ and for each pair $\theta_1, \theta_2 \in \Theta$ with $\theta_1 \neq \theta_2$, $f'(\theta_1, \theta_2) = f(\theta_1, \theta_2)$. By construction, (Θ^2, f') implements v and does not necessitate renegotiation.

Next, we prove that $V = V^{\mathrm{NR}}$ implies the terminal property and, hence, that the contractual setting is comprehensive. Presuming that the terminal property fails, we will find a value function that is an element of V but is not an element of V^{NR}. Because the terminal property does not hold, there is a state θ and a default public action $\underline{x}$, such that for every $\underline{x}' \in X$ satisfying $u((\underline{x}, y^{\underline{x}}), \theta) = \tilde{u}(\underline{x}', \theta)$, it is the case that $y(\underline{x}', \theta) \neq (\underline{x}', \theta)$. Define payoff outcome w in association with $\underline{x}$, so that $w(\theta') \equiv u((\underline{x}, y^{\underline{x}}), \theta')$ for every $\theta' \in \Theta$. Define mechanism (Θ^2, f) so that f is the constant function $f(\theta_1, \theta_2) = w$ for all (θ_1, θ_2). Clearly, this mechanism implements $v \equiv w$, but there is no other mechanism that implements this value function, because there does not exist $w^{\theta\theta} \in \hat{W}^{\theta}$ for which $w^{\theta\theta}(\theta) = w(\theta)$. ***Q.E.D.***

Note that there are two possible reasons for failure of the RPP. First, it may be that the contractual setting is not comprehensive. Failure of comprehensiveness means that there are payoffs that could only be achieved with non-minimal renegotiation activity. Second, it may be that the technology of renegotiation (such as third party charges for contract modifications) is inconsistent with the terminal property.

[15] One can imagine a stronger version of renegotiation-proofness that extends this requirement to out-of-equilibrium contingencies. [6], for example, work with a stronger version.

3.2. Reformulation in Terms of Payoff Vectors

In this subsection, we reformulate Theorem 1 by expressing the comprehensiveness and terminal properties in terms of payoff vectors before and after renegotiation. This formulation is also used in the following section.

The reformulation requires two assumptions to guarantee that we can invert the renegotiation operation. First, we assume that whenever the renegotiated payoff equals the payoff vector that the players would have gotten with minimal renegotiation, then the renegotiation function specifies minimal renegotiation.

Assumption 1: *For every* $\theta \in \Theta$ *and every* $\underline{x} \in X$*, if* $u((\underline{x}, y^{\underline{x}}), \theta) = \tilde{u}(\underline{x}, \theta)$*, then it is the case that* $y(\underline{x}, \theta) = (\underline{x}, \underline{t})$.

This assumption clarifies "minimal renegotiation activity;" players do not engage in non-minimal renegotiation, unless it is to alter payoffs.

Next, we assume that the renegotiation function, y, implies a functional relation between the disagreement payoff and the renegotiated payoff in any state.

Assumption 2: *Fix* $\theta \in \Theta$*. If* $\tilde{u}(\underline{x}, \theta) = \tilde{u}(\underline{x}', \theta)$*, then* $u((\underline{x}, y^{\underline{x}}), \theta) = u((\underline{x}', y^{\underline{x}'}), \theta)$.

We next define a function, H, to represent this relation. Let

$$Q(\theta) \equiv \{z \in \mathbf{R}^2 \mid \text{there exists } \underline{x} \in X \text{ with } z = \tilde{u}(\underline{x}, \theta)\},$$

which is the set of possible disagreement payoffs in state θ. For every $z \in Q(\theta)$, let

$$H(z, \theta) \equiv u((\underline{x}^z, y^{\underline{x}^z}), \theta) = \tilde{u}(y_x(\underline{x}^z, \theta), \theta) + y_t(\underline{x}^z, \theta),$$

where $\underline{x}^z$ is such that $\tilde{u}(\underline{x}^z, \theta) = z$. Thus, if z is the disagreement payoff in state θ, then $H(z, \theta)$ is the renegotiated payoff. Note that comprehensiveness implies that $H(Q(\theta), \theta) \subset Q(\theta)$ for every $\theta \in \Theta$.

