Conditioning Institutions and Renegotiation

Garey Ramey and Joel Watson*

March 1999, June 2006

Abstract

We propose a theory of contracting in long-term relationships, emphasizing the role of social institutions in conditioning players' joint selection of equilibria. Players adopt a social conditioning system in order to place boundaries on their recurrent negotiation and thereby sustain a desirable joint selection of equilibrium. Social conventions have value because players cannot freely reinterpret the labels attached to histories, in contrast to labels that the players might assign internally. We present examples of social conventions that are useful for sustaining cooperative interaction. Our model combines an explicit bargaining technology with a renegotiation concept, *coherent equilibrium*, that builds on internal consistency.

1 Introduction

Language, custom, social convention, legal sanction, and other institutions evidently play an important role in shaping long-term contractual relations between trading partners. Current theories of contracting, however, have yet to adequately capture the connections between such institutions and the partners' strategic interaction. The bulk of game-theoretic models abstract entirely from the social backdrop of relationships, while evolutionary models of social conventions largely sidestep explicit consideration of strategic choices. There are several noteworthy exceptions to this portrayal of the literature, emphasizing institutional means of *direct external enforcement*.[1] We

*Ramey: Department of Economics, UC San Diego; Watson: Department of Economics, UC San Diego, and Cowles Foundation, Yale University. Internet: http://weber.ucsd.edu/∼jwatson/. We appreciate the comments of Vincent Crawford, Mark Machina, George Mailath, Steven Matthews, John McMillan, David Pearce, Phil Reny, Joel Sobel, William Thomson, and Chris Woodruff, as well as seminar participants at Harvard/MIT, Johns Hopkins, Michigan, NYU, Penn, Pittsburgh, UC San Diego, and Yale. We thank the National Science Foundation for financial support under grant SBR-9630270.

[1] For example, the work of Bendor and Mookherjee (1991), Ellickson (1991), Kandori (1992), Greif (1993), and Greif, Milgrom, and Weingast (1994) shows how the threat of community sanctions helps partners maintain cooperation. Milgrom, North, and Weingast (1990) and others examine more centralized enforcement institutions. See also North (1993).

suggest, however, that institutions also shape contractual relations through their affect on how parties negotiate — an influence that is equally powerful and ubiquitous.

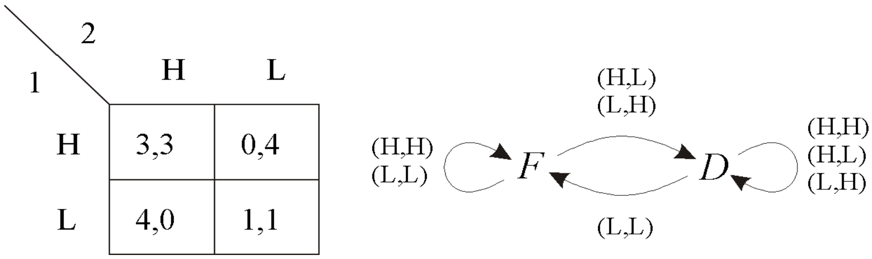
This paper offers a new approach to modeling institutions in a strategic context. In particular, we identify the role that institutions play in structuring long-term relationships by providing the players with a framework for *organizing histories*. This framework, or *conditioning system*, shapes the way in which players communicate with each other or with outside observers; it provides a system of events on which the players can condition their behavior over time. We show that conditioning systems exert a powerful influence on strategic interaction, despite that they have no direct effect on available actions or payoffs.

The basis of our analysis is a new theory of joint decision-making in long-term relationships, called *coherent equilibrium*, which combines an explicit model of recurrent negotiation with a selection criterion based on internal and external consistency. The essential idea is that consistency is determined relative the *state of the relationship*, which is governed by a conditioning system. If a social convention places little restriction on how players interpret history, then the players' freedom to reinterpret history gives wide latitude for selecting jointly favorable outcomes. This undermines incentives to cooperate, to the extent that cooperation is sustained by punishments that are jointly unfavorable. More stringent conventions, however, compress history into a smaller number of states, limiting players' reinterpretive options. Correspondingly, the scope for selecting favorable outcomes is restricted, and cooperation is made easier to sustain. Thus, the institutional environment, as reflected by conditioning systems, can play a central role in determining the outcome of strategic interaction. Our analysis demonstrates how, in an environment of multiple systems (representing multiple institutions), particular conditioning systems emerge as valued. In particular, we show that the "standard" system of repeated game histories is generally of little value.

The modeling exercise herein is an attempt to generalize and embellish the models of Ramey and Watson (2002, 2004), which investigate how specific legal institutions (courts versus arbitrators in one study, the WTO for international agreements in the other) provide a conditioning system for contracting parties. Here we take the idea to one abstract limit, based on some simplifying assumptions regarding the way negotiation is treated. There are alternative ways of conceptualizing the notion of a conditioning system, some which may be much better suited for applications than is the rather intricate construction reported here. We believe, however, that the present analysis is useful in laying out, in general terms, key ingredients of the theoretical idea and demonstrating the scope of the theory.

A simple example will demonstrate the intuition behind our theory. Consider the repeated game whose stage game is pictured in Figure 1, and restrict attention to strongly symmetric equilibria.[2] In symmetric games, strong symmetry can be viewed

[2] These are equilibria in which the players are supposed to take the same action, contingent on the history.



1 \ 2	H	L
H	3,3	0,4
L	4,0	1,1

Figure 1: Example of conditioning institution.

as capturing equal bargaining power between the players. We impose this restriction here for simplicity; our formal model addresses bargaining power explicitly. Assume the discount factor is at least 1/2.

Suppose there are two institutions that provide ways of evaluating histories. First, there is an internal system, whereby the players keep track of the full history and interpret it as they wish. Second, there is a social convention which defines a language to assess histories of interaction. The language reflects the ways in which the society describes and judges relationships. In this example, we suppose society has two labels to describe relationships: "fine" (F) and "dysfunctional" (D). A relationship is called dysfunctional if (H,L) or (L,H) has been played in the past, and since then (L,L) has not been played. Otherwise, the relationship is described as fine. The transitions between social designations are diagrammed in the figure. This language is used by members of society to communicate about the long-term relationship. Importantly, the players are not able to manipulate the labels that the social convention assigns to their relationship following any given history.

The partners in this game may attribute value to one conditioning system over the other by essentially adopting it as the basis for their long-term interaction. Adoption amounts to conditioning only on one of the systems, in terms of both private behavior and joint decisions (negotiation). Suppose, for instance, that the players condition on the social convention. One can verify that there is an equilibrium of the game in which the players select (H,H) in the F state and (L,L) in the D state; call this equilibrium e^*. This equilibrium facilitates cooperation. Furthermore, this equilibrium is *uniformly optimal over all equilibria that condition on the social descriptor*, meaning that it is best in both the D and F states. There is no other equilibrium, conditioned so, that yields a higher continuation value in either state. In this sense, the equilibrium is strongly favored by the players in both states. Thus, in the context of the social convention, the players never have the joint incentive to abandon it in favor of another equilibrium.

On the other hand, the internal conditioning system is more problematic. Although there are equilibria in which the players cooperate on the equilibrium path, none has the uniform optimality property with respect to the internal system. Every cooperative equilibrium relies on a low-value continuation in some contingency; but the flexibility of the conditioning system allows the players to re-establish cooperation in such a contingency by selecting a new equilibrium. The only equilibrium that does not exhibit a conflict of this sort is the bad one in which the players select (L,L) forever.

This simple analysis demonstrates the sense in which the social convention has value to the players. By adopting the social conditioning system as the basis for their recurrent negotiation, the players avoid problems arising from the incentive to renegotiate: in every contingency, they jointly prefer e^*. In contrast, renegotiation hinders cooperation in the context of the internal conditioning system; there, the only equilibrium which does not yield a conflict is the low-value one. Since the social convention yields a greater value after every history, one might expect the players to embrace it. Our general theory formalizes the intuition from this example, both on the level of negotiation within a conditioning system and on the level of comparisons between multiple conditioning systems.[3]

Our concept of conditioning institutions admits a large range of interpretations. Within relationships, conditioning systems may reflect the mechanisms that players use to record histories and communicate them with one another. The language for communication can be more or less precise; our theory demonstrates that cooperation is undermined when language becomes excessively precise. Conditioning systems may also capture the information and attitudes of external observers, such as neighbors or colleagues. We show that social attitudes may alter strategic outcomes, even though they have no direct effect on players' actions or payoffs. The main idea is that players can benefit from their ability to condition behavior on social attitudes, to the extent that attitudes are influenced by behavior but are not directly manipulable by the players. Courts, arbitrators, and other mechanisms of legal enforcement and dispute resolution fit into our framework as well. In our setting, the conditioning system reflects how the official status of the relationship, in the eyes of enforcement authorities, is affected by the players' actions. We show that a carefully designed conditioning system can sustain cooperative behavior, even in the absence of external sanctions.

Our study of conditioning systems is couched in terms of an explicit model of negotiation in long-term relationships. By directly incorporating joint decisions, we thus depart from standard analyses of repeated games. Our theory addresses different layers of negotiation between the players. For the most basic level of interaction, we propose an equilibrium concept that formalizes how the players' negotiation is resolved according to bargaining powers and outside options. This concept, called *negotiation equilibrium*, relates the division of value in a relationship to the technology of negotiation. Thus, the intuition of Abreu, Pearce and Stacchetti (1993) that equilibrium selection should be sensitive to players' bargaining power is strongly

[3] Our model admits the possibility that players possess both the social convention and the internal conditioning system. In the example above, players would continue to use the social convention to condition equilibria and joint selection, and the internal conditioning system would be irrelevant.

reflected in our equilibrium concept. In terms of the example discussed above, bargaining power and disagreement options yield outcomes that are similar in flavor to the strongly symmetric equilibria. Otherwise, negotiation equilibrium is analogous to subgame perfect equilibrium in a standard repeated game.

We model meta-level joint selection over negotiation equilibria in much the same way as game theorists have examined renegotiation-proofness criteria on the selection of a subgame perfect equilibrium. In this regard, we build on Bernheim and Ray (1989) and Farrell and Maskin (1989) in utilizing notions of dynamic consistency. Our work differs from theirs due the explicit modeling of negotiation, the study of conditioning systems, and the non-stationary environment created by arbitrary conditioning systems. We capture a generalized notion of consistency in a concept called *pivotal equilibrium*, which is applied within and between conditioning systems to yield our notion of coherent equilibrium. Not only does coherence formalize an intuitive and novel idea, but it also has some very attractive technical features. In particular, coherent equilibria exist generally, and we show further that coherent equilibrium outcomes are unique for a plausible specification of the bargaining environment.

Section 2 presents our model of long-term contractual relationships. The coherence concept is developed in Section 3. Institutional interpretations are offered in Section 4. Section 5 presents an extension of the coherence concept incorporating a form of backward induction in the meta-level joint decision problem. Section 6 contains some general comments.

2 A Model of Long-Term Contractual Relationships

In this section we present a general model of a long-term relationship. We depart from the literature that examines standard repeated games, since repeated games are not equipped to explicitly account for recurrent negotiation. Instead, we consider a variant of a repeated game that includes a joint decision each period, in addition to a non-cooperative stage game. The joint decision models negotiation between the players over spot-contractible transfers and, at a meta-level, renegotiation over equilibria. Thus, we include a bargaining institution in the specification of the game. Parameters of the negotiation problem are tied in a realistic way to the technology of the ongoing relationship. Bargaining power and disagreement options are specified accordingly. Our equilibrium concept incorporates the Nash solution for joint decisions, which is in line with methods of the contract theory literature.

Description of the Game

Assume there are n players who interact over an infinite number of discrete periods. Each period is divided into the *negotiation phase* and the *action phase*, which occur in this order. During the negotiation phase, the agents make a joint decision d

which is selected from a set D. The joint decision is interpreted as a spot contract, established through bargaining between the players. The set D contains a *default decision*, $\underline{d}$, which defines the physical outcome of the negotiation phase if the players fail to reach an agreement on D. We shall model behavior in the negotiation phase using a cooperative bargaining solution, under the supposition that each player can unilaterally induce the default outcome. Players' bargaining weights are given by non-negative numbers $\pi_1, \pi_2, \ldots, \pi_n$, which sum to one. These bargaining weights are exogenously defined by the technology of the relationship and are fixed over time. We think of meta-level renegotiation over equilibria as also occuring in the negotiation phase.

In the action phase, the players simultaneously and independently select private actions — in standard repeated game parlance, this is the "stage game." The actions available to the players in the action phase depend on whether default was the outcome of the negotiation stage in the current period. In particular, default implies that the players are limited to the single *default action profile* $\underline{a}$. Otherwise, A_i is the set of actions available to player i, for $i = 1, 2, \ldots, n$. We write $A \equiv A_1 \times A_2 \times \cdots \times A_n$ as the set of action profiles and, assuming $\underline{a} \notin A$, we let $A' \equiv A \cup \{\underline{a}\}$. For each i, there is a payoff function $u_i : A' \to \mathbf{R}$ defining the stage game payoff for player i. We define $u(a) \equiv \sum_{i=1}^n u_i(a)$ as the total payoff. Also, it is convenient to let $\underline{u}_i \equiv u_i(\underline{a})$ and $\underline{u} \equiv u(\underline{a})$.

