UC Berkeley

Other Recent Work

Title

A Model of Pre-Game Communication

Permalink

<https://escholarship.org/uc/item/902697n7>

Author

Rabin, Matthew

Publication Date

1991-04-01

Peer reviewed

UNIVERSITY OF CALIFORNIA AT BERKELEY

Department of Economics

Berkeley, California 94720

Working Paper No. 91-164

A Model of Pre-Game Communication

Matthew Rabin

University of California at Berkeley

April 1991

Key words: Pre-game communication, Pareto-efficiency

JEL Classification: 026

I thank Vince Crawford, Joe Farrell, Ben Hermalin, Joel Sobel and especially Eddie Dekel for helpful comments on a previous draft.

Abstract

I present a model of players making repeated, simultaneous, cheap-talk statements prior to the play of a two-player, complete-information game. Building from one basic set of assumptions about how players reach voluntary agreements, I define related non-equilibrium and equilibrium solution concepts, as well as solution concepts with "some equilibrium."

I use the model to examine each of two common hypotheses for complete-information games: First, that communication guarantees players will coordinate on Nash-equilibrium outcomes, and second, that within the equilibrium framework, communication guarantees players will coordinate on Pareto-efficient equilibria. I reject the strong versions of each of these hypotheses, but provide a variety of alternative results. With regard to the communication-yields-efficiency hypothesis, a simple result obtains if players can communicate for a long time: In every equilibrium of every game, each player gets a payoff at least as great as that of his worst Pareto-efficient Nash equilibrium.

A Model of Pre-Game Communication⁰

I. Introduction

Many game theorists have over the years motivated basic solution concepts with informal arguments about the effects of non-binding, pre-game communication. One common claim has been that communication will help players coordinate on a Nash equilibrium. Many analysts have conjectured further that it solves coordination problems more generally: sometimes it is argued--and frequently it is assumed--that only Pareto-efficient equilibria are likely in communication-rich environments. (Throughout the paper, I shall use "Pareto-efficient" equilibria to mean those Nash equilibria that are Pareto-efficient among the set of Nash equilibria.)

In contrast to these informal conjectures, communication was until recently omitted from formal game theory. Only Farrell [1988] and Rabin [1990] consider whether communication provides foundations for equilibrium analysis (and their results yield at best weak support for the communication-yields-equilibrium hypothesis). There has been more research studying the effects of communication within the equilibrium framework, but almost no general analysis of the communication-yields-efficiency hypothesis for complete-information games, where the efficiency assumption is widely espoused, and widely applied in specific models.¹

⁰ I thank Vince Crawford, Joe Farrell, Ben Hermalin, Joel Sobel, and especially Eddie Dekel for helpful comments on a previous draft.

¹ Most research on cheap talk has either examined specific economic models, or has focused on incomplete-information games. Specific cheap-talk models

In this paper, I model a form of pre-game communication in two-player, complete-information games. I assume that players make repeated, simultaneous statements before they play the game, allowing them to make voluntary agreements to play particular equilibria during the game. Roughly, I assume that if the players propose the same equilibrium in a given period, they have an agreement to play that equilibrium. If not, they have no agreement. I then construct equilibrium and non-equilibrium solution concepts based on the assumptions about communication.

With this model, I "test" the communication-yields-equilibrium and communication-yields-efficiency hypotheses. I reject the strongest versions of these hypotheses, but present alternative, weaker results about the effects of pre-game communication.

The main results of the paper relate to my equilibrium concept, Negotiated Equilibrium (NGE), with which I investigate the communication-yields-efficiency hypothesis. A NGE is simply a Nash equilibrium (of the complete game, with negotiations) that is consistent with certain assumptions about how players negotiate.²

Some implications of NGE are illustrated in Example 1. This battle-of-the-sexes game represents a common situation, and one that is central to understanding the role of communication in coordinating actions:

include Farrell and Gibbons [1989], Matthews and Postlewaite [1989], and Farrell [1987]. Examples of cheap talk in more general incomplete-information games include Crawford and Sobel [1982], Farrell [forthcoming], Matthews et al [1990], Myerson [1989], and Rabin [1990]. (Myerson [1989] also includes the case of complete-information games.)

² To study what communication achieves by itself--without extra-communicational assumptions about which equilibria are focal--I assume that players might reach voluntary agreements to play any Nash equilibrium. In a previous draft of this paper, I also defined solution concepts that incorporate focal points by maintaining the basic assumptions but limiting the range of equilibria over which the players negotiate, where any subset of Nash equilibria can be chosen as a focal set for negotiations.

Two people wish to coordinate to avoid bad outcomes, but disagree over which superior outcome to play. The mixed-strategy equilibrium of this game--with payoffs of $(2/3, 2/3)$ --is Pareto-dominated by each of the two pure-strategy Nash equilibria--with payoffs $(2, 1)$ and $(1, 2)$ --which are not themselves Pareto-rankable. Next to the normal form of the game is a representation of the three Nash equilibria in expected-payoff space.

One possibility is that the players confidently believe they will play the same efficient equilibrium even if they do not reach an agreement, so that there will be no meaningful negotiations. Because I allow for this possibility, each of the two pure-strategy equilibria will be an outcome consistent with a NGE.

Alternatively, the players may fear that they will play the inefficient equilibrium unless they reach an agreement. In this case, they would try to reach an agreement to play one of the efficient equilibria. Though each player will want to negotiate an agreement to his preferred equilibrium, he would rather concede than reach no agreement. There are, in fact, NGE in which one player concedes for sure to the other player's preferred equilibrium, yielding efficient outcomes. But there is also a symmetric, mixed-strategy equilibrium where each player proposes each of the efficient equilibria with positive probability, doing so each period until either their proposals match or they run out of time. Farrell [1987] first formulated this type of equilibrium in a model of a firm's decision whether to enter a market, and argued that this equilibrium was the most natural in his model. While I shall not treat it as the uniquely natural equilibrium in this game, this symmetric equilibrium is one of the NGE.

The outcome in this equilibrium is that the players agree to each of the efficient equilibria with positive probability, but they also ultimately

disagree with positive probability.³ Indeed, no matter how long the negotiations last, the symmetric equilibrium puts enough weight on the bad equilibrium to yield expected payoffs to the players of less than (1,1). Thus--because this outcome is Pareto-dominated by each of the efficient Nash equilibria--NGE does not guarantee efficiency in this game.

This possibility is more general: In a large class of games, even if players negotiate for a long time, they might not coordinate on an efficient equilibrium. Moreover, the inefficiencies consistent with NGE can in many games be quite severe: abundant communication seems not to justify the conclusion that only Pareto-efficient equilibria will be played.

Yet a happier fact can be observed if we examine Example 1 again. For lengthy negotiations, every NGE in this game yields a payoff to each player that either exceeds 1 or is arbitrarily close to 1. This contrasts with the payoffs of 2/3 for each player possible without communication: While full efficiency cannot be guaranteed, the worst payoffs possible with communication are better than the worst payoffs without communication. Thus, the potential efficiency gains from communication illustrated by Farrell [1987] do not rely on ad hoc assumptions of symmetry or on selecting a particular type of mixed-strategy equilibrium. Rather, the efficiency gains in Example 1 inhere in the basic assumptions about how players use language.

Observe that NGE guarantees in Example 1 that each player does at least as well as in his worst Pareto-efficient Nash equilibrium. The main finding of this paper is that this result is completely general: If players can negotiate for a long time, then in every NGE of every two-player, complete-information

³ As made clear by other papers (e.g., Forges [1986]), an almost immediate consequence of natural models of communication is that the players might play a non-Nash correlated equilibrium in the post-communication game.

game, each player's expected payoff is at least as great as the payoff from his worst Pareto-efficient Nash equilibrium. While my model does not vindicate the strongest (and most widely applied) version of the communication-yields-efficiency hypothesis, it does provide a general, simple, alternative efficiency result.

In pure-coordination games, for instance, NGE implies full efficiency. Perhaps more importantly, in nearly-pure-coordination games NGE implies near-efficiency. In games where each player has strong preferences over which efficient equilibrium is played, however, NGE often excludes only extremely inefficient equilibria. Indeed, in the pure bargaining setting of the Nash bargaining game, NGE involves no restrictions on the set of equilibrium outcomes.

There is a simple intuition for this efficiency result. Suppose a player could preemptively concede, and say to the other player, "Let's not argue; you choose an equilibrium." Then--because one player's least-favorite Pareto-efficient equilibrium is the other player's most-favorite equilibrium--his payoff would be precisely that of his worst Pareto-efficient equilibrium. A player might choose to negotiate tough if he thought he could get a better payoff, but the option to concede preemptively would provide a lower bound on his payoff.

Because players cannot really concede preemptively in a model with simultaneous statements, coordination problems may arise. Neither player may know, for instance, whether the other will concede. With enough periods of negotiations, however, such coordination problems become minimal. With lengthy negotiations, the only reason player 1 (say) would be unable to concede is if player 2 were consistently proposing a more generous agreement--in which case player 1 can do better by not trying to concede. Either way, the result holds.

The preemptive-concession argument won't work with any payoff greater than that of the player's worst Pareto-efficient equilibrium: any "concession" which demands that the other player accept a payoff worse than that of his favorite equilibrium might reasonably be rejected. It turns out that in most games there exist NGE in which each player receives a payoff equal to that of his worst Pareto-efficient Nash equilibrium.

If we turn to investigating the communication-yields-equilibrium hypothesis, we can no longer assume equilibrium behavior.⁴ I present below a non-equilibrium solution concept, Negotiated Rationalizability (NGR), which allows for unlimited strategic uncertainty, much as does rationalizability. I maintain the assumption that players negotiate over Nash-equilibrium agreements, but allow non-equilibrium behavior during negotiations, and--if the players do not reach an agreement--allow non-equilibrium behavior during the game itself.

NGR does not yield equilibrium behavior; indeed, it is rarely more restrictive than rationalizability. Communication seems neither to motivate equilibrium analysis, nor to have much power without it.⁵

This conclusion, however, may be too pessimistic. Rabin [1989] formalizes an approach to making assumptions in between the extremes of equilibrium analysis (which assumes no strategic uncertainty) and the standard non-equilibrium approach (which assumes unlimited strategic uncertainty). I consider in Section IV some tentative implications of this "in-between analysis", and show that the effects of communication can appear when we consider weaker-than-equilibrium restrictions on strategic uncertainty. For

⁴ Except for Farrell [1988], Rabin [1990], and Zapater [1991], all research on cheap talk has assumed that players will play an equilibrium.