The following definition puts comprehensiveness and the terminal property in payoff terms. We thank Joel Sobel for suggesting this formulation.

Definition 7: *We say that* **Sobel's Condition is satisfied** *if, for every* $\theta \in \Theta$ *and every* $z \in Q(\theta)$*,* $H(H(z, \theta), \theta) = H(z, \theta)$.

This condition has a simple interpretation. Consider any state, θ. From a given disagreement payoff z, renegotiation would lead to payoff $z' = H(z, \theta)$. We know, from the comprehensiveness assumption, that there is a public action $\underline{x}' \in X$, such that $z' = \tilde{u}(\underline{x}', \theta)$. When Sobel's condition is satisfied, the default public action, $\underline{x}'$, would not be renegotiated in state θ. The next result follows directly from Theorem 1.

Corollary 1: *Under Assumptions 1 and 2, the Renegotiation-Proofness Principle is valid if and only if Sobel's Condition is satisfied.*

We next provide two corollaries that add to the intuition of Theorem 1 and Corollary 1. The next assumption and those in the next section are stated using function H.

Assumption 3 (Individual Rationality): *For every* $\theta \in \Theta$ *and for every disagreement payoff,* $z \in Q(\theta)$*, it is the case that* $H(z, \theta) \geq z$ *(in the vector sense).*

Assumption 3 is the standard assumption that the players would not accept less than they could get by refusing to renegotiate. Define:

$$V^{\text{Eff}} \equiv \{v \in V \mid \text{there is no state } \theta \in \Theta \text{ for which } \tilde{u}(\underline{x}, \theta) > v(\theta) \text{ for some } \underline{x} \in X.\}.$$

Value functions in V^{Eff} yield efficient payoffs in all states.

Corollary 2: *Under Assumptions 1-3,* $V^{\text{Eff}} \subset V^{\text{NR}}$.

In words, any value function that specifies an efficient payoff vector in every state can be implemented with minimal renegotiation in every state. This follows from the fact that, if $z \in Q(\theta)$ is efficient in state θ then, by individual rationality, $H(z, \theta) = z$. Of course, in many important settings, theorists and practitioners are interested in achieving *inefficient* payoffs in some states. The example we provide in Section 5, which illustrates the failure of the RPP, has this flavor.

Our final corollary is useful for discussing the class of renegotiation functions introduced in Subsection 2.4. We use the following definition.

Definition 8: *We say that* **Condition PR (Pareto ranked) is satisfied** *if, for every state* $\theta \in \Theta$*, there is a set* $Z^\theta \subset Q(\theta)$*, such that (i) no two elements of* Z^θ *are Pareto-ranked and (ii) for every* $z \in Q(\theta)$*,* $H(z, \theta) \in Z^\theta \cup \{z\}$.

By "Pareto ranked", we mean that there are vectors $z, z' \in Z^\theta$, such that $z > z'$.

Corollary 3: *Under Assumptions 1-3, Condition PR implies Sobel's Condition and, hence, validity of the Renegotiation-Proofness Principle.*

To understand this corollary, consider any state θ and a vector $z \in Q(\theta)$. Condition PR implies that either $H(z, \theta) = z$ or $H(z, \theta) \in Z^\theta$. In the former case, Sobel's Condition clearly holds at point z. In the latter case, part (i) of Condition PR and individual rationality imply Sobel's Condition at point z.

Returning to the class of renegotiation functions described in Subsection 2.4, first consider the special case of $\alpha = \beta = 0$, where renegotiation is frictionless (there are no institutional or technological constraints). In this case, all implementable value functions belong to V^{Eff}, so the RPP is valid. In other words, the RPP is valid in the free-renegotiation setting that is popular in the contract theory literature. This case illustrates the importance of the comprehensiveness assumption, which states that every payoff vector that can be achieved via non-minimal renegotiation can also be reached with minimal renegotiation (by specifying an appropriately chosen public action).