One interpretation of the default action is that the default decision in the negotiation phase induces delay, keeping the players from productive interaction in the current period. Another interpretation is that default induces players to coordinate on a myopic equilibrium profile in the current action phase. We elaborate on the latter interpretation in section 3.[4]

Before discussing long-run payoffs, we put more structure on the set D. In particular, we assume that D specifies transfers between the players, with the default decision implying that no transfer is made. That is, each of the players has the power to induce default and veto any immediate transfer. Accordingly, we define:

$$D \equiv \{\underline{d}\} \cup \left\{ m \in \mathbf{R}^n \mid \sum_{i=1}^n m_i \leq 0 \right\},$$

where m refers to the transfer. Note that the players can specify no transfer ($m = (0, 0, \ldots, 0)$) and the players can also actively select the default decision.

Player i's payoff in period t is given by $m_i^t + u_i(a^t)$, where $a^t \in A$ is the action profile chosen and m^t is the vector of transfers made in the period (which is zero in

[4] Our methodology also covers the following variation: default in the negotiation phase induces the relationship to be severed, yielding exogenous outside options. Ramey and Watson (1997) and Den Haan, Ramey, and Watson (1999) explore particular labor/macro settings with long-term contractual relationships, where default induces severance.

the case of default). Player i's payoff in the entire game is given by

$$\sum_{t=1}^{\infty}[m_i^t + u_i(a^t)]\delta^{t-1},$$

where δ is the common discount factor. Note that we assume transferable utility.[5]

Conditioning Systems

Institutional conventions serve to organize the players' evaluation of histories, which shapes their negotiation and action choices. The notion of a convention is formalized in terms of a mapping from past actions to current assessments of the relationship, forming a partition of the histories that we refer to as a *conditioning system*. In summarizing the history of the players' interaction from the perspective of an institution, a conditioning system incorporates any limitations in observing, processing, or forming opinions about past action choices. Multiple conditioning systems exist if there are several institutions rendering judgments of the relationship. Conditioning systems impose no direct constraints on contract negotiation, available actions, or payoffs within the relationship. However, a conditioning system may still have a powerful effect on the players' selection of equilibria to the extent that it is isolated as the basis for conditioning behavior. In this section we develop our formal model of conditioning systems; interpretations of the model and examples are discussed in Section 4.

A conditioning system consists of a set of *states of the relationship* X and a *transition function* μ. The former defines the institutional descriptors of the relationship; the latter defines the transition of the state from one period to the next as a function of the players' behavior. The set of states takes the form

$$X \equiv \bigcup_{t=1}^{\infty} X^t,$$

where $X^1, X^2, \ldots$ are finite, disjoint sets describing the possible states of the relationship in the periods of the game. That is, X^t is the set of states for period t (and only period t). The set X^1 has exactly one member, which is the initial state. We make the following assumption on the state transition function.

Assumption 1 *Transition of the state from the current period to the next does not depend on transfers made in the current period.*

[5] There are two reasons we assume this payoff structure. First, transferable utility seems appropriate for an archetypical model of contract. Second, transferable utility makes our analysis much simpler and cleaner than it would be otherwise, because it allows us to interpret the players' joint decision problem as a standard bargaining problem. Our concepts extend in a natural way to the standard repeated game setting, on which we will elaborate in future work.

According to this assumption, the conditioning system pays no attention to transfers made during negotiation. For example, if states reflect the assessments of external observers, then Assumption 1 can be interpreted to mean the players conduct negotiations in secret. If states represent the players' internal assessments, then this assumption implies that, in negotiation, they can always separate the continuation of the game from the current transfer. We make this assumption for simplicity, as it guarantees the players' joint decision problem can be viewed as a standard bargaining problem with a well-behaved set of alternatives.[6]

Under Assumption 1, the transition of the state is described by a function $\mu : A' \times X \to X$. If x is the state in the current period, then action profile a induces a transition to state $x' = \mu(a, x)$ at the start of the next period.[7] Note that for each $x^t \in X^t$, $\mu(A', x^t) \subset X^{t+1}$. We require the transition function to yield a precedence relation over states that has a tree structure. Thus, we assume:

Assumption 2 *For each $t > 1$ and $x^t \in X^t$, there is exactly one state $x^{t-1} \in X^{t-1}$ with the property $x^t \in \mu(A', x^{t-1})$.*

Given a conditioning system (X, μ), define $R(x)$ to be the set of points that are reachable from x (including x itself). Formally:

Definition 1 *For each $x \in X$, $R(x)$ is defined by: $y \in R(x)$ if and only if either (i) $y = x$ or (ii) there exist action profiles $a^1, a^2, \ldots, a^K \in A'$ and states $z^1, z^2, \ldots, z^{K+1} \in X$ such that $z^1 = x$, $z^{K+1} = y$, and $\mu(a^k, z^k) = z^{k+1}$ for $k = 1, 2, \ldots, K$.*

Where it is necessary to make the dependence on the conditioning system clear, we write $R(\cdot; X, \mu)$.

States are separated into equivalence classes by the following criterion. Two states are called *equivalent* if they induce isomorphic mappings from action sequences in their continuations to future states. In other words, by re-labeling future states, continuations from two equivalent states "look the same." Formally:

Definition 2 *States $x, y \in X$ are called* **equivalent** *if there is a one-to-one mapping $\eta : R(x) \to R(y)$ such that $\eta(x) = y$ and for all $z \in R(x)$ and each $a \in A'$, $\eta(\mu(a, z)) = \mu(a, \eta(z))$.*

An important special case of conditioning system is the space of full histories with the transition function linking histories in the standard repeated-game manner. One can think of this as the "standard" conditioning system. Under Assumption 1, though, the state does not describe the spot-contractible transfers in each period. Thus, we formally speak of the *standard conditioning system* as the finest partition

[6] Relaxing Assumption 1 brings non-standard bargaining problems to the fore; we aim to tackle such issues in future work.

[7] Function μ is extended to the domain of subsets of A' in the usual way, so that $\mu(A', x) = \{\mu(a, x) \mid a \in A'\}$.

of histories satisfying Assumption 1. Specifically, this is the conditioning system in which the state fully describes the action profiles chosen in each action phase through a period in the game. We denote the standard system by (X_s, μ_s) and reserve the term "history" for use with this system. We have $X_s = \bigcup_{t=0}^{\infty}(A')^t$ and $\mu_s(a, x) \equiv xa$ for all $x \in X$ and $a \in A'$, where xa denotes x appended with a. One can easily verify that the standard system satisfies Assumptions 1 and 2. In addition, all states are equivalent in the standard conditioning system.

Note that every conditioning system (X, μ) is related to the standard system by a function mapping the history (in X_s) to the state in X.

Definition 3 *Consider any conditioning system (X, μ). The* **translator** *from (X_s, μ_s) to (X, μ) is the function $\beta : X_s \rightarrow X$ defined so that $\beta(y; X, \mu)$ is the state in X that results following action sequence y.*

The translator is well-defined and unique.

The Conditioning Environment

Several alternative conditioning systems may compose the environment in which the players interact. For example, suppose a social convention defines how the relationship is judged, according to the conditioning system (X', μ'). In addition, the standard system (X_s, μ_s) may also be available, whereby the players keep a full account of the history of their relationship. It is important to specify the entire collection of conditioning systems serving as the backdrop for the relationship. We call this collection the *conditioning environment* and we denote it by $\mathcal{C}$.

Conditioning systems can be compared on the basis of refinement.

Definition 4 *(X, μ) is a* **refinement** *of (X', μ') if, for $x, y \in X_s$, $\beta(x; X, \mu) = \beta(y; X, \mu)$ implies $\beta(x; X', \mu') = \beta(y; X', \mu')$.*

A conditioning system is called the *finest in $\mathcal{C}$* if it is a refinement of every other conditioning system in $\mathcal{C}$. Note that, since (X_s, μ_s) refines every other conditioning system, if it is included in $\mathcal{C}$ then it is the finest in $\mathcal{C}$.

Assumption 3 *$\mathcal{C}$ comprises a finite number of conditioning systems and contains a finest conditioning system (X_f, μ_f).*

Strategies and Values

We use a generalized notion of strategy to investigate behavior in this game. The players' decisions in negotiation phases are described by a function $\theta : X_s \rightarrow D$, which can be interpreted as their *joint strategy.* The players' noncooperative behavior in the action phase of each period is described by a function $\sigma : X_s \rightarrow \Delta^u A$, where Δ^u

denotes the set of uncorrelated probability distributions.[8] A strategy specification (θ, σ) is called a *regime*.

Note that the players are not directly constrained by any particular conditioning system, in that their behavior is a function of the standard system. However, the players may effectively condition on a more coarse system. We shall say that a regime (θ, σ) is (X, μ)-*measurable* if for every $x, x' \in X_s$, we have $\theta(x) = \theta(x')$ and $\sigma(x) = \sigma(x')$ whenever $\beta(x; X, \mu) = \beta(x'; X, \mu)$. In this case, we can regard the regime as a function of X.[9]

Next we define values in continuations of the game. Given a regime (θ, σ) and a history $x \in X_s$, player i's continuation value conditional on x is defined as

$$v_i(x) \equiv \mathcal{E}_x \sum_{t=\tau(x)}^{\infty} [m_i^t + u_i(a^t)]\delta^{t-\tau(x)},$$

where $\mathcal{E}_x$ denotes the expected value and $\tau(x)$ is defined to be the period in which x occurs ($x \in X_s^{\tau(x)}$). The players' joint continuation value is given by $v(x) \equiv \sum_{i=1}^n v_i(x)$. To make the dependence on the regime explicit, we sometimes write $v_i^{(\theta,\sigma)}$ and $v^{(\theta,\sigma)}$. Also, if the regime is (X, μ)-measurable for a particular conditioning system (X, μ), then we can regard the values as functions of X, rather than X_s.

Negotiation Equilibrium

In this subsection, we define an equilibrium concept incorporating renegotiation over the items that are spot-contractible in the negotiation phase (default and immediate transfers). The concept captures how negotiation over dividing the value of the relationship is resolved. We begin by combining best response behavior by the players in each action phase with a unilateral default condition for the negotiation phase.

A regime (θ, σ) is said to be *dynamically incentive compatible* if

$$u_i(\sigma(x)) + \delta \sum_{a \in A} v_i(\mu_s(a, x))\sigma(x)(a) \geq$$

$$u_i(a_i', \sigma_{-i}(x)) + \delta \sum_{a_{-i} \in A_{-i}} v_i(\mu_s((a_i', a_{-i}), x))\sigma_{-i}(x)(a_{-i}),$$

for every $x \in X_s$, each player i, and each $a_i' \in A_i$. That is, no player has an incentive to deviate in the action phase in any history, conditional on the regime's continuation values. Given θ, call θ' *feasible by unilateral deviation* if, for every $x \in X_s$, either $\theta(x) = \theta'(x)$ or $\theta'(x) = \underline{d}$. Any player can unilaterally induce the outcome given by

[8] Assumption 1 implies that private actions in a period do not depend on transfers made earlier in the period, because actions are conditioned on the state, and the state does not record the transfer.

[9] We avoid cumbersome notation by not indexing x, X, or μ, except in special cases which include the standard system (X_s, μ_s) and the finest system (X_f, μ_f). To avoid creating confusion regarding from which conditioning system a particular state x is drawn, we shall generally make this explicit.

(θ', σ) if players have coordinated on (θ, σ). A regime (θ, σ) is said to be *dynamically individually rational* if $v_i^{(\theta,\sigma)}(x) \geq v_i^{(\theta',\sigma)}(x)$ for each player i, each history x, and every θ' that is feasible by unilateral deviation. In words, no player obtains less than what he could get by inducing default.

Definition 5 *A regime (θ, σ) is called an* **equilibrium** *if it is dynamically incentive compatible and individually rational.*

Next, we define a refinement of the equilibrium notion to capture how bargaining weights influence joint decisions. Consider the joint decision problem at history x, given an equilibrium (θ, σ) describing play in the subsequent action phase and in future periods. Interaction in the negotiation phase can be interpreted as a bargaining problem, wherein the players negotiate over the set of continuation values associated with the various joint decisions. In fact, since the players can transfer utility and unilaterally induce default, this is a standard bargaining problem with a well-behaved set of alternatives and disagreement point. The disagreement point is given by the default decision, which from history x yields a continuation value of

$$w_i(x) \equiv \underline{u}_i + \delta v_i(\mu_s(\underline{a}, x))$$

for player i and a total continuation value of

$$w(x) \equiv \underline{u} + \delta v(\mu_s(\underline{a}, x)).$$

Note that in these expressions, continuation values v and v_i are defined by (θ, σ). The implicit set of alternatives in the negotiation phase — that is, the set of attainable continuation values — includes this default value, since the default decision can be unilaterally or jointly selected by the players. The set of alternatives also includes what the players can obtain by avoiding default in the current period. They can achieve any joint value less than or equal to

$$\hat{v}(x) \equiv u(\sigma(x)) + \delta \sum_{a \in A} v(\mu_s(a, x))\sigma(x)(a).$$

To obtain values that are strictly less than this, the players specify transfers that in sum are negative. By making the appropriate transfer in the current period, the players can divide their joint value in any way desirable.

To determine the outcome of this well-defined bargaining problem, we employ the Nash bargaining solution.[10]

Definition 6 *A regime (θ, σ) is called a* **negotiation equilibrium** *if it is an equilibrium which satisfies (i) $v(x) = \max\{w(x), \hat{v}(x)\}$ and (ii) $v_i(x) = w_i(x) + \pi_i[v(x) - w(x)]$ for each player i and every $x \in X_s$.*

[10] Most standard cooperative solution concepts yield the same value.

Condition (i) means the players always make a decision in the negotiation phase that maximizes the value of the relationship conditional on future behavior (including the action profile in the current period) determined by the regime. Condition (ii) means the players divide the surplus in the negotiation phase according to their individual bargaining powers and threat of default.