⁵ By contrast, Rabin [1990] shows (in a different context) that a non-equilibrium communication concept can restrict play in meaningful ways.

example, full efficiency is guaranteed in pure-coordination games if we assume that either player has correct beliefs about the other player's behavior, even if the other player has incorrect beliefs. I also provide an example where communication has no apparent effect when employing either equilibrium or standard non-equilibrium analysis, but can have an effect when using in-between analysis.⁶

A related approach to non-equilibrium analysis--used, for example, in Brandenburger and Dekel [1987]--is to consider not the outcomes possible in a game, but rather each player's (possibly incorrect) expectations of his own payoffs. Using this approach, my model of communication implies--even without the equilibrium assumption--that each player expects to do at least as well as his least-favorite NGE. Recall that there is no comparable result without communication: If we do not assume equilibrium, a player might expect to do worse than in his worst Nash equilibrium. In terms of players' expectations of their own payoffs, therefore, non-equilibrium analysis and equilibrium analysis are more similar in games with communication than in games without communication.

This paper represents a particular perspective on modeling communication, and on approaches to non-cooperative game theory more generally. One reflection of this approach is that I present a range of solution concepts built up from one basic set of assumptions, and related to each other in easy-to-interpret ways. NGR combines the basic assumptions about negotiations with the assumption that players are rational. NGE is constructed by adding

⁶ I have come to speculate--based on research on communication and other areas--that we will rarely find "pure non-equilibrium" foundations for equilibrium analysis. That is, I doubt we can often use solution concepts without any restrictions on strategic uncertainty to support equilibrium analysis; research into the foundations of equilibrium analysis may involve considering how a little bit of equilibrium can go a long way.

the equilibrium assumption to NGR, while the in-between solution concepts are constructed by adding "some" equilibrium.⁷ Among the the advantages of this approach is its potential for empirical and experimental research on communication. There is no natural non-equilibrium or "in-between" version of most communication equilibrium concepts. This makes it impossible both to test the equilibrium hypothesis while maintaining the underlying theory of negotiations, and to test the theory of communication without maintaining the equilibrium assumption.

The results of this paper depend on assuming that negotiations consist of repeated, simultaneous statements--which is (at best) only one of many natural ways to model negotiations. Unilateral communication, or bilateral, alternating-offers negotiations are reasonable alternatives. The diversity of models is especially troubling because different models of negotiations can yield very different results. I will discuss in Section V some alternative models, and plausible general results about pre-game communication.

In Section II, I formulate the basic assumptions of negotiations and define the non-equilibrium solution concept; in Section III, I develop the equilibrium version; in Section IV, I define various "in-between" solution concepts. I conclude in Section V.

⁷ In a previous draft of this paper, I also defined solution concepts that incorporate focal points by maintaining the basic assumptions on how players negotiate, while limiting the range of equilibria over which they negotiate. Any subset of Nash equilibria can be chosen as a focal set for negotiations, and a solution concept can be formulated with any "amount" of equilibrium. See the as yet non-existent paper, "Focal Points and Pre-Game Communication."

II. Negotiated Rationalizability

I consider a two-player, complete-information, normal-form game G preceded by non-binding, cheap-talk negotiations. Assuming complete information means that players use communication only to make claims about intended moves, rather than to reveal exogenous private information. It is most natural to assume that G is a one-shot, simultaneous-move game; applying the theory to a sequential-moves game would only be appropriate if the players were unable to communicate after the first move of the game. I assume that there is a finite number, N_i , of pure strategies for each player i , and a finite number, N , of Nash equilibria.

The sets of pure strategies for players 1 and 2 are A_1 and A_2 . $\mathcal{A}_1$ and $\mathcal{A}_2$ are the sets of mixed strategies for the two players, where each $a \in \mathcal{A}_i$ is a vector of weights on the (ordered) pure strategies in A_i : $a = (q_1^1(a), q_2^1(a), \dots, q_{N_i}^1(a))$. $U_1(a_1, a_2)$ and $U_2(a_1, a_2)$ are the players' payoffs as a function of their moves. $V_1(a_1, a_2)$ and $V_2(a_1, a_2)$ are the expected payoffs from mixed strategies a_1 and a_2 derived using the functions U_1 and U_2 :

$$V_1(a_1, a_2) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} q_i^1(a_1) \cdot q_j^2(a_2) U_1(a_i^1, a_j^2) \quad \text{and} \quad V_2(a_1, a_2) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} q_i^1(a_1) \cdot q_j^2(a_2) U_2(a_i^1, a_j^2).$$

Before they play the "action game" G , the players can communicate in each of T periods, yielding the "full game," $\Gamma(G, T)$. In every round, each player makes a statement m from some set of statements M , where M is (for notational convenience) the same for both players and for each period. For a given game G , there are enough elements in M to accomodate all of the definitions below. A strategy in the full game consists of a player's choice of statements followed by his strategy in the action game, where statements and action-game strategies can be functions of the statements made in previous periods. I let

S_1 and S_2 be the strategy sets (including mixed strategies) in the game $\Gamma(G, T)$.

The payoffs to each of the players do not depend directly on the statements made and there is no cost of delay from bargaining: this communication is "cheap talk". Together, these assumptions mean that the payoffs to the game $\Gamma(G, T)$ are simply the payoffs--as defined above--to the actions induced in G .

I let $\mathcal{E}$ be the set of Nash equilibria to the game G . Let $(a_1^E, a_2^E) \in A_1 \times A_2$ be the actions which constitute the equilibrium $E \in \mathcal{E}$, and let $\pi_1(E)$ and $\pi_2(E)$ be the expected payoffs for the players from this equilibrium. For ease of reference, I say an equilibrium is superior to another if it strictly Pareto-dominates it. I shall represent the set of weakly Pareto-dominant Nash equilibria of the game G by $\mathcal{E}^*$: an equilibrium E is in $\mathcal{E}^*$ if and only if no other equilibrium is superior to it. For each $E \in \mathcal{E}$, $\mathcal{E}(E)$ is the set of equilibria in $\mathcal{E}$ which are superior to E , and $\mathcal{E}^*(E)$ is the set of equilibria in $\mathcal{E}^*$ which are superior to E .

Because of the cheap-talk structure, some behavioral assumption about how language focuses expectations must underlie claims about the effects of this form of pre-game communication. I attempt to capture some natural implications of negotiations in the next several definitions. These definitions will be used to construct a class of negotiating languages, which specify those combinations of messages that produce non-binding agreements.

The simplest example of a negotiating language is "Pareto talk." This language consists solely of proposals to play each of the Pareto-efficient Nash equilibria. If both players mention the same equilibrium in a given period, they have an agreement to play that equilibrium; if not, then they have no agreement. They can keep proposing each period until they reach an agreement, or until they run out of negotiating periods. (Pareto talk is a

natural extension to more general games of the language used in Farrell [1987].)

A similar language is one which additionally allows either player to be "silent". When incorporated into a solution concept, silence by one player will mean that any proposal by the other player will be interpreted as an agreement to play the corresponding equilibrium.⁸

Finally, I consider negotiating languages in which the players can make agreements to non-Pareto-efficient equilibria. These are especially likely as interim agreements from which the players can continue to negotiate, presumably to better equilibria.

As these examples suggest, the central feature of negotiating languages will be sets of rules for forming and changing agreements. To formalize such agreements, I first say, for each equilibrium $E \in \mathcal{E}$, that each statement $m \in M$ has an E-meaning. Essentially, the E-meaning of words will define how players can negotiate new agreements after an agreement to play the equilibrium $E \in \mathcal{E}$ has previously been reached. The $\emptyset$ -meaning will correspond to the meaning of a statement when no agreement has yet been reached. The E-meaning of a message will be either some equilibrium $F \in \mathcal{E}$, or the empty set; So I define the set $\mathcal{F} = \mathcal{E} \cup \{\emptyset\}$, where $\emptyset$ is interpreted as silence. I denote the set of statements that have the E-meaning $F \in \mathcal{F}$ as $Q(E;F)$, and $Q(\emptyset;F)$ will be the set of proposals when no agreement has thus far been reached.

Definition 1 defines an agreement in terms of the E-meanings of the

⁸ Silence can be interpreted as a statement equivalent to the statement, "Whatever equilibrium you propose, that will be our agreement. In a previous draft, I allowed "menu statements" which generalized this type of message, including such statements as "If you propose this period that we play equilibrium E or equilibrium F, then we have an agreement to the one you propose; anything else, we have no agreement." All essential results of this paper go through allowing this more general vocabulary.

players' statements. The players have an agreement to play E after one period of negotiation if one player proposes that they play E, and the other player either also proposes this equilibrium or remains silent. The players can then make new agreements to equilibria superior to E according to the same rules.

Definition 1:

The players have an agreement to play E after period t if and only if either:

There exists $F \in \mathcal{F}$ such that there is an agreement to play F after period $t-1$ and:

A1) At least one player makes a proposal $m \in Q(F;E)$ in period t ; and

A2) Both players make a proposal $m \in Q(F;D)$ for $D = E$ or $D = \emptyset$,

or:

B1) there was an agreement to play E after period $t-1$, and

B2) conditions A1-A2 do not both hold with respect to D for any equilibrium $D \in \mathcal{E}$, $D \neq E$.

Observe that only "Nash agreements" are explicitly modeled. Farrell [1988] assumes--more realistically--that a player might propose something more vague than a Nash equilibrium--i.e., proposing to restrict play to actions among some non-singleton sets of strategies. My approach of restricting players to Nash agreements means that I may be exaggerating the coordination effects of communication; I suspect that most results would hold if such vague agreements were added, if combined with more complicated assumptions about negotiations.

Definition 2 specifies negotiating languages for which the set of E-meanings, $\{Q(E,F)\}$, is sufficiently complete and unambiguous for purposes of the theory I will develop. It says that for every $E \in \mathcal{E}$, the set of E-meanings

is a partition of M . Combined with Definition 1, this means that at every point in negotiations, every statement has a well-specified meaning regarding further agreements.

Definition 2:

The language Q is complete with respect to G iff:

- 1) For all $m \in M$ and $E \in \mathcal{F}$, there exists $F \in \mathcal{F}$ such that $m \in Q(E;F)$; and
- 2) For all $E, F, F^* \in \mathcal{F}$, $Q(E;F) \cap Q(E;F^*) = \emptyset$.

Definition 2 does not mean that every two messages $m, m' \in Q(E,F)$ need to have exactly the same meanings to the players; rather the messages need only share the formal agreements-oriented meanings as specified in Definition 1. For instance, statement m by a player might be a proposal to play equilibrium F together with an announcement that, in lieu of an agreement, he'll negotiate tough from now on, with m' being a proposal to play F and an announcement that he shall be conciliatory.⁹

Definition 3 provides characteristics of complete negotiating languages sufficient to establish the main results of the paper. Condition 1 says that after any non-Pareto-efficient agreement, all superior Pareto-efficient equilibria remain potential agreements. Condition 2 says that after any non Pareto-efficient agreements, only superior (but perhaps not Pareto-efficient)

⁹ These languages might appear restrictive in the sense that every statement corresponds to some form of discussion about which equilibrium in G will be played. This does not mean, however, that a player must "actively" discuss G : He could talk about the weather. The assumptions mean only that any statements not directly relevant to G must have some coherent interpretation for negotiations. The most natural interpretation, of course, is that seemingly irrelevant statements imply consent by silence, just as does literal silence. Furthermore, if both players make irrelevant statements, both can be interpreted as silence--and it is consistent with my model that no progress in negotiations is implied in such a case.

equilibria can be proposed as new agreements. Condition 3 says that if at any stage the players reach an agreement to a Pareto-efficient equilibrium, it is common knowledge that this agreement won't be changed.