Next consider the case of $\alpha > 0$ and $\beta = 0$. Here, Condition PR holds, and thus, the RPP is again valid. To see this, first note that the players only pay a fixed joint expenditure to alter the default public action in any way. In a given state θ, the players will pay the renegotiation cost if and only if it does not exceed the joint value of changing the public action. It follows that every default decision yielding a payoff that is sufficiently close to the Pareto frontier will not be renegotiated. In this range, $z = H(z, \theta)$. On the other hand, public actions that would yield lower payoffs will be renegotiated, leading to payoffs in the following set:

$$Z^\theta \equiv \left\{ (\phi_1, \phi_2) \,\middle|\, \phi_1 + \phi_2 = \max_{x \in X} [\tilde{u}_1(x, \theta) + \tilde{u}_2(x, \theta)] - \alpha \right\}.$$

Clearly, Z^θ has no Pareto-ranked points and $H(z, \theta) = z$ for every $z \in Z^\theta$. In the special case of $\alpha = \infty$ (where renegotiation is not possible), we have $z = H(z, \theta)$ for every $z \in Q(\theta)$, and so the RPP trivially holds.

Finally, suppose that $\alpha = 0$ and $\beta > 0$. In this important case, Sobel's Condition fails, implying that the RPP is not valid. Here is the intuition behind failure of Sobel's Condition. Suppose that default public action $\underline{x}$ is specified and that it is inefficient in state θ. Let x^* be an efficient public action, which maximizes $\tilde{u}_1(x,\theta) + \tilde{u}_2(x,\theta)$ by choice of x. With the proportional renegotiation cost, the players will renegotiate to select x^*, and they will obtain payoff vector z' given by

$$z' = \tilde{u}(\underline{x},\theta) + \pi(1-\beta)\sum_{i=1,2}\left[\tilde{u}_i(x^*,\theta) - \tilde{u}_i(\underline{x},\theta)\right].$$

The comprehensiveness assumption implies the existence of another public action, $\underline{x}'$, satisfying $\tilde{u}(\underline{x}',\theta) = z'$. If $\underline{x}'$ were specified as the default public action, the players would renegotiate to select x^* and, factoring in the proportional cost, the players obtain the payoff vector,

$$\tilde{u}(\underline{x}',\theta) + \pi(1-\beta)\sum_{i=1,2}\left[\tilde{u}_i(x^*,\theta) - \tilde{u}_i(\underline{x}',\theta)\right],$$

which strictly exceeds z'. Thus, a mechanism that yields payoff z' in state θ necessitates renegotiation.

4. Monotonicity

In this section, we evaluate the RPP's underlying theme that renegotiation opportunities constrain the set of implementable value functions. We show that this insight is more general than is the RPP itself. In particular, in the contracting environments we study, the set of implementable value functions is always increasing in the cost of renegotiation. We start with two additional assumptions on the renegotiation function, stated in terms of H. These assumptions are satisfied by standard theories of bargaining, including the class of renegotiation functions described in Section 2.4.

Assumption 4 (Continuity): *For every state θ, $H(\cdot,\theta)$ is continuous (in z).*

Assumption 5 (Monotonicity): *Fix any state $\theta \in \Theta$ and any two disagreement payoff vectors, $z, z' \in Q(\theta)$. For any $i = 1,2$ and $j \neq i$, if $z_i \geq z'_i$ and $z_j = z'_j$, then $H_j(z,\theta) \leq H_j(z',\theta)$.*

In words, Assumption 5 states that, if the disagreement payoff shifts in player i's favor, then player j's renegotiated payoff weakly decreases. See [21] for a discussion of this type of assumption in bargaining theory.

To formally state our result, we compare different renegotiation functions on the basis of their implied renegotiation cost. Because the renegotiated payoff depends on the renegotiation function, we now explicitly identify y as a parameter in function H, by writing $H(\cdot,\cdot;y)$.

Definition 9: *Renegotiation function $\hat{y}$* **represents higher renegotiation costs** *than does function y if, for every state $\theta \in \Theta$ and every $z \in Q(\theta)$, it is the case that $H(z,\theta;y) \geq H(z,\theta;\hat{y})$.*

In other words, one renegotiation function is costlier than is another if, in every state and for every public action, the renegotiated payoff vector is weakly lower under the costlier renegotiation activity.

To state our main result, we write the set of implementable value functions as $V(y)$, which makes explicit the dependence of the implementable set on the renegotiation function.