Let E^N denote the set of negotiation equilibria. For each conditioning system (X, μ), let $E^N(X, \mu)$ denote the negotiation equilibria that are (X, μ)-measurable. These negotiation equilibria are the ones conditioned on (X, μ). Note that all negotiation equilibria are (X_s, μ_s)-measurable. We write a generic negotiation equilibrium as $e = (\theta, \sigma)$. For any $E \subset E^N$, define the associated set of value functions as $V(E) \equiv \{v^e \mid e \in E\}$. We utilize the metric on value functions defined by the weighted sup norm, where

$$\|v\| \equiv \sup_{x \in X} v(x)/\tau(x).$$

Recall that $\tau(x)$ is defined as the period in which x occurs. Statements about compactness refer to the metric space defined by $V(E^N)$ and the weighted sup norm.

Theorem 1 *For each conditioning system (X, μ), the set $E^N(X, \mu)$ is nonempty and $V(E^N(X, \mu))$ is compact. In addition, for each $e \in E^N$ and $x \in X_s$, $v_i(x) = \underline{u}_i/(1 - \delta) + \pi_i[v(x) - \underline{u}/(1 - \delta)]$ and $v(x) \geq \underline{u}/(1 - \delta)$.*

The third statement of the theorem implies that players have the same rankings over equilibrium continuation values. That is, for any two negotiation equilibria e and f, and for any states x and y, $v_i^e(x) > v_i^f(y)$ if and only if $v^e(x) > v^f(y)$. That the players share rankings over equilibria is a direct consequence of their individual ability to impose default outcomes in each period.[11]

3 Coherent Equilibrium

The negotiation equilibrium concept embodies a joint decision of narrow scope (transfers and whether to default in each period). Players may be thought to also engage in meta-level joint decision-making *over* negotiation equilibria. That is, they jointly select a long-term contract, and re-evaluate the contract over time. In this section, we propose a selection criterion to model the resolution of the meta-level decision problem.

[11] This intuitive conclusion highlights one of the differences between our modeling approach and that of others. As Abreu, Pearce, and Stacchetti (1993) point out, notions of bargaining power are missing from most analyses of "renegotiation-proofness" in the repeated game literature. These authors advocate building bargaining power into selection criteria and they do so by assuming that, in a symmetric game, the players select an equilibrium that yields symmetric payoffs from the beginning of the game. We take the further step of modeling bargaining power and disagreement outcomes as embedded in the technology of joint decision-making. The players' power to disrupt the relationship in the short run implies that they share the continuation value in fixed proportions.

Motivation

Our new concept is based on the notion of *internal consistency*, developed by Bernheim and Ray (1989) and Farrell and Maskin (1989).[12] We depart from the literature by considering the conditioning environment, which yields a non-stationary setting. In addition, due to the explicit modeling of recurrent negotiation, our analysis has a different theme than in much of the related game-theory literature. For example, the renegotiation-proofness concepts proposed by Bernheim and Ray (1989) and Farrell and Maskin (1989) rely on the ability of the players to punish each other while maintaining a high total continuation value. This form of punishment amounts to having a deviant player compensate the other players after the deviation, to a degree that would deter cheating in the first place. If the set of equilibria offers adequate scope to vary a player's share of the continuation value, high-value cooperation may be sustained in this fashion. In a negotiation equilibrium, however, punishments must involve a decrease in the *total* value of the relationship, not just in the value of a single player. Thus, variation of players' shares cannot occur if players have fixed bargaining power and the threat of default each period.

Although players have the same preferences over negotiation equilibria, conditional on the history, the players' preferences do depend on the states (of various conditioning systems) in which they find themselves at any given time. We use the following notation as a shorthand way of describing the players' joint preferences over equilibria. Given equilibria $e, f \in E^N$, we write $e \succeq f$ if $v^e(x) \geq v^f(x)$ for every $x \in X_s$. The expression $e \succ f$ signifies that $v^e(x) \geq v^f(x)$ for every $x \in X_s$ and $v^e(y) > v^f(y)$ for some $y \in X_s$. Finally, $e \sim f$ means that $v^e(x) = v^f(x)$ for every $x \in X_s$. We say that equilibria e and f *are in conflict* if neither $e \succeq f$ nor $f \succeq e$. In words, e and f are in conflict if there exist histories $x, y \in X_s$ such that the players strictly prefer e to f from history x and f to e from history y (that is, $v^e(x) > v^f(x)$ and $v^e(y) < v^f(y)$).

We propose a selection criterion akin to internal consistency. The criterion also embodies a form of the related notion of *external consistency.* To develop intuition, consider a simple example. Take as given a conditioning system (X, μ) and a set E of (X, μ)-measurable negotiation equilibria. Suppose we have equivalent states, x and y, such that $y \in R(x; X, \mu)$. Think of y as being reached from x only if one of the players cheats the other. In addition, suppose there is an equilibrium e that achieves a level of cooperation at x under the threat of punishment if a player cheats. That is, the continuation payoff from state x is high, while the continuation payoff from y is low: $v^e(x) > v^e(y)$.

Since the states are equivalent, we can find another negotiation equilibrium f specifying the same behavior following y that e prescribes from x. Therefore, $v^f(y) = v^e(x)$. It seems reasonable to assert that if e is viewed by the players as viable from

[12] Related concepts are studied by Bergin and MacLeod (1993), Ray (1994), and van Damme (1989). Blume's (1994) analysis adds costs of renegotiation. Other notions of renegotiation-proofness are examined by Asheim (1991) and Pearce (1989,1991).

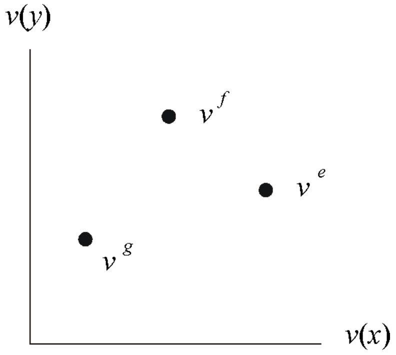


Figure 2: Non-viable level.

state x then f should be viewed as viable in state y. But suppose the players select among viable equilibria and do so to maximize their continuation value (since they agree on the rankings of negotiation equilibria). Then in state y the players should choose f over e. As a result, e cannot be considered viable in the first place. This is the intuition behind the notion of internal consistency.

Continuing with the example, it may be impossible to achieve the continuation payoff $v^e(x)$ without specifying the punishment that e requires at y. Thus, we also have $v^f(x) < v^e(x)$. As Figure 2 illustrates, in this example equilibria e and f are in conflict. This observation leads to the following reinterpretation of the example. Imagine the players viewing a set of equilibria as viable. That is, a theory summarizing the players' joint decision process selects a subset of negotiation equilibria $F \subset E$. The intuition discussed above suggests a reasonable condition of the theory: that if $e \in F$ then $f \in F$ as well. In addition, the players, free in the selection process, can at any time choose between elements of F. The conflict between e and f makes their selection incompatible with state-contingent optimization. Thus, neither e nor f can be elements of F.

There are also cases in which conflicts between equilibria represent failure of external consistency. For example, one may have a theory which does not require the internal consistency condition noted above, in which case the conclusion $F = \{e\}$ may be allowed. That is, the players only view e as viable. It may be the case that such a theory also allows the prediction $F = \{f\}$. In other words, there are two manifestations of the theory: (i) the players will view *only* e as viable and (ii) the players view *only* f as viable. However, one may argue that such a theory violates external consistency in the sense that the players should be able to freely select among manifestations of the theory in each state, which is incompatible with the conflict between e and f.

In summary, we assert that failures of consistency amount to conflicts between equilibria. Therefore, under consistency, equilibria without conflicts represent lower bounds on what can be achieved by the players. We propose a selection criterion that is based directly on the conflict notion, which is well-suited to the non-stationarities

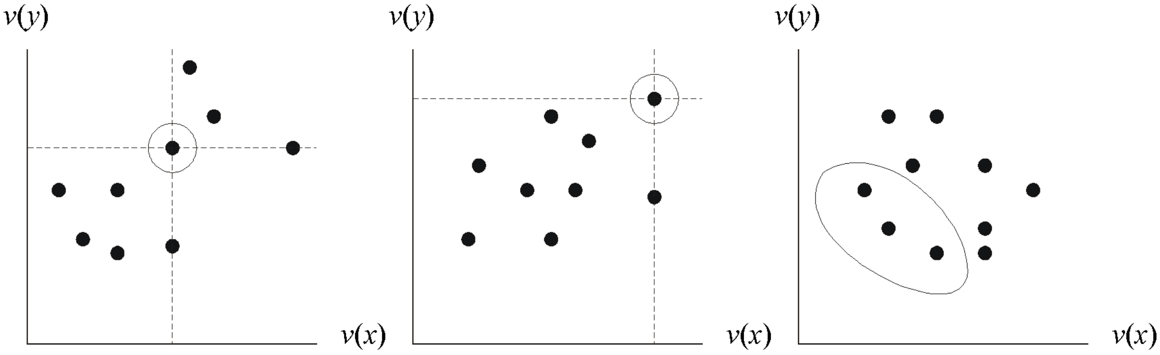


Figure 3: Examples of pivotal equilibria.

exhibited by our model (there may be non-equivalent states).

Pivotal Criterion

Take as given a reference set of equilibria $E \subset E^N$. Equilibria without conflicts belong to the set

$$\hat{E} \equiv \{e \in E \mid \text{for every } e' \in E, \text{ either } e' \succeq e \text{ or } e \succeq e'\}.$$

As argued above, in the least these equilibria provide lower bounds on the value functions predicted by a "reasonable" selection criterion. Lower bounds are also furnished by the set of minimal equilibria:

$$\underline{E} \equiv \{e \in E \mid \text{there is no } f \in E \text{ such that } e \succ f\}.$$

The greatest lower bound is thus given by the maximal elements of $\underline{E} \cup \hat{E}$.

Definition 7 *Given $E \subset E^N$, an equilibrium e is called* **pivotal** *if $e \in \underline{E} \cup \hat{E}$ and there is no equilibrium $e' \in \underline{E} \cup \hat{E}$ with $e' \succ e$. Let $P(E)$ denote the set of pivotal equilibria.*

A pivotal equilibrium achieves the greatest values possible without conflict. That is, if e is pivotal and f is a negotiation equilibrium satisfying $f \succ e$, then it must be that f has a conflict. Thus, for a selection criterion based on the full force of the conflict idea, $P(E)$ is exactly the prediction.

Figure 3 depicts a few different cases of pivotal equilibria. The value functions of pivotal equilibria are circled in the pictures. In the first diagram, the players are able to select a non-minimal equilibrium; it is the greatest equilibrium with no conflicts (since there are no points to the lower-right or upper-left regions of the circled point). In the second diagram, there is a single maximum point (with no conflicts) and thus it is selected. In the third diagram, there is no equilibrium without conflicts; therefore, the pivotal set consists of the minimal equilibria.

Lemma 1 *If $E \neq \emptyset$ and $V(E)$ is compact then $P(E)$ is nonempty. Furthermore, either (a) $e, f \in P(E)$ implies $e \sim f$, or (b) $P(E) = \underline{E}$.*

Note that, regarding pivotal equilibrium, $\underline{E}$ is relevant only when $\hat{E}$ is empty.

We stress that there are strong and weak ways of viewing our pivotal criterion. Under the strong viewpoint, one assumes our theory verbatim, in which case only pivotal equilibria are selected. Under the weak viewpoint, one seeks a theory incorporating the idea of consistency, which may be weaker than the "no conflicts" condition. In this case, pivotal equilibria serve to provide lower bounds on what the players achieve.

Definition of Coherence

We now formally define our concept of how players make their meta-level joint decision over negotiation equilibria. The conditioning environment, $\mathcal{C}$, influences equilibrium selection to the extent that the players can isolate individual conditioning systems for consideration. In other words, in their recurrent discussion over negotiation equilibria, the players may focus on (X, μ)-measurable equilibria, for some particular $(X, \mu) \in \mathcal{C}$. Conflicts play a role in the selection problem at two levels: (a) in the context of an individual conditioning system, which the players have isolated in their negotiation, and (b) selection *between* conditioning systems. We apply the pivotal criterion to resolve both aspects of the selection problem.

Selection *within* conditioning systems produces the set

$$P_{\mathcal{C}} \equiv \bigcup \{P(E^N(X, \mu)) \mid (X, \mu) \in \mathcal{C}\},$$

which comprises the equilibria that are pivotal with respect to individual conditioning systems. Selection *among* conditioning systems is then given by $P(P_{\mathcal{C}})$, which picks out the equilibria that are pivotal among this set.

Definition 8 *The set of* **coherent equilibria** *is defined as $Q_{\mathcal{C}} \equiv P(P_{\mathcal{C}})$.*

Theorem 2 *There exists a coherent equilibrium.*

Observe that, in the special case in which the conditioning environment consists of a single conditioning system (X, μ), we have $Q_{\mathcal{C}} = P(E^N(X, \mu))$; that is, the coherent equilibria are those pivotal with respect to $E^N(X, \mu)$.

Natural Default Setting

We have defined coherent equilibria for arbitrary specifications of the default decision. We now specialize the model to a particular bargaining environment in which the stage game is played even when players default in a period. That is, the players select actions from A after default just as they do after avoiding default. We assume, however, that default triggers a transition of the state that is independent of the actual actions chosen in the period. This captures the idea that, when there is no agreement, the players view the actions in the current period as having no consequence on the future.

In this setting, we must include a theory of how the players behave following the default decision. We adopt the view that they coordinate on a focal action profile. Since the default transition is not influenced by the players' actions, players can coordinate on any Nash equilibrium of the stage game following disagreement. It seems plausible that the players coordinate on the "best one-shot Nash equilibrium" of the game. In other words, default triggers the profile $\alpha^* \in \Delta^u A$ which maximizes u over all one-shot Nash equilibria. Then $\underline{u}_i = u_i(\alpha^*)$ for each player i.