Definition 3:

A complete negotiating language Q is a Progressive Negotiating Language (PNL) iff:

- 1) For every $E \in \mathcal{F}$, and for every $F \in \mathcal{E}^*(E)$, $Q(E;F)$ is non-empty;
- 2) For every $E \in \mathcal{E} \setminus \mathcal{E}^*$, and for every $F \in \mathcal{F}$ such that $Q(E;F) \neq \emptyset$, $F \in \mathcal{E}(E)$ or $F = \emptyset$; and
- 3) For every $E \in \mathcal{E}^*$, $Q(E;E) = M$.

The simple example of Pareto-talk meets the conditions of a PNL trivially. The only relevant restriction is that if players agree to a Pareto-efficient equilibrium, they cannot change that agreement. While it is perhaps artificial to say they can't change the agreement--even by mutual consent--it is clearly realistic to suppose they won't change the agreement: any change would hurt one player or the other. (But saying the players can't change the agreement obviates the need for extra weak-dominance or sequential-rationality arguments.)

The definition involves more substantial restrictions for non-Pareto-efficient interim agreements. Condition 2 means that after an interim agreement, players can only make better agreements. By assuming players can't reach worse agreements, PNLs incorporate a bias towards "progressive bargaining." Note that because M does not change as a function of past statements, this means that the interpretation of statements change depending on agreements previously reached, not that the physical message

space is altered.

Such changes in interpretation seems realistic at a general level. Suppose that one player proposes his favorite equilibrium while the other player is silent; this might form an agreement by silent consent at the beginning of negotiations, but not after an agreement already exists to an equilibrium which the silent player prefers. Silence may carry an implicit message that a player is not going to concede any ground on a previous agreement.

An even more restrictive implication of Definition 3 is that even if both players propose an equilibrium worse than that of a previous agreement, the previous agreement is still in place. It seems intuitive that neither player is likely to negotiate to make himself worse off. Yet, without adding subsidiary assumptions restricting strategies in complicated ways to capture this intuition, abandoning these assumptions would allow a player to give up an agreement for a worse agreement because he thinks this will in the end make him better off. While it would be more convincing to derive "progressive bargaining" as the outcome of more primitive assumptions, I have found no tractable way to do so.

Definition 3 also implies that neither player can "block" all agreements: If there is not yet an agreement to a Pareto-efficient equilibrium, neither player can prevent with probability 1 all further agreements. A player can, however, be disagreeable in several ways: he can demand over and over again his best available equilibrium; he can (for PNLs that allow silence) not discuss anything about the game, so that a new agreement will only be reached if the other player makes a proposal; he can vary his proposals, trying to trick the other player into always making conflicting proposals. Yet, for all such behavior, the other player can always make a proposal which will with positive probability create an agreement.

This "no blocking" assumption is what validates the preemptive-concession argument in the Introduction, and will help guarantee the efficiency results; it essentially says that the only way one player can be kept from conceding is if the other player is making more generous proposals.

There is also what might be called a "no forcing" assumption incorporated into PNLs: A player can repeatedly demand his best remaining feasible equilibrium, and thus prevent any other agreement from being reached--neither player can be forced to make concessions, or to compromise. Thus, the sense in which all agreements--at least agreements to efficient equilibria--are completely voluntary can be stated formally: For every Pareto-efficient equilibrium, each player can choose a strategy such that no agreements to other equilibria can be reached, and against which the optimal strategy for the other player is to agree to that equilibrium. This assumption is important for the inefficiency results of the paper.¹⁰

I will now incorporate these assumptions into the solution concepts T,Q-Negotiated Rationalizability, ($NGR(T,Q)$), defined for each positive integer T , denoting the length of negotiations, and each PNL Q . These solution concepts will specify explicitly both the players' strategies and their beliefs, as opposed to specifying only strategies. Making beliefs explicit allows me to use some of the general results of Rabin [1989], and clarifies many of the arguments below. Of course, results about strategies and beliefs can be converted into results about strategies alone simply by ignoring the beliefs.

A plan $\theta = (\sigma, \mu)$ for player i consists of his strategy, $\sigma \in S_i$, and his beliefs, $\mu \in S_j$, about what the other player will do. The set of all possible

¹⁰ This assumption is also important for guaranteeing that the solution concepts will be Pareto-indeterminate, as will be defined later.

plans for player i is Θ_i . A prediction p_i is a probability distribution over Θ_i , with the weight p_i puts on the plan θ denoted by $p_i(\theta)$. Each prediction p_i induces a mixed strategy $p_i^S \in S_i$: $p_i^S = \sum_{(\sigma, \mu)} p_i(\sigma, \mu) \cdot \sigma$. The set of all predictions for player i is $\mathcal{P}_i$.

T,Q-negotiated rationalizability combines the rules of agreement-formation from Definitions 1, 2, and 3 with the rationality of the players.

Definition 4:

For a game $\Gamma(G, T)$, and a PNL Q , define the sets $(\mathcal{L}_1(0), \mathcal{L}_2(0)) \subseteq (\mathcal{P}_1, \mathcal{P}_2)$ as follows.

The prediction p_i is in $\mathcal{L}_i(0)$ if and only if for every plan (σ_i, μ) such that $p_i(\sigma, \mu) > 0$, it is true that:

- 1) σ_i is a best response to μ ;
- 2) if there is an agreement to play E after period T , σ_i induces action a_i^E in the game G ; and
- 3) if there is no agreement after period T , σ_i induces some strategy which is rationalizable in G .

Then, for $r \geq 1$, $(\mathcal{L}_1(r), \mathcal{L}_2(r))$ is defined inductively as follows. The prediction p_i is in $\mathcal{L}_i(r)$ if and only if

- 1) $p_i \in \mathcal{L}_i(r-1)$, and
- 2) For every plan (σ, μ) such that $p_i(\sigma, \mu) > 0$, there exists $p_j \in \mathcal{L}_j(r-1)$, $j \neq i$, with $\mu = p_j^S$.

Let T,Q-Negotiated Rationalizability $= (\mathcal{L}_1^*, \mathcal{L}_2^*) = \bigcap_{r=0}^{\infty} (\mathcal{L}_1(r), \mathcal{L}_2(r))$.

(I let the sets $(\mathcal{L}_1^S, \mathcal{L}_2^S)$ represent the strategies alone, so that the set of outcomes consistent with the theory is $\mathcal{L}_1^S \times \mathcal{L}_2^S$.)

While this solution concept incorporates behavioral assumptions beyond

common knowledge of rationality and of the structure of the game, it is important that it still be internally consistent. Rabin [1989] incorporates consistency criteria for behavioral solution concepts into the idea of a Public Theory. Among the characteristics of public theories are that they are always subsets of rationalizability, and they always contain an equilibrium. Public theories are behavioral analogues of rationalizability. I give the formal definition of a public theory in the appendix, and prove that for all T , and for all PNLs Q , T, Q -negotiated rationalizability is a public theory.

T, Q -negotiated rationalizability has a simple interpretation: using the focal meaning of language and the definitions of agreements presented, it is "rationalizability as if agreements were binding." That is, it corresponds to the implications of common knowledge of rationality in a hypothetical game where agreements are legally enforceable.

Of course, by assumption agreements are not enforceable. The reason the solution concept remains consistent with rationality is that (by assumption) all agreements are Nash equilibria, which are self-enforcing. A solution concept which looked similar to this, but which allowed non-Nash agreements, would likely not be consistent with rationality, and not be a public theory.

The "if" part of Definition 4 means that abiding by agreements is the only restriction (beyond rationalizability) on behavior during G . If an agreement is not reached, the players can play practically anything: no agreement, no restrictions. If we assume that play is restricted even if the players reach no agreements, then we could not claim that it was communication alone that was guaranteeing coordination.

What does T, Q -negotiated rationalizability imply about behavior in the action game G ? In particular, when is equilibrium behavior guaranteed?

If G has a unique equilibrium, and if Q does not allow silence, then

T,Q-negotiated rationalizability predicts that the players will play the equilibrium for sure. But this result is by brute force: the assumptions combine to mean that there is essentially only one word, and this word leads to an agreement.

It turns out that when there is more than one equilibrium or when the language includes silence, T,Q-negotiated rationalizability is no more restrictive than rationalizability. In such cases, players might have rational but unfortunate beliefs about each other's behavior, and not coordinate on an agreement. This can happen for arbitrarily large number of periods, and even in pure-coordination games. In Example 2, for instance, one of the T,Q-negotiated rationalizable solutions is for each player, each period, to think he knows which proposal the other player will make, and for these beliefs to be absolutely wrong. Then, each period, the players will make conflicting proposals, and never reach an agreement.

One might think that some sort of weak-dominance argument--not used in constructing the solution concepts of this paper--could guarantee equilibrium outcomes. In some games, this is in fact the case. Consider Example 3, for instance, where there are many rationalizable outcomes, but a unique Nash equilibrium. Then without weak dominance, and if the language includes silence, it is consistent with the solution concept for each player to remain silent, believing that the other player will propose the equilibrium. Then--if neither player specifically proposes it--they may not play the equilibrium. But since the equilibrium Pareto-dominates all other outcomes in the game, each player would likely propose the equilibrium just to make sure. Applying weak-dominance arguments in this game would guarantee equilibrium behavior.

This would generalize to all games in which there is a unique equilibrium, and in which this equilibrium Pareto-dominates all other rationalizable

outcomes in the game.¹¹ But the set of games for which such a weak-dominance argument holds is quite small, and includes very few interesting economic games. Indeed, Example 4 (Example 2 of Farrell [1988]) shows that the weak-dominance argument can work the other way: even when each player plans to play his equilibrium strategy, revealing this fact is weakly dominated by remaining silent, or trying to trick the other player into believing that they are not going to play it. Whether or not we think the Nash-equilibrium outcome is likely in this game, it is clearly not more likely with communication than without.

Most damning for the communication-yields-equilibrium conjecture, there are no instances--even with weak dominance arguments--in which communication yields equilibrium behavior when there are multiple strict equilibria, which is where equilibrium analysis has been most forcefully challenged. Communication does not seem a compelling justification for equilibrium analysis: it rarely guarantees equilibrium behavior in the action game without assuming it in the communication stage.