Theorem 2: *Suppose Assumptions 1–5 hold. If $\hat{y}$ represents higher renegotiation costs than does y, then $V(y) \subset V(\hat{y})$.*

That is, any increase in the cost of renegotiation widens the scope of implementability.

To illustrate Theorem 2, we use the class of renegotiation functions developed in Subsection 2.4. It is easy to verify that, for this class of renegotiation functions, the renegotiated payoff decreases in

parameters α and β. More formally, suppose y is defined by parameters α and β, $\hat{y}$ is defined by parameters $\hat{\alpha}$ and $\hat{\beta}$ and assume that $\hat{\alpha} \geq \alpha$ and $\hat{\beta} \geq \beta$. Then, $\hat{y}$ represents higher renegotiation costs than does y and, thus, $V(y) \subset V(\hat{y})$.

We prove Theorem 2 with the help of two lemmas. Note that we shall now write the set of payoff outcomes as $W(y)$, to make explicit its dependence on the renegotiation function.

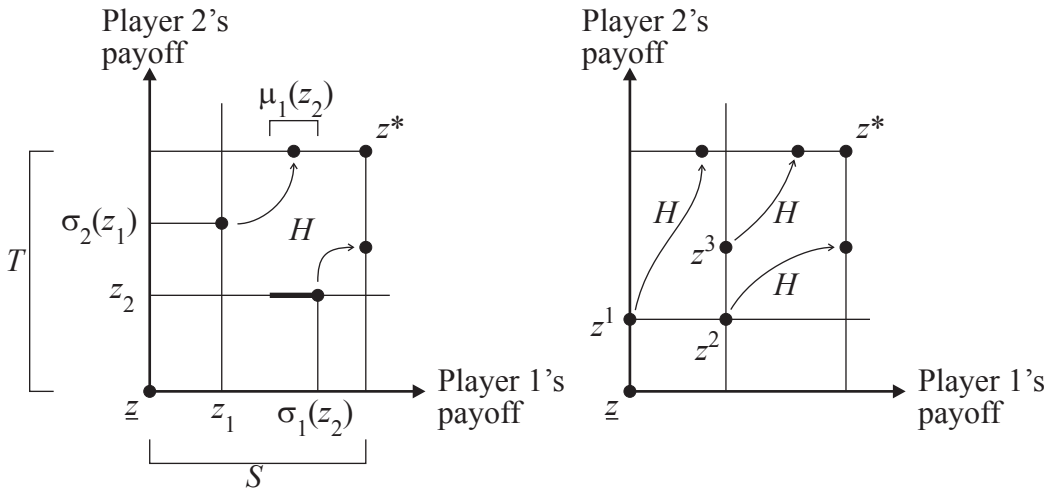
Lemma 2: *If $\hat{y}$ represents higher renegotiation costs than does y, then, for every $w \in W(y)$, there exists $\hat{w} \in W(\hat{y})$ such that $w(\theta) \geq \hat{w}(\theta)$ for each $\theta \in \Theta$.*

Proof: Define $\hat{w}$ using the same default public actions that are specified to define w. In each state, every default public action leads to lower renegotiated payoffs under $\hat{y}$ than under y, by Definition 9. *Q.E.D.*

Lemma 3: *If $\hat{y}$ represents higher renegotiation costs than does y, then it is the case that $H(Q(\theta), \theta; y) \subset H(Q(\theta), \theta; \hat{y})$ for every $\theta \in \Theta$.*

Proof: Fix a state, θ, and a vector of renegotiated payoffs, $z^* \in H(Q(\theta), \theta; y)$, and let the vector of disagreement payoffs, $\underline{z} \in Q(\theta)$, satisfy $H(\underline{z}, \theta; y) = z^*$. We will show that there is another disagreement payoff vector, $\hat{z} \in Q(\theta)$, such that $H(\hat{z}, \theta; \hat{y}) = z^*$. Figure 3 illustrates our construction.

Figure 3. Components of the monotonicity proof.