Definition 9 *The relationship is said to be in the* **natural default setting** *if* $\underline{u}_i = u_i(\alpha^*)$ *for each* i.

The next theorem demonstrates that, in the natural default setting, the coherent equilibria have some interesting and intuitive properties. Two new terms are used. We say that histories $x, y \in X_s$ are *completely equivalent on* $\mathcal{C}$ if for each conditioning system $(X, \mu) \in \mathcal{C}$, $\beta(x; X, \mu)$ is equivalent to $\beta(y; X, \mu)$. In words, complete equivalence means that histories x and y are equivalent regardless of which conditioning system is used. Also, we say $\mathcal{C}$ has the *access property* if the following is true for each $(X, \mu) \in \mathcal{C}$: for every $x, y \in X$, there exists $z \in X$ such that z is equivalent to x and $z \in R(y)$. The access property means that every equivalence class can be reached from every state.

Theorem 3 *In the natural default setting, there is a unique coherent equilibrium value function. That is, if* e *and* f *are coherent equilibria, then* $e \sim f$*. Further, if* $\mathcal{C}$ *has the access property and* x *and* y *are completely equivalent histories, then* $v^e(x) = v^e(y)$.

The theorem establishes that coherent equilibria are essentially *unique* in that all coherent equilibria have the same value function. Thus, our concept of coherent equilibrium delivers both existence and uniqueness for a broad class of relationships.[13] Theorem 3 also confirms invariance of the coherence concept. The criterion yields the same value for all completely equivalent histories. The access property is sufficient for this result; it is not a necessary condition, as discussed in Section 5.

4 Institutions and Conditioning Systems

Conditioning systems can be understood as reflecting languages, customs, or societal ideals which can be adopted by the players in interpreting their own past actions. It

[13] As the proof of Theorem 3 demonstrates, uniqueness is related to whether all minimal equilibria have the same value function. This is the case in the natural default setting, where the minimum is given by α^* played each period.

is useful to distinguish between two types of conditioning systems. *External systems* are associated with third parties, such as nearby individuals who assess the workings of the relationship. A broader social convention is an external system; its means for assessing the relationship may be reinforced by other members of the society. On the other hand, *internal conditioning systems* are those under the direct control of the players in the relationship, whereby the players use their own descriptors of the past. Since in most settings, players are free to make arbitrary distinctions between histories, we regard the standard conditioning system as the most appropriate model of the players' internal system.

In describing how actions determine future states, and correspondingly in negotiating the selection of equilibria, players are influenced by the precision with which conditioning systems allow them to distinguish between histories. The greatest degree of precision is associated with the standard conditioning system. Unfortunately for the players, the internal system is of little use to them. To see this, observe that, with the standard system, every history is a distinct state and all states are equivalent. Another way of thinking about this is that, for any two histories x and y, the sets of continuation equilibria from x and y are identical. Players are therefore free to reinterpret the meanings of histories in a manner that induces the most attractive continuation equilibrium; as a result, there are too many conflicts between equilibria to sustain a high-value equilibrium. Players cannot sustain cooperation requiring punishment for misdeeds, because they would be tempted to restart cooperation following an episode of cheating, which destroys cooperation in the first place. Formally, the intuition is confirmed by the following general theorem.

Theorem 4 *Consider a relationship in the natural default setting. If $\mathcal{C} = \{(X_s, \mu_s)\}$ — that is, if the conditioning environment consists only of the standard system — then every coherent equilibrium e satisfies $v_i^e(x) = \underline{u}_i/(1-\delta)$, for each $x \in X_s$ and each player i.*

This negative result is easily proved using the value characterization of Theorem 3, since all states are equivalent in the standard system.

Combining Theorems 3 and 4, we conclude that *any* conditioning environment yields the players higher state-contingent values than does the environment consisting of only the standard system. In fact, the standard system has no value even in conjunction with other conditioning systems:

Theorem 5 *Consider a relationship in the natural default setting, under conditioning environment $\mathcal{C}$, and let $\mathcal{C}' = \mathcal{C} \cup \{(X_s, \mu_s)\}$. The coherent equilibrium value function in environment $\mathcal{C}$ is identical to the one in environment $\mathcal{C}'$.*

As the previous two theorems indicate, the players attribute value only to conditioning systems that are more coarse than the standard one. The key idea is that under the standard system, players have complete freedom to redefine labels attached to histories. Coarser conditioning systems tied to social conventions do not allow such

2 1	H	L	R
H	2,2	0,4	0,1
L	4,0	1,1	1,0
R	1,0	0,1	1-γ,1-γ

$\gamma > 0$

Figure 4: Partnership stage game.

freedom, however, since the labels have intrinsic meaning in the eyes of external parties, and the players have no control over this intrinsic meaning.

An example of a coarse conditioning system that supports a cooperative outcome can be constructed along the lines of familiar "grim strategies." Consider the stage game pictured in Figure 4. Here "H" and "L" refer to high and low effort, respectively, which the players can exert in their relationship each period. Action "R" is a "reconciliation" action which will be considered below. The common discount factor δ is assumed to be greater than 2/3. Suppose the players' bargaining weights are $\pi_1 = \pi_2 = 1/2$. Also assume the natural default setting, where $\underline{u}_1 = \underline{u}_2 = 1$ and so $\underline{u} = 2$.

Suppose by social convention, two labels are reserved to describe the relationship: "good" (G) and "bad" (B). These labels can be interpreted, for example, as value judgments concerning players' conduct within their relationship. A state at period t is a description of the labels imputed in periods 1 through $t-1$, where we assume the label in period 1 is G. For any state x and label $L \in \{B, G\}$, denote by xL the state formed when L is appended to x. The players may also devise their own labels by appealing to the standard conditioning system, but this does not affect the coherent equilibrium. For example, players might freely revise their own value judgments concerning their past behavior, but they continue to be constrained by social value judgments.

Consider the "grim" convention in which G means the players have always played H. The relationship is labeled bad (B) if someone played other than H at some point in the past. In the case of default, assume the designation from the current period carries over into the next. In the grim conditioning system (X_g, μ_g), the states are defined by

$$X_g^t \equiv \{G\}^t \cup \left(\bigcup_{k=1}^{t-1} \{G\}^k \times \{B\}^{t-k} \right),$$

with $X_g \equiv \bigcup_{t=1}^{\infty} X_g^t$. That is, the state is a description of the history of the relationship using the language of the grim social convention. Once the relationship is described as bad, it is judged bad forever after. Thus, for every state x ending with designation B, we have $\mu_g(a, x) = xB$.

There are two types of states for period t. One type involves the G designation for all periods through t. The other type involves G for some number k of periods, followed by B in the remaining periods. It is not difficult to see that there are thus two equivalence classes of states, given by

$$Y \equiv \bigcup_{t=1}^{\infty} \{G\}^t$$

and $Y' \equiv X_g \setminus Y$.

With the conditioning system and technology of the relationship — defined above — in place, we now turn to locate the unique coherent equilibrium in this setting. Note that once the designation is B, the transition of the state from period to period does not depend on the players' behavior. Therefore, in any period under the B designation, the players are only able to sustain the unique one-shot Nash equilibrium (L,L,). As a result, every negotiation equilibrium e satisfies $v^e(x) = \underline{u} = 2/(1-\delta)$ for each $x \in Y'$. Next observe that, since 4 is the greatest joint payoff possible in the stage game, we have $v(x) \leq 4/(1-\delta)$ for every $x \in X_g$. Finally, note that there is a negotiation equilibrium $e^* = (\theta^*, \sigma^*)$ specifying $\theta^*(x) = (0,0)$ for all $x \in X_g$, $\sigma^*(x)((\text{H,H})) = 1$ for all $x \in Y$, and $\sigma^*(x)((\text{L,L})) = 1$ for all $x \in Y'$. In words, this regime entails no spot transfers in the negotiation phase (this divides the continuation value according to the equal bargaining weights); players select (H,H) with probability 1 when the state is in Y; and (L,L) is played when the state is in Y'.[14]

One can readily verify that $v^{e^*}(x) = 4/(1-\delta)$ for every $x \in Y$ and so e^* yields the highest state-contingent values over all negotiation equilibria. We can thus say that e^* is *uniformly best.* It follows that e^* is pivotal with respect to $E^N(X_g, \mu_g)$, and therefore it is the coherent equilibrium. Note that this joint payoff exceeds the joint payoff in the best one-shot Nash equilibrium, which is $2/(1-\delta)$ when repeated over time. In this example, players benefit from greater coarseness of labeling.

The latter finding does not mean, however, that language should communicate nothing at all. Indeed, consider the most coarse system (X_0, μ_0), which we call the *null conditioning system*, specifying $X_0^t = \{x_0^t\}$ for every t. In this system, since there is a single possible state in each period, there is no way to condition actions on past behavior. Therefore, only stage game Nash equilibria can be supported in each period, which implies the value $\underline{u}/(1-\delta)$ in each state. The following theorem generalizes the result.

Theorem 6 *Consider a relationship in the natural default setting, under conditioning environment $\mathcal{C}$, and let $\mathcal{C}' = \mathcal{C} \cup \{(X_0, \mu_0)\}$. The coherent equilibrium value function in environment $\mathcal{C}$ is identical to the one in environment $\mathcal{C}'$.*

[14] Note that we are defining the regime on X_g; the extension to X_s is made via $\beta^{-1}(\cdot; X_g, \mu_g)$.

In order for a social convention to have strictly positive value, its conditioning system must have non-equivalent states that are associated with meaningful distinctions among past behavior. Further, the system must restrict the players' flexibility in designing cooperative equilibria conditional on certain states. In terms of the coherence concept, a valuable conditioning system provides a framework in which there is a uniformly best equilibrium: one that is ranked above the others, independent of the state.

To build more intuition and further demonstrate our theory, we proceed by analyzing two variations of the above example. We continue to use the stage game pictured in Figure 4, under the natural default setting and with equal bargaining weights. We analyze separately two conditioning environments; they are related to the partially-formalized example presented in the Introduction and differ on the basis of how transitions between designations G and B occur. Given Theorem 5, in both variations we constrain attention to a single conditioning system (X, μ), which represents a social convention. Since all equilibria are (X, μ)-measurable in this environment, we define regimes and continuation values as functions of X.

The "Reconciliation" Convention: In this social convention, the term G means that either (i) the relationship had the G designation in the previous period and (H,H) was played then, or (ii) (R,R) was played in the previous period. That is, the society has language to describe whether a relationship is in "good standing," defined as the condition in which deviations from cooperation have been reconciled. Mathematically, in this case we have a conditioning system (X_r, μ_r), where

$$X_r^t \equiv \{G\} \times \{B, G\}^{t-1}$$

and $X_r \equiv \bigcup_{t=1}^{\infty} X^t$. The transition function $\mu_r : (\{\text{H,L,R}\}^2 \cup \{\underline{a}\}) \times X_r \to X_r$ is defined as follows. Take any $x \in X_r$. If the last term of x is G, then

$$\mu_r((\text{H,H}), x) = \mu_r((\text{R,R}), x) = \mu_r(\underline{a}, x) = xG.$$

For every other action profile a in the stage game, we have $\mu_r(a, x) = xB$. On the other hand, if the last term of x is B, then $\mu_r((\text{R,R}), x) = xG$; for every other action profile a, we have $\mu_r(a, x) = xB$.

We shall demonstrate that this social convention is of value if and only if the reconciliation cost γ is sufficiently large. The intuition runs as follows. Note that, once society judges the relationship bad, the only way for the players to return to good social standing is to take the reconciliation action profile (R,R). However, if γ is very large — meaning action (R,R) yields a large negative payoff — then the players would never wish to reconcile. That is, when γ is large there is no negotiation equilibrium in which (R,R) is played. Therefore, for all intents and purposes, this conditioning system works just like the grim system discussed above. There is a uniformly best equilibrium e^* that sustains cooperation under the G designation. This equilibrium is coherent.

On the other hand, if γ is small then the players can easily regain good social standing when under the B designation. In fact, their ability to manipulate the social descriptor at low cost *undermines* their prospect of sustaining cooperation. Consider, for example, a negotiation equilibrium in which the players select (H,H) in period t, following some history that ends with the G designation. To support (H,H), the players must rely on a punishment starting in period $t+1$ in the event one or both players selects L in period t. However, if γ is small then, in such a contingency, the players could agree to a new equilibrium in which (R,R) is played in period $t+1$ and cooperation is re-started at $t+2$. One would expect — and we verify this below — that the new equilibrium is in conflict with the old one. As a result, the only coherent equilibrium is the one yielding the value $1/(1-\delta)$ for each player. The case of γ small thus has the same flavor as the case of the standard conditioning system: when players can easily reinterpret history in the context of the conditioning system, conflicts render cooperation impossible to sustain.

We next provide the formal details for this example. Note that since the conditioning environment consists of the single system (X_r, μ_r), to find the coherent equilibria we simply need to compute $P(E^N(X_r, \mu_r))$. We consider separately three ranges of the cost γ. The following fact proves useful in the analysis of the cases. For each $x \in X_r$ ending with the G designation, there is a negotiation equilibrium e satisfying $v^e(x) = 4/(1-\delta)$. In other words, we can construct an equilibrium in which the players select (H,H) on the equilibrium path in each period following x. To support cooperative behavior, we specify that if a player deviates in the future (inducing designation B) then the players revert to (L,L) thereafter.[15] The bound on the discount factor implies that the players prefer not to unilaterally deviate from this prescription.