In the next section, I define a solution concept which corresponds to the intersection of the set of Nash equilibria and the set of T,Q-negotiated rationalizable outcomes. After adding the equilibrium hypothesis, it is no longer interesting to ask whether communication yields equilibrium behavior. Rather, I ask what pre-game communication implies about efficiency among the set of equilibrium outcomes.

I shall return in Section IV to the question of how communication relates

¹¹ This is the bilateral-communication analogue to Proposition 2 of Farrell [1988], in which he examines when unilateral communication guarantees an equilibrium outcome. My condition is more restrictive than Farrell's: the equilibrium outcome in my case must be the best for both players, as opposed to just the player who can speak.

to the equilibrium assumption itself, and investigate what communication accomplishes outside the realm of both pure equilibrium analysis and "pure non-equilibrium analysis". I find that assuming "some" equilibrium means that communication can have some of the effects it does with equilibrium analysis.

III. Negotiated Equilibrium

In this section I develop a solution concept which includes all of the Nash equilibria consistent with NGR.¹² Definition 5 combines Nash equilibrium with T,Q-negotiated rationalizability. Note that I define an equilibrium as predictions in which the beliefs of each player correspond to the behavior of the other player, not just when behavior happens to be mutually optimal.

Definition 5:

The strategies (σ_1, σ_2) are a T,Q-Negotiated Equilibrium, $(NGE(T,Q))$ if and only if (σ_1, σ_2) is a Nash equilibrium in $\Gamma(G,T)$, and $\exists (p_1, p_2) \in (\mathcal{L}_1^*, \mathcal{L}_2^*)$ such that, for $i = 1, 2$, $p_i^S = \sigma_i$, and $\forall (\sigma, \mu)$ such that $p_i(\sigma, \mu) > 0$, $\mu = \sigma_j$.

The play in the action game G induced by negotiations, rather than the details of negotiations (i.e., strategies in $\Gamma(T,G)$), are of primary concern. Let $\mathcal{A}^*$ be the set of probability distributions over the outcome of the game G ,

¹² While I do not use any notion of perfection, the inefficiency results below are proven by constructing subgame-perfect equilibria, and because the efficiency result is true for Nash equilibrium, it is true a fortiori for perfect equilibrium. Thus, all the main implications of this section would hold if equilibrium-perfection were incorporated. Indeed, it seems likely that we could abandon some of the restrictions incorporated into PNLs in exchange for strengthening the equilibrium concept, and still get all the results of this section (but not Section IV).

with $\pi_1(a)$ and $\pi_2(a)$ the expected payoffs of $a \in \mathcal{A}^*$. (Observe that $\mathcal{A}^* \neq \mathcal{A}_1 \times \mathcal{A}_2$, because it allows for correlation in the moves of the players.) For each strategy pair (σ_1, σ_2) in the game $\Gamma(G, T)$, let $x(\sigma_1, \sigma_2) \in \mathcal{A}^*$ be the probability distribution of play which it induces in G .

Let $C \subseteq \mathcal{A}^*$ be the set of outcome distributions consistent with the set of correlated equilibria using publicly observable randomizing devices (i.e., C represents probability distributions over Nash equilibria). It is clear that any Nash equilibrium in $\Gamma(T, G)$ induces such a correlated equilibrium in G :

Lemma 1:

For any Nash equilibrium (σ_1, σ_2) of $\Gamma(G, T)$, $x(\sigma_1, \sigma_2) \in C$.

Proof:

Choose (σ_1, σ_2) . Suppose $x(\sigma_1, \sigma_2) \notin C$. Then there exists some sequence of messages that occurs with positive probability such that the players choose non-Nash strategies in G . Thus (σ_1, σ_2) could not be an equilibrium. Q.E.D.

I represent player i 's least favorite Pareto-efficient equilibrium by $\pi_i^* = \min_{E \in \mathcal{E}^*} \{\pi_i(E)\}$, for $i = 1, 2$. The Pareto-Meet (PM) as feasible payoffs that are superior to each player's worst Pareto dominant payoff:

$$PM = \{(\pi_1, \pi_2) \mid \pi_i \geq \pi_i^* \text{ for } i=1,2, \text{ and } \exists x \in C: (\pi_1, \pi_2) = (\pi_1(x), \pi_2(x))\}.$$

The Pareto-meet corresponds to those payoffs lying in the space bordered above by the Pareto frontier of C , and below by (π_1^*, π_2^*) . I label the set of Nash equilibria of G with payoffs in the PM as $\mathcal{E}^{PM}$. The PM in Example 1 corresponds to the shaded area in the graph. (Note that when the Pareto frontier consists of a unique point--as in pure-coordination games--the PM consists only of that point.)

Theorem 2 is the first inefficiency result: All $E \in \mathcal{E}^{PM}$ are T,Q-negotiated equilibrium outcomes, for all T and all PNLs Q. Thus, inefficiency is possible whenever $\mathcal{E}^{PM} \neq \mathcal{E}^*$.

Theorem 2:

For every T, PNL Q, and $E \in \mathcal{E}^{PM}$, there exists $(\sigma_1, \sigma_2) \in \text{NGE}(T, Q)$ such that $x(\sigma_1, \sigma_2) = (a_1^E, a_2^E)$.

Proof:

Choose $E \in \mathcal{E}^{PM}$. I will construct a Nash equilibrium $(p_1, p_2) \in (\mathcal{L}_1(0), \mathcal{L}_2(0))$ in which each player i will play a_i^E with probability 1.

For each player i , choose some $E_i \in \mathcal{E}^*$ such that $\pi_i(E_i) = \text{Max}_{F \in \mathcal{E}^*} \{\pi_i(F)\}$.

Let σ_i be such that player i proposes some $m \in Q(\emptyset; E_i)$ each period in which the players have not yet reached an agreement, and that player i plays a_i^E in G if they reach no agreement after T periods. Then I claim that (σ_1, σ_2) constitutes a Nash equilibrium. Consider the possible deviations the players could make. The only payoff-relevant deviation for player i is to make a statement which would yield him the payoff $\pi_i(E_j)$. But $\pi_i(E) \geq \pi_i(E_j) \forall E \in \mathcal{E}^{PM}$. And, since E is an equilibrium, neither player would wish to deviate during the action game G.

Then let $\theta_1 = (\sigma_1, \sigma_2)$ and $\theta_2 = (\sigma_1, \sigma_2)$. Let $p_1 \in \mathcal{P}_1$ and $p_2 \in \mathcal{P}_2$ be such that $p_1(\theta_1) = 1$, $p_2(\theta_2) = 1$. Since $(p_1, p_2) \in (\mathcal{L}_1(0), \mathcal{L}_2(0))$ is an equilibrium, it survives the iterated dominance, and thus $(p_1, p_2) \in (\mathcal{L}_1^*, \mathcal{L}_2^*)$. This proves the theorem. Q.E.D.

The theorem is proven by constructing an equilibrium in which both players bargain "tough", demanding each period their best equilibrium. When they fail

to reach an agreement, they play the appropriate equilibrium with payoffs in the PM. This is an equilibrium because the only deviation by a player that would change his payoffs would be to concede to the other player, which would yield him a worse payoff than does his equilibrium strategy.

Consider Example 5, for instance. Among the Nash equilibria in this game are E, yielding payoffs (1,4); F, yielding payoffs (3,3); H, yielding (4,1); and J, yielding (2,2). (There are many mixed-strategy equilibria as well.) While equilibrium J is in $\mathcal{E}^{PM}$, it is not in $\mathcal{E}^*$, because it is strictly Pareto-dominated by equilibrium F.

One T,Q-negotiated equilibrium (for all T and PNLs Q) is where player 1 demands H each period, player 2 demands E each period, and if they do not reach an agreement--which they will not--they play equilibrium J.

Because this equilibrium is inefficient, we might ask: If the players know they will play it, why do they not agree to play F instead? Why does neither player propose this? It is worth considering ways in which we might exclude this equilibrium, and perhaps guarantee efficiency in this game.

First, observe that there is no weak-dominance argument that players should be less tough in bargaining. Indeed, if there were a small probability that the other player would concede all the way, or be silent, then hanging tough would be strictly preferred. Or if it is believed that the other player will interpret an attempt at compromise as a sign of weakness or concession, and then assume that they will play his best equilibrium even without agreement, such attempts would be counterproductive.

In Example 1, $\mathcal{E}^{PM} = \mathcal{E}^*$, so that Theorem 2 does not apply. However, Theorem 3 shows that the types of inefficient mixed-strategy equilibria discussed in Example 1 are NGE.

Theorem 3:

Choose $E, F \in \mathcal{E}^*$, and $D \in \mathcal{E}$, such that, for $i = 1, 2$, $\pi_i(D) < \text{Min}[\pi_i(E), \pi_i(F)]$. Choose any PNL Q .

Then for every T there exists $(\sigma_1, \sigma_2) \in \text{NGE}(T, Q)$ such that for $i = 1, 2$, $\pi_i(x(\sigma_1, \sigma_2)) < \text{Min}[\pi_i(E), \pi_i(F)]$.

Proof:

Choose E, F, D , and Q as the theorem states. Consider (σ_1, σ_2) such that, for each period t in which the players do not yet have an agreement, player 1 says $m \in Q(\emptyset; E)$ with probability $\alpha_1(t)$, and $m \in Q(\emptyset; F)$ with probability $1 - \alpha_1(t)$; player 2 says $m \in Q(\emptyset; E)$ with probability $1 - \alpha_2(t)$, and $m \in Q(\emptyset; F)$ with probability $\alpha_2(t)$. If the players reach an agreement, they play it; if they do not at the end of T periods, they play D .

From these strategies, we can define the continuation payoffs, $v_i(t)$, for player i if there is no agreement at the beginning of period t . Note that $v_i(T+1) = \pi_i(D)$. Then, by induction, let

$$\alpha_2(t) = [\pi_1(E) - v_1(t+1)] / [\pi_1(E) + \pi_1(F) - 2v_1(t+1)],$$

$$\alpha_1(t) = [\pi_2(F) - v_2(t+1)] / [\pi_2(E) + \pi_2(F) - 2v_2(t+1)],$$

$$v_1(t) = (1 - \alpha_2(t)) \cdot \pi_1(E) + \alpha_2(t) \cdot v_1(t+1), \text{ and}$$

$$v_2(t) = \alpha_1(t) \cdot \pi_2(E) + (1 - \alpha_1(t)) \cdot v_2(t+1).$$

It can be verified that this represents a Nash equilibrium, and it is consistent with $\text{NGR}(T, Q)$. It can be shown by algebra that if $v_i(t+1) < \text{Min}[\pi_i(E), \pi_i(F)]$, then $v_i(t) < \text{Min}[\pi_i(E), \pi_i(F)]$. Then since $v_i(T+1) = \pi_i(D) < \text{Min}[\pi_i(E), \pi_i(F)]$, by induction we get that $v_i(t) < \text{Min}[\pi_i(E), \pi_i(F)]$ for all t , including $t = 0$. But since $v_i(0) = \pi_i(x(\sigma_1, \sigma_2))$, this completes the proof.

