The Renegotiation-Proofness Principle and Costly Renegotiation

Jim Brennan and Joel Watson*

Preliminary Draft, December 2001
Complete Version, May 2002

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 modelled 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* generally valid. However, we prove a general monotonicity result that confirms the RPP's "renegotiation is bad" message. Our monotonicity theorem establishes that the set of implementable state-contingent payoffs increases with the costs of renegotiation.

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

*University of California, San Diego. Internet: http://weber.ucsd.edu/~jwatson/. We thank Matthew Jackson, Ilya Segal, Joel Sobel, Chris Woodruff, and colleagues at UC San Diego for their comments. The research reported herein is supported by Watson's NSF grant SES-0095207.

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] 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 into 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 renegotiation is bad for contracting. Unfortunately, the RPP is an ambiguous "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,

[1] See Maskin and Moore (1999) for the basic "mechanism design with ex post renegotiation" methodology. The mechanism-design approach is used in applied settings by, among others, Bernheim and Whinston (1998), Che and Hausch (1999), Jackson and Palfrey (2001), Noldeke and Schmidt (1995), Rubinstein and Wolinsky (1992), Segal and Whinston (2002), and Watson (2002).

[2] Hart and Tirole (1988) and Dewatripont (1988) utilize the RPP in models with incomplete information. The RPP fails in Reiche's (1999) 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 commonly-heard statements about "whether a mechanism can replicate what the players could achieve by renegotiating" 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 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 "free-renegotiation" intuition that underlies the literature's current understanding of the RPP.

Third, we prove a general monotonicity result: that higher renegotiation costs imply a larger set of implementable state-contingent payoffs. This result elucidates the "renegotiation is bad" intuition 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 Schwartz and Watson (2001), who study the implications of costly contracting and renegotiation and who identify real costs associated with the legal system. Further, it complements the work of Watson (2002), who demonstrates the importance of explicitly modeling the technology of trade in a mechanism-design framework, and the work of Bull and Watson (2001), who take a similar line in the analysis of evidence disclosure.

Before indulging in the details of our modeling exercise, we provide a graphical

[3]In a related paper, Watson (2002) explains why models of mechanism design with ex post renegotiation are invalid 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.

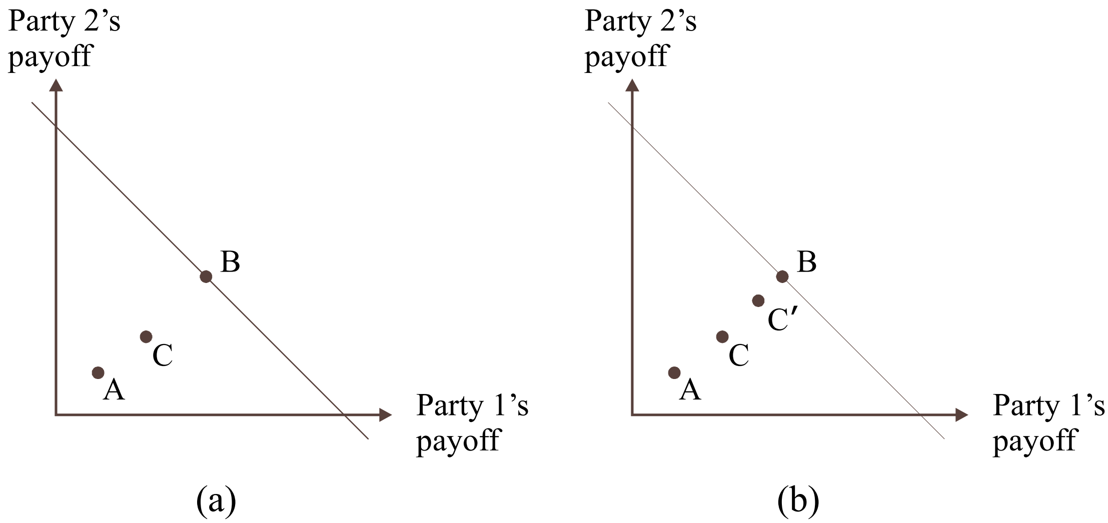

Figure 1: Renegotiation in payoff space for a given state.