Let

$$S \equiv \{z_1 \in \mathbf{R} \mid \underline{z}_1 \leq z_1 \leq z_1^*\},$$

and let

$$T \equiv \{z_2 \in \mathbf{R} \mid \underline{z}_2 \leq z_2 \leq z_2^*\}.$$

Note that $S \times T \subset Q(\theta)$ by the comprehensiveness assumption. Additionally, define correspondence $\mu_2 \colon S \Rightarrow T$ as:

$$\mu_2(z_1) \equiv \{z_2 \in T \mid H_2((z_1, z_2), \theta; \hat{y}) = z_2^*\}.$$

Define $\mu_1 : T \Rightarrow S$ analogously. It is easy to verify that correspondences μ_1 and μ_2 are well-defined and compact-valued. (Both facts rely on continuity of H and the individual-rationality and monotonicity assumptions.) Next, define $\sigma_1 : T \to S$ and $\sigma_2 : S \to T$ by $\sigma_1(z_2) \equiv \max \mu_1(z_2)$ and $\sigma_2(z_1) \equiv \max \mu_2(z_1)$. Both are well-defined given the properties of μ_1 and μ_2.

Construct a sequence of disagreement payoffs, $\{z^k\}$, as follows. First, let $z^1 = (\underline{z}_1, \sigma_2(\underline{z}_1))$ and let $z^2 = (\sigma_1(z_2^1), z_2^1)$. Proceeding inductively, for any odd k, let $z^k = (z_1^{k-1}, \sigma_2(z_1^{k-1}))$, and for any even k, let $z^k = (\sigma_1(z_2^{k-1}), z_2^{k-1})$. The right picture in Figure 3 depicts the first three elements of the sequence $\{z^k\}$. Since $S \times T$ is compact and since $\{z^k\}$ is increasing—that is, the sequences $\{z_1^k\}$ and $\{z_2^k\}$ are each increasing—we know that $\{z^k\}$ converges to some point $\hat{z} \in S \times T$.

Consider subsequences $\{o^n\} \equiv \{z^1, z^3, z^5, z^7, ...\}$ and $\{e^n\} \equiv \{z^2, z^4, z^6, z^8, ...\}$. Note that, since they are subsequences of $\{z^k\}$, $\{o^n\}$ and $\{e^n\}$ converge to $\hat{z}$. Furthermore, by continuity of H, $\{H(o^n, \theta; \hat{y})\}$ and $\{H(e^n, \theta; \hat{y})\}$ must converge to $H(\hat{z}, \theta; \hat{y})$. In addition, we have that $\{H(o^n, \theta; \hat{y})\} \subset S \times \{z_2^*\}$ and $\{H(e^n, \theta; \hat{y})\} \subset \{z_1^*\} \times T$. Because sets $S \times \{z_2^*\}$ and $\{z_1^*\} \times T$ are closed, we know that $H(\hat{z}, \theta; \hat{y})$ is in the intersection of these two sets, which means that $H(\hat{z}, \theta; \hat{y}) = z^*$. We conclude that $z^* \in H(Q(\theta), \theta; \hat{y})$. *Q.E.D.*

Proof of Theorem 2: For any $v \in V(y)$ and any mechanism (Θ^2, f) that implements it, we can easily find another mechanism, $(\Theta^2, \hat{f})$, that implements v when $\hat{y}$ is the renegotiation function instead. Letting $w^{\theta_1\theta_2} \equiv f(\theta_1, \theta_2)$ for all $\theta_1, \theta_2 \in \Theta$, define $\hat{f}$ as follows. First, note that $w^{\theta\theta}(\theta) \in H(Q(\theta), \theta; y)$ for every θ, which further implies (using Lemma 3) the existence of another outcome $\hat{w}^{\theta\theta} \in W(\hat{y})$, such that $\hat{w}^{\theta\theta}(\theta) = w^{\theta\theta}(\theta)$. Define $\hat{f}(\theta, \theta) \equiv \hat{w}^{\theta\theta}$. Second, consider any pair $\theta_1, \theta_2 \in \Theta$, such that $\theta_1 \neq \theta_2$. For this message profile, define $\hat{f}(\theta_1, \theta_2) \equiv \hat{w}$ for that outcome $\hat{w}$ identified in Lemma 2 with $w = f(\theta_1, \theta_2)$. Clearly, honest reporting is an equilibrium in every state, with mechanism, $(\Theta^2, \hat{f})$, under renegotiation function, $\hat{y}$. This mechanism implements v by construction. Thus, $V(y) \subset V(\hat{y})$. *Q.E.D.*

5. An Example

In this section, we present an example to illustrate how the RPP fails in some settings. The example involves a hold-up problem in a bilateral contractual relationship, where the parties' ability to renegotiate interferes with their incentives to make relationship-specific investments.[16] We show that, with moderate renegotiation costs, there is a contractually-specified mechanism that induces investment and implements a particular value function. Importantly, this value function can *only* be implemented by a mechanism that necessitates renegotiation.