Q.E.D.

Theorem 3 says that if there is a Nash equilibrium in G that is Pareto-dominated by each of two Pareto-efficient equilibria, then there exists a $NGE(T, Q)$ that is Pareto-dominated by those two equilibria. For an even larger class of games than implied by Theorem 2, communication does not guarantee efficiency. The strongest implication of Theorem 3 is the following:

Corollary 1:

If $\exists D \in \mathcal{E}$ such that, for $i = 1, 2$, $\pi_i(D) < \pi_i^*$, then $\forall$ PNLs Q , $\forall T$, $\exists (\sigma_1, \sigma_2) \in NGE(T, Q)$ such that for $i = 1, 2$, $\pi_i(x(\sigma_1, \sigma_2)) < \pi_i^*$.

Proof:

Choose $E, F \in \mathcal{E}^*$ such that $\pi_1(E) = \pi_1^*$ and $\pi_2(F) = \pi_2^*$. Q.E.D.

Corollary 1 says that if there exists an equilibrium in G which is Pareto-dominated by the payoffs (π_1^*, π_2^*) , then for all PNLs Q , and for all T there exists a T, Q -negotiated equilibrium dominated by (π_1^*, π_2^*) . In Example 5, for instance, Corollary 1 means that, for all T and Q , there exists a T, Q -negotiated equilibrium with payoffs of less than $(1, 1)$.

To consider the efficiency effects of lengthy negotiations, I now define the solution concept Negotiated Equilibrium (NGE), which includes those outcomes that are consistent with $NGE(T, Q)$, for some PNL Q , as the number of periods of communication T becomes arbitrarily large.

If we choose some ordering on the set $\mathcal{E}$, then an element of the convex hull of Nash equilibria, c , can be represented by the weights it puts on each Nash equilibrium: $c = (r_1, r_2, \dots, r_N)$. Then we can define the Euclidean distance between two correlated equilibria c and $c^* = (r_1^*, r_2^*, \dots, r_N^*)$: $D(c, c^*) = \sum_{i=1}^N (r_i - r_i^*)^2$.

Definition 5:

A correlated equilibrium $c \in C$ is a Negotiated Equilibrium (NGE) for game G if and only if for every PNL Q and $\varepsilon > 0$, there exists a $T^* > 0$ such that:
For all $T > T^*$ there exists $x \in \text{NGE}(T, Q)$ such that $D(x, c) < \varepsilon$.

The following are infinite-period corollaries to Theorems 2 and 3.
Corollary 2 follows immediately from Theorem 2:

Corollary 2:

For all $E \in \mathcal{E}^{\text{PM}}$, (a_1^E, a_2^E) is a NGE.

Corollary 3 follows from Case I in the proof of Theorem 3.

Corollary 3:

Choose $E, F \in \mathcal{E}^*$, and $D \in \mathcal{E}$, such that, for $i = 1, 2$, $\pi_i(D) < \text{Min}[\pi_i(E), \pi_i(F)]$. Then $\exists x \in \text{NGE}$ such that for $i = 1, 2$, $\pi_i(x) = \text{Min}[\pi_i(E), \pi_i(F)]$.

Proof:

Choose $Q = \text{Pareto talk}$. Then the continuation payoff $v_i(t)$ is bounded above by $\text{Min}[\pi_i(E), \pi_i(F)]$. Further, $v_i(0)$ increases monotonically with T , and, for fixed T , $v_i(t) > v_i(t+1) \forall t \leq T$. Therefore, $\lim_{T \rightarrow \infty} \text{Min}_{t \leq T} \{[v_i(t) - v_i(t+1)]\} = 0$. Algebra shows that this condition implies that $\lim_{T \rightarrow \infty} \text{Max}_{t \leq T} \{v_i(t)\} = \text{Min}[\pi_i(E), \pi_i(F)]$, which by monotonicity implies that $\lim_{T \rightarrow \infty} v_i(0) = \text{Min}[\pi_i(E), \pi_i(F)]$. This proves the corollary. Q.E.D.

Each of these corollaries show that even arbitrarily lengthy negotiations

generally do not guarantee full efficiency. With lengthy negotiations, however, some efficiency is guaranteed. I state the main result of the paper:

Theorem 4:

If c is a NGE, then $(\pi_1(c), \pi_2(c)) \in PM$.

Proof:

Choose PNL Q . Let $\mathcal{E}^*(0) = \mathcal{E}^*$. For all $k > 0$, let $\mathcal{E}^*(k) = \{E \in \mathcal{E} \mid \mathcal{E}^*(E) \subseteq \mathcal{E}^*(k-1) \cup \dots \cup \mathcal{E}^*(0)\}$. Let $\hat{\pi}_i$ be the worst possible payoff for player i in the game.

Choose $\mu \in \mathcal{L}_j^S$. Let $v_i(t, \mu, E)$ = the continuation payoff to player i if the players have an agreement to play E at period t , if player i chooses an optimal strategy against μ . If the players have reached an agreement to $E \in \mathcal{E}^*(1)$, player i can guarantee that he will reach some new agreement F , where $\pi_i(F) \geq \pi_i^*$, each period with probability at least $(1/2)^N$; the highest probability of no agreement is $1 - (1/2)^N$ each period. Then $v_i(t, \mu, E) \geq [1 - (1 - (1/2)^N)^{T-t}] \cdot \pi_i^* + [(1 - (1/2)^N)^{T-t}] \cdot \hat{\pi}_i$. Choose $t = \mathcal{J}(T(N-1)/N)$, where $\mathcal{J}(X)$ represents the greatest lesser integer of X . Then $v_i(t, \mu, E) \geq [1 - (1 - (1/2)^N)^{T/N}] \cdot \pi_i^* + [(1 - (1/2)^N)^{T/N}] \cdot \hat{\pi}_i$.

Now choose $t' = \mathcal{J}(T(N-2)/N)$. Then $v_i(t', \mu, E) \geq [1 - (1 - (1/2)^N)^{T/N}] \cdot v_i(\mathcal{J}(T(N-1)/N), \mu, E) + [(1 - (1/2)^N)^{T/N}] \cdot \hat{\pi}_i$. By induction and substitutions, $v_i(0, \mu, \emptyset) \geq [1 - (1 - (1/2)^N)^{T/N}] \cdot \pi_i^* + [(1 - (1/2)^N)^{T/N}] \cdot \hat{\pi}_i$.

But then $\lim_{T \rightarrow \infty} v_i(0, \mu, \emptyset) \geq 1 \cdot \pi_i^* + 0 \cdot \hat{\pi}_i = \pi_i^*$. But since $\pi_i(c) \geq v_i(0, \mu, \emptyset)$ for all μ , this proves the theorem. Q.E.D.

The gist of the proof is most apparent if we imagine that the negotiating language includes only discussions about the Pareto-efficient equilibria.

Then, for any statement by one player--a specific proposal or silence--the other player can make a statement that yields an agreement to an efficient equilibrium. Moreover, for any beliefs the other player has, he can choose a proposal which will guarantee such an agreement with positive probability. If there are many possible such statements, this probability may be small; but as the number of periods becomes large, a player can choose a strategy which guarantees that with probability 1 some agreement will eventually be reached. Since any such agreement means they will play some Pareto-efficient equilibrium, this means that player 1 must get a payoff of at least π_1^* . (This argument must be iterated to take into account interim, non-Pareto-efficient equilibria.)

Between the "negative" Theorems 2 and 3 and the "positive" Theorem 4, I have largely characterized the set of outcomes consistent with negotiated equilibrium. For all games, NGE implies first, that all correlated equilibria with payoffs outside the Pareto-meet are excluded, and, second, all Nash equilibria with payoffs in the Pareto-meet are included. Furthermore, for most games, NGE includes inefficient correlated equilibria with payoffs inside the Pareto-meet.

IV. Pre-Game Communication and "In-Between" Analysis.

The foundations of equilibrium analysis is a topic of continuing research in game theory. Almost all game theory focuses on either equilibrium analysis--where all strategic uncertainty is ruled out--or "pure non-equilibrium analysis"--where unlimited strategic uncertainty is allowed. Yet this dichotomy is artificial; it seems solution concepts in between are of interest.

In this section, I try to combine NGR with "some" equilibrium. This allows us to consider the related questions of how much equilibrium is needed for the results of Section II, and to consider whether communication narrows the difference between non-equilibrium and equilibrium analysis.

One approach to formulating in-between solution concepts is to measure the distance between each player's beliefs and the other player's actual behavior, and impose restrictions on this distance. Equilibrium analysis assumes the distance must be zero for both players, while rationalizability and public theories place no restrictions on the distance. "In-between analysis" could examine those predictions for which the beliefs of each player are within some fixed distance from the other player's behavior. However, any combination of choosing a metric and fixing a maximum distance would be fairly arbitrary: There is no natural way of saying when beliefs are "close to" or "far from" actual behavior.

There is one less arbitrary "in-between" theory: From any public theory, we can define the set of player-assessed predictions as those predictions in which at least one player is exactly correct in his beliefs.¹³ This approach is less restrictive than equilibrium analysis, which requires both players to be correct, but is more restrictive than "pure non-equilibrium", in which neither player is required to be correct. It is non-arbitrary because the definition does not rely on the precise distance metric used.

Definition 7:

For a given T and PNL Q , the strategies (σ_1, σ_2) are a T,Q-Player-Assessed Negotiated Outcomes ($PANO(T, Q)$) if and only if $\exists (p_1, p_2) \in (\mathcal{L}_1^*, \mathcal{L}_2^*)$ such that,

¹³ This approach was developed in Rabin [1989].

for $i = 1$ or 2 , $j \neq i$, each of the following two conditions hold:

- 1) $\mu = p_j^S$ for all (σ, μ) such that $p_i(\sigma, \mu) > 0$, and
- 2) $(\sigma_1, \sigma_2) = (p_1^S, p_2^S)$.

From this definition, we can formulate the set of player-assessed predictions consistent with some PNL as the number of periods of negotiations becomes arbitrarily long.

Definition 8:

A probability distribution $y \in \mathcal{A}^*$ is a player-assessed negotiated outcome (PANO) if and only if $\forall \varepsilon > 0$, $\forall T^* > 0$, $\exists T > T^*$ and PNL Q such that $\exists (\sigma_1, \sigma_2) \in \text{PANO}(T, Q)$ such that $D(y, x(\sigma_1, \sigma_2)) < \varepsilon$.