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). 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;

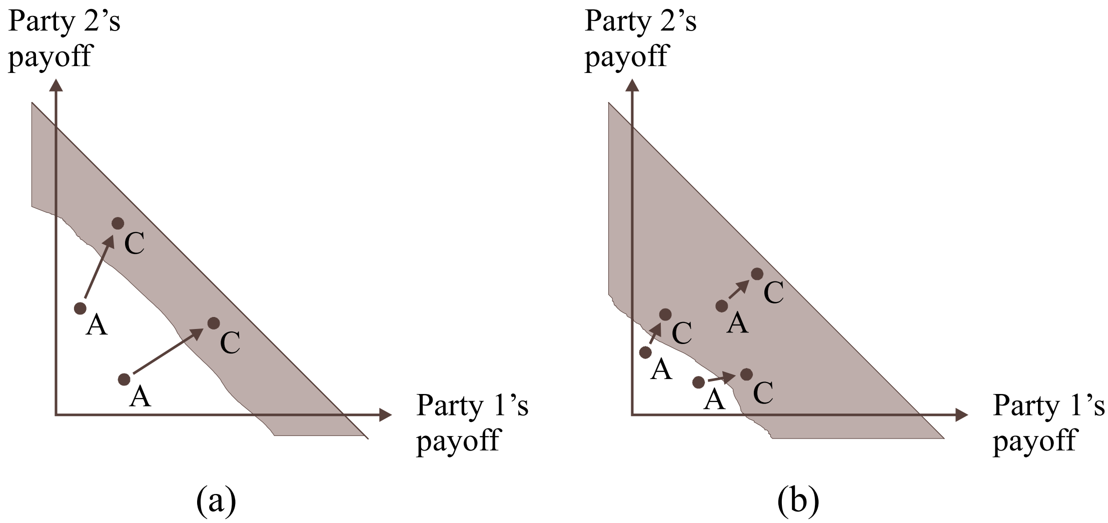

Figure 2: Illustration of the monotonicity result.

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

Figure 2 portrays our 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 labelled 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 third 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 $u : D \times \Theta \to \mathbf{R}^2$. We write $u(d|\theta)$ as 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 the 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 payoffs outcomes associated with the set of physical outcomes:

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

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, made by the external enforcer. In other words, the players' verifiable actions are modelled, 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 particular verifiable action as a function of other verifiable events, such as 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 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 modelled by Nash equilibrium, so the players' contracting problem is one of "Nash implementation" (Maskin 1999). 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 (Green and Laffont 1977; Myerson 1979) 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

[4] In the literature, the specific mechanics of trade and enforcement are usually not explicitly studied. See Watson (2002) regarding the appropriate modeling of trade and enforcement.

[5] Where there are multiple equilibria in an induced game, the players' contract specifies which equilibrium they will play. Jackson (1999) discusses the issue of multiple equilibria and describes several general definitions of implementability.

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: *A mechanism* (Θ^2, f) *is* **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: *A mechanism* (Θ^2, f) **implements** *value function* v *if it is incentive-compatible and* $v(\theta) = w^{\theta\theta}(\theta)$ *for all* $\theta \in \Theta$.

Furthermore, we say that a 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.[7] 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."[8] Suppose the players agree to a mechanism (Θ^2, g) in Phase 1; the 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 make the

[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]See Watson (2002) regarding the appropriate theoretical approach to modeling trade technology and external enforcement.

[8]Papers in the "mechanism design with ex post renegotiation" literature include Maskin and Moore (1999) and Segal and Whinston (2002).

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 $u(d'|\theta) > u(d|\theta)$ (in the vector sense[9])—then the players have an incentive to alter what their mechanism prescribes, instructing the external enforcer to make a different public decision than d. That 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.[10]

To model ex post renegotiation, theorists have used the following clever trick (following Maskin and Moore 1999). 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 the payoff vector

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

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

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

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.

[9]Note that "$>$" means weakly greater for both coordinates and strictly greater for at least one coordinate.

[10]When trade decisions are explicitly modelled, it is more appropriate to call this "interim renegotiation;" see Watson (2002). There may also be "ex ante renegotiation"—occuring before players send messages—which we discuss in the Conclusion but do not study here.