Case 1: $\gamma < (2-\delta)/\delta$. Let $\underline{e}$ denote the negotiation equilibrium in which the players avoid default, but make no transfer, in the negotiation phase and always select (L,L) in the action phase. We have $v^{\underline{e}}(x) = \underline{u}/(1-\delta) = 2/(1-\delta)$ for every $x \in X_r$. We shall demonstrate that $v^{\underline{e}}$ is the coherent equilibrium value function. First note that Theorem 1 implies $e \succeq \underline{e}$ for each $e \in E^N(X_r, \mu_r)$. This means $\underline{e}$ is a minimal equilibrium and that all minimal equilibria have the same value function. Then consider any equilibrium e with the property that $v^e(x') > \underline{u}/(1-\delta)$ for some $x' \in X_r$. There must be a state y such that $\sigma^e(y)((\text{H,H})) > 0$; that is, e supports the play of (H,H) in the action phase from y. We shall construct another negotiation equilibrium f that conflicts with e. Figure 5 illustrates the steps taken.

Observe that y must end with the G label; otherwise, all action profiles other than (R,R) induce the designation B in the following period (and the same continuation payoff), which makes it impossible to motivate a player to select H. In addition, it must be that $4 + \delta v_i^e(yB) \leq 2/(1-\delta)$, for otherwise player i has the strict incentive to choose L in state y, inducing the B label in the following period. On the right side of this inequality is the maximum possible continuation payoff for player i; this

[15] Actions in histories which are not successors of x can be specified to produce an equilibrium.

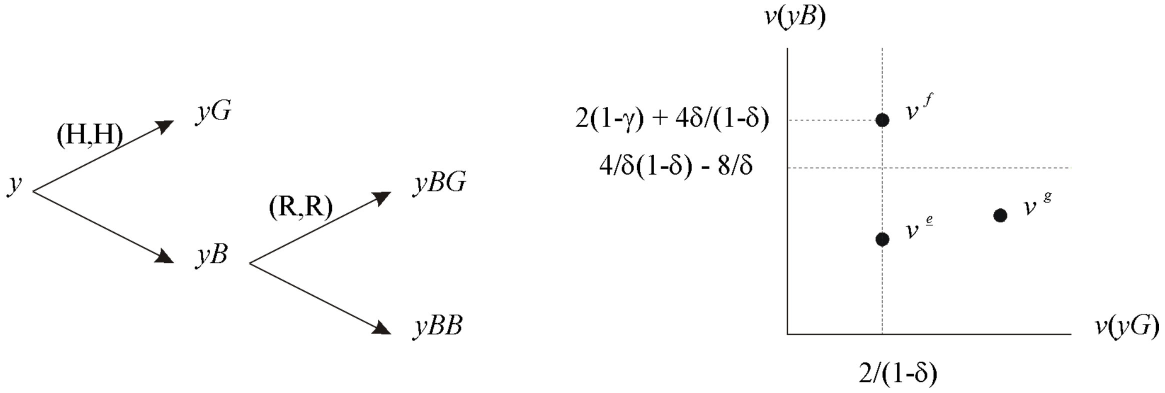


Figure 5: Conflict in the reconciliation example.

bound comes from the fact that $v \leq 4/(1-\delta)$ and, from Theorem 1, $v_i = [v - 2/(1-\delta)]/2 + 1/(1-\delta) = v/2$. In terms of total value,

$$8 + \delta v^e(yB) \leq 4/(1-\delta). \tag{1}$$

It must also be the case that

$$v^e(yG) > 2/(1-\delta). \tag{2}$$

To see this, note that $v^e(yB) \geq \underline{u}/(1-\delta) = 2/(1-\delta)$. Since L strictly dominates H in the stage game, (2) must hold if players have the incentive to play H in state y.

As noted above, we can find an equilibrium f that yields $v^f(yBG) = 4/(1-\delta)$. Also, we can specify that f prescribe (L,L) in every period following state yG as well as in all periods following state yBB. Thus,

$$v^f(yG) = v^f(yBB) = 2/(1-\delta). \tag{3}$$

Furthermore, we can assume that f prescribes (R,R) in the period starting in state yB. In this regard, one can easily check incentive compatibility at yB; it follows from $\delta \geq 2/3$ and $\gamma < (2-\delta)/\delta$. Thus, we have $v^f(yB) = 2(1-\gamma) + 4\delta/(1-\delta)$. The bound on γ then implies

$$8 + \delta v^f(yB) > 4/(1-\delta). \tag{4}$$

From Equations 1, 2, 3, and 4, we have $v^e(yG) > v^f(yG)$ and $v^e(yB) < v^f(yB)$. Therefore, e and f are in conflict.

Case 2: $\gamma \in [(2-\delta)/\delta, \delta/(1-\delta)]$. In this case, one can easily verify that the following strategy specifies an equilibrium. In any state ending with G, the players select (H,H). In any state ending with B, the players choose (R,R). The players always avoid default and agree to no transfer in the negotiation phase. Call this equilibrium $\overline{e}$. The equilibrium has the *uniform best* property mentioned above: it is maximal in the set $E^N(X_r, \mu_r)$ in that $\overline{e} \succeq e$ for every $e \in E^N(X_r, \mu_r)$. To see this, note that the value of $\overline{e}$ in a state ending with B is greater than the value of playing (L,L) perpetually. This greater value prevails because, while γ is large enough to deter cheating under a G designation, γ is small enough to give the players the joint incentive to reconcile when in a state ending with B. The only incentive compatible way for the players to achieve more than $\underline{u}$ in a state ending with B is to choose (R,R) in such a state. But then $\overline{e}$ supports the greatest such value. We conclude that $\overline{e}$ is pivotal and thus it is the coherent equilibrium.

Case 3: $\gamma > \delta/(1-\delta)]$. This case is very similar to the example of a grim conditioning system and to Case 2. There is an equilibrium $\overline{e}$ that is defined as in Case 2, except that it specifies (L,L) in every state ending in B. In this case, the reconciliation cost is so great that the players have no joint incentive to play (R,R), regardless of the continuation payoff that it would generate. As in Case 2, $\overline{e}$ is uniformly best and thus is pivotal on X_s.

In summary, the reconciliation convention can facilitate the cooperative outcome, but only if the reconciliation cost is great enough. If the cost is too small then the players find it too easy to satisfy society's definition of a relationship in good standing, leading to a situation in which the convention has no value. In other words, the players attribute value to the social convention insofar as it is costly for the players to manipulate their social designation.[16]

The "Forgetting" Convention: Here there are K distinct B designations, B^1 through B^K. When a relationship loses good standing, it is called "bad" for K periods, at which point it regains the G designation. In other words, this social convention consists of a language to describe whether a relationship is good or bad, as well as the number of periods (up to K) in which the relationship has been bad. A state in period t is a sequence of labels from $\{G, B^1, \ldots, B^K\}$ for the periods through t, with the following constraints. First, G is the designation in the first period. Second, if B^k is the designation in some period t', and if $k < K$, then B^{k+1} is the designation in period $t'+1$. Third, if B^K is the designation in some period t', then G is the designation in period $t'+1$. The convention specifies a transition that maintains designation G if (H,H) is played in the current period; otherwise, the transition is from G to B^1. From a state ending in B^k, with $k < K$, the label in the next period is automatically B^{k+1} regardless of the players' actions. From a state ending in B^K, the next label is G.

With the forgetting convention, all negotiation equilibria specify (L,L) in bad states. Cooperation in the good state can be sustained if K is sufficiently large. To be precise, if $3\delta - \delta^{K+1} \geq 2$ then the coherent equilibrium entails play of (H,H) in all states ending with the good designation. This equilibrium yields the value $4/(1-\delta)$ in states ending with G. On the other hand, if $3\delta - \delta^{K+1} < 2$ then the coherent equilibrium prescribes (L,L) in every state. The social convention in this example is

[16] Ramey and Watson (1997) consider how this issue arises in the context of dispute resolution systems.

valuable if the society demands a long enough waiting period after misdeeds before allowing a relationship to be called good again.

Stability of Institutions

In the analysis above, the conditioning environment is taken as fixed. That is, social institutions are considered static and a primitive in the players' contractual relationship. In this subsection, we briefly consider whether the conditioning system is "stable" in the sense that it best serves society, as well as individual relationships. In other words, we ask whether there might be pressure to change or abandon a social convention. To conduct a truly satisfactory inquiry, we should address the process by which the social convention is created and maintained, and how it may evolve over time. While a full analysis of this issue is beyond the scope of this paper, we offer a modest result here in the hopes of stimulating further research on institutional design and evolution. More broad intuition and indications for further research appear in the next section.

In a society with many active relationships, the pressure to change the convention is related to the aggregate benefit of doing so over time. One can imagine that, given some convention, there may be a time at which a large fraction of the relationships in society would benefit from developing a new conditioning system. We look for conditioning systems that resist such pressure. For simplicity, we suppose the conditioning environment consists of just one system.

Take as given a stage game. Let u^* be the supremum continuation value in a coherent equilibrium, over all states and conditioning systems. That is, u^* is the greatest value obtainable when one is allowed to arbitrarily select the conditioning system. We call a conditioning system (X, μ) *superior* if it induces a coherent equilibrium e^* such that $v^{e^*}(x) = u^*$ for every state x on the equilibrium path of e^*. If society adopts the convention represented by such a conditioning system, then at no time is there pressure for the members of society to alter it. The following theorem establishes the existence of such a conditioning system. It also demonstrates that the ideal can be achieved by a simple system of the "grim" form. Call a conditioning system *simple* if (a) it specifies two labels, G and B, in each period; (b) states are sequences of labels over time; and (c) the B label is absorbing in the sense that, regardless of play, a B designation in period t implies a B designation in period $t + 1$.

Theorem 7 *For any long-term contractual relationship, if δ is sufficiently large then there is a conditioning system that is simple and superior.*

That the conditioning system is simple means the language used by the players need not be complicated. However, we conjecture that more sophisticated, reconciliation-style conventions would be required in settings with heterogeneous relationships and/or noise in the course of play.

5 An Extension of the Coherence Concept

In this section we develop a version of the coherence concept incorporating a form of backward induction on meta-level negotiation.[17] Our extension is based on the idea that, at each point in the game, certain states of the relationship may not be relevant to the players' joint selection problem. Specifically, if a state y is unreachable from state x then it ought not influence the equilibrium selection at x. We also capture the notion that the players essentially face the same selection problem at any two equivalent states; that is, equivalent states pose the players with the same "configuration" of the contract environment. Thus, in evaluating the players' joint selection of equilibrium, we focus on unions of equivalence classes (while still keeping individual states as the unit of analysis). To formalize our extended definition, we first must define generalized versions of the concepts and operators of Section 3 in order to make reference to a given set of histories $Y \subset X_s$. The analysis then parallels the exposition of Section 3.

We write $e \succeq_Y f$ if $v^e(x) \geq v^f(x)$ for every $x \in Y$; the relations $\succ_Y$ and $\sim_Y$ are similarly defined. Further, we say that equilibria e and f *are in conflict on* Y if neither $e \succeq_Y f$ nor $f \succeq_Y e$. Given a set of equilibria $E \subset E^N$, equilibria without conflicts are those in

$$\hat{E} \equiv \{e \in E \mid \text{for every } e' \in E, \text{ either } e' \succeq_Y e \text{ or } e \succeq_Y e'\},$$

while the set of minimal equilibria is

$$\underline{E} \equiv \{e \in E \mid \text{there is no } f \in E \text{ such that } e \succ_Y f\}.$$

Definition 10 *Given $E \subset E^N$ and $Y \subset X_s$, an equilibrium e is called* **pivotal on** *Y if $e \in \underline{E} \cup \hat{E}$ and there is no equilibrium $e' \in \underline{E} \cup \hat{E}$ with $e' \succ_Y e$. Let $P(E;Y)$ denote the set of equilibria that are pivotal on Y.*

Lemma 2 *If $E \neq \emptyset$ and $V(E)$ is compact then $P(E;Y)$ is nonempty. Furthermore, either (a) $e, f \in P(E;Y)$ implies $e \sim_Y f$, or (b) $P(E;Y) = \underline{E}$.*

Next we redefine operators P and Q. Given $(X,\mu) \in \mathcal{C}$ and $E \subset E^N$, let $E(X,\mu)$ denote the set of equilibria in E that are (X,μ)-measurable. For any set $Y \subset X_s$, define

$$P_{\mathcal{C}}(E;Y) \equiv \bigcup\{P(E(X,\mu);Y) \mid (X,\mu) \in \mathcal{C}\}$$

and

$$Q_{\mathcal{C}}(E;Y) \equiv P(P_{\mathcal{C}}(E;Y);Y).$$

Lemma 3 *Given $E \subset E^N$ and $Y \subset X_s$, if $V(E)$ is compact and if $E(X,\mu) \neq \emptyset$ for some $(X,\mu) \in \mathcal{C}$, then $Q_{\mathcal{C}}(E;Y)$ is nonempty.*

[17] Note that standard backward induction is captured in the equilibrium and negotiation equilibrium concepts.

With the augmented definitions in hand, we turn to address backward induction on meta-level negotiation. We begin by making the following additional technical assumption.

Assumption 4 *The conditioning system has a finite number of equivalence classes.*

Given (X, μ), let $\Lambda(X, \mu)$ be defined as the set of unions of equivalence classes for this conditioning system. That is, $Y \in \Lambda(X, \mu)$ if for some equivalence classes $Y^1, Y^2, \ldots, Y^K \subset X$ it is the case that $Y = Y^1 \cup Y^2 \cup \cdots \cup Y^K$. We call a set $Y \subset X$ *isolated* if $x \in R(Y; X, \mu)$ implies $x \in Y$; in this case, we can write $Y = R(Y; X, \mu)$. An isolated set has the property that no states outside of it can be reached from states in the set. Define $\Lambda^I(X, \mu)$ to comprise the sets in $\Lambda(X, \mu)$ which are isolated. The following facts are easily derived: $Y, Z \in \Lambda^I(X, \mu)$ implies $Y \cup Z \in \Lambda^I(X, \mu)$ and $Y \cap Z \in \Lambda^I(X, \mu)$.