PANO can impose restrictions on outcomes that NGR does not. In Example 1, for instance, in any PANO each player must get payoffs of at least $1/2$, and the total payoffs must be at least $3/2$. (The claims in the next few paragraphs are an implication of Lemma 2 presented shortly.) While these are not as high as the minimal payoffs guaranteed by NGE, they contrast with the payoffs of zero that are consistent with NGR. In terms of minimal total payoffs, PANO gets more than half way from NGR to NGE.

Of course, the approach of examining player-assessed predictions is not limited to NGR; it can be applied to rationalizability without communication. If we wish to consider how communication affects behavior, the appropriate comparison is between PANO and player-assessed rationalizable outcomes (PARO). In Example 1, calculations show that the minimal payoff for a player consistent with PARO are $1/3$, and the minimal total payoffs consistent with PANO are 1. Comparing PANO to PARO, we see that with player-assessed

analysis--as with equilibrium analysis--communication seems to add to efficiency.

The implications of player-assessed analysis are more striking in Example 2. Here, the only payoffs consistent with PANO are (1,1), so that PANO and NGE yield the same result. That is, if we assume only that one player has correct beliefs about the other player's strategy--the other player may have very incorrect beliefs--full efficiency is guaranteed. This example suggests that one does not need full equilibrium analysis to get the full effects of communication in pure-coordination games. By contrast, the payoffs in PARO and Nash equilibrium can be as low as (1/2,1/2) in Example 2: Comparing PANO and Nash equilibrium, we see that a weaker solution concept with communication can guarantee greater efficiency than can a stronger solution concept without communication.

Lemma 2 compares outcomes with and without communication for simple battle-of-the-sexes games, such as both Examples 1 and 2. It shows that, when applying player-assessed analysis in such games, the worst payoffs possible with communication are higher than without communication.

Lemma 2:

Consider a game where $A_1 = \{a_1, a_2\}$ and $A_2 = \{b_1, b_2\}$ such that $(\pi_1(a_1, b_1), \pi_2(a_1, b_1)) = (t, 1)$, $(\pi_1(a_2, b_2), \pi_2(a_2, b_2)) = (1, t)$, and $(\pi_1(a_1, b_2), \pi_2(a_1, b_2)) = (\pi_1(a_2, b_1), \pi_2(a_2, b_1)) = (0, 0)$, where $t \geq 1$.

Then the payoffs in every PARO yield each player at least $1/(1+t)$, and yield total payoffs of at least 1, and the payoffs in every PANO yield each player at least $1/t$, and yield total payoffs of at least $(t+1)/t$.

Proof:

By examination of the best-response correspondences and the structure of the game. Q.E.D.

Example 6 (to which Lemma 2 does not apply) well illustrates one reason to consider in-between analysis when modeling communication. As always, NGR includes the same predictions as does rationalizability. Furthermore, because there is a unique equilibrium, NGE predicts the same thing as Nash equilibrium. By standard criteria, communication apparently has no effect in this game.

Comparing PANO to PARO indicates, however, that this conclusion is misleading. While PARO permits total payoffs of less than zero, PANO guarantees that the total payoffs of the two players will be at least zero. Communication essentially gives each player the option of focusing expectations on the equilibrium strategies, (U,L). Each player might choose not to propose the equilibrium (assuming the PNL includes silence) if he plans not to play his equilibrium strategy, and wants to try to outsmart the other player in choosing from the strategy space $\{M,D\} \times \{C,R\}$. But communication rules out the possibility of a player choosing his equilibrium strategy when he expects the other player not to play his equilibrium strategy. Because such outcomes are the most inefficient ones in this game, communication can contribute to efficiency in a natural way.

An approach which bears some resemblance to player-assessed analysis--used, for example, in Brandenburger and Dekel [1987]--is to consider not the set of outcomes possible in a game, but rather each player's expectation of his own payoff. This might be the appropriate consideration, for instance, if we want to know whether a player prefers to play a game rather than receive some

alternative payoff. Or, it would be a helpful approach if we wished to engage in a type of non-paternalistic welfare analysis which considers players' perceptions of likely outcomes, rather than our own judgement of likely outcomes.¹⁴

In fact, I have already proven a result concerning players' expectations of their payoffs: In the proof of Theorem 4, I showed that for any beliefs of a player, there exists a best response which guarantees that player a payoff at least as great as his worst Pareto-dominant outcome. The proof did not require that those beliefs correspond to equilibrium behavior by the other player. Thus, I have already proven that communication guarantees that each player expects to do at least as well as his worst NGE; i.e., at least π_i^* .

Theorem 5:

$\forall T, \text{ PNLs } Q, \text{ let } \Theta_i(T, Q) = \{\theta \in \Theta_i \mid \exists p_i \in \mathcal{L}_i^* \text{ such that } p_i(\theta) > 0\}.$

Then $\lim_{T \rightarrow \infty} \min_{\text{PNLs } Q, \theta \in \Theta_i(T, Q)} \pi_i(\theta) \geq \pi_i^*.$

Proof:

See the proof of Theorem 4.

Thus, the worst payoff a player can expect without the equilibrium condition is the same as the worst he can expect with the equilibrium condition. Observe that this result contrasts to games without communication, where a player might expect to do worse than in his worst Nash equilibrium (as

14 The difference between this approach and the player-assessed approach is that PANO and PARO include each player's expectations of both his and the other player's payoffs. As a prediction-making heuristic, player-assessed analysis uses players' beliefs to help guide our predictions, giving no special role to what a player predicts about himself, rather than the other player.

can happen in Example 6). In terms of players' expectations of their own payoffs, non-equilibrium analysis seems more similar to equilibrium analysis in games with communication than in games without communication.¹⁵

If we return to player-assessed analysis, we can see that Theorem 4 also implies (via Theorem 5) a simple result about PANO for games in general:

Corollary 4:

If $x \in A^*$ is a PANO, then either $\pi_1(x) \geq \pi_1^*$ or $\pi_2(x) \geq \pi_2^*$.

Proof:

The payoffs in any PANO must, by definition, correspond to at least one of the player's beliefs. Therefore the result follows immediately from Theorem 5.

Q.E.D.

While NGE implies that players' payoffs must lie in the PM, PANO implies only that they cannot lie in the region which is strictly dominated by the PM. Figure 1 shows (for a given set of Pareto-efficient Nash-equilibrium payoffs) the payoffs consistent with NGE (in black) and the payoffs consistent with PANO (in black and grey).

This corollary implies, in turn, PANO guarantees full efficiency in games where all the Pareto-efficient equilibria have the same payoffs. Corollary 5, which follows immediately from corollary 4, formalizes this statement.

Corollary 5:

¹⁵ There is, however, no comparable result that a player cannot expect to do better than in any NGE. In Example 6, for instance, if neither player proposes the equilibrium, each player may believe that he can outsmart the other, and each player may expect a payoff for himself of 5.

Suppose $\forall E \in \mathcal{E}^*$, for $i = 1, 2$, $\pi_i(E) = \pi_i^*$. Then $\forall x \in \mathcal{A}^*$ such that x is a PANO, for $i = 1, 2$, $\pi_i(x) = \pi_i^*$.

Corollary 5 means that not only is the efficiency result stronger for pure-coordination games than for other games--implying full efficiency--but that the efficiency is guaranteed with a weaker solution concept--PANO rather than NGE.

V. Discussion and Conclusion

Before concluding with some conjectures about how the results of this paper might generalize, I wish to discuss aspects of the approach taken in this paper. First, if I "fail" to guarantee efficiency with my solution concept, it is largely because such a guarantee is not my goal. Rather, I am trying to capture approximately how much efficiency is implied by the logic and practice of everyday languages.¹⁶

There is not a consensus among game theorists about what constitutes an internally consistent solution concept, especially if it incorporates

¹⁶ The normative approach has received some attention. Banks and Calvert [1989], for instance, essentially state that mechanism design is the goal of their paper. Furthermore, both they and Myerson [1989] assume the existence of a welfare-promoting, disinterested mediator. Considering the effects of an efficiency-oriented mediator can certainly be relevant, even to positive analysis (it may be reasonable, for instance, in studying the design of channels of communication in organizations). But most situations in which communication matters do not involve the help of such mediators.

In any event, in complete-information games, the question of whether communication can achieve efficiency is uninteresting; it can. But so can it be achieved without communication--as in Harsanyi and Selten's [1989] solution concept. In some sense, the question must be whether one thinks the restrictions are realistic.

behavioral assumptions; perhaps another reason I fail to guarantee efficiency is my somewhat conservative approach. I essentially construct the solution concepts in two steps. First, I formulate a "language correspondence," a mapping from what is said during negotiations to behavior in the game. The correspondence I choose is clearly compatible with rationality and common knowledge: There are some sequences of messages which the players interpret as agreements to play particular Nash equilibria, with all other sequences of messages implying no restrictions (beyond rationalizability) on behavior.

Given this language correspondence, we can think of the model in terms of a "hypothetical game": the game as if agreements were binding. This hypothetical game reflects, I believe, a psychological reality: Two people who fully intend to abide by non-binding agreements will likely behave in negotiations much as if the agreements were binding.

The second step is to define a solution concept on the hypothetical game which incorporates elements of rationality, internal consistency, and assumptions about strategic uncertainty. Given my language correspondence, NGR corresponds exactly to rationalizability on this hypothetical game, and NGE corresponds exactly to Nash equilibrium. I also use the solution concept of player-assessed rationalizability, as defined in Rabin [1989].

The general approach here is the same as that of Rabin [1989] Rabin [1990], and differs substantially from a much used alternative approach to modeling communication. Farrell [forthcoming], and subsequently Myerson [1989] and Matthews et al [1990], explore the effects of language in terms of "neologisms": equilibria are tested for the existence of any unused messages which could be sent by some types of a player, and, if interpreted literally and rationally, yield those types payoffs higher than in the equilibrium. Equilibria for which such credible neologisms (as variously defined) exist are

deemed implausible. While this alternative is currently applied mostly in incomplete-information games, I believe it could be applied as well to complete-information games. Moreover, I believe the approach would yield substantially different results from my model.

Several aspects of the "neologism approach" are problematic, or at least controversial. First, messages are interpreted by considering a deviation by some players from a would-be equilibrium which other players do not deviate from--presumably because they believe all players will adhere to it. Thus (as I contentiously interpret it), excluding an equilibrium often relies on a logic that requires different players to have different beliefs about who will send which messages. I believe this is an inconsistent method of ruling out equilibria.¹⁷ A second problem with the neologism approach is the a priori assumption that not all messages are used with positive probability in any equilibrium (thus guaranteeing that neologisms always exist). It is hard to reconcile this assumption with a well-defined message space. A third problem in the neologism approach--presumably related to the first two--is non-existence: Often no equilibria are consistent with the assumptions of neologism-based solution concepts.¹⁸

My approach adopts the rich-language assumption more directly than does the neologism approach: language has focal meaning to the players, not when used as deviations, but a priori, before negotiations begin. Within the equilibrium framework, meaningful communication is not used as a test for breaking up equilibria, but rather as an assumption which determines equilibria.