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. To some extent, 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 represent renegotiation activity in the fundamentals of a mechanism-design problem and to represent noncontractibility of renegotiation behavior as a constraint on mechanism design. For example, we could define the "outcome" to include, in addition to those things that the external enforcer can compel, any payments made to an attorney. The contracted game form would then describe (a) how the players are to send messages and (b) the manner in which the players will renegotiate. To capture the idea that the opportunity to renegotiate is noncontractible, we should restrict attention to game forms that contain a fixed "ex post" renegotiation phase, whereby the outcome is a particular function of the renegotiation behavior. With this theoretical foundation, it is trivially true that "a mechanism can replicate renegotiation," because renegotiation is defined as a component of the mechanism.

Rather than modeling renegotiation as a component of the game form, we follow Watson's (2002) lead and develop a variation of this modeling approach in which the renegotiation activity is a component of the outcome. That renegotiation activity is noncontractible will imply a constraint on the set of outcomes that can be specified in a mechanism. This approach lends itself more easily to standard mechanism-design analysis (as does the literature's current trick for the costless renegotiation setting) than does the tack described in the preceding paragraph.

To represent the renegotiation opportunity, we suppose that interaction in Phase 4 of the contractual relationship can be divided into two sub-phases, the resolution of which are given by x and y. The variable x represents what the external enforcer can compel; that is, the external authority either directly controls x, or at least x is verifiable and can be compelled by the threat of external punishment. From this point on, we call x the *public decision*. Think of the public decision as representing trade and transfers. Let X be the set of possible public decisions.

The variable y represents the players' renegotiation activity.[11] It is noncon-

[11] One can think of y as a strategy profile in some non-cooperative renegotiation game that is

tractible, due to either technological or institutional limitations, and, therefore, the external enforcer does not control y. Let Y be the set of all possible renegotiation activity. An element of Y, for instance, 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.'" Since the players know the state when they renegotiate, y is conditioned on θ. We let the function $\gamma : \Theta \to Y$ represent the renegotiation activity as a function of the state, and we let Γ be the set of all such functions.

We next add structure to the x and y components of interaction in Phase 4. Specifically, we assume that there is a *default public decision* $\underline{x}$ that the players know will be chosen by the external enforcer if they fail to renegotiate it. Furthermore, we summarize the (perhaps complicated) renegotiation activity in terms of (a) the renegotiated public decision and (b) transfers and expenditures made by the players during the renegotiation process. Thus, we define $Y \equiv X \times \mathbf{R}^2_-$, 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. We assume that the total transfer is nonpositive; if it is negative then this is the players' joint renegotiation expenditure. We write $y = (x, t)$. Also, given a state-contingent schedule of renegotiation activity γ, we write $\gamma(\theta) = (\gamma_x(\theta), \gamma_t(\theta))$, where $\gamma_x : \Theta \to X$ and $\gamma_t : \Theta \to \mathbf{R}^2_-$.

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

The physical outcome in Phase 4 is defined by a default public decision $\underline{x}$ and a specification of renegotiation activity γ. 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 decision, and any transfers and expenditures made during renegotiation.

We make the following assumption jointly on the set of public decisions and the payoff function:

Assumption 1 (Comprehensiveness): *Take any state $\theta \in \Theta$, any default public decision $\underline{x} \in X$, and any renegotiation activity $(x, t) \in Y$ such that $\tilde{u}(x|\theta) + t \geq \tilde{u}(\underline{x}|\theta)$. Then there exists a public decision $\underline{x}' \in X$ such that $\tilde{u}(\underline{x}'|\theta) = \tilde{u}(x|\theta) + t$.*

played after messages are sent. We use a simpler "cooperative game theory" representation.

This assumption 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 decision). This is assumed for every payoff vector that weakly exceeds (in the vector sense) the payoff from some arbitrary public decision.[12]

As an example, suppose the default public decision $\underline{x}$ 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 decision 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 decision. 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 requires the existence of another public decision $\underline{x}'$ 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." As this example demonstrates, comprehensiveness relies on a fairly broad range of contractible items; the assumption is essential for the Renegotiation-Proofness Principle.

Standard mechanism-design analysis can be employed for the 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 decision $\underline{x}$, the players' renegotiation activity is restricted to some set $\hat{Y}(\underline{x}|\theta) \subset Y$. We are especially interested in how $\hat{Y}$ 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 decision "sell 500 bushels at \$11 each" may require a nonneglible expenditure.