Two risk-neutral parties, a buyer (player 1) and a seller (player 2) contract to trade one unit of an intermediate good. After forming their initial contract, the players simultaneously and independently make private investment decisions—each choosing between high and low effort. Contingent on their investment choices, the players pay private effort costs or obtain private benefits. To be precise, if both parties choose high effort, then each pays three immediately. If both choose low effort, then each pays nothing. Finally, if player i exerts high effort, while player j exerts low effort, then player i pays three, whereas player j obtains a private gain of 4.2. Think of high effort as a reliance investment. The cost of three measures a party's opportunity cost of making this investment, whereas the gain of 4.2 reflects the benefit of selfishly using the other party's investment. The players do not observe each others' effort choices.

The investments influence the state of the relationship, which is the buyer's value of the intermediate good. In particular, if both parties exert high effort, then the buyer's value will be 20 with probability $.9$, and it will be 10 with probability $.1$. If either party exerts low effort, then the buyer's value will be 10 for sure. The realization of the buyer's value, θ, is commonly observed by the players, though it is not verifiable to the enforcement authority (the court). We assume that the seller's cost of delivery is zero.[17]

[16] Notable early papers in the hold-up literature include [22–25] and [26]. Our example here is along the lines of the example in [12].

[17] Thus, player 1's investment is an "own investment," whereas player 2's is a "cross investment" (which [3] call "cooperative").

The public action, x, has two components, k and c, with $k \in \{\mathrm{T}, \mathrm{N}\}$ and $c = (c_1, c_2) \in \mathbf{R}^2_-$. The variable, k, indicates whether the intermediate good is traded; "T" means trade. For $i = 1, 2$, c_i is a court-enforced transfer to player i. If $c_1 + c_2 < 0$, then the court imposes a penalty. The payoffs from the public action are given by $\tilde{u}(x, \theta)$, which is defined as follows: $\tilde{u}(\mathrm{T}, 20) = (20, 0) + c$, $\tilde{u}(\mathrm{N}, 20) = (0, 0) + c$, $\tilde{u}(\mathrm{T}, 10) = (10, 0) + c$ and $\tilde{u}(\mathrm{N}, 10) = (0, 0) + c$. Note that player 1 (the buyer) obtains his value of the intermediate good if trade occurs, plus his transfer; player 2 obtains his transfer. The costs and benefits incurred at the investment phase are sunk and not included here.

Suppose that the parties must employ an attorney to alter the public action, and the attorney charges them a fee equal to the minimum of 10 and three-fifths of the gain from doing this. Thus, if in state θ the players want to renegotiate from default decision $\underline{x} = (\underline{k}, \underline{c})$ to public action $x = (k, c)$, then they must pay

$$\eta(x, \underline{x}, \theta) \equiv \min\{10, (.6)[\tilde{u}_1(x, \theta) + \tilde{u}_2(x, \theta) - \tilde{u}_1(\underline{x}, \theta) - \tilde{u}_2(\underline{x}, \theta)]\}.$$

The set of feasible renegotiation activity is, therefore,

$$Y(\underline{x}, \theta) \equiv \{(x, t) \mid t_1 + t_2 \leq \eta(x, \underline{x}, \theta)\}.$$

Regarding the renegotiation function, y, we suppose that the players maximize their joint value and equally divide the surplus of renegotiation; that is, they have equal bargaining weights.[18]

Clearly, efficiency requires both players to exert high effort in the investment phase and for the intermediate good to be traded. If the parties do not give any money away, this would yield to them a joint payoff of

$$-3 - 3 + (.9)(20) + (.1)(10) = 13.$$

Note that the players can get no more than 11.2 if only one of them invests high, and they get only 10 if neither invests high.