To put different conditioning systems on the same playing field, we can describe the isolated sets of any conditioning system in terms of the standard system. For example, for any set $Z \in \Lambda^I(X, \mu)$, $\beta^{-1}(Z; X, \mu)$ is the subset of X_s corresponding to Z. It is easy to see that $\beta^{-1}(Z; X, \mu)$ is itself isolated in this case. Let

$$\lambda(X, \mu) \equiv \{\beta^{-1}(Z; X, \mu) \mid Z \in \Lambda^I(X, \mu)\}.$$

Then define the collection $\Omega(\mathcal{C})$ to be the set generated by $\{\lambda(X, \mu) \mid (X, \mu) \in \mathcal{C}\}$, using the operations of union and intersection.

Lemma 4 *$\Omega(\mathcal{C})$ is a finite collection of sets.*

With the dependence on $\mathcal{C}$ understood, we write Ω.

We incorporate backward induction as follows. Take a negotiation equilibrium e and suppose the players are evaluating e in the context of a set of histories $Y \in \Omega$. Furthermore, suppose there is another $Z \in \Omega$ that is a proper subset of Y. The players know they will be re-evaluating and defending equilibrium e in the context of Z at a later contingency, because at that point they will have entered a smaller isolated set under some conditioning system. That is, the "configuration" of the conditioning environment will have changed once Z is reached. Therefore, at Y the players need only defend e against alternatives that specify the same values for states in Z. The comparison set is given by:

$$\Gamma(e, Y) \equiv \{f \in E^N \mid f \sim_Z e \text{ for every } Z \in \Omega \text{ satisfying } Z \subset Y \text{ and } Z \neq Y\}.$$

This leads to our extended definition.

Definition 11 *A regime $e = (\theta, \sigma)$ is called a* **$*$-coherent equilibrium** *if $e \in E^N$ and $e \in Q_{\mathcal{C}}(\Gamma(e, Y); Y)$ for each $Y \in \Omega$.*

Theorem 8 *There exists a $*$-coherent equilibrium.*

The proof of theorem 8 identifies an inductive procedure for constructing the set of coherent equilibria. One first analyzes the smallest members of Ω, refining the set of equilibria on the basis of the pivotal requirement. Then one studies larger and larger sets, repeating the application of the pivotal condition. Since Ω is finite, the procedure stops after a finite number of steps. Existence at each stage of the procedure relies on compactness and a guarantee that equilibria can be found which are simultaneously pivotal on disjoint sets of histories.

Obviously the set of $*$-coherent equilibria coincides with the set of coherent equilibria if $\Lambda^I(X,\mu) = \{X\}$ for each $(X,\mu) \in \mathcal{C}$. In this case, there are no isolated sets of equivalence classes other than the entire set of histories. It is easy to see that the definitions also coincide if for any two negotiation equilibria e and f, the following holds: for every $Y \in \Omega(\mathcal{C}) \setminus X_s$ and $x \in Y$, $v^e(x) = v^f(x)$. Thus, the definitions coincide for all of the examples of the preceding section. More generally, the relationship between the two definitions is complicated; moreover, the conclusions of Theorem 3 do not hold generally for the $*$-coherence notion.[18] Preliminary study reveals some strong, yet reasonable, conditions under which Theorem 3 extends to $*$-coherent equilibrium. However, a thorough investigation is left for future work.

6 Conclusion

We have proposed a new perspective on contracting in long-term relationships, emphasizing the role of social institutions in a context of renegotiation by the players. We describe how social convention provides players with a system for classifying and codifying the history of their relationship. The social descriptors have intrinsic value in a conditioning environment, even when the players have the means for conditioning arbitrarily on history. Players adopt the social conditioning system because doing so establishes boundaries on the players' recurrent negotiation that are useful in every contingency of their relationship. The basis of our model is a new theory of recurrent negotiation, combining a bargaining technology with a selection criterion akin to internal consistency. We show that coherent equilibria always exist. Further, the coherent equilibrium value function is unique in a wide range of settings. We also demonstrate the existence of simple social conventions with attractive stability properties.

We conclude with several comments on interpretation, application, and extension of our theory.

1. One might wonder about an alternative theory based on the idea that the social convention suggests a *strategy* to the players, instead of a language for organizing

[18] We fail to obtain uniqueness in environments where non-trivial isolated states create nonstationarities. For example, outside the setting of Theorem 3, one can construct examples in which α^* cannot be sustained in the stage game. In this case, there may be several minimal equilibria for some sets of states under consideration. As noted in Figure 3, one can then get several different pivotal equilibria and thus $*$-coherent equilibria with different value functions.

history. Indeed, there is a sense in which a valuable conditioning system induces a particular strategy to be adopted by the players. However, if one views a convention only on the level of a strategy, one is still left to provide a theory of how the convention is internalized into the decision-making process of individual players. Our model provides a sensible framework for developing an understanding of how social convention influences strategic interaction, in which the novelty and strength of the theory comes from the interpretation of convention as a conditioning system. In addition, it seems natural that the practical value of a convention lies in its ability to shape the way people think, communicate, and evaluate history. It is less plausible that conventions actually prescribe strategies of play. Our approach is even more compelling in settings involving heterogeneity of relationships, where the social convention can provide a general guideline for behavior but cannot reasonably define a strategy for each game played in the population. Such settings represent an interesting area for future research. In addition, it would be worthwhile to study the implications of introducing noise in the action phase and randomness in the state transition; adding these factors would certainly alter the optimal conditioning system.

2. The size of the society affects the credibility of the social convention. We believe that a large society is important because it renders the conditioning system exogenous to an individual relationship. For example, if the "society" consists of only the two players in a relationship, then they could invent the conditioning system themselves, and reinvent it at any time. As noted above, this situation indicates the standard conditioning system, which has no value due to the players' ability to manipulate the interpretation of history. A large society therefore lends credibility to the convention. Communication between members of society (across relationships) may also be important, because it confirms the convention.

3. It may be useful to think of the convention as manipulable at a cost. That is, it may be possible for a society to actively change its conditioning system by bringing its members together in large-scale negotiation. If the cost of bringing a large number of people together is prohibitive, then the social convention is likely not to be overthrown on a whim. In this context, the convention is a form of *community enforcement*, where third parties to a relationship reaffirm the conditioning system for the benefit of the players in the relationship. However, it is not an example of direct external enforcement, since third parties do not impose direct costs.

4. In some settings, the social convention may rely on verifiability of information about a relationship to outside parties. That is, players may find a label credible only if it represents the actual judgment of a third party. In this case, the convention can be interpreted as a dispute resolution system. Players condition their behavior on the judgment of an intermediary, whose designation for the relationship depends on evidence about past behavior that is provided by the players. For example, the "reconciliation" example in the previous section can be interpreted as an example of intermediation, where the third party keeps track of whether the relationship is in dispute and offers to certify dispute resolution at a cost. Here the value of a third

party goes beyond what can be enforced externally. Certification can be valuable on the basis of the implied conditioning system. Ramey and Watson (1997) pursue this application further. Among other things, they compare the benefits of costly certification with the benefits of monitoring and external enforcement of transfers. They also investigate the optimal dispute resolution systems, including whether it pays a sub-population of agents to bypass a general system by creating their own specific system. This direction of research also seems promising at the level of the present model, in particular with regard to noise and heterogeneity.

A Proofs

Proof of Theorem 1:

First we prove that $E^N \neq \emptyset$. There are two cases to consider. Case (1): There is a Nash equilibrium α of the stage game (A, u), such that $u(\alpha) \geq \underline{u}$. In this case, specify $\sigma(x) = \alpha$ for all $x \in X_s$. With θ defined as follows, one can easily check that (θ, σ) is a negotiation equilibrium. For each x, $\theta(x)$ specifies no default and a transfer $m \in M$ to satisfy $u_i(\alpha) + m_i = \underline{u}_i + \pi_i[u(\alpha) - \underline{u}]$. Case (2): No such Nash equilibrium exists. Here, let α be any Nash equilibrium of the stage game. Specify $\sigma(x) = \alpha$ and $\theta(x) = \underline{d}$, for every $x \in X_s$. It is obvious that (θ, σ) is a negotiation equilibrium.

In both of these cases, we have constructed a negotiation equilibrium specifying the same behavior after every history. Thus, the equilibrium is (X, μ)-measurable for every $(X, \mu) \in \mathcal{C}$, proving $E^N(X, \mu) \neq \emptyset$.

Next we prove the value characterization. Take any negotiation equilibrium and let v_i and v denote the associated value functions. Define $s(x) \equiv v(x) - w(x)$ for each $x \in X_s$. By (i) in the definition of negotiation equilibrium, we have $v_i(x) = w_i(x) + \pi_i s(x)$. Also, $w_i(x) = \underline{u}_i + \delta v_i(\mu_s(\underline{a}, x))$, so

$$v_i(x) = \underline{u}_i + \pi_i s(x) + \delta v_i(\mu_s(\underline{a}, x)).$$

Applying this identity inductively, we have for each positive integer K,

$$v_i(x) = \sum_{k=1}^{K} \left[\underline{u}_i + \pi_i s(x^k)\right] \delta^{k-1} + \delta^K v_i(x^{K+1}),$$

where $\{x^1, \ldots, x^{K+1}\}$ is defined by $x^1 = x$ and $x^{k+1} = \mu_s(\underline{a}, x^k)$. Each player has the option of inducing default perpetually, which means $v_j \geq \underline{u}_j/(1-\delta)$ for each j. This further implies that v_i is uniformly bounded. Thus, letting $K \to \infty$, we have $v_i(x) = \underline{u}_i/(1-\delta) + \pi_i S(x)$, where $S(x) \equiv \sum_{k=1}^{\infty} s(x^k)\delta^{k-1}$. Summing over i, we have $v(x) = \underline{u}/(1-\delta) + S(x)$. Using this expression to substitute for $S(x)$ in the previous expression completes the characterization of the value function.

Our last step involves proving that $V(E^N(X, \mu))$ is compact for each $(X, \mu) \in \mathcal{C}$. For any negotiation equilibrium e, write (θ^e, σ^e) as the regime and define $\rho^e(x) = 1$ if $\theta^e(x) = \underline{d}$ and $\rho^e(x) = 0$ otherwise. We shall refer to the following as *Fact (*)*:

> Take as given a conditioning system (X, μ). For any sequence $\{f^k\} \subset E^N(X, \mu)$ and any t, we can find a subsequence $\{\overline{f}^j\}$ and functions $\overline{\rho}^t : X_s^t \to \{0, 1\}$ and $\overline{\sigma}^t : X_s^t \to \Delta^u A$, such that $\rho^{\overline{f}^j}(x) = \overline{\rho}^t(x)$ for every j and every $x \in X_s$, and $\sigma^{\overline{f}^j}(x) \to \overline{\sigma}^t(x)$ as $j \to \infty$, for all $x \in X_s$.

This is true because X_s^t is finite and $\{0,1\} \times \Delta^u A$ is compact.

Fix $(X, \mu) \in \mathcal{C}$ and take any sequence $\{e^k\} \subset E^N(X, \mu)$. Form the subsequence $\{g^k\}$ inductively as follows. Define $h^{0,k} \equiv e^k$, for all k. Using Fact $(*)$, we can find a subsequence of $\{h^{0,k}\}_{k=1}^{\infty}$ with the properties specified for $t = 1$. Let $\{h^{1,j}\}_{j=1}^{\infty}$ be such a subsequence. Then, for any $t \geq 1$, given $\{h^{t,k}\}_{k=1}^{\infty}$, define $\{h^{t+1,j}\}_{j=1}^{\infty}$ to be a subsequence of $\{h^{t,k}\}_{k=2}^{\infty}$ that has the properties identified by Fact $(*)$ for $t+1$. We define $\{g^k\}$ by $g^k \equiv h^{k,1}$ for each k. By construction, $\{g^k\}$ is a subsequence of $\{e^k\}$.

Define $\overline{\rho}$ and $\overline{\sigma}$ by $\overline{\rho}(x) = \overline{\rho}^t(x)$ and $\overline{\sigma}(x) = \overline{\sigma}^t(x)$ for each $x \in X_s^t$ and each t. We have that $\rho^{g^k}(x) \to \overline{\rho}(x)$ and $\sigma^{g^k}(x) \to \overline{\sigma}(x)$, for every $x \in X_s$. This implies that $\theta^{g^k}(x) \to \overline{\theta}(x)$ for all $x \in X_s$, for some function $\overline{\theta} : X \to D$. Letting $\overline{v}$ be the value function associated with $(\overline{\theta}, \overline{\sigma})$, it is obvious that $v^k \to \overline{v}$. The weighted sup norm is critical here, so that small weight is placed on differences between v^k and $\overline{v}$ for sufficiently large t.

Finally, we must show that $(\overline{\theta}, \overline{\sigma}) \in E^N(X, \mu)$. Individual rationality is satisfied due to the convergence of the value functions. Incentive compatibility requires that $\overline{\sigma}(x)$ specify a Nash equilibrium of the static game induced by the stage game with continuation values defined by $\overline{v}_i$. This is implied by upper hemi-continuity of the Nash equilibrium correspondence, along with the fact that g^k prescribes an equilibrium conditional on each x. We know $(\overline{\theta}, \overline{\sigma})$ is (X, μ)-measurable since it is the limit of (X, μ)-measurable regimes. *Q.E.D.*

Proof of Lemma 1:

We divide the proof into four claims.