The second constraint on γ is behavioral: γ must be consistent with an appropriate theory of bargaining behavior. In other words, the selection of an element in $\hat{Y}$ depends on one's theory of negotiation. At this point, we will not adopt any particular bargaining theory. However, we do assume that the bargaining theory identifies a single element of $\hat{Y}(\underline{x}|\theta)$ for every state θ and every default public decision $\underline{x}$. We let $y^*(\underline{x}|\theta)$ denote the prediction of the bargaining theory in state θ, given default public decision $\underline{x}$.

Our analysis hereinafter takes $y^*(\cdot|\cdot)$ as fundamental. That is, we suppose there is a function $y^* : X \times \Theta \to Y$ that gives the renegotiation activity as a function

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

of the default public decision and the state. This function implicitly represents the constraints of $\hat{Y}$ as well as the theory of negotiation. We call y^* the *renegotiation function*. Where we need to separate the x and t components of y, we write $y^*(\underline{x}|\theta) \equiv (y^*_x(\underline{x}|\theta), y^*_t(\underline{x}|\theta))$. Also, with regard to a specific default decision $\underline{x}$ and a state θ, we speak of $\tilde{u}(\underline{x}|\theta)$ as the *disagreement payoff* and we call

$$\tilde{u}(y^*_x(\underline{x}|\theta)|\theta) + y^*_t(\underline{x}|\theta)$$

the *renegotiated payoff*.

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

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

The set D^* gives precisely the set of contractible physical outcomes for the setting of mechanism design with ex post renegotiation. Thus, a setting of mechanism design with ex post renegotiation, given by $\langle \Theta, D, u, y^* \rangle$, 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.

2.4 Assumptions on the Renegotiation Function

We make five straightforward assumptions on the renegotiation function y^*. We start by assuming that, whenever the renegotiated payoff equals the payoff vector that the players would have gotten with minimal renegotiation, then the function specifies minimal renegotiation.

Assumption 2: *For every $\theta \in \Theta$ and every $\underline{x} \in X$, if $\tilde{u}(y^*_x(\underline{x}|\theta)|\theta) + y^*_t(\underline{x}|\theta) = \tilde{u}(\underline{x}|\theta)$ 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 3: *Fix any state θ. If $\tilde{u}(\underline{x}|\theta) = \tilde{u}(\underline{x}'|\theta)$, then*

$$\tilde{u}(y^*_x(\underline{x}|\theta)|\theta) + y^*_t(\underline{x}|\theta) = \tilde{u}(y^*_x(\underline{x}'|\theta)|\theta) + y^*_t(\underline{x}'|\theta).$$

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

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

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.[13]

The rest of our assumptions are stated using the function H.

Assumption 4 (Individual Rationality): *In every state $\theta \in \Theta$ and for every disagreement payoff $z \in Q(\theta)$, it is the case that $H(z|\theta) \geq z$ (in the vector sense).*

Assumption 4 is the standard assumption that the players would not accept less than they could get by refusing to renegotiate.

Assumption 5 (Continuity): *For every state θ, $H(\cdot|\theta)$ is continuous (in z).*

Our final assumption is monotonicity of the renegotiated payoff with respect to the disagreement payoff.

Assumption 6 (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 4 states that, if the disagreement payoff shifts in player i's favor, then player j's renegotiated payoff weakly decreases. See Thomson (1987) for a discussion of this type of assumption in bargaining theory.

2.5 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 decision. For example, suppose that public decision $\underline{x}$ is about to be enforced in state θ. If there is another public decision $x \in X$ for which

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

[13] Assumption 3 simplifies the analysis, but versions of our results would hold if we relaxed this assumption.

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

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

Here, $I_{X \setminus \{\underline{x}\}}$ is the indicator function that equals 0 when $x = \underline{x}$ and equals 1 otherwise.

Facing $\hat{Y}$ in state θ and with default public decision $\underline{x}$, the players' available renegotiation surplus is

$$r(\underline{x}|\theta) = \max_{(x,t) \in \hat{Y}(\underline{x}|\theta;\alpha,\beta)} \sum_{i=1}^{2} \left[\tilde{u}_i(x|\theta) + t_i - \tilde{u}_i(\underline{x}|\theta) \right]. \qquad (2)$$

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 $\tilde{u}_i(\underline{x}|\theta) + \pi_i r(\underline{x}|\theta)$. The assumptions stated in the previous subsection all hold.