¹⁷ Similar criticisms of this general type of deviations-based argument have been made (and disputed) in other contexts (see, for instance, the discussion regarding the "Stiglitz critique" in Cho and Kreps [1987]).

¹⁸ Farrell [forthcoming] and Matthews et al [1990] have an existence problem. Myerson [1989] does not, but (as he shows) there would be an existence problem if it were not for the mediator.

Each of the problems with the neologism approach are avoided. Because players consider language directly, no precarious deviations-based arguments are needed. My approach uses a well-defined message space. The results should hold in any situation in which the language is rich enough to express all of the strategically relevant ideas, irrespective of whether "neologisms" always exist. Finally, all of the solution concepts in this paper are non-empty for all games.

The most obvious difference in approaches does not concern equilibrium analysis: Because I construct my equilibrium concepts by adding the equilibrium assumption to well-defined non-equilibrium concepts, it is easy to modify or remove the equilibrium assumption. The neologism approach, on the other hand, precludes non-equilibrium analysis: There are simply no natural non-equilibrium versions of any of the neologism-based equilibrium concepts.

This difference may be central in experimental and other empirical studies of communication. Consider a quote from Vince Crawford's survey of experiments involving pre-game communication (Crawford [1990]):

Much of the behavior observed in these experiments appears to have been driven by strategic uncertainty. By contrast, most of the theoretical analyses of explicit bargaining reviewed in Section II [of his paper] either focus on the signaling role of cheap talk, or study the coordination role within an equilibrium selected by refinements that are largely insensitive to strategic uncertainty. The growing experimental literature in this area offers the possibility of developing a theoretical and practical understanding of how strategic uncertainty and cheap talk interact (with or without private information) to determine bargaining outcomes.

As I have argued, my approach can help study how strategic uncertainty and cheap talk interact. Note that we cannot adequately test this interaction simply by comparing a particular non-equilibrium solution concept to an equilibrium solution concept; the two concepts must come from the same

underlying behavioral theory, so that we can interpret them as differing solely in terms of the amount of strategic uncertainty.¹⁹

An important distinction between NGE and other communicational equilibrium concepts can be formulated purely in terms of the outcomes consistent with the solution concepts, irrespective of the issue of strategic uncertainty, or broader issues of "approach." All existing models of communication which rule out all inefficient equilibria in complete-information games also rule out some efficient equilibria. By contrast, NGE includes all Pareto-efficient Nash equilibria as potential outcomes. I shall refer to this property as Pareto-Indeterminacy.²⁰

There are at least two reasons to focus on Pareto-indeterminate solution concepts. First, many papers that assume Pareto-efficiency also assume Pareto-indeterminacy.²¹ If we wish to provide formal foundations for such papers, our solution concepts must be Pareto-indeterminate.

Second, it simply seems wrong to argue that something inherent in language compels us to rule out efficient equilibria in a simultaneous-communication

¹⁹ It is also evident that the "in-between" approaches could in principle be useful in empirical studies--allowing us to measure the "amount" of equilibrium, rather than simply testing the null hypothesis that there is no strategic uncertainty. Implementing this, however, probably would require the use of in-between measures different than those in this paper.

²⁰ A better formal definition would be that a solution concept is Pareto-indeterminate if, for every Pareto-efficient Nash equilibrium, the solution concept includes some correlated equilibrium which weakly Pareto-dominates it. This allows for solution concepts for which the only reason players never play some efficient Nash equilibrium is because they would play a superior correlated equilibrium instead.

²¹ Consider especially the renegotiation literature: Bernheim and Ray [1989], Farrell and Maskin [1989], and related approaches, assume both Pareto-efficiency (or the related idea of non-Pareto-rankability among feasible continuation equilibria) and Pareto-indeterminacy. In order for their solution concepts not to degenerate into an argument for stationary strategies in supergames, there must be a range of efficient continuation equilibria which are feasible at every point. (Pearce [1987] assumes Pareto-efficiency, but does not rely on Pareto-indeterminacy.)

model. For instance, in Example 1 both Banks and Calvert [1989] and Farrell [1987] include only symmetric outcomes. This is not necessarily an unreasonable restriction, but it is clearly not due solely to communication (nor do the authors claim so). If it is common knowledge that player 1 will repeatedly demand they play the (2,1) equilibrium, and that he'll play it even without an agreement, then no natural assumption about language will prevent this outcome.

Indeed, all public theories I have thus far examined that are Pareto-indeterminate are also non-Pareto-efficient. This gives rise to a more challenging question about efficiency: When negotiations consist of repeated, simultaneous statements, can we formulate a reasonable solution concept that is both Pareto-efficient and Pareto-indeterminate?

My conjecture is No. I believe that in many games (such as Example 1), the natural strategic logic of simultaneous communication leads to either the inclusion of inefficient equilibria or the exclusion of some efficient equilibria. Indeed, I will make a stronger conjecture: I believe that Theorems 2 and 3--which place an upper bound on the amount of efficiency guaranteed--essentially hold for all Pareto-indeterminate solution concepts in simultaneous-communication models.

I conclude by discussing possible alternative structures of negotiations, and speculate on some general implications of negotiations in two-player, complete-information games. I shall make some big conjectures without extensive attempts to justify them to the reader.

A simple, common, and natural way to model pre-game communication is to assume that only one player can communicate. It is considered likely by many game theorists that the player who can speak will have all of the bargaining power. Thus, a natural conjecture for unilateral communication is that

whichever player can speak can get his best equilibrium outcome.²² This, of course, would yield a Pareto-efficient solution concept. It also implies that--unlike simultaneous communication--unilateral communication does not lead naturally to solution concepts that are Pareto-indeterminate.

Of course, the true structure of cheap-talk negotiations is often neither simultaneous statements nor unilateral statements. For instance, a natural and commonly used structure is one in which players alternate in making statements. In this case, capturing the right behavioral assumptions can be subtle. At issue is how much emphasis to give the "last-mover advantage": Can the last player to speak during negotiations simply announce that he plans to play his best equilibrium, and get the other player to go along? If so, then alternating-offers negotiations (without discounting, and with a finite number of rounds) will yield results essentially equivalent to the unilateral-statement model.

Alternatively, the last-mover advantage may be less dramatic. The player making the penultimate statement may, for instance, be able to announce his intention to play his favorite equilibrium no matter what the other player's final statement is. If we consider such a statement to be as credible as a last-statement proposal (but no more so), then perhaps the natural assumption in alternating-offers models is to assume that mutual consent is needed to yield agreements. Under this assumption, the results of the model will probably resemble those of the simultaneous-statement model.

This example indicates that bilateral, simultaneous communication and unilateral communication can perhaps be viewed as two extreme forms of pre-game negotiations, with most natural models lying in between. This

²² For an argument against this conclusion, see the "Aumann critique" in Farrell [1988].

suggests, in turn, that if solution concepts derived from the simultaneous-statement and the unilateral-statement structures share certain properties, then these properties may be true for all natural solution concepts based on pre-game communication.

One such generalization is suggested by the results of this paper, and by conjectures made above. I have argued that the simultaneous-statements model can be consistent with Pareto-indeterminacy, or Pareto-efficiency, but not both. Unilateral communication seems consistent with Pareto-efficiency, but not Pareto-indeterminacy. Thus, we might conclude that for any cheap-talk negotiating structure, and for any behavioral assumptions which use the language-correspondence approach, the resulting solution concept will not be both Pareto-efficient and Pareto-indeterminate. Moreover, I believe the inefficiency results--Theorems 2 and 3--essentially hold for all Pareto-indeterminate equilibrium concepts in all such structures.

To close on a positive note, I conjecture that in any situation with lots of communication, Theorem 4 holds for any natural equilibrium concept--Pareto-indeterminate or otherwise. That is, in every communication-rich environment, in every equilibrium of every reasonable model of cheap-talk negotiations, each player will get a payoff at least as high as that in his worst Pareto-efficient Nash equilibrium.

Recall the intuition provided in the Introduction: that with enough communication, a player's lowest payoff ought be determined by the fact that if he wanted, he could fully concede, and beg the other player to play the other player's favorite equilibrium. This intuition does not seem to rely on the simultaneous-statements assumption--and indeed seems hampered by it. With convenient circularity, we might say that any structure of pre-game negotiations that does not provide adequate opportunities for such concession

is not "lots" of communication.

These conjectures indicate that the basic equilibrium results of this paper are somewhat general: In two-player, complete-information games, every equilibrium outcome in every reasonable model with ample opportunity for communication before the game will be a NGE outcome.

Appendix

Definition of a Public Theory:

The set of predictions $(P_1, P_2) \subseteq (\mathcal{L}_1, \mathcal{L}_2)$ is a Public Theory if for $i = 1, 2$, for $j \neq i$,

- 1) $\forall p_i \in P_i, \forall \theta_i = (\sigma, \mu)$ such that $p_i(\theta_i) > 0$, σ is a best response to μ ,
- 2) $\forall p_j \in P_j, \exists p_i \in P_i$ such that $\forall \theta_i = (\sigma, \mu)$ where $p_i(\theta_i) > 0$, $\mu = p_j^S$,
- 3) $\forall \mu$ such that $\exists \theta_i = (\sigma, \mu)$ with $p_i(\theta_i) > 0$ for some $p_i \in P_i, \exists p_j \in P_j$ such that $p_j^S = \mu$,
- 4) If $p_i \in P_i$ and $p_i^* \in P_i$, then $\forall \lambda \in [0, 1], \exists p_i^{**} \in P_i$ such that $\forall \theta_i \in \Theta_i, p_i^{**}(\theta_i) = \lambda \cdot p_i(\theta_i) + (1-\lambda) \cdot p_i^*(\theta_i)$.

Condition 1 is simply that players are rational: in any plan given positive weight by the theory, each player is maximizing his expected utility given his beliefs.

Condition 2 says that for every prediction about player 2's strategy, there is a prediction that player 1 with probability 1 believes that same prediction. This means that a game theorist cannot predict player 2 will probabilistically take certain actions while simultaneously assuming that for sure player 1 won't make such a prediction.

Condition 3 says that for any beliefs that we allow player 1 to have about player 2's strategy, we must include that same prediction about player 2 in our theory. Unlike equilibrium theory, Condition 3 does not dictate that a public theory exclude predictions in which a player's beliefs are incorrect. Rather, it insists that the theory include the prediction that a player's beliefs are correct.

Condition 4 is the convexity assumption. Note that convexity applies over predictions, and not the set of responses by a player for given beliefs. By imposing convexity of predictions for a player, we essentially impose convexity on the sets of permissible beliefs (about what that player will do) for both the game theorist and the other player. If by including each of two beliefs in a theory we mean that reasonable players or analysts might have those beliefs, it would be awkward to rule out a convex combination of predictions as being unreasonable.