(1) First we show that $V(\hat{E})$ is compact. Suppose $\hat{E} \neq \emptyset$. Take any sequence in $\hat{E}$. By compactness of $V(E)$, there exists a subsequence $\{f^l\}$ and an equilibrium $f \in E$ such that $v^{f^l} \to v^f$. Presume $f \notin \hat{E}$. Then there exists $g \in E$ such that f and g are in conflict. That is, for some $y, y' \in X_s$, $v^f(y) > v^g(y)$ and $v^g(y') > v^f(y')$. Let

$$\varepsilon \equiv \frac{1}{2} \min \left\{ v^f(y) - v^g(y), v^g(y') - v^f(y') \right\}.$$

By the definition of the weighted sup norm, convergence of v^{f^l} to v^f implies convergence of $v^{f^l}(x)$ to $v^f(x)$ for all $x \in X_s$. Thus, we can find L such that

$$v^{f^L}(y) > v^f(y) - \varepsilon > v^g(y)$$

and

$$v^{f^L}(y') < v^f(y') + \varepsilon < v^g(y').$$

This means f^L and g are in conflict, which is a contradiction.

(2) Next we prove that $\hat{E}$, if nonempty, attains a maximum using the order of dominance; that is, there exists $f \in \hat{E}$ such that $f \succeq e$ for all $e \in \hat{E}$. We start by

enumerating the (countable) set X_s, defining $\{x^k\} \equiv X_s$. Let $\overline{v}(x) \equiv \sup\{v^e(x) \mid e \in \hat{E}\}$, for each $x \in X_s$. This is finite. Define $\{f^k\}$ inductively as follows.

First, to define f^1, we find a sequence $\{e^l\} \subset \hat{E}$ such that $v^{e^l}(x^1) \to \overline{v}(x^1)$. Since $V(\hat{E})$ is compact, this sequence has a subsequence whose value functions converge. Let f^1 be such that v^{f^1} is the limit of the subsequence of value functions. Since x^1 has positive weight in the sup norm, it must be that $v^{f^1}(x^1) = \overline{v}(x^1)$.

Next, suppose $f^1, f^2, \ldots, f^{k-1} \in \hat{E}$ and $v^{f^l}(x^j) = \overline{v}(x^j)$ for each $l \in \{1, 2, \ldots, k-1\}$ and each $j \in \{1, 2, \ldots, l\}$. Define f^k as follows. Let

$$G \equiv \{e \in \hat{E} \mid v^e(x^j) = \overline{v}(x^j) \text{ for all } j = 1, 2, \ldots, k-1\}$$

and let $\phi \equiv \sup\{v^e(x^k) \mid e \in G\}$. Obviously $V(G)$ is nonempty (since $f^{k-1} \in G$) and compact. Therefore we can find $f^k \in G$ satisfying $v^{f^k}(x^k) = \phi$. (f^k is defined so that v^{f^k} is the limit of an appropriately defined sequence, using the method employed in the previous paragraph.) Further, it must be that $\phi = \overline{v}(x^k)$. To see this, note the implications of $\phi < \overline{v}(x^k)$: there would exist $g \in G$ such that $v^g(x^k) > v^{f^k}(x^k)$ yet $v^g(x^l) < \overline{v}(x^l) = v^{f^k}(x^k)$ for some $l \in \{1, 2, \ldots, k-1\}$ (since $g \in G$); but this would contradict $f^k \in \hat{E}$.

The sequence $\{f^k\}$ has two noteworthy properties. First, $f^k \in \hat{E}$ for every k. Second, for every l and each $k \geq l$, $v^{f^k}(x^l) = \overline{v}(x^l)$. Since $V(\hat{E})$ is compact, there exists $f \in \hat{E}$ and a subsequence of $\{f^k\}$ such that the value functions of the subsequence converge to v^f. By construction, $v^f(x) = \overline{v}(x)$ for every $x \in X_s$, which means $f \succeq e$ for all $e \in \hat{E}$.

(3) We now show that $\underline{E} \neq \emptyset$. The method of prove uses a construction similar to the one from claim (2). We form a sequence of sets $\{F^k\} \subset E$ as follows. Enumerate X_s as in the proof of claim (2). Define $F^0 \equiv E$. For $k \geq 1$, define

$$F^k \equiv \{e \in F^{k-1} \mid v^e(x^k) \leq v^f(x^k) \text{ for all } f \in F^{k-1}\}.$$

By construction, F^k is nonempty and $V(F^k)$ is compact. Let $\{f^k\}$ be any sequence with $f^k \in F^k$ for each k. Compactness of $V(E)$ implies the existence of an equilibrium $f \in E$ such that $v^{f^k} \to v^f$. We claim that $f \in \underline{E}$. Suppose not. Then there exists $g \in E$ such that $f \succ g$. Let l equal the smallest k satisfying $v^g(x^k) < v^f(x^k)$. It must be that $g \in F^{l-1}$. Also, by the definition of F^l, we have $v^e(x^l) = v^f(x^l)$ for all $e \in F^l$. But this contradicts the construction of $\{F^k\}$.

(4) Finally, we note that $e \succeq f$ for every $e \in \hat{E}$ and every $f \in \underline{E}$. To see this, suppose it were the case that $e \not\succeq f$. If $f \succ e$ then f would not be an element of $\underline{E}$. Otherwise, e and f are in conflict, which would contradict $e \in \hat{E}$.

These claims imply existence. Furthermore, if $\hat{E} \neq \emptyset$ then we have situation (i); otherwise, we have situation (ii). *Q.E.D.*

Proof of Theorem 2:

Theorem 1 and Lemma 1 imply $P_{\mathcal{C}} \neq \emptyset$, $\underline{P_{\mathcal{C}}} \neq \emptyset$, and that $\hat{P}_{\mathcal{C}}$ contains only a finite number of negotiation equilibria. These facts are sufficient for the result. *Q.E.D.*

Proof of Theorem 3:

Let $\underline{e} = (\underline{\theta}, \underline{\sigma})$ denote the negotiation equilibrium in which α^* is played after every history. That is, for all $x \in X_s$, $\underline{\sigma}(x) = \alpha^*$ and $\underline{\theta}(x)$ specifies a transfer $m \in M$ to satisfy $u_i(\alpha^*) + m_i = \underline{u}_i + \pi_i[u(\alpha^*) - \underline{u}]$. Obviously $\underline{e}$ is a minimal equilibrium for every conditioning system, so that $f \succeq \underline{e}$ for every $f \in E^N$. We proceed by noting three facts.

> *Fact 1: Given $E \subset E^N$, if $\hat{E} \neq \emptyset$ then $P(E) \subset \hat{E}$ and $e \sim f$ for all $e, f \in P(E)$.*

To see this, note that if $\hat{E} \neq \emptyset$ then $e \succeq f$ for every $e \in \hat{E}$ and $f \in \underline{E}$. The fact then follows from the definition of pivotal.

> *Fact 2: In the natural default setting, $E^N\widehat{(X,\mu)} \neq \emptyset$ for every $(X,\mu) \in \mathcal{C}$.*

This is true since $\underline{e} \in E^N(X,\mu)$ and $\underline{e}$ has no conflicts in E^N. These two facts imply that at most $|\mathcal{C}|$ (finite) value functions are represented in $P_{\mathcal{C}}(E^N)$. We know $P_{\mathcal{C}}(E^N) \neq \emptyset$ as well.

> *Fact 3: In the natural default setting, $P_{\mathcal{C}}\widehat{(E^N)} \neq \emptyset$.*

To demonstrate this fact, take any $g \in P(E^N(X_f, \mu_f))$. Facts 1 and 2 imply that g has no conflicts in $E^N(X_f, \mu_f)$. By the definition of (X_f, μ_f), every equilibrium in $P_{\mathcal{C}}(E^N)$ is (X_f, μ_f)-measurable. Thus, $g \in P_{\mathcal{C}}\widehat{(E^N)}$.

Recall that $Q_{\mathcal{C}}(E^N) = P(P_{\mathcal{C}}(E^N))$. Facts 3 and 1 imply a unique coherent equilibrium value function.

To prove the second part of the theorem, take any coherent equilibrium e and let $x, y \in X_s$ be completely equivalent histories. Assume $\mathcal{C}$ has the access property. We shall establish that $v^e(x) = v^e(y)$ by proving three claims below. Before addressing the claims, note that there is a conditioning system $(X', \mu') \in \mathcal{C}$ such that $e \in P(E^N(X', \mu'))$. Facts 1 and 2 above imply:

> *Fact 4: e has no conflict in $E^N(X', \mu')$.*

Let $x' \equiv \beta(x; X', \mu')$ and $y' \equiv \beta(y; X', \mu')$ be the states in X' corresponding to x and y. Since e is (X', μ')-measurable, we can represent this regime as a function of X'. We shall consider other (X', μ')-measurable equilibria, which we also write as functions of X'.

Claim 1: If $x \notin R(y)$ and $y \notin R(x)$ then $v^e(x) = v^e(y)$.

Proof: Presume that $v^e(x) \neq v^e(y)$ and we will establish a contradiction. We can find another (X', μ')-measurable negotiation equilibrium f such that $v^f(x) = v^e(y)$ and $v^f(y) = v^e(x)$. To see this, observe that f can be specified so that (a) in the continuation following x', f prescribes exactly the behavior that e prescribes following y', and (b) in the continuation following y', f prescribes exactly the behavior that e prescribes following x'. Since x' and y' are equivalent, this specification is well-defined. There is also no problem regarding the players' incentives following x' and y', since x' and y' are not ordered by precedence. We can specify that f prescribes the same behavior as does e on all other states in periods greater than or equal to $\max\{\tau(x'), \tau(y')\}$. Using backward induction to find incentive-compatible and individually rational decisions on predecessor states (in $X'^1 \cup X'^2 \cup \cdots X'^{t-1}$), we complete the specification of f, which is an (X', μ')-measurable negotiation equilibrium by construction. Note that we have defined f on X', which extends to X_s using $\beta^{-1}(\cdot; X', \mu')$. By construction, e and f are in conflict, which contradicts Fact 4.

Claim 2: If $y \in R(x)$ then $v^e(x) \geq v^e(y)$.

Proof: If $x = y$ then the claim is obvious. Suppose $x \neq y$. Presume that $v^e(x) < v^e(y)$ and look for a contradiction. Since x' and y' are equivalent, we can find an (X', μ')-measurable negotiation equilibrium f that specifies in the continuation from x' what e specifies from y'. We have $v^f(x') = v^e(y')$. It cannot be that f and e are in conflict, for this would contradict that e is pivotal on $E^N(X', \mu')$ (using Facts 1 and 2 above). Since $v^f(x') > v^e(x')$ it must be that $f \succeq e$. In addition, there must be an equilibrium $g \in E^N(X', \mu')$ such that g and f are in conflict.

There are two cases to consider. First, take the case in which $X'^{\tau(x')}$ is a singleton set. Here, conditioning system (X', μ') makes no distinctions between histories prior to the time in which state x' occurs. It should be obvious that, due to the natural default setting, the equilibrium values of predecessors of x' are positively related to $v(x')$.[19] Thus, the conflict between f and g occurs on states in $R(x')$. We can then find another (X', μ')-measurable negotiation equilibrium h that specifies in the continuation from y' exactly what g specifies from x'. We observe that h and e are in conflict (using states in $R(y')$), which contradicts Fact 4.

Next take the case in which $X'^{\tau(x')}$ contains more than one state. Then there must be a state z' that is equivalent to x' and y' and also satisfies $z' \notin R(x')$ and $x' \notin R(z')$.[20] We thus have that $z' \notin R(y')$ and $y' \notin R(z')$ as well. There are two subcases to consider. (i) $v^e(z') \geq v^e(y')$. Here we can use a construction as in the proof of Claim 1 to find an (X', μ')-measurable negotiation equilibrium h with $v^h(z') = v^e(x')$ and $v^h(x') = v^e(y')$. (ii) $v^e(z') < v^e(y')$. Here we construct h so that $v^h(z') = v^e(y')$ and $v^h(y') = v^e(x')$. In both cases h and e are in conflict, contradicting Fact 4.

[19] Only one-shot Nash equilibria can be supported in previous periods and natural default implies that the return in such a period must be $\underline{u}$.

[20] Here we use the access property: that all equivalence classes of states are reachable from every state.

Claim 3: If $y \in R(x)$ then $v^e(x) \leq v^e(y)$.

Proof: Presume $v^e(x) > v^e(y)$ and we shall establish a contradiction. Define $Y \equiv \{z' \in R(x') \mid z' \notin R(y'), y' \notin R(z')\}$. If Y is empty then, as noted in the proof of the previous claim, only one-shot Nash equilibria can be supported in periods prior to $\tau(y')$. Furthermore, y' is reached regardless of the play. Due to natural default, the payoff in each of the first $\tau(y') - 1$ periods is exactly $\underline{u}$. Since $\underline{u}/(1-\delta)$ is a lower bound on the value of every state, it must be that $v^e(x') \leq v^e(y')$, a contradiction.