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

$$\hat{Y}(\underline{x}|\theta;\alpha,0) = \{(x,t) \in Y \mid x = \underline{x} \text{ or } t_1 + t_2 \leq -\alpha\},$$

then we have the case in which the players only must pay a lump sum α to make any change in the specified public decision. 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 decision, 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. As an illustration, consider the example that follows the comprehensiveness assumption in Subsection 2.4. The default public decision $\underline{x}$ is "trade 50 units of the intermediate good at \$20 per unit;" the public decision x^* is "trade 62 units of the intermediate good at \$19 per unit." In state θ, $\tilde{u}(\underline{x}|\theta) = (240, 350)$, whereas $\tilde{u}(x^*|\theta) = (300, 410)$. Suppose that, to change their contractually-specified public decision, the players must obtain the services of an attorney who (because of his bargaining power) extracts 50 percent of the gain. If the default public decision is $\underline{x}$, then the players would renegotiate to select x^* and, subtracting the attorney's share, they obtain the payoff vector $(270, 380)$. Further, consider the public decision $\underline{x}'$: "trade 62 units of the intermediate good at \$19 per unit and each donate 30 to charity." If $\underline{x}'$ were specified as the default public decision, then renegotiation would entail selection of x^* with attorney's fees of 15 for each player. In other words, the attorney helps strike the "charity" line from the players contract, at a fee of 15 per player.

3 The Renegotiation-Proofness Principle

In this section, we use the concept of minimal renegotiation to formally evaluate the Renegotiation-Proofness Principle. Our definitions are written in terms of payoff outcomes. Recall that an arbitrary mechanism-design problem with renegotiation function y^*, $\langle \Theta, D, u, y^* \rangle$, is equivalent to an unconstrained mechanism-design problem $\langle \Theta, D^*, u \rangle$, where D^* is a subset of D. Incorporating the D^* constraint, the set of payoff outcomes is:

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

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

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

Note that, in the definition of $\hat{W}^\theta$, $H(\tilde{u}(\underline{x}|\theta)|\theta) = \tilde{u}(\underline{x}|\theta)$ means that $y^*(\underline{x}|\theta) = (\underline{x}, \underline{t})$.

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 3: *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 4: *The* **Renegotiation-Proofness Principle (RPP)** *is said to be* **valid** *if* $V = V^{\mathrm{NR}}$.[14]

Our first theorem characterizes the conditions under which the RPP is valid. Note that comprehensiveness and individual rationality (Assumptions 1 and 4) imply that, for every $\theta \in \Theta$, $H(Q(\theta)|\theta) \subset Q(\theta)$.

Definition 5: *We say that* **Sobel's Condition**[15] **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 the payoff $z' = H(z|\theta)$. We know, from the comprehensiveness assumption, that there is a public decision $\underline{x}' \in X$ such that $z' = \tilde{u}(\underline{x}'|\theta)$. When Sobel's condition is satisfied, the default public decision $\underline{x}'$ would not be renegotiated in state θ. The condition is essential to the RPP because, if payoff outcome z' were desired in state θ, it could be achieved with minimal renegotiation activity.

Theorem 1: *The Renegotiation-Proofness Principle is valid if and only if Sobel's Condition is satisfied.*

To prove Theorem 1, we use the following result.

Lemma 1: *If Sobel's Condition is satisfied then, for every* $w \in W$ *and every* $\theta \in \Theta$, *there is a payoff outcome* $\hat{w}^\theta \in \hat{W}^\theta$ *such that* $\hat{w}^\theta(\theta) = w(\theta)$.

Proof: That $w \in W$ implies the existence of some $\underline{x} \in X$ such that $w(\theta) = H(\tilde{u}(\underline{x}|\theta)|\theta)$. Comprehensiveness implies the existence of some $\underline{x}' \in X$ such that $\tilde{u}(\underline{x}'|\theta) = w(\theta) \equiv z'$. Sobel's condition then implies that $H(z'|\theta) = z'$. From Assumption 2, we know that $y^*(\underline{x}'|\theta) = (\underline{x}', \underline{t})$. Define payoff outcome $\hat{w}^\theta$ by

$$\hat{w}^\theta(\theta') \equiv H(\tilde{u}(\underline{x}'|\theta')|\theta')$$

for every $\theta' \in \Theta$. Clearly, $\hat{w}^\theta(\theta) = w(\theta)$ and $\hat{w}^\theta \in \hat{W}^\theta$. *Q.E.D.*

Proof of Theorem 1: First we prove that Sobel's condition implies $V = V^{\mathrm{NR}}$. 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

[14] This notion of renegotiation-proofness requires only that minimal renegotiation activity follow truthful reports. One can imagine a stronger version of renegotiation-proofness that extends this requirement to out-of-equilibrium contingencies. Rubenstein and Wolinsky (1992), for example, work with a stronger version.