Value function $v\colon \Theta \to \mathbf{R}^2$ gives the payoff vector from the public action and renegotiation activity; it does not include the private costs and benefits incurred in the investment phase. The players' contractual objective is to implement a value function that achieves the highest joint payoff, once the investment-phase gains and losses are factored in.

For the players to each have the incentive to invest high, they must implement a value function that satisfies

$$-3 + (.9)v_i(20) + (.1)v_i(10) \geq 4.2 + v_i(10)$$

for $i = 1, 2$. The left side of this inequality is player i's expected payoff from the investment phase if both players invest high, whereas the right side is what player i would get by deviating to invest low. Rearranging terms, this is $v_i(20) - v_i(10) \geq 8$. Clearly, no efficient value function has this property.

High effort can only be achieved with a value function that represents an inefficient outcome in state 10. Define function v^* by $v^*(20) = (10, 10)$ and $v^*(10) = (2, 2)$. It is easy to verify that the following contractual mechanism implements value function v^*. The players send to the court reports of the buyer's value. If the message profile is $(20, 20)$—that is, both players report value 20—then the default decision is trade ($k = \mathrm{T}$), and a transfer of 10 from the buyer to the seller ($c_1 = -10$, $c_2 = 10$). Note that this would not be renegotiated. For the other three message profiles, the default decision is no trade ($k = \mathrm{N}$) and no transfer ($c_1 = c_2 = 0$). Note that, in state 10, this would be renegotiated to specify trade; however, attorney's fees reduce the renegotiation surplus from 10 to four, so each player gets two.

By writing the contract just described and, thus, implementing value function v^*, the players achieve a joint payoff of

$$-3 - 3 + (.9)(20) + (.1)(4) = 12.4,$$

which is the highest payoff that can be reached. Importantly, v^* is the optimal value function; increasing $v_i(10)$ would annul investment incentives, whereas decreasing $v_i(10)$ just reduces joint value.

[18] One can easily verify that our technical assumptions are satisfied in this example.

Furthermore, there is no way of implementing v^* without renegotiation. That is, there is no mechanism that implements v^* and does not necessitate renegotiation.

6. Conclusion

We have developed a modeling framework to explicitly capture the constraints on implementation imposed by the non-contractible opportunity for parties to renegotiate. Our framework has allowed us to rigorously characterize conditions under which the Renegotiation-Proofness Principle is valid. We have demonstrated that the RPP generally will not apply in settings of moderate renegotiation costs. However, our monotonicity result authenticates the intuition underlying the RPP that renegotiation opportunities have negative consequences. Our result establishes that the implementable set increases with the cost of renegotiation.

Our analysis emphasizes the need to properly incorporate institutional and technological constraints into mechanism-design analysis ([8,13]). We encourage further research in this direction. For instance, it may be instructive to study costly contracting and recontracting over multiple dimensions, where there are non-linearities in the contracting costs due to bundling opportunities and sequencing. As in [27,28] and [12], it may be useful to defer contracting on some items to a later date.

We wish to point out that our analysis can be applied in a straightforward manner to the case of *ex ante* renegotiation, where the players have the opportunity to renegotiate following realization of the state, but before messages are sent to the external enforcer. One could easily restate the RPP for *ex ante* renegotiation, which would be valid if and only if *ex ante* versions of comprehensiveness and the terminal property hold.[19] Likewise, our monotonicity result applies in the *ex ante* case.[20]

Acknowledgments

UC San Diego. Internet: http://weber.ucsd.edu/∼jwatson/. We thank Matthew Jackson, David Levine, Ilya Segal, Joel Sobel, Chris Woodruff and colleagues at UC San Diego for their comments. The research reported herein was supported by NSF grants SES-0095207 and SES-1227527.