Next suppose Y is nonempty. It must be that $v^e(z') = \underline{u}/(1-\delta)$ for each $z' \in Y$. Too see this, note that such a value can always be supported at any state z' by specifying play according to $\underline{e}$ on $R(z')$. We can find an (X', μ')-measurable negotiation equilibrium h with $v^h(z') = \underline{u}/(1-\delta)$ and $v^h(y') = v^e(x')$ (using the construction method employed above). If $v^e(z') > \underline{u}/(1-\delta)$ then h and e are in conflict, which contradicts Fact 4. By similar reasoning, we conclude that $v^e(y') = \underline{u}/(1-\delta)$. To see this, note that there exists a state $z' \in Y$ that is equivalent to x' and y'. This follows from the fact that every equivalence class can be reached from every state ($\mathcal{C}$ has the access property). Then if $v^e(y') > \underline{u}/(1-\delta)$, we can find an (X', μ')-measurable negotiation equilibrium h satisfying $v^h(y') = \underline{u}/(1-\delta) < v^e(y')$ and $v^h(z') = v^e(x') > \underline{u}/(1-\delta) = v^e(z')$. Here h and e are in conflict, contradicting Fact 4.

Continuing with the case of Y nonempty, we have shown that $v^e(z'') = \underline{u}/(1-\delta)$ for every $z'' \in X'^{\tau(y')}$. In words, the continuation value from period $\tau(y')$ does not depend on play in previous periods. This implies that only one-shot Nash equilibria can be supported prior to period $\tau(y')$. As in the first case discussed in this proof, it must be that $v^e(x') \leq v^e(y')$, a contradiction.

The three claims imply $v^e(x) = v^e(y)$. *Q.E.D.*

Proof of Theorem 5:

Let e be a coherent equilibrium in the setting with $\mathcal{C}$. From Facts 1,2, and 3 in the proof of Theorem 3, e has no conflict in $P_{\mathcal{C}}(E^N)$. As Theorem 4 indicates, $P(E^N(X_s, \mu_s)) = \{f \in E^N \mid f \sim \underline{e}\}$, where $\underline{e}$ is defined as in the previous proof. This implies that $P_{\mathcal{C}'}(E^N) = P_{\mathcal{C}}(E^N) \cup \{f \in E^N \mid f \sim \underline{e}\}$. Obviously we have $e \succeq \underline{e}$, which implies the result. *Q.E.D.*

Proof of Theorem 6:

The proof follows the same argument as the proof of Theorem 5. *Q.E.D.*

Proof of Theorem 7:

For δ sufficiently large, one can find a grim-trigger style equilibrium that achieves the maximal continuation value from every history on the equilibrium path and uses the best one-shot Nash equilibrium in the action phase as the punishment for deviations. It is easy to construct a simple and superior conditioning system that emits this equilibrium as maximal. *Q.E.D.*

Proof of Lemma 2:

This proof involves exactly the same arguments as does the proof of Lemma 1. One need only use Y in place of X_s throughout and also add the subscript Y where appropriate (such as $\succ_Y$). *Q.E.D.*

Proof of Lemma 3:

That $V(E)$ is compact implies $V(E(X,\mu))$ is compact, for each $(X,\mu) \in \mathcal{C}$. Since $E(X,\mu) \neq \emptyset$ for some conditioning system (X,μ), Lemma 2 implies that $P_{\mathcal{C}}(E;Y) \neq \emptyset$. In addition, for $e, f \in P(E(X,\mu);Y)$, either $e \sim_Y f$ or e and f are in conflict on Y (as members of $\underline{E}(X,\mu)$). The finiteness of $\mathcal{C}$ implies that $P_{\mathcal{C}}(E;Y)$ contains a minimum point and has a finite number of elements without conflicts. These facts are sufficient for the claim. *Q.E.D.*

Proof of Lemma 4:

Finiteness follow from the assumption that $\mathcal{C}$ is finite and every conditioning system has a finite number of equivalence classes. *Q.E.D.*

Proof of Theorem 8:

We prove existence by characterizing the set of coherent equilibria using an inductive construction. Define collections $\{\Omega^k\}_{k=0}^K$ by induction as follows. First, $\Omega^0 \equiv \emptyset$. Then for each k, given Ω^k, define

$$\Omega^{k+1} \equiv \{Y \in \Omega \mid Z \in \Omega^k \text{ for each } Z \in \Omega \text{ satisfying } Z \subset Y \text{ and } Z \neq Y\}.$$

We let K be the first integer for which $\Omega^{K+1} = \Omega^K$, which exists since Ω has a finite number of members. Note that $\Omega^k \subset \Omega^{k+1}$ for each k, and $\cup\{Y \in \Omega^K\} = X_s$. Let $E^0 \equiv E^N(X_f, \mu_f)$ and for each $k \in \{0, 1, \ldots, K\}$ define

$$E^k \equiv \{e \in E^N \mid e \in Q_{\mathcal{C}}(\Gamma(e,Y);Y) \text{ for each } Y \in \Omega^k\}.$$

By definition, E^K is the set of coherent equilibria. Furthermore, $E^{k+1} \subset E^k$, for each k. Note also that $E^k \subset E^N(X_f, \mu_f)$ by the definition of $Q_{\mathcal{C}}$.

We need to prove that $E^K \neq \emptyset$. Since $E^0 = E^N(X_f, \mu_f)$, existence of a coherent equilibrium is then implied by:

Lemma 5 *If $E^k \neq \emptyset$ then $E^{k+1} \neq \emptyset$.*

To prove this lemma, we begin by establishing another useful lemma.

Lemma 6 *Given $e, f \in E^N(X_f, \mu_f)$ and $Y, Y' \in \Omega$, if $e \sim_{Y \cap Y'} f$ then there exists $g \in E^N(X_f, \mu_f)$ such that $g \sim_Y e$ and $g \sim_{Y'} f$.*

Proof of Lemma 6: Instead of presenting all of the components formally, we avoid some complications in notation by sketching some of the intuitive steps. We start with a partially specified regime g' that incorporates e's prescription of behavior for histories in Y and f's prescription for histories in $Y' \setminus Y$. Given $e \sim_{Y \cap Y'} f$, we thus have $g' \sim_Y e$ and $g' \sim_{Y'} f$. In addition, the conditions of a negotiation equilibrium are satisfied by g' for every $x \in Y \cup Y'$. We can then examine a truncated game in which play ends at period T, with the players obtaining "continuation values" given by $v_i^{g'}(x^T)$ for $x^T \in Y \cup Y'$ and zero if $x^T \notin Y \cup Y'$. We can find a regime g^T that conforms to g' on $Y \cup Y'$ and also satisfies the conditions of a negotiation equilibrium for $x \in \bigcup_{t=1}^{T-1} X_s^t$. Such a regime can be found by using a backward induction procedure starting from period $T-1$, where a static equilibrium in the action phase is selected conditional on each state and continuation values. The specification of g^T for $x \in (\bigcup_{t=T}^{\infty} X_s^t) \setminus (Y \cup Y')$ is arbitrary. One can then follow the line of argument used in the proof of Theorem 1 to find a subsequence of $\{g^T\}_{T=1}^{\infty}$, written $\{h^k\}$, such that $\theta^{h^k}(x) \to \theta^g(x)$ and $\sigma^{h^k}(x) \to \sigma^g(x)$, for all x and some regime g. Just as in the other proof, we obtain $g \in E^N(X_f, \mu_f)$. By construction, we have $g \sim_Y e$ and $g \sim_{Y'} f$.

We continue by proving Lemma 5. Suppose $E^k \neq \emptyset$ and take any negotiation equilibrium $e \in E^k$. We have $e \in Q_{\mathcal{C}}(\Gamma(e, Y); Y)$ for each $Y \in \Omega^k$. In addition, it is the case that $V(\Gamma(e; Y))$ is compact, for each $Y \in \Omega$, which follows from compactness of $V(E^N)$. Let $Y^1, Y^2, \ldots, Y^L$ be such that

$$\{Y^1, Y^2, \ldots, Y^L\} \equiv \Omega^{k+1} \setminus \Omega^k.$$

We know L is finite since Ω has a finite number of elements. By Lemma 3, for each $l \in \{1, 2, \ldots, L\}$ we can find h^l such that $h^l \in Q_{\mathcal{C}}(\Gamma(e, Y^l); Y^l)$. We can then find $f^l \in E^N(X_f, \mu_f)$ such that $f^l \sim_{Y^l} h^l$ and $f^l \sim_Z e$ for all $Z \in \Omega^k$. To see this, let $Y = \cup\{Z \in \Omega^k\}$ and $Y' = Y^l$. For each $Z \in \Omega^k$ we have $Z \cap Y^l \in \Omega^k$ (otherwise $Y^l \notin \Omega^{k+1} \setminus \Omega^k$) and thus $h \sim_Z e$ for each $h \in \Gamma(e, Y^l)$. (Note that here we are using the fact that $Y^l, Z \in \Omega$ implies $Y^l \cap Z \in \Omega$.) Therefore $h^l \sim_{Y \cap Y'} e$, so that Lemma 6 implies the existence of the specified f^l.

We have found equilibria $f^1, f^2, \ldots, f^L$ such that $f^l \in Q_{\mathcal{C}}(\Gamma(e, Y^l); Y^l)$ and $f^l \sim_{\cup\{Z \in \Omega^k\}} e$, for each l. We define the equilibria $g^1, g^2, \ldots, g^L$ inductively as follows. Let $g^1 \equiv f^1$. For $l \geq 2$ we find $g^l \in E^N(X_f, \mu_f)$ such that $g^l \sim_{Y^l} f^l$ and $g^l \sim_{Z^l} g^{l-1}$, where $Z^l \equiv \cup\{Z \in \Omega^{k+1} \mid Z \neq Y^l\}$. To see that such a g^l exists for each l, presume $g^{l-1} \sim_Z e$ for every $Z \in \Omega^k$ (which is immediate for $l = 2$). Then note that $Z^l \cap Y^l \subset \cup\{Z \in \Omega^k\}$, since $Z \cap Y^l \in \Omega^k$ for every $Z \in \Omega^{k+1}$ with $Z \neq Y^l$. This implies that $g^{l-1} \sim_{Z^l \cap Y^l} f^l$, at which point Lemma 6 establishes the existence of the specified g^l.

By construction, we have $g^L \sim_Z e$ for each $Z \in \Omega^k$, which implies $g^L \in Q_{\mathcal{C}}(\Gamma(e, Y); Y)$ for each $Y \in \Omega^k$. In addition, $g^L \sim_{Y^l} f^l$ for each $l = 1, 2, \ldots, L$. Thus, $g^L \in Q_{\mathcal{C}}(\Gamma(e, Y^l); Y^l)$ for each l. These facts imply that $g^L \in E^{k+1}$, proving Lemma 5. *Q.E.D.*

References

[1] Abreu, D. and D. Pearce, "A Perspective on Renegotiation in Repeated Games," in *Game Equilibrium Models* (R. Selten, ed.), volume 2, Springer-Verlag, 1991.

[2] Abreu, D., D. Pearce, and E. Stacchetti, "Renegotiation and Symmetry in Repeated Games," *Journal of Economic Theory* 60 (1993): 217-240.

[3] Asheim, G., "Extending Renegotiation-Proofness to Infinite Horizon Games," *Games and Economic Behavior* 3 (1991): 278-294.

[4] Bendor, J. and D. Mookherjee, "Norms, Third Party Sanctions, and Cooperation," *Journal of Law, Economics, and Organization* 6 (1990): 33-63.

[5] Bergin, J. and B. MacLeod, "Efficiency and Renegotiation in Repeated Games," *Journal of Economic Theory* (1993).

[6] Bernheim, D. and D. Ray, "Collective Dynamic Consistency in Repeated Games," *Games and Economic Behavior* 1 (1989): 295-326.

[7] Blume, A., "Intraplay Communication in Repeated Games," *Games and Economic Behavior* 6 (1994): 181-211.

[8] Den Haan. W., G. Ramey, and J. Watson, "Liquidity Flows and Fragility of Business Enterprises," UC San Diego Working Paper 99-07, 1999.

[9] Ellickson, R., *Order without Law: How Neighbors Settle Disputes*, Cambridge: Harvard University Press, 1991.

[10] Farrell, J. and E. Maskin, "Renegotiation in Repeated Games," *Games and Economic Behavior* 1 (1989): 327-360.

[11] Greif, A., "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83 (1993): 857-882.

[12] Greif, A., P. Milgrom, and B. Weingast, "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild," *Journal of Political Economy* 102 (1994): 745-776.

[13] Kandori, M., "Social Norms and Community Enforcement." *Review of Economic Studies* 59 (1992): 63-80.

[14] Klimenko, M., G. Ramey, and J. Watson, "Recurrent Trade Agreements and the Value of External Enforcement," UC San Diego Discussion Paper 2001-01 (2004).

[15] Matsushima, H., “Long-Term Partnership in a Repeated Prisoners’ Dilemma with Random Matching,” *Economics Letters* 34 (1990): 243-248.

[16] Milgrom, P., D. North, and B. Weingast, “The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs,” *Economics and Politics* 2 (1990): 1-23.

[17] Nash, J. F., “The Bargaining Problem,” *Econometrica* 18 (1950): 155-162.

[18] North, D.C., *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, 1990.

[19] Pearce, D., “Repeated Games: Cooperation and Rationality,” Cowles Foundation Discussion Paper 983, 1991.

[20] Pearce, D., “Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation,” Cowles Foundation Working Paper, Yale, 1989.

[21] Ramey, G. and J. Watson, “Contractual Fragility, Job Destruction, and Business Cycles,” *Quarterly Journal of Economics* 112 (1997): 873-911.

[22] Ramey, G. and J. Watson, “Contractual Intermediaries,” *Journal of Law, Economics, and Organization* 18 (2002): 362-384.

[23] Ray, D., “Internally Renegotiation-Proof Equilibrium Sets: Limit Behavior with Low Discounting,” *Games and Economic Behavior* 6 (1994): 162-177.

[24] van Damme, E., “Renegotiation-Proof Equilibrium in Repeated Prisoners’ Dilemma,” *Journal of Economic Theory* 47 (1989): 206-217.