[15] We thank Joel Sobel for suggesting that we formulate the condition in this way.

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 Sobel's condition. Presuming that Sobel's Condition is not satisfied, we will find a value function that is an element of V but not an element of V^{NR}. Because Sobel's Condition is not satisfied, there is a state θ and a payoff vector $z \in Q(\theta)$ such that

$$H(H(z|\theta)|\theta) \neq H(z|\theta).$$

Let $\underline{x}$ satisfy $\tilde{u}(\underline{x}|\theta) = z$ and define w^* by

$$w^*(\theta') = H(\tilde{u}(\underline{x}|\theta')|\theta'),$$

for every $\theta' \in \Theta$. Let $\underline{x}'$ satisfy $\tilde{u}(\underline{x}'|\theta) = H(z|\theta)$. We then have $w^*(\theta) = H(z|\theta)$ but $y^*(\underline{x}'|\theta) \neq (\underline{x}', \underline{t})$, so non-minimal renegotiation activity occurs in state θ with default public decision $\underline{x}'$. (We are justified in focusing on one public decision $\underline{x}'$ because of Assumptions 2 and 3.) Define a mechanism by $f(\theta_1,\theta_2) \equiv w^*$ for all $\theta_1, \theta_2 \in \Theta$. This mechanism implements the value function $v^* \equiv w^*$, which is not an element of V^{NR}. *Q.E.D.*

We next provide two corollaries that add to the intuition of Theorem 1. Define

$$\begin{aligned} V^{\mathrm{Eff}} \;\equiv\; & \{v \in V \mid \text{there is no state } \theta \in \Theta \text{ for which} \\ & \tilde{u}(\underline{x}|\theta) > v(\theta) \text{ for some } \underline{x} \in X.\}. \end{aligned}$$

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

Corollary 1: $V^{\mathrm{Eff}} \subset V^{\mathrm{NR}}$.

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 (Assumption 4), $H(z|\theta) = z$. 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 demonstrates the failure of the RPP, has this flavor.

Our second corollary is useful for discussing the class of renegotiation functions introduced in Subsection 2.5. We use the following definition.

Definition 6: *We say that* **Condition PR 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 2: *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 the point z. In the latter case, part (i) of Condition PR and individual rationality (Assumption 4) imply Sobel's Condition at point z.

Returning to the class of renegotiation functions described in Subsection 2.5, 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 decision).

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 decision 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 decision. 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 decisions that would yield lower payoffs will be renegotiated, leading to payoffs in the 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 and so the RPP is not valid. Here is the intuition behind failure of Sobel's Condition. Suppose that a default public decision $\underline{x}$ is specified and that it is inefficient in state θ. Let x^* be an efficient public decision, 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 the payoff vector z' given by

$$z'_i = \tilde{u}_i(\underline{x}|\theta) + \pi_i(1 - \beta)[\tilde{u}_i(x^*|\theta) - \tilde{u}_i(\underline{x}|\theta)],$$

for $i = 1, 2$. The comprehensiveness assumption implies the existence of another public decision $\underline{x}'$ satisfying $\tilde{u}(\underline{x}'|\theta) = z'$. If $\underline{x}'$ were specified as the default public decision, the players would renegotiate to select x^* and, factoring in the proportional cost, player i obtains

$$\tilde{u}_i(\underline{x}'|\theta) + \pi_i(1-\beta)[\tilde{u}_i(x^*|\theta) - \tilde{u}_i(\underline{x}'|\theta)],$$

which exceeds z_i', for $i = 1, 2$. 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.

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 the parameter y^* in the function H.

Definition 7: *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 words, one renegotiation function is costlier than is another if, in every state and for every public decision, 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: *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 2.5. It is easy to verify that, for this class of renegotiation functions, the renegotiated payoff decreases in the 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 $\hat{\alpha} \geq \alpha$ and $\hat{\beta} \geq \beta$. Then $\hat{y}^*$ represents higher renegotiation costs than does y^* and thus $V(y^*) \subset V(\hat{y}^*)$.