References

1. Maskin, E.; Moore, J. Implementation and renegotiation. *Rev. Econ. Stud.* **1999**, *66*, 39-56.
2. Bernheim, B.D.; Whinston, M. Incomplete contracts and strategic ambiguity. *Am. Econ. Rev.* **1998**, *88*, 902-932.
3. Che, Y.-K.; Hausch, D. Cooperative investments and the value of contracting. *Am. Econ. Rev.* **1999**, *89*, 125-147.
4. Jackson, M.; Palfrey, T. Voluntary implementation. *J. Econ. Theory* **2001**, *98*, 1-25.
5. Nöldeke, G.; Schmidt, K. Option contracts and renegotiation: A solution to the hold-up problem. *Rand* **1995**, *26*, 163-179.
6. Rubinstein, A.; Wolinsky, A. Renegotiation-proof implementation and time preferences. *Am. Econ. Rev.* **1992**, *82*, 600-614.
7. Segal, I.; Whinston, M. The mirrlees approach to mechanism design with renegotiation (with applications to hold-up and risk-sharing). *Econometrica* **2002**, *70*, 1-45.
8. Watson, J. Contract, mechanism design, and technological detail. *Econometrica* **2007**, *75*, 55-81.

[19] Specifically, this requires defining a new renegotiation mapping to select a renegotiated mechanism, given the state and the "default mechanism," with the possibility of expenditures at the *ex ante* stage. Then, the RPP will be valid whenever the *ex ante* renegotiation function has the appropriate terminal property.

[20] Holding the *ex post* renegotiation technology and environment—defined by states, preferences and physical outcomes—fixed, if the costs of *ex ante* renegotiation activity increase, then so must the set of implementable value functions.

9. Hart, O.; Tirole, J. Contract renegotiation and coasian dynamics. *Rev. Econ. Stud.* **1988**, *55*, 509-540.
10. Dewatripont, M. Commitment through renegotiation-proof contracts with third parties. *Rev. Econ. Stud.* **1988**, *55*, 377-389.
11. Reiche, S. Foundation of Incomplete contracting in a model of asymmetric information and renegotiation. manuscript, 1999.
12. Schwartz, A.; Watson, J. The law and economics of costly contracting. *J.L. Econ. & Org.* **2004**, *20*, 2-31.
13. Buzard, K.; Watson, J. Contract, renegotiation, and hold up: Results on the technology of trade and investment. *Theoretical Economics* **2012**, *7*, 283-322.
14. Bull, J.; Watson, J. Evidence disclosure and verifiability. *J. Econ. Theory* **2004**, *118*, 1-31.
15. Maskin, E. Nash Equilibrium and welfare optimality. *Rev. Econ. Stud.* **1999**, *66*, 23-38.
16. Jackson, M. A crash course in implementation theory. *Soc. Choice Welfare* **2001**, *18*, 655-708.
17. Green, J.; Laffont, J.-J. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica* **1977**, *45*, 427-438.
18. Myerson, R. Incentive compatibility and the bargaining problem. *Econometrica* **1979**, *47*, 61-73.
19. Aghion, P.; Dewatripont, M.; Rey, P. Renegotiation design with unverifiable information, *Econometrica* **1994**, *62*, 257-282.
20. Watson, J. Contract and game theory: Basic concepts for settings of complete information. manuscript, June 2001.
21. Thomson, W. Monotonicity of bargaining solutions with respect to the disagreement point. *J. Econ. Theory* **1987**, *42*, 50-58.
22. Grossman, S.; Hart, O. The Costs and benefits of ownership: A theory of vertical and lateral integration. *J. Polit. Econ.* **1986**, *94*, 691-719.
23. Grout, P.A. Investment and wages in the absence of binding contracts: A Nash bargaining approach. *Econometrica* **1984**, *52*, 449-460.
24. Klein, B.; Crawford, R.A.; Alchian, A. Vertical integration, appropriable rents, and the competitive contracting process. *J. Law Econ.* **1978**, *21*, 297-326.
25. Williamson, O. Transactions-cost economics: The governance of contractual relations. *J. Law Econ.* **1979**, *22*, 233-261.
26. Hart, O.; Moore, J. Incomplete contracts and renegotiation. 13 *Econometrica* **1988**, *56*, 755-785.
27. Battigalli, P.; Maggi, G. Rigidity, Discretion, and the costs of writing contracts. *Am. Econ. Rev.* **2002**, *92*, 798-817.
28. Battigalli, P.; Maggi, G. Costly contracting in a long-term relationship. *Rand* **2008**, *39*, 352-377,