Theorem 1:

$\forall G, \forall T, \forall \text{PNLs } Q, \text{NGR}(T, Q)$ is a public theory.

Proof:

Because G contains an equilibrium, by construction, $(\mathcal{L}_1(0), \mathcal{L}_2(0))$ contains an equilibrium pair of plans. Moreover, by construction, this equilibrium pair will be contained in $(\mathcal{L}_1^*, \mathcal{L}_2^*)$, so that $(\mathcal{L}_1^*, \mathcal{L}_2^*)$ is non-empty.

By Definition 3.1 of Rabin [1989], $(\mathcal{L}_1(0), \mathcal{L}_2(0))$ are convex covered sets. By Definition 3.2, $(\mathcal{L}_1^*, \mathcal{L}_2^*)$ is a reduction of $(\mathcal{L}_1(0), \mathcal{L}_2(0))$. Because $(\mathcal{L}_1^*, \mathcal{L}_2^*)$ is non-empty, Theorem 3.1 implies that it is a public theory. Q.E.D.

References

- Banks, Jeffrey S. and Randall L. Calvert, "Communication and Efficiency in Coordination Games", Department of Political Science, University of Rochester, Working Paper No. 196, August 1989.
- Bernheim, B. Douglas, "Rationalizable Strategic Behavior," Econometrica 52, 1007-1028, July 1984.
- Bernheim, B. Douglas and Debraj Ray, "Collective Dynamic Consistency in Repeated Games," Games and Economic Behavior 1, 295-326, December 1989.
- Brandenburger, Adam and Eddie Dekel, "Rationalizability and Correlated Equilibria," Econometrica 55, 1391-1402, November 1987.
- Cho, In-Koo and D. M. Kreps, "Signaling Games and Stable Equilibria," Quarterly Journal of Economics 102, 179-221, May 1987.
- Crawford, Vincent P., "Explicit Communication and Bargaining Outcomes," AER Papers and Proceedings, 213-219, May 1990.
- Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," Econometrica 50, 1431-1451, November 1982.
- Farrell, Joseph, "Cheap Talk, Coordination, and Entry," RAND Journal of Economics 18, 34-39, Spring 1987.
- Farrell, Joseph, "Communication, Coordination, and Nash Equilibrium," Economic Letters 27, 209-214, 1988.
- Farrell, Joseph, "Meaning and Credibility in Cheap Talk Games," in M. Demster (Ed.), Mathematical Models in Economics, Oxford: Oxford University Press, forthcoming.
- Farrell, Joseph and Robert Gibbons, "Cheap Talk Can Matter in Bargaining," Journal of Economic Theory 48, 221-237, 1989.
- Farrell, Joseph and Eric Maskin, "Renegotiation in Repeated Games," Games and Economic Behavior 1, 327-360, December 1989.
- Forges, Francoise, "An approach to Communication Equilibria," Econometrica 54, 1375-1386, November 1986.
- Harsanyi, John C. and Reinard Selten, A General Thoery of Equilibrium Selection in Games, Cambridge, MA: The MIT Press, 1988.
- Kohlberg, E. and J. F. Mertens, "On the Strategic Stability of Equilibria," Econometrica 54, 1003-1038, 1986.
- Matthews, S. and A. Postlewaite, "Pre-Play Communication in Two-Person Sealed-Bid Double Auctions," Journal of Economic Theory 48, 238-263, 1989.
- Matthews, Steven A., Okuno-Fujiwara, Masahiro, and Andrew Postlewaite,

"Refining Cheap-Talk Equilibria," Discussion Paper No. 892, Center for Mathematical Studies in Economics and Management Science, June 1990.

Myerson, Roger, "Multistage Games With Communication," Econometrica 54, 323-58, 1986.

Myerson, Roger B., "Credible Negotiation Statements and Coherent Plans," Journal of Economic Theory 48, 264-303, June 1989.

Pearce, David, "Rationalizable Strategic Behavior and the Problem of Perfection," Econometrica 52, 1029-1050, July 1984.

Pearce, David G., "Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation," mimeo, June 1987.

Rabin, Matthew, "Predictions and Solution Concepts In Non-Cooperative Games," PhD Dissertation, MIT, June 1989.

Rabin, Matthew, "Communication Between Rational Agents," Journal of Economic Theory 51, 144-170, June 1990.

Zapater, Iñigo, "Generalized Communication Between Rational Agents," Brown University Economics Department Working Paper No. 91-10, March 1991.

Example 1

		Player 2	
		L	R
Player 1	U	2,1	0,0
	D	0,0	1,2

Example 2

		Player 2	
		L	R
Player 1	U	1,1	0,0
	D	0,0	1,1

Example 3

		Player 2		
		L	C	R
Player 1	U	3,3	1,0	1,0
	M	0,1	2,-2	-2,2
	D	0,1	-2,2	2,-2

Example 4

		Player 2		
		L	C	R
Player 1	U	-2, -2	1, -3	1, -3
	M	-3, 1	2, -2	-2, 2
	D	-3, 1	-2, 2	2, -2

Example 5

		Player 2			
		B ₁	B ₂	B ₃	B ₄
Player 1	A ₁	4, 1	0, 0	0, 0	0, 0
	A ₂	0, 0	1, 4	0, 0	0, 0
	A ₃	0, 0	0, 0	2, 2	0, 0
	A ₄	0, 0	0, 0	0, 0	3, 3

Example 6

		Player 2		
		L	C	R
Player 1	U	3, 3	1, -2	1, -2
	M	-2, 1	5, -5	-5, 5
	D	-2, 1	-5, 5	5, -5

Figure 1

May 1, 1991

**Working Paper Series
Department of Economics
University of California, Berkeley**

Individual copies are available for \$3.50 in the USA or Canada, \$6.00 to Europe/South America, \$7.00 to Japan/Middle East. Papers may be obtained from the Institute of Business and Economic Research: send requests to IBER, 156 Barrows Hall, University of California, Berkeley CA 94720. Prepayment is required. Make checks or money orders payable to "The Regents of the University of California."

- 89-129 "The Impact of Corporate Restructuring on Industrial Research and Development." Bronwyn H. Hall. December 1989.
- 89-130 "Converters, Compatibility, and the Control of Interfaces." Joseph Farrell and Garth Saloner. December 1989.
- 90-131 "And Now Won/Dollar Negotiations? Lessons From the Yen/Dollar Agreement of 1984." Jeffrey A. Frankel. January 1990.
- 90-132 "One Money for Europe? Lessons from the US Currency Union." Barry Eichengreen. January 1990.
- 90-133 "Does Foreign Exchange Intervention Matter? Disentangling the Portfolio and Expectations Effects for the Mark." Kathryn Dominguez and Jeffrey A. Frankel. January 1990.
- 90-134 "The Constrained Asset Share Estimation (CASE) Method: Testing Mean-Variance Efficiency of the U.S. Stock Market." Charles Engel, Jeffrey A. Frankel, Kenneth A. Froot, and Anthony Rodrigues. January 1990.
- 90-135 "International Nominal Targeting (INT): A Proposal for Overcoming Obstacles to Policy Coordination." Jeffrey A. Frankel. January 1990.
- 90-136 "One-Sided Patience with One-Sided Communication Does Not Justify Stackelberg Equilibrium." Eddie Dekel and Joseph Farrell. February 1990.
- 90-137 "Japanese Finance: A Survey." Jeffrey A. Frankel. March 1990.
- 90-138 "Government Policy, Individual Actions, and Safety: Some Evidence for Motor Vehicles." Theodore E. Keeler. April, 1990.
- 90-139 "Adverse Selection, Short-Term Contracting, and the Underprovision of On-the-Job Training." Benjamin Hermalin. February 1990.
- 90-140 "Why Legal Restrictions on Private Contracts Can Enhance Efficiency." Philippe Aghion and Benjamin Hermalin. April 1990.
- 90-141 "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency." Benjamin E. Hermalin and Michael L. Katz. May 1990.
- 90-142 "The Financial System and the Economic Crisis of the Interwar Years." Barry Eichengreen. June 1990.

- 90-143 "Nonstandard Methods in Mathematical Economics." Robert M. Anderson. June 1990.
- 90-144 "Before the Accord: U.S. Monetary-Financial Policy 1945-51." Barry Eichengreen and Peter M. Garber. June 1990.
- 90-145 "The Impact of Permanent and Temporary Import Surcharges on the U.S. Trade Deficit." Barry Eichengreen and Lawrence H. Goulder. July 1990.
- 90-146 "Trends and Cycles in Foreign Lending." Barry Eichengreen. July 1990.
- 90-147 "Relaxing the External Constraint: Europe in the 1930s." Barry Eichengreen. July 1990.
- 90-148 "The Effects of Competitive Pressures on Executive Behavior." Benjamin E. Hermalin. September 1990.
- 90-149 "The 1933 World Economic Conference as an Instance of Failed International Cooperation." Barry Eichengreen and Marc Uzan. October 1990.
- 90-150 "Costs and Benefits of European Monetary Unification." Barry Eichengreen. October 1990.
- 90-151 "Is Europe an Optimum Currency Area?" Barry Eichengreen. October 1990.
- 90-152 "Major Fiscal Trends in the 1980s and Implications for the 1990s." George Break. October 1990.
- 90-153 "Historical Research on International Lending and Debt." Barry Eichengreen. December 1990.
- 91-154 "Risktaking, Capital Markets, and Market Socialism." Pranab Bardhan. January 1991.
- 91-155 "Is Inequality Harmful for Growth? Theory and Evidence." Torsten Persson and Guido Tabellini. January 1991.
- 91-156 "The Origins and Nature of the Great Slump, Revisited." Barry Eichengreen. March 1991.
- 91-157 "The Making of Exchange Rate Policy in the 1980s." Jeffrey Frankel. March 1991.
- 91-158 "Exchange Rate Forecasting Techniques, Survey Data, and Implications for the Foreign Exchange Market." Jeffrey Frankel and Kenneth Froot. March 1991.
- 91-159 "Convertibility and the Czech Crown." Jeffrey Frankel. March 1991.
- 91-160 "The Obstacles to Macroeconomic Policy Coordination in the 1990s and an Analysis of International Nominal Targeting (INT)." Jeffrey A. Frankel. March 1991.
- 91-161 "Highway Safety, Economic Behavior, and Driving Environment." Theodore E. Keeler. March 1991.
- 91-162 "Can Informal Cooperation Stabilize Exchange Rates? Evidence from the 1936 Tripartite Agreement." Barry Eichengreen and Caroline R. James. March 1991.
- 91-163 "Reneging and Renegotiation." Matthew Rabin. April 1991.
- 91-164 "A Model of Pre-game Communication." Matthew Rabin. April 1991.