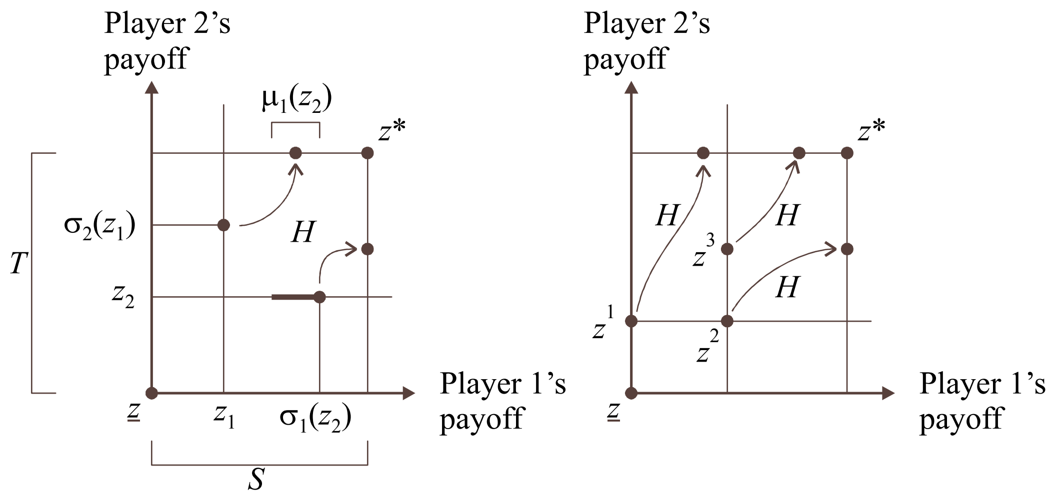

Figure 3: Components of the monotonicity proof.

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 decisions that are specified to define w. In each state, every default public decision leads to lower renegotiated payoffs under $\hat{y}^*$ than under y^*, by the definition of "higher renegotiation cost." *Q.E.D.*

Lemma 3: *If $\hat{y}^*$ represents higher renegotiation costs than does y^*, then $H(Q(\theta)|\theta; y^*) \subset H(Q(\theta)|\theta; \hat{y}^*)$ 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. 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. Also define the correspondence $\mu_2 : S \Rightarrow T$ 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 the 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 the 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 $\hat{w}^{\theta\theta} \in W(y^*)$ 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 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

[16]Notable early papers in the hold-up literature include Klein, Crawford, and Alchian (1978), Williamson (1979), Grout (1984), Grossman and Hart (1986) and Hart and Moore (1988). Our example here is along the lines of the example in Schwartz and Watson (2000).

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 3 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 3 whereas player j obtains a private gain of 4.2. Think of high effort as a reliance investment. The cost of 3 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]

The public decision 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 decision 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.

We assume that renegotiation is costly. More precisely, the parties must employ an attorney to alter the public decision. The attorney charges them a fee equal to the minimum of 10 and six-tenths of the gain from changing the default public decision. In other words, if the gain of changing the default decision exceeds 50/3, then the attorney charges a flat fee of 10; if the gain is less than 50/3, then the attorney charges .6 times the gain. In technical terms, if in state θ the players want to renegotiate

[17] Thus, player 1's investment is an "own investment," whereas player 2's is a "cross investment" (which Che and Hausch (1999) call "cooperative").

from default decision $\underline{x} = (\underline{k}, \underline{c})$ to public decision $x = (k, c)$, 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 thus

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

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.

The value function $v : \Theta \to \mathbf{R}^2$ gives the payoff vector from the public decision 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 v that achieves the highest joint payoff, once the investment-phase gains and losses are factored in.

For example, suppose the players select a mechanism that implements the value function defined by $v(20) = (10, 10)$ and $v(10) = (5, 5)$. Then, if both players exert high effort in the investment phase, they each expect a total payoff of

$$-3 + (.9)(10) + (.1)(5) = 6.5.$$

Interestingly, this value function does not give the players the incentive to invest high; in particular, anticipating that the other player will invest high, a player obtains

$$4.2 + 5 = 9.2$$

by exerting low effort. 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 $i = 1, 2$. Rearranging terms, this is

$$v_i(20) - v_i(10) \geq 8.$$

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

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 ($k = \text{T}$) 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 ($k = \text{N}$) 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 4, so each player gets 2.

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. 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 modelling framework to explicitly capture the constraints on implementation imposed by the noncontractible 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 "renegotiation is bad" intuition that underlies the RPP. Our result establishes that the implementable set increases with the cost of renegotiation. Our results emphasize the need to properly incorporate institutional and technological constraints into mechanism-design analysis, which we unabashedly refer to as "Watson's program" (as this paper continues the line of Watson 2002; see also Bull and Watson 2001). We encourage further research in this direction.

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 an "ex ante Sobel Condition" holds.[19] Likewise, our monotonicity result applies in the ex ante case.[20] An ex ante version of comprehensiveness is required.

We encourage researchers to study real renegotiation costs and further the theory of contract under institutional and technological constraints.

[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 fixed-point 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.

References

[1] Bernheim, B.D. and M. Whinston, "Incomplete Contracts and Strategic Ambiguity," *American Economic Review* 88 (1998): 902-932.

[2] Bull, J. and J. Watson, "Evidence Disclosure and Verifiability," manuscript, 2001.

[3] Che, Y.-K. and D. Hausch, "Cooperative Investments and the Value of Contracting," *American Economic Review* 89 (1999): 125-147.

[4] Dewatripont, M., "Commitment Through Renegotiation-Proof Contracts with Third Parties," *Review of Economic Studies* 55 (1988): 377-389.

[5] Green, J., and J.-J. Laffont, "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods." *Econometrica* 45 (1977): 427-438.

[6] Grossman, S. and O. Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *The Journal of Political Economy* 94 (1986): 691-719.

[7] Grout, P.A.,"Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," *Econometrica* 52 (1984): 449-460.

[8] Hart, O. and J. Moore, "Incomplete Contracts and Renegotiation," *Econometrica* 56 (1988): 755-785.

[9] Hart, O. and J. Tirole, "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies* 55 (1988): 509-540.

[10] Klein, B., R.A. Crawford, and A. Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics* 21 (1978): 297-326.

[11] Jackson, M., "A Crash Course in Implementation Theory," manuscript, July 1999.

[12] Jackson, M. and T. Palfrey, "Voluntary Implementation," *Journal of Economic Theory* 98 (2001): 1-25.

[13] Maskin, E., "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies* 66 (1999): 23-38.

[14] Maskin, E. and J. Moore, "Implementation and Renegotiation," *Review of Economic Studies* 66 (1999): 39-56.

[15] Myerson, R., "Incentive Compatibility and the Bargaining Problem," *Econometrica* 47 (1979): 61-73.

[16] Noldeke, G. and K. Schmidt, "Option Contracts and Renegotiation: A Solution to the Hold-Up Problem," *RAND Journal of Economics* 26 (1995): 163-179.

[17] Reiche, S., "Foundation of Incomplete Contracting in a Model of Asymmetric Information and Renegotiation," manuscript, 1999.

[18] Rubenstein, A. and A. Wolinsky, "Renegotiation-Proof Implementation and Time Preferences," *American Economic Review* 82 (1992): 600-614.

[19] Schwartz, A. and J. Watson, "Economic and Legal Aspects of Costly Recontracting," UC San Diego Economics Working Paper 2000-05, April 2000.

[20] Schwartz, A. and J. Watson, "The Law and Economics of Costly Contracting," UC San Diego Economics Working Paper 2001-20, December 2001.

[21] Segal, I. and M. Whinston, "The Mirrlees Approach to Mechanism Design with Renegotiation (with Applications to Hold-Up and Risk-Sharing)," *Econometrica* 70 (2002): 1-45.

[22] Thomson, W., "Monotonicity of Bargaining Solutions with Respect to the Disagreement Point," *Journal of Economic Theory* 42 (1987): 50-58.

[23] Watson, J., "Contract and Game Theory: Basic Concepts for Settings of Complete Information," manuscript, June 2001.

[24] Watson, J., "Contract, Mechanism Design, and Technological Detail," UC San Diego Economics Working Paper 2002-04, February 2002.

[25] Williamson, O., "Transactions-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics* 22 (1979): 233